Exploring the Genetics of Unresolved Inherited Retinal Diseases in Ireland.

A thesis submitted to

Trinity College Dublin, the University of Dublin for the degree of Doctor of Philosophy.

May 2023



Laura Anne Whelan

B.A. (mod.) Human Genetics

14310322

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish.

This thesis includes unpublished and published work of collaborators which are duly acknowledged in the text wherever included.

SIGNED:



Publication History

Appendix 1

Laura Whelan *et al*. Findings From a **Genotyping Study of Over 1000 People with Inherited Retinal Disorders in Ireland.** *Genes.* 2020 January, **11(1)**, 105.

* Components of this publication appear in section 3.

Appendix 2

Mubeen Khan, Stéphanie S. Cornelis, Marta del Pozo-Valero, **Laura Whelan**, *et al*. **Resolving the dark matter of *ABCA4* for 1,054 STGD1 probands through integrated genomics and transcriptomics.** *Genetics in Medicine.* 2020 April, **22**, 1235–1246.

* Components of this publication appear section 5.

Appendix 3

Zeinab Fadaie, **Laura Whelan**, *et al*. ***BBS1*** **branchpoint variant is associated with non-syndromic retinitis pigmentosa.** *Journal of Medical Genetics.* 2021 May, **0**, 1–7.

* Components of this publication appear section 6.

Appendix 4

Kirk A. J. Stephenson1 , Julia Zhu, Niamh Wynne, Adrian Dockery, Rebecca M. Cairns, Emma Duignan, **Laura Whelan**, *et al*. **Target 5000: a standardized all-Ireland pathway for the diagnosis and management of inherited retinal degenerations.** *Orphanet Journal of Rare Diseases*. 2021 May, **16**:200.

* Components of this publication appear section 7.

Appendix 5

Adrian Dockery, **Laura Whelan**, *et al*. **Next-Generation Sequencing Applications for Inherited Retinal Diseases.** *International Journal of Molecular Sciences.* 2021 May, 22(11)

* Components of this publication appear section 1.

Appendix 6

Zeinab Fadaie**, **Laura Whelan****, *et al*. **Whole Genome Sequencing and *in vitro* Splice Assays Reveal Genetic Causes for Inherited Retinal Diseases.** *Npj Genomic Medicine.* 2021 November, 6, 97.

* Components of this publication appear section 4.

** These authors contributed equally.

Appendix 7

Kirk A.J. Stephenson*, Julia Zhu*, Adrian Dockery, **Laura Whelan** *et al*. **Clinical and Genetic Re-Evaluation of Inherited Retinal Degeneration Pedigrees following Initial**

Negative Findings on Panel-Based Next Generation Sequencing. *International Journal of Molecular Sciences.* 2022 January, 23(2), 995.
* Components of this publication appear throughout.
** These authors contributed equally.

The following publications are not discussed in detail in herein, however, this research was performed in parallel with the primary objectives of this thesis.

Appendix 8
Hedwig M. Velde, Janine Reurink, Sebastian Held, Catherina H. Z. Li, Suzanne Yzer, Jaap Oostrik, Jack Weeda, Lonneke Haer-Wigman, Helger G. Yntema, Susanne Roosing, Laurenz Pauleikhoff, Clemens Lange, **Laura Whelan** *et al.* **Usher syndrome type IV: clinically and molecularly confirmed by novel *ARSG* variants.** *Human Genetics.* 2022 February, 2022 Nov;141(11):1723-1738.

The following manuscripts are under peer review at the time of submission of this thesis.

Laura Whelan *et al.* **Enrichment of *ABCA4* c.4539+2028C>T in Stargardt disease patients in Ireland - a detailed genotype-phenotype analysis.**

Suzanne E. de Bruijn, Kim Rodenburg, Kornelia Neveling, Jordi Corominas, Tamar Ben-Yosef, Janine Reurink, Hannie Kremer, **Laura Whelan** *et al.* **Optical genome mapping and revisiting short-read genome sequencing data reveals previously overlooked structural variants disrupting retinal disease-associated genes.**

Ann O'Connell, Julia Zhu, Kirk AJ Stephenson, **Laura Whelan** *et al.* **MFRP-associated retinopathy and nanophthalmos in two Irish probands - A Case Report**

Funding

Research carried out throughout this doctoral programme was supported by grant awards from Fighting Blindness Ireland (FB Irl; FB16FAR, FB18CRE, FB20DOC), The Health Research Board of Ireland (HRB; POR/2010/97) in conjunction with Health Research Charities Ireland (HRCI; MRCG-2013-8, MRCG-2016-14), the Irish Research Council (IRC; GOIPG/2017/1631) and Science Foundation Ireland (SFI; 16/1A/4452).

Collaborations

The vital role of national and international collaboration in the success of this research endeavour cannot be overstated. Formal acknowledgements are provided for each

section. Patients were primarily assessed and counselled at the Department of Ophthalmology, the Royal Victoria Eye and Ear Hospital, Dublin, Ireland by Mr. Paul Kenna, Ms. Emma Duignan and Dr. Niamh Wynne; the Mater Misericordiae University Hospital Clinical Ophthalmic Genetics Unit, Dublin, Ireland by Mr. David Keegan, Dr Kirk Stephenson, Dr. Julia Zhu, Dr. James O' Byrne and Ms Jacqueline Turner and the Belfast Health and Social Care Trust, Northern Ireland by Ms. Giuliana Silvestri and Ms. Claire Kirk. The bioinformatic pipeline employed in the target-capture arm of this study was designed by Dr. Matthew Carrigan and Dr. Adrian Dockery. Investigation and validation were carried out in collaboration with Dr. Adrian Dockery, Dr. Ciara Shortall, Ms. Ella Kopčić and Ms. Iris JM Post. Supervision was carried out by Dr. Adrian Dockery and Prof. G. Jane Farrar. smMIPs based sequencing of *ABCA4* and WGS was carried out at RadboudUMC, Nijmegen, The Netherlands in collaboration primarily with Dr. Mubeen Khan, Dr. Zeinab Fadaie, Dr. Susanne Roosing and Prof. Frans Cremers. Several undergraduate students aided in the interpretation of candidate variants discovered as part of Target 5000 through optimisation of amplification reactions, sanger sequencing and phase determination (Ben McClinton, Emer King, Rachel Dalton, Clara O' Connor, Ella Kopčić and Anna Ridgeway).

Summary

Target 5000 was established in 2009, following a meeting of Irish clinical specialists, scientists, and patient groups with a primary goal to develop a national care pathway for inherited retinal disease (IRD) patients in Ireland. Participants first undergo a clinical assessment, which includes clinical history and analysis with multimodal retinal imaging, electrophysiology, and visual field testing. If suitable for recruitment, a sample is taken and used for genetic analysis. First tier genetic testing is conducted by use of a retinal gene panel target capture sequencing approach. With over 1000 participants from 710 pedigrees now screened, there is a positive candidate variant detection rate of approximately 70% (495/710). Where an autosomal recessive inheritance pattern is observed, an additional 9% (64/710) of probands have tested positive for a single candidate variant. Many novel variants have also been detected as part of this endeavour. However, it is imperative that we determine the efficacy of more comprehensive sequencing methods in order to provide the 30% of individuals who remain unresolved following preliminary testing with an accurate genetic diagnosis. This doctoral thesis presents data from a pilot study on whole genome sequencing as a diagnostic tool for 100 previously unresolved individuals, with a positive diagnostic yield of 24%. In addition

another pilot study on cost effective whole-gene sequencing of *ABCA4* for 162 Irish Stargardt/cone-rod dystrophy cases is presented, with a positive diagnostic yield of 61% and 59% for pre-screened and unscreened cases respectively. Lastly, three case studies are presented illustrating the application of knowledge gained regarding pathogenic splice altering variants from both pilot studies when applied to existing target capture data. The results presented within this thesis highlight the significant value of a panel target capture sequencing strategy as a preliminary diagnostic measure, with remaining elusive cases undergoing more extensive genetic testing in addition to functional analyses of putative pathogenic variants. The Target 5000 team continues to progress its goal of providing patients with accurate genetic diagnoses and pioneers a multidisciplinary clinical and research care pathway approach with patient-centric care at its core.

" ...and a little more knowledge might light our way" – Yoda in Star Wars: Episode III

Formal Acknowledgements

3. Target capture NGS of the target 5000 cohort.

Conceptualization, P.H., D.K., P.K. and G.J.F.; Data curation, L.W. and A.D.; Formal analysis, L.W. and A.D.; Funding acquisition, G.J.F.; Investigation, L.W., A.D., N.W., J.Z., K.S., G.S., J.T., J.J.O., D.K., P.F.K. and G.J.F.; Methodology, A.D., M.C. and G.J.F.; Project administration, G.J.F.; Resources, G.J.F.; Software, A.D. and M.C.; Supervision, A.D., M.C. and G.J.F.; Validation, L.W. and A.D.; Visualization, L.W. and A.D.; Writing—original draft, L.W. and A.D.; Writing—review & editing, L.W. and A.D. All authors have read and agreed to the published version of the manuscript.

4. Whole genome sequencing and in vitro splice assays reveal genetic causes for inherited retinal diseases.

Z.F. and L.W. are co-first authors. Z.F., L.W., and S.R. performed the data analysis and experiments. C.G., J.C., and G.D.N.A. provided infrastructure and bioinformatic expert input. T.B.-Y., L.I.v.d.B., C.B.H., N.W., E.S.D., and P.F.K. collected clinical cases and performed clinical examinations of patients. Z.C. and L.d.R. contributed to the experiments. S.R., F.P.M.C., T.B.-Y., A.D., and G.J.F. contributed significantly to the design of the study, provided infrastructure, and strategic support. Z.F., L.W., F.P.M.C., and S.R. wrote the manuscript. All authors read and approved the manuscript.

5. Entire ABCA4 gene sequencing of 162 Irish Stargardt disease/cone-rod dystrophy probands.

Conceptualization, L.W., A.D., F.P.M.C. and G.J.F.; methodology, L.W., A.D., K.A.J.S., J.Z., F.P.M.C, D.J.K., P.F.K. and G.J.F. formal analysis, L.W., A.D., K.A.J.S., J.Z., F.P.M.C., D.J.K., P.F.K. and G.J.F.; investigation, L.W., A.D., K.A.J.S., J.Z., E.K., M.K., Z.C., I.J.M.P., N.W., J.J.O.B., E.S.D., S.R., D.J.K., P.F.K., F.P.M.C. and G.J.F; resources, S.R., D.J.K., P.F.K., F.P.M.C. and G.J.F; data curation, L.W., A.D., K.A.J.S., J.Z. M.K., and Z.C. ; writing—original draft preparation, L.W., A.D. and G.J.F. ; writing—review and editing, L.W., A.D., K.A.J.S., J.Z., E.K., M.K., Z.C., I.J.M.P., N.W., J.J.O.B., E.S.D., S.R., D.J.K., P.F.K., F.P.M.C. and G.J.F; visualization, L.W., K.A.J.S. and P.F.K.; supervision, S.R., F.P.M.C. and G.J.F.; funding acquisition, S.R., F.P.M.C. and G.J.F.

6.1 Enrichment of ABCA4 c.4539+2028C>T in Stargardt disease patients in Ireland - a detailed genotype-phenotype analysis.

Conceptualization, L.W., A.D., F.P.M.C. and G.J.F.; methodology, L.W., A.D., K.A.J.S., J.Z., F.P.M.C, D.J.K., P.F.K. and G.J.F. formal analysis, L.W., A.D., K.A.J.S., J.Z., F.P.M.C., D.J.K., P.F.K. and G.J.F.; investigation, L.W., A.D., K.A.J.S., J.Z., E.K., M.K., Z.C., I.J.M.P., N.W., J.J.O.B., E.S.D., S.R., D.J.K., P.F.K., F.P.M.C. and G.J.F; resources, S.R., D.J.K., P.F.K., F.P.M.C. and G.J.F; data curation, L.W., A.D., K.A.J.S., J.Z. M.K., and Z.C. ; writing—original draft preparation, L.W., A.D. and G.J.F. ; writing—review and editing, L.W., A.D., K.A.J.S., J.Z., E.K., M.K., Z.C., I.J.M.P., N.W., J.J.O.B., E.S.D., S.R., D.J.K., P.F.K., F.P.M.C. and G.J.F; visualization, L.W., K.A.J.S. and P.F.K.; supervision, S.R., F.P.M.C. and G.J.F.; funding acquisition, S.R., F.P.M.C. and G.J.F.

6.2 A synonymous variant in PROM1 causes recessive early onset retinal dystrophy through disrupted splicing.

Conceptualization, L.W., A.D., E.S.D. and G.J.F.; methodology, L.W., A.D., E.S.D., P.F.K. and G.J.F. formal analysis, L.W., A.D., E.S.D. P.F.K. and G.J.F.; investigation, L.W., A.D., C.S., E.K., E.S.D. P.F.K., and G.J.F; resources, E.S.D., P.F.K. and G.J.F; data curation, L.W., A.D., and E.S.D.; writing—original draft preparation, L.W., A.D. and E.S.D and G.J.F. ; writing—review and editing, L.W., A.D., C.S., E.K., E.S.D. P.F.K., and G.J.F; visualization, L.W. and E.S.D.; supervision, A.D., E.S.D. and G.J.F.; funding acquisition, E.S.D., P.F.K. and G.J.F.

6.3 A BBS1 branchpoint variant is associated with non-syndromic retinitis pigmentosa

ZF performed data analyses, sequencing analyses, RNA expression studies. LW performed data analyses and segregation analyses. CHZL, LIvdB, CBH, CR, AKL and PK collected clinical cases and performed clinical examinations of patients. CG and JC provided infrastructure, and bioinformatic expert input. JF, FPMC, AD and JME provided strategic support, expert input, and supervision for the project. ZF, LW, FPMC and SR contributed significantly to design of the study. ZF, LW, PK and SR wrote the manuscript.

P.H. = Peter Humphries, D.J.K. = David J. Keegan, P.F.K. = Paul F. Kenna, G.J.F = G. Jane Farrar, L.W. = Laura Whelan, A.D. = Adrian Dockery, N.W. = Niamh Wynne, J.Z. = Julia Zhu, G.S. = Giuliana Silvestri, J.T. = Jaqueline Turner, J.J.O. = James J O Byrne, M.C. = Matthew Carrigan. Z.F. = Zeinab Fadaie, S.R. = Susanne Roosing, C.G. = Christian Gilissen, J.C. = Jordi Corominas, G.D.N.A. = Galuh D.N. Astuti, T.B-Y = Tamar Ben-Yosef, L.I.v.d.B = L Ingeborgh van den Born, C.B.H = Carel B Hoyng, E.S.D = Emma S. Duignan. Z.C. = Zelia Corradi, L.dR. = Laura De Rooij, F.P.M.C = Frans P.M. Cremers, E.K. = Ella Kopčić, I.J.M.P. = Iris J.M. Post, CHZL = Catherina H. Z. Li, C.R. = Charlie Rowlands, R.M. = Roly Megaw, A.K.L. = Anne K. Lampe, J.M.E. = Jamie M. Ellingford.

Acknowledgements

Ar scáth a chéile a mhaireann na daoine.

Professor G. Jane Farrar

I'd like to express my sincere gratitude to you Jane. This entire process has in parts been fun, challenging, exciting, difficult and everything in between and you have been supportive every step of the way. From an academic point of view, thank you for always encouraging my ideas and ambitions. Thank you for always helping me see things from different perspectives – it's made me a better scientist. I'm so grateful for everything you have taught me, and I'll carry it with me for the rest of my career. On a personal level, thank you for always being just a phone call away. Your care and understanding whenever I needed it meant that I couldn't have asked for a better supervisor and role model.

Dr Adrian Dockery

Where do I even begin?! You've taught me everything I know about sequencing and without your willingness to share your knowledge and teach I wouldn't be writing this thesis, so for that I am thankful. Taking on a PhD was scary, but you've always made me feel like I had something to work on or to contribute and that I was in the right place. Above all you made the process *fun*. Working with you in the lab everyday was like hanging out with a friend, raving about something we're both really interested in. It rarely felt like a job. Thank you for reminding me to go home, and for every guilt free weekend I didn't spend in the lab. Thanks for every chocolate bar, can of coke and gin and tonic when needed. Thank you for helping me finish this out even after you had taken on your new role. You've contributed more than you know to the scientist that I am, and I'll be eternally grateful for that.

The Farrar Lab

A huge thank you to members of the Farrar laboratory past and present. Thank you to my lab big sister Dr Laura Finnegan for advising and guiding me from day one. Thank you to Dr Ciara Shortall for teaching me how to be good scientist during my undergraduate thesis project and continuing to do so when you joined team Target 5000. Thank you to Ella Kopčić for being a conscientious scientist and a privilege to teach. I know you'll do brilliant things throughout the rest of your PhD and beyond. Thank you to Dr Daniel Maloney for your help and advice with cells and RNA. Thank you to Dr Matthew Carrigan for all your work on Target 5000 prior to this thesis. To Dr Naomi Chadderton and Dr Sophia Millington Ward – I sincerely appreciate your scientific expertise and for always being there when I needed advice. Thank you to Dr Arpad Palfi for being my first Farrar lab supervisor and for making me do all those megapreps many moons ago.

Clinical Teams
A very special thank you to the brilliant and dedicated Target 5000 clinical teams. Mr Paul Kenna, Mr David Keegan, Ms Giuliana Silvestri, Ms Emma Duignan, Dr Jim O'Byrne, Dr Kirk Stephenson, Dr Niamh Wynne, Dr Julia Zhu, Jackie Turner, Claire Kirk, and Hillary Dempsey. Thank you all for your expertise and input. This project and thesis would not be possible without all of you.

The Nijmegen team
Thank you to Professor Frans Cremers and Dr Susanne Roosing for allowing me to join their research groups for a summer in Nijmegen. This period of my PhD was such a huge learning opportunity. Thank you to Dr Mubeen Khan, Dr Zeinab Fadaie, Dr Suzanne de Bruijn and Dr Janine Reurink for teaching me everything I know about splice variants, midigene analysis and whole genome sequencing. I had such a wonderful experience in The Netherlands and that was made possible by all of you.

The Smurfit Institute of Genetics
Thank you to all of my genetics lecturers for fostering an interest in such an exciting field. Thank you to Brenda Campbell, Phillip Campbell, Dr David Noone, and Rachel Sullivan for their technical help throughout this degree. Thank you to Professor Pete Humphries and Professor Aoife McLysaght for examining my transfer report and for providing valuable feedback and encouragement.

TCD Postgraduate Workers' Alliance
Thanks to my friends Thomas Dineen and Conor Reddy and other members of the TCDPWA for their masses of hard work in fighting for the establishment of working rights for TCD postgraduates and postgraduates nationally.

Honourable Mentions
The entire cast and crew of The Mandalorian, especially Grogu, Firehouse Pizza and whoever invented Coca Cola Zero™.

Sarah and Olivia – thank you both for always being there for me.. It's so great to have friends who you know will always be there and will always be rooting for you no matter how long it's been since you've seen each other. I'm so excited to see where we all end up in the future.

Madeleine – I am so proud of who you have become. I've always been in awe of your perseverance. I've always admired your ability to take risks and they consistently pay off. You've continually reminded me to put my happiness first and to do what I love. I'm so glad you're following your dreams.

Derv – Thank you for being my soul sister, for always understanding me and for always listening when I needed you. I've missed you every day I haven't been able to see you. I'm so proud of you and of everything you've achieved. The invisible string that ties us together is at times a lifeline.

Kristina – Thank you for being an ever-present source of excitement, joy, and positivity. There is no better person to turn to for encouragement and a smile. Thank you for always making life fun, even if it's just dancing in the kitchen at 10pm. Thank you for always being an enabler and energiser, for always being the kind of person I'd like to be. I'm lucky to have you as a friend.

Ciarán – Thank you for always being present on this journey with me. Thank you for every piece of advice you've given me. Thank you for understanding my frustrations and celebrating my successes. I'm so excited for you and all the good things you have coming your way. I hope we keep visiting Gaoith Dobhair and have people wonder why we never talk about anything but genetics even when we're ninety.

Joseph – Thank you for being my partner in all of this. Thank you to your mam and dad, Eileen, and Pat for always giving us somewhere lovely to escape to when needed. Thank you for always being there and always listening, Thank you for the reassurances and the encouragement. Thank you for believing in me, for telling me that I could do anything I set my mind to. Thank you for choosing to do this with me.

Mam – Thank you for always showing the importance of education and encouraging my interests. Thanks for reading everything I've ever written. Thanks for always being on the end of the phone for good and bad calls. Thank you for every piece of advice and for reminding me to eat the elephant one bite at a time.

Dad – Thank you for always having an interest and listening and for always reminding me I work too much and to take time off. Thanks for all the drives to school and lifts home from supervised study – I hope you think that's all paid off now.

Catherine – Thank you for being the best big sister to look up to. Thank you for encouraging me to stay in college as long as possible – think I'm doing a pretty good job of that. Thank you for always being a voice of reason and for always only being a message away.

Neasa – This one's for you best girl. Thank you for always putting a smile on my face. I am so proud to be your aunty and godmother. Follow your dreams and interests wherever they take you. You can do anything you set your mind to.

Nana, Da and Billy - I miss you all every day. I hope this work makes you proud. I know you'd have been front row at my viva or graduation or would have loved to keep hearing about life in Dublin. I'll have a cocktail for each of you when this is done

Table of Contents

1. INTRODUCTION 16

1.1 The retina and visual transduction. 16
1.2 Inherited retinal diseases 17
1.3 A brief history of DNA sequencing 20
1.4 Next-generation sequencing and its application to IRDs currently 22
1.5 The splicing process and its contribution to IRD causation 27
1.6 Target 5000 31
1.7 Aims and objectives of the doctoral thesis. 32

2. MATERIALS AND METHODS 32

2.1 Ethics 33
2.2 Patient sites and clinical assessment 33
2.3 Sample acquisition and extraction 34

3. TARGET CAPTURE NGS OF THE TARGET 5000 COHORT 35

3.1 Introduction 35
3.2 Methods 35
3.2.1 DNA isolation 35
3.2.2 Target capture sequencing and variant confirmation 35
3.2.3 Data analysis and variant interpretation of target capture NGS data 36
3.2.4 CHM deletion confirmation 37
3.2.5 Sequencing of RPGR ORF15 40
3.2.6 Single-molecule molecular inversion probe-based sequencing of ABCA4 40
3.3 Results 41
3.3.1 Clinical presentation and positive candidate detection rates 41
3.3.2 Retinitis pigmentosa 42
3.3.3 Stargardt disease and other macular dystrophies 46
3.3.4 Usher syndrome 49
3.3.5 Other IRDs encompassed by Target 5000 51
3.3.6 Novel variants 55
3.3.7 RPE65 58
3.3.8 FLVCR1 58
3.3.9 CHM 60
3.4 Discussion 63

4. WHOLE GENOME SEQUENCING AND *IN VITRO* SPLICE ASSAYS REVEAL GENETIC CAUSES FOR INHERITED RETINAL DISEASES. 68

4.1 Introduction 68
4.2 Methods 69
4.2.1 Patient cohort 69
4.2.2 DNA acquisition and preliminary genetic analysis 69
4.2.3 Variant prioritization approach 71
4.2.4 Variant frequency, pathogenicity prediction and midigene splice assays 71
4.3 Results 75
4.3.1 Patient characterization 75
4.3.2 Identification of coding and noncoding candidate variants after WGS data analysis 83
4.3.3 Splice defects due to NCSS variants in CDHR1, HGSNAT, and USH2A 93
4.3.4 Splice defects due to deep-intronic variants in ABCA4, EYS, GRM6, and USH2A 95
4.4 Discussion 104

5. ENTIRE *ABCA4* GENE SEQUENCING OF 162 IRISH STARGARDT DISEASE/CONE-ROD DYSTROPHY PROBANDS 110

5.1 Introduction 110
5.2 Methods 112
5.2.1 smMIPs design 112
5.2.2 Data analysis and variant classification 113
5.2.3 Data processing 114
5.2.4 CNV analysis 114
5.3 Results 115
5.3.1 Patient cohort 115

5.3.2 Candidate variant detection rate *115*
5.3.3 Intronic variants are a significant cause of STGD1 and CRD in Ireland *124*
5.4 DISCUSSION 125

6. CASE STUDIES FURTHER ILLUSTRATING THE DISEASE-CAUSING POTENTIAL OF TRADITIONALLY OVERLOOKED SPLICE ALTERING VARIANTS. 131

6.1 ENRICHMENT OF *ABCA4* C.4539+2028C>T IN STARGARDT DISEASE PATIENTS IN IRELAND - A DETAILED GENOTYPE-PHENOTYPE ANALYSIS 131
6.1.1 Introduction *131*
6.1.2. Materials and methods *132*
6.1.2.1 Participants 132
6.1.2.2 Clinical assessment 134
6.1.2.3 Sequencing procedures 134
6.1.2.4 Variant categorisation 135
6.1.2.5 Statistical analyses and data visualisation. 135
6.1.3 Results *136*
6.1.3.1 *ABCA4* c.4539+2028C>T is enriched in the Irish STGD1 population 136
6.1.3.2 Clinical characterisation of individuals carrying ABCA4 c.4539+2028C>T 140
6.1.3.2.1 Visual acuity 140
6.1.3.2.2 Fundus autofluorescence and optical coherence tomography. 143
6.1.3.2.3 Electroretinography 152
6.1.3.2.4 Age at onset 156
6.1.3.3 Variant Assessment in accordance with the ACMG-AMP guidelines 157
6.1.4 Discussion *158*
6.2 A SYNONYMOUS VARIANT IN *PROM1* CAUSES RECESSIVE EARLY ONSET RETINAL DYSTROPHY THROUGH DISRUPTED SPLICING 163
6.2.1 Introduction *163*
6.2.2 Materials and Methods *163*
6.2.2.1 Patient details 163
6.2.2.2 Clinical evaluation 163
6.2.2.3 Genetic analyses 164
6.2.3 Results *165*
6.2.3.1 Clinical findings 165
6.2.3.2 Genetic findings and in silico analyses 170
6.2.3.3 Functional analyses 172
6.2.4 Discussion *175*
6.3 A BBS1 BRANCHPOINT VARIANT IS ASSOCIATED WITH NON-SYNDROMIC RETINITIS PIGMENTOSA 178
6.3.1 Introduction *178*
6.3.2 Material and methods *178*
6.3.2.1 Clinical evaluations 178
6.3.2.2 Genetic analysis and variant prioritisation 179
6.3.2.3 In vitro splice assay 179
6.3.3 Results *180*
6.3.2.1 Clinical findings 180
6.3.2.1 Genetic findings 182
6.3.4 Discussion *183*

7. DISCUSSION 186

7.1 THE EVOLUTION AND EXPANSION OF TARGET 5000 186
7.2 CHALLENGES FACED IN THE INTERPRETATION OF IRD VARIANTS 187
7.3 GENOMIC SERVICES IN IRELAND WITH A FOCUS ON IRD PATIENTS 192
7.4 THE HEALTH ECONOMICS OF IRDS IN IRELAND 194
7.5 THE PERSONAL IMPACT OF GENETIC TESTING FOR IRDS 196
7.6 CURRENT GENE BASED THERAPIES/CLINICAL TRIALS FOR IRD. 198
7.7 FUTURE PERSPECTIVES ON GENETIC DIAGNOSES FOR IRD 202
7.8 BEYOND GENOMICS 207

8. CONCLUSION 211

REFERENCES 212

List of abbreviations

AAV	– Adeno-associated virus
ACMG	– American College of Medical Genetics and Genomics
AF	– Allele frequency
AMD	– Age-related macular degeneration
AMP	– Association for Molecular Pathology
AMP	– Association for Molecular Pathology
AON	– Antisense oligonucleotide
BBS	– Bardet Biedl syndrome
BWA	– Burrows-Wheeler aligner
CNV	– Copy number variant
CRD	– Cone-rod dystrophy
CRT	– Central retinal thickness
CSS	– Canonical splice site variant
DIV	– Deep-intronic variant
ERG	– Electroretinography
FAF	– Fundus autofluorescence
HGP	– Human genome project
IRD	– Inherited retinal disease
LCA	– Leber congenital amaurosis
LHON	– Leber hereditary optic neuropathy
LOVD	– Leiden open variant database
MES	– MaxEntScan
NCSS	– Non-canonical splice site variant
NGS	– Next-generation sequencing
NNS	– NNSPLICE
OCT	– Optical coherence tomography
PCARP	– Posterior column ataxia with retinitis pigmentosa
PCR	– Polymerase chain reaction
PPC	– Photoreceptor precursor cell
ROI	– Republic of Ireland
RP	– Retinitis pigmentosa
RPE	– Retinal pigment epithelium
smMIPs	– Single molecule molecular inversion probes
SNV	– Single nucleotide variant
SSFL	– SpliceSiteFinder-like
STGD1	–Stargardt disease
SV	– Structural variant
TCS	– Target capture sequencing
VUS	– Variant of unknown significance
WES	– Whole exome sequencing
WGS	– Whole genome sequencing

1. Introduction

1.1 The retina and visual transduction.

The eye is a highly complex and multi-faceted organ. The anterior components of the eye receive light and focus it on the back of the eye. The retina is located in the posterior of the eye and is an outpost of the central nervous system, marking the beginning of the visual signalling pathway. The retina consists of a variety of neuronal cell layers made up of distinct neurons, including photoreceptors, bipolar cells, ganglion cells and interneurons (Figure 1.1). Photons of light are first detected by the 11-cis-chromophore (Figure 1.2), which binds to opsin proteins found within the photoreceptor cells, of which there are two types, rods and cones. The vast majority of photoreceptors are rods, with approximately 90 million distributed throughout the human retina. Cone cells, of which there are between 4-6 million reported in the literature, are found predominantly in the macula, and are the only photoreceptor cell type the fovea[1].

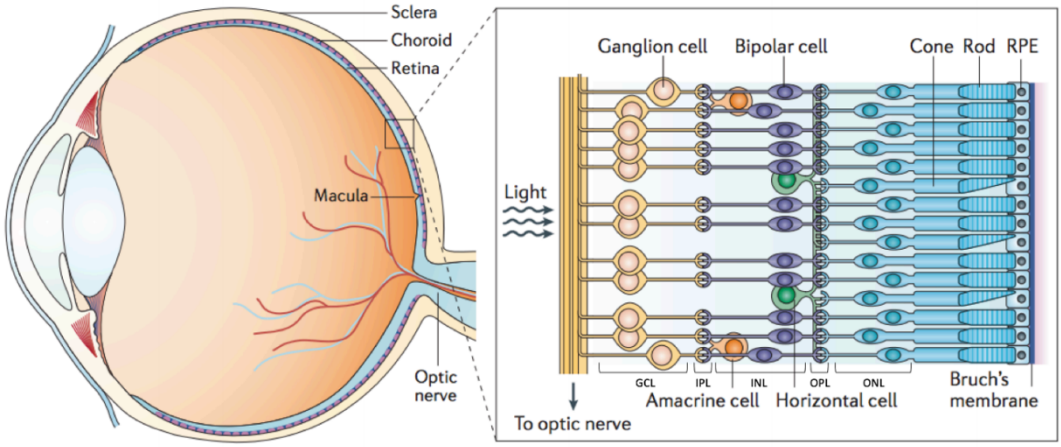


Figure 1.1 | The eye and retinal cell types. The retina is the innermost layer of the eye. The retina consists of three neuronal cell layers: The photoreceptor layer, a middle layer consisting of amacrine, bipolar and horizontal cells and an innermost layer of ganglion cells. These cells facilitate the transmission of neuronal signalling from photoreceptors to the visual cortex. GCL: ganglion cell layer; IPL: inner plexiform layer; INL: inner nuclear layer; OPL: outer plexiform layer; ONL: outer nuclear layer (adapted from Kimbrel and Lanza, 2015[2]).

When the 11-*cis*-chromophore binds to an opsin protein a signal is induced in the rod and cone photoreceptors. Rods have a highly responsive opsin, rhodopsin, which facilitates perception of dim light, peripheral and monochromatic vision. Cone opsins are not as sensitive but are capable of detecting light of a variety of wavelengths facilitating trichromatic vision[3]. Their concentration at the centre of the retina, the fovea, allows for a high level of visual acuity. While multiple rods converge on a single bipolar cell, only one cone converges on a singular bipolar cell, allowing for the perception of fine detail[4]. Once

signals have been transduced from photoreceptors to bipolar cells, they are then transmitted to the ganglion cells[5], which produce different outputs based on the signals received.

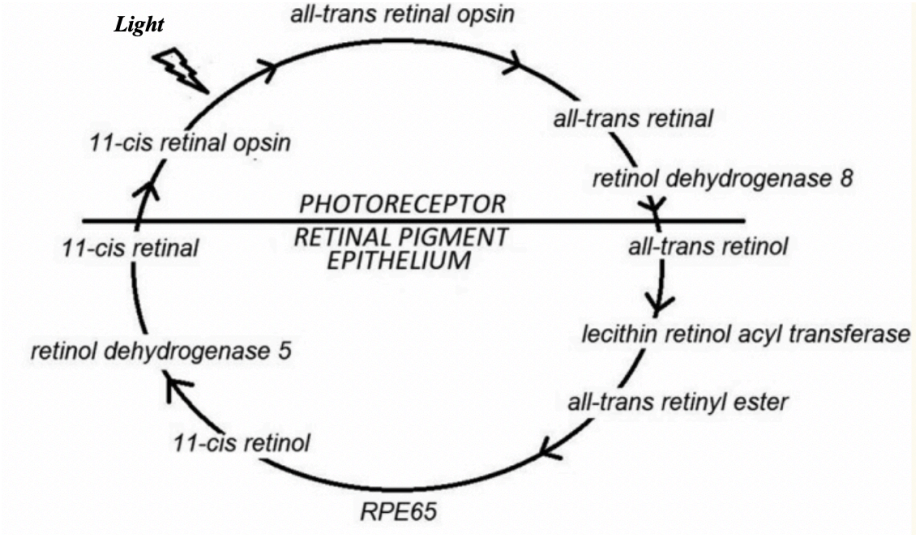


Figure 1.2 | The classical visual cycle. A photon of light facilitates conversion of 11-cis-retinal to all-trans-retinal, resulting in hyperpolarisation of the photoreceptor. The change in action potentials trigger nerve impulses that the brain interprets as vision. Following this conformational change and release from the opsin, all-trans-retinal is reduced to all-trans-retinol and transferred to the RPE. It is converted to all-cis-retinal via esterification and by isomerohydrolase RPE65. Finally it is oxidised to 11-cis-retinal, returns to the photoreceptors to bind with the opsins, starting the cycle again (adapted from Tanaka and Tachabana 2013[6]).

Horizontal cells and amacrine cells are interneurons which interfere with the main axis of signal transduction[7]. Horizontal cells process and integrate the light signal from photoreceptors, while amacrine cells regulate transduction of bipolar cell information to ganglion cells, primarily through lateral inhibition[6]. The signal is further processed by the ganglion cell layer, the axons of which extend posteriorly to the optic disc and exit the eye via the optic nerve. The signal is transmitted to the thalamus and then to the visual cortex where an image is formed and processed[8]. Although this description greatly over-simplifies the visual transduction process, it is clear that disturbances to the genes that encode any key component of the retina have the potential cause an inherited retinal disease (IRD).

1.2 Inherited retinal diseases

IRDs are a subset of ocular diseases that have been extensively studied over the last four decades, revealing a vast landscape of phenotypic diversity and genetic heterogeneity. At least 50 different IRD subtypes have been described to date[9]. The diversity of these diseases persist in terms of age at onset. IRDs can manifest at developmental stages, they may congenital (Leber congential amourosis (LCA) for

example, or become apparent in later decades of life. Some IRDs are stationary but for others the rate at which visual function declines is also highly variable. The areas of the retina affected also vary. Some IRDs manifest in the peripheral retina, for example retinitis pigmentosa (RP), due to the loss of rod photoreceptor cells (Figure 1.3). Other IRDs will first affect the central retina, for example cone dystrophies due to the loss of cone photoreceptor cells in the macula and fovea (Figure 1.4). Other tissues within the eye can also be affected in addition to the retina. For example the choroid in patients with choroideremia or the optic nerve in patients with dominant optic atrophy. Some cases may display widespread disease from the onset, while others will begin in one primary region and eventually affect secondary regions.

This clinical variablilty results in a myriad of IRD associated symptoms. IRDs predominately affecting the rod photoreceptors may first manifest as nyctalopia and peripheral field loss. This makes it difficult for patients to navigate in low light environments. This can then be followed by cone involvement and central vision loss. Patients may notice islands of scotoma in their visual field which eventuallyforms a ring in the periphery leading to severe visual contriction. IRDs predominately affecting the cone photorecptors result in decreased visual acuity, color vision disturbance and photophobia. This can be followed by subsequent rod dysfunction and peripheral vision loss.

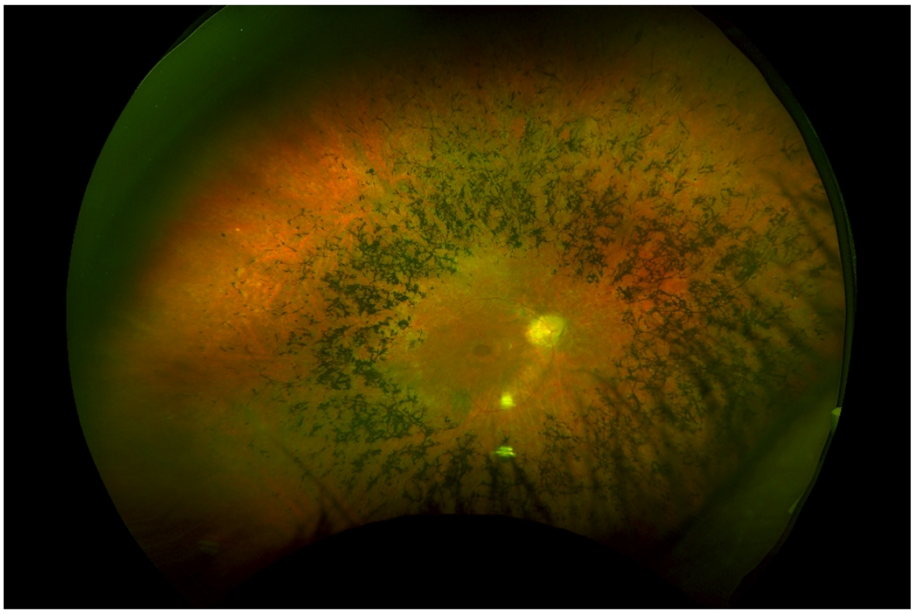


Figure 1.3 | A typical retinitis pigmentosa fundal phenotype. There is disc pallor, vessel attenuation and bone-spicule pigments in the mid-periphery with retinal atrophy (photo from gene.vision[10]).

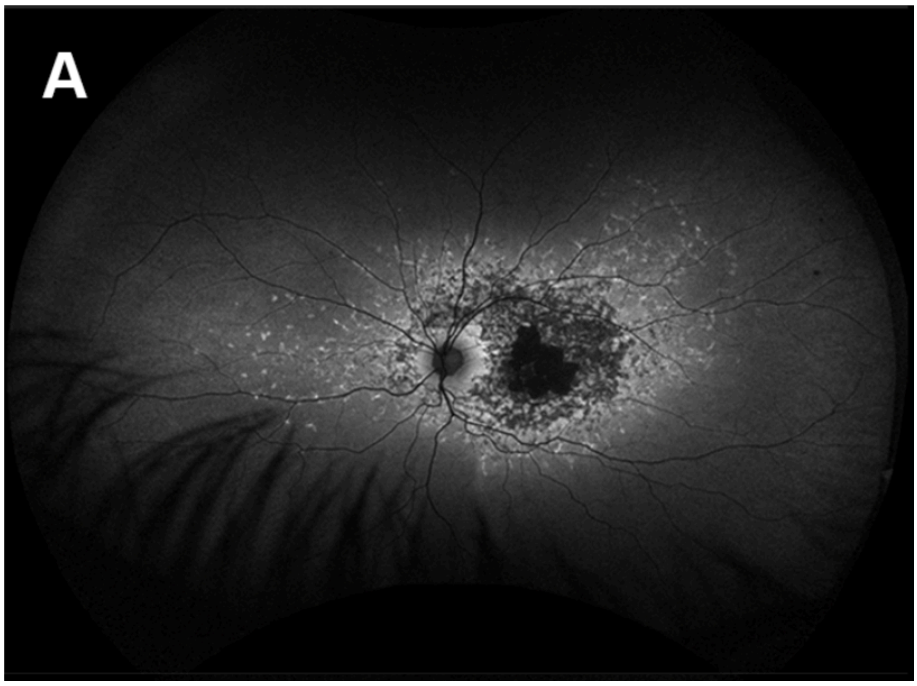


Figure 1.4 | Fundus autofluorescence imaging of a patient with *ABCA4* associated disease. The patient has a central dark area of atrophy confined to the posterior pole. Flecks are seen up to the mid-periphery (adapted from gene.vision[11]).

In the majority of IRDs, ocular manifestations occur in isolation. For example, many subtypes of retinitis pigmentosa or Stargardt disrease. When disease feautures in other organs are present in addition to the ocular phenotype, such diseases are referred to as syndromic IRDs. 80 conditions with a retinal phenotype in addition to 200 genes associated with retinal health have been described[12]. Some of the more comparatively common forms of syndromic IRD in Ireland include Usher Syndrome and Bardet-Biedl syndrome (BBS)[13,14], albeit both are still rare orphan disorders. Usher syndrome is sub-divided into four categories and systemic features include bilateral sensorineural hearing loss in additional to RP. BBS is a multi-systemic disease and additional features include obesity, polydactyly, renal abnormalities and learning difficulties among other manifestations.

Underlying this clinical heterogeneity is a highly complex genetic foundation with some sources citing over 300 genes implicated as causative to date. The RetNet database lists 281 genes[15]. The Genomics England PanelApp[16], a crowdsourcing tool housing gene panels shared by the scientific community, lists 413 genes under retinal disorders of which 253 genes have been categorised as ‘’green’’ i.e. are clinically actionable. Blueprint Genetics, a widely used commercial sequencing company uses 351 genes in its retinal dystrophy panel. The aforementioned clinical diversity relies not only on the gene that is affected, but also on the specific type of pathogenic variant(s) and their location within the gene. For example, pathogenic variation in the *RP1* gene causes either RP or cone-rod

dystrophy (CRD) dependent upon mutation type and its location[17]. Another example is Best disease, caused by pathogenic variants in *BEST1.* This disease can be dominantly or recessively inherited based on the location and type of the variation within the gene[18]. Additionally, multiple phenocopies exist whereby one clinical presentation may not be readily distinguishable from another despite completely different underlying genetic causes. An RP diagnosis epitomises this, with pathogenic variation in over 70 genes implicated as causative of to date[19]. Other IRDs are less heterogeneous at the gene level, autosomal recessive Stargardt Disease (STGD1), for example, which is caused by pathogenic variations in *ABCA4* in approximately 95% of cases[20–23]. To add to this complexity, modifier loci have been identified and the lines between Mendelian and polygenic forms of IRD continue to be blurred[24,25]. These intricacies highlight the need for an accurate and comprehensive clinical and genetic diagnostic approach – DNA sequencing.

1.3 A brief history of DNA sequencing

The last four decades represent an age of technological revolution. DNA sequencing is one of the many scientific advances to have been discovered and developed exponentially during this time. Starting with base pairs, to kilobases and now to millions and billions of bases across multiple fully sequenced genomes, DNA sequencing represents a *tour de force* in biological research. Early sequencing was pioneered by Fred Sanger, who strongly believed that uncovering the chemical structure of biological molecules was imperative to truly facilitate greater understanding[26].

Initial endeavours in DNA sequencing set the scene for a challenging road ahead. Wu and colleagues used a primer extension methods to describe 12 bases of bacteriophage lamda[27]. This was followed by a 2 year-long study by Gilbert and Maxam in 1973 reporting 24 bases of the lactose-repressor binding site. Sanger and Coulson and Maxam and Gilbert developed the chain terminator procedure and the chemical cleavage procedure respectively in the mid 1970s[28,29]. Shotgun sequencing was pioneered by Staden in 1979[30]. This led to the de novo assembly of bacteriophage lamda in 1982[31]. Smith, Hood and Applied Biosystems created automated, fluorescence-based Sanger sequencing machines by 1987[32].

With a focus on our own genomes, the Human Genome Project (HGP) was established in 1990 and relied on cloning large fragments of DNA into bacterial artificial chromosomes and subsequent Sanger sequencing. The cumbersome nature of this approach spurred attempts to improve on costs and efficiency. The 1990s saw these efforts

begin to come to fruition with improvements including but not limited to the number of sequencing reactions occurring at once[33], DNA purification methods[34], double-stranded DNA sequencing, capillary electrophoresis[35] and the adoption of industrial processes such as automation and quality control. The first draft of the HGP's human genome was achieved in 2001 with a final sequence obtained in 2004[36,37]. In parallel a private project by Celera Genomics and Craig Ventor was carried out[38].

Following the completion of the HGP, next-generation sequencing (NGS) was born. Also known as second generation or massively parallel sequencing, a key difference from early first generation sequencing is the ability to multiplex samples. As opposed to the bacterial cloning methods mentioned previously, NGS utilises a "sequencing by synthesis" approach that involves *in vitro* amplification of the template. An approach developed by Balasubramanian and Klenerman in 1998 and implemented by Solexa relied on polymerase-mediated incorporation of fluorescently labelled deoxynucleotides and subsequent imaging[39–41]. In 2005, Escherichia coli was resequenced using NGS by Shendure, Porreca, Mitra and Church[42]. Additionally Mycoplasma genitalium was assembled de novo by 454, Margulies and Rothberg. Resequencing of phiX174 and a human BAC by Solexa was also achieved at this time[43].

The first commercial NGS platform was released by 454 in same year. While Applied Biosciences monopolised the sequencing arena throughout the HGP era, a multitude of sequencing companies including 454, Complete Genomics, Ion Torrent and Solexa were competing heavily throughout the early 2000s. This NGS race and the advancements it produced led to a substantial decrease in the cost of sequencing between 2007 and 2012[44]. Since then, Illumina has become the sequencing platform of choice for many clinical and research laboratories globally.

Third generation sequencing, defined herein as sequencing that does not require amplification, is no longer on the horizon, but rather is readily implemented in many laboratories in both academia and industry[20]. Two leading approaches include PacBio sequencing and nanopore sequencing. The former relies on real-time optical observation of polymerase-mediated synthesis[45,46]. The latter works off the principle that when single-stranded DNA passes through a narrow channel it will produce a pattern dictated by the flow of ions[47].

It is clear that DNA sequencing has revolutionised biomedical research and will likely greatly transform clinical medicine in years to come. As gene-based therapeutics become more widely available, DNA sequencing will be an essential tool to identify

patients who will benefit from these medicines. IRDs serve as a model of its potential use in clinical and research settings.

1.4 Next-generation sequencing and its application to IRDs currently

Over the past decade, IRDs have become routinely genetically diagnosed through various DNA sequencing approaches. This is due in part to a high level of Mendelian inheritance associated with these conditions, as opposed to more complex polygenic diseases such as heart disease, autism or schizophrenia. IRDs represent ideal candidates for diagnosis via NGS due to a wealth of data generated to suggest primary manifestation of disease due to specific, singular pathogenic variants. It is important to introduce the field of sequencing for IRDs in its current state comprehensively, given the wide-reaching and varied approaches that are currently being implemented.

The NGS procedure of choice for IRDs has traditionally been a targeted capture sequencing (TCS) approach, whereby a fully customisable panel of genes or subsets of their components are sequenced. The primary reasons for adopting this strategy include cost-effectiveness in terms of sequencing, data storage and analysis. Maximal coverage of captured regions is also beneficial as well as minimising the risk of detecting secondary findings by only sequencing and analysing IRD-associated genes. Disadvantages of TCS include the possibility of having to design multiple panels for various diseases or re-design and re-sequence upon the discovery of new IRD associated genes or variants. Heterozygous structural or copy number variants (CNVs) have proven difficult for TCS to detect, given that breakpoints can lie in regions not covered by the panel.

The use of whole exome sequencing (WES) has become increasingly popular in IRD diagnostics in recent years given decreasing cost. WES allows the detection of variation in all discovered exons of the genome. Common uses focus on extracting panels of sequence data from IRD associated genes, leaving the "excess" data to be interrogated if the need arises in the future. This circumvents the need for resequencing when new disease associations are identified. However, WES is not without its disadvantages. It is limited to only detecting intra-exon variation or variation that lies near these exonic regions. It is not capable of detecting variants located deep within the intronic regions of DNA. Targeted sequencing is capable of this if known deep-intronic variants (DIVs) considered in the panel design. Whole genome sequencing (WGS) also allows for the identification of such variation.

WGS is the most comprehensive DNA sequencing procedure, facilitating detection of variation across the entire genome. Lionel *et al.* investigated 103 cases of

diverse genetic disorders comparing WGS to targeted panel sequencing[48]. Not only was the solve rate superior when WGS was used, 41% versus 24%, 18 diagnoses were made based on structural variants or intronic variants that were not captured by the targeted sequencing method[48]. Similar findings have been obtained in IRD focused studies[49–52]. Although targeted sequencing has the ability to target a large number of IRD associated genes, results to date suggest that approximately 43% of cases remain unresolved and that studies are employing whole genome sequencing to address this. (Table 1.1). WGS enables the discovery of some non-coding single nucleotide variants, structural variants and CNVs that are not readily detected by targeted sequencing and WES. Disadvantages include the production a vast amount of data that requires sophisticated downstream processing. This includes but is not limited to computational power and specialised scientific staff. Another important consideration for both WGS and WES is the discovery of secondary findings not directly related to the phenotype that is being investigated. Figure 1.5 provides a comparison of target capture sequencing, WES and WGS in terms of coverage, time/cost effectiveness and depth.

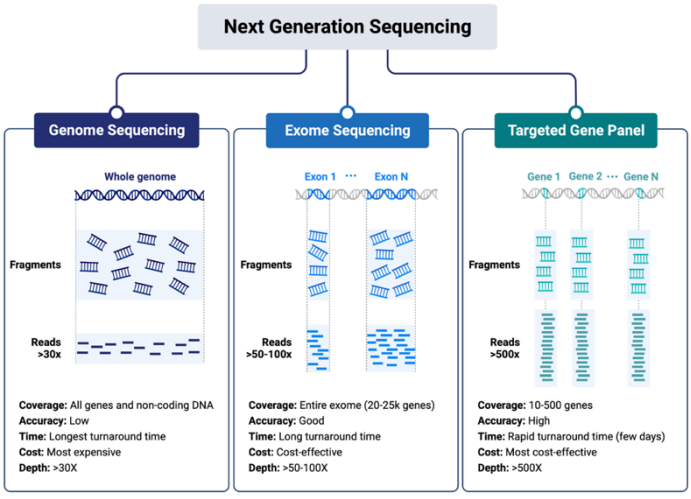


Figure 1.5 | A Comparison of various forms of next-generation sequencing (Image from BioRender[53]). WGS is optimal in terms of coverage, while TCS is optimal in terms of turnaround time, cost effectiveness and sequencing depth. WES provides a viable intermediate.

Whole sequencing of indidvidual genes is a cost effective method currently being employed when the disease is primarily caused by pathogenic variation in a singular gene. *ABCA4* related disease exemplifies the use of this approach. A recent study by Khan *et al.*, utilised fully automated, high throughput, cost-effective and comprehensive sequencing of the entire *ABCA4* gene using single-molecule molecular inversion probes (smMIPs) on over 1000 unresolved STGD1 and CRD probands[54]. Advantages of smMIPs include reduced costs, detection of low frequency targets, high multiplexing, ease of new

probe addition and the ability to successfully sequence with a low quantity/concentration of starting material[55]. In combination with *in vitro* functional analysis of putative splice altering variants this study identified SVs and DIVs in 25% of biallelic cases. Patient samples from Target 5000 were included in this study and additional samples were also sequenced with this approach. This is discussed in detail in section 4. A similar approach has also been employed for the *USH2A* gene[56].

Studies carried out on IRD cohorts globally report a diagnostic yield of between 25 and 100%[57–61] (Table 1.1). This range can be expected due to a number of factors. Firstly, deep-phenotyping to accurately categorise IRDs greatly assists with genotype-phenotype assessment. Assessor variability will exist between clinics and studies, contributing greatly to the variability of diagnostic yield. Secondly, there are differences between the numbers of genes and regions sequenced in these cohorts, where smaller gene panels will not capture some of the known genes associated with IRDs[58]. Lastly, the employment of additional analyses such as WES or WGS in some other studies has increased the detection of candidate variants in previously unresolved cases[52,61–63]. Recent studies have indicated that non-coding variants can be causative of some forms of IRD. This suggests that much of the remaining missing heritability associated with IRDs may be due to insufficient investigation of the non-coding portions of these IRD genes. As mentioned previously, limitation of exploration to the coding segments of genes overlooks disease caused by CNVs, some structural variants (SVs) as well as regulatory variants[64,65].

It is clear that a wide variety of DNA sequencing approaches are currently being employed for the genetic diagnosis of IRDs globally. In Ireland, genetic testing for IRD patients is not routinely carried out as part of clinical care and takes place predominantly in the research setting followed by clinical confirmation. The importance of identifying the genetic cause of IRDs is becoming increasingly clear, allowing for more accurate prognosis of disease progression and risk assessment for other family members. Importantly, a genetic diagnosis can also facilitate a better understanding of the disease for those affected. Furthermore, multiple clinical trials to treat these diseases are ongoing, with the first gene therapy for IRD, Luxturna™, having been approved for use by both the US Food and Drug Administration and the European Medicines Agency for biallelic *RPE65* IRDs[66]. These trials and treatments generally require that a causative variant has been identified in order to participate. Given that a large proportion of cases remain unresolved following targeted sequencing approaches of the exons of IRD associated genes, it is imperative that we investigate alternate sources of disease aetiology - one such example is aberrant splicing.

Table 1.1 | Overview of several sequencing studies on inherited retinal disease cohorts. LCA: Leber Congenital Amaurosis, IRD: Inherited Retinal Disease, RP: Retinitis Pigmentosa, MD: Macular Dystrophy, CRD: Cone-rod dystrophy, TCS: Targeted capture sequencing, WGS: Whole genome sequencing. N.A. = not available.

Country	Author	Year	Pedigrees	Solve Rate (%)	Cohort Details	TCS	WES	WGS
Australia	Thompson[67]	2017	34	90	LCA	Yes	\	\
Brazil	Motta[68]	2018	559	72	IRD	Yes	Yes	\
China	Liu[69]	2020	800	60	RP	Yes	Yes	\
China	Gao[70]	2019	1243	72	RP	Yes	\	\
China	Liu[71]	2020	182	48	IRD	Yes	Yes	\
China (Han)	Huang[72]	2017	98	41	RP	\	Yes	\
China	Dan[73]	2020	76	57	IRD	Yes	Yes	\
China	Wang[74]	2018	319	39	IRD	Yes	Yes	\
Finland	Avela[75]	2019	53	77	IRD	Yes	\	\
Germany	Weisschuh[76]	2020	1785	69	IRD	Yes	\	\
Germany	Birtel[77]	2018	251	74	MD/CRD	Yes	\	\
Iran	Tayebi[78]	2019	50	72	IRD	Yes	\	\
Ireland	Whelan[13]	2020	710	70	IRD	Yes	\	\
Israel	Sharon[50]	2020	2420	56	IRD	Yes	Yes	Yes
Japan	Koyanagi[79]	2019	1204	30	RP	Yes	\	\
Japan	Numa[80]	2020	220	45	RP	Yes	\	Yes
Korea	Surl[81]	2020	50	78	LCA	Yes	Yes	\
Korea	Kim[82]	2019	86	44	IRD	Yes	\	\
Mexico	Zenteno[83]	2019	143	66	IRD	Yes	\	\
Norway	Holtan[84]	2020	650	32	IRD	Yes	\	\
Poland	Wawrocka[85]	2018	18	39	CRD	Yes	Yes	\
Polynesian and Māori	Vincent[86]	2017	16	44	IRD	Yes	\	\
Spain	Perea-Romero[49]	2021	3951	53	IRD	Yes	Yes	Yes
Spain	Martin-Merida[87]	2019	877	38	RP	Yes	\	\
Spain	Gonzàlez-Duarte[88]	2019	73	85	IRD	Yes	\	\
Spain	Diñeiro[89]	2020	100	45	IRD	Yes	\	\
Taiwan	Chen[90]	2020	60	53	IRD	Yes	\	\
Tunisia	Habibi[91]	2020	73	68	IRD	\	Yes	\

UAE	Khan[92]	2020	71	100	Pediatric IRD	Yes	Yes	\
UAE	Patel[93]	2018	75	82	IRD	Yes	Yes	\
UK	Jiman[94]	2020	106	49	Syndromic IRD	Yes	\	\
UK	Shah[95]	2020	655	43	IRD	Yes	\	\
UK	Carss[52]	2017	722	56	IRD	\	Yes	Yes
UK	Lenassi[96]	2020	201	64	Pediatric IRD	Yes	\	\
UK	Patel[58]	2019	277	25	Pediatric IRD	Yes	\	\
UK	Taylor[97]	2017	85	79	Pediatric IRD	Yes	\	\
USA and Canada	Goetz[98]	2020	5385	62	IRD	Yes	\	\
USA	Stone[59]	2017	1000	76	IRD	\	Yes	\
USA	Bryant[99]	2018	69	64	IRD	\	Yes	\
UK	Pontikos[100]	2020	3195	N.A.	IRD	Yes	Yes	Yes
Netherlands	Haer-Wigman[62]	2017	266	51	Inherited eye disorders	Yes	Yes	\
UAE	Méjécase[101]	2020	74	90	Inherited eye disorders	Yes	\	\
Portugal	Peter[102]	2023	230	76	IRD	Yes	Yes	Yes

1.5 The splicing process and its contribution to IRD causation

Splicing is an orchestral process (Figure 1.6), requiring a significant number of proteins interacting and operating in harmony to produce the mature mRNA transcript. Firstly, Small nuclear ribonucleoproteins bind to key splice site sequences as the splicing process begins. U1 binds to the splice donor site, SF1 binds to the branch point (A) sequence and U2AF binds to the splice acceptor site. SF1 is displaced by U2 which then interacts with U1 and the U4-U5-U6 complex. This facilitates release of U1 and U4. The spliceosome is now active and the 5' end of the intron is cleaved. This allows for the formation of the lariat loop structure and finally cleavage of the 3' end of the intron. The U2-U5-U6-lariat complex is released and the exons are joined[103].

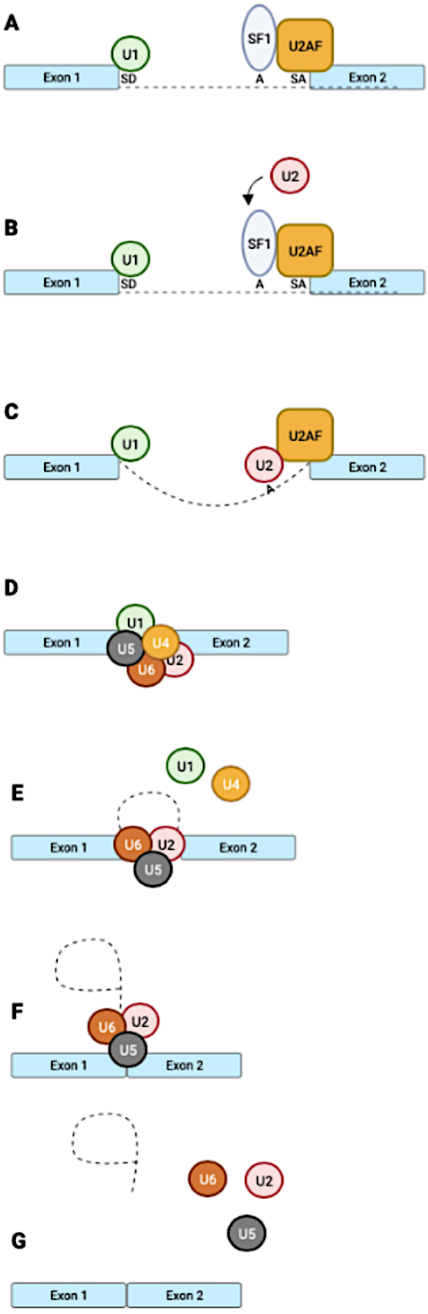


Figure 1.6 | Stages of the splicing process. A: Small nuclear ribonucleoproteins (snRNPs) bind to key splice site sequences as the splicing process begins. U1 binds to the splice donor (SD) site, SF1 binds to the branch point (A) sequence and U2AF binds to the splice acceptor (SA) site. B/C/D: SF1 is displaced by U2 which then interacts with U1 and the U4-U5-U6 complex. E: This facilitates release of U1 and U4. F: The spliceosome is now active and the 5' end of the intron is cleaved. This allows for the formation of the lariat loop structure and finally cleavage of the 3' end of the intron. G: The U2-U5-U6-lariat complex is released, and the exons are joined. Introns are represented by a dotted black line (Image created using BioRender[53]).

Specific sequences facilitate the accurate execution of this process, with GT and AG bases at the +1/+2 and -1/-2 canonical splice site (CSS) positions being highly conserved, alongside a branchpoint sequence located between 9-400 bases upstream of the acceptor site[104–106]. Furthermore, less conserved sequences constitute the non-canonical splice site (NCSS) regions located between the +3 and +6 positions and the -3 and -14 positions, as well as the first and endmost three bases of an exon[107]. Genetic variation in the key motifs outlined have the potential to disrupt the splicing process. Should this occur, it's possible that a truncated or aberrant protein product is created that may cause a form of inherited disease. These variants typically negatively affect canonical splice sites (CSS) and the NCSS regions or create/activate cryptic splice sites leading to exon skipping, truncation, elongation or pseudoexon inclusion. Diagrams highlighting the complexity of splice-altering variants can be found in Figure 1.7. It should also be noted that the alterations illustrated in Figure 1.7 do not only occur in isolation, but that multiple mutant transcripts may be produced by a single variant, resulting in a combination of splice defects. In recent years, a multitude of studies have demonstrated the contribution of pathogenic splice-altering variants to the causation of IRD[51,56,65,108–115]. Therefore, it became imperative that this source of disease manifestation be investigated in the Irish care pathway for IRD patients – Target 5000.

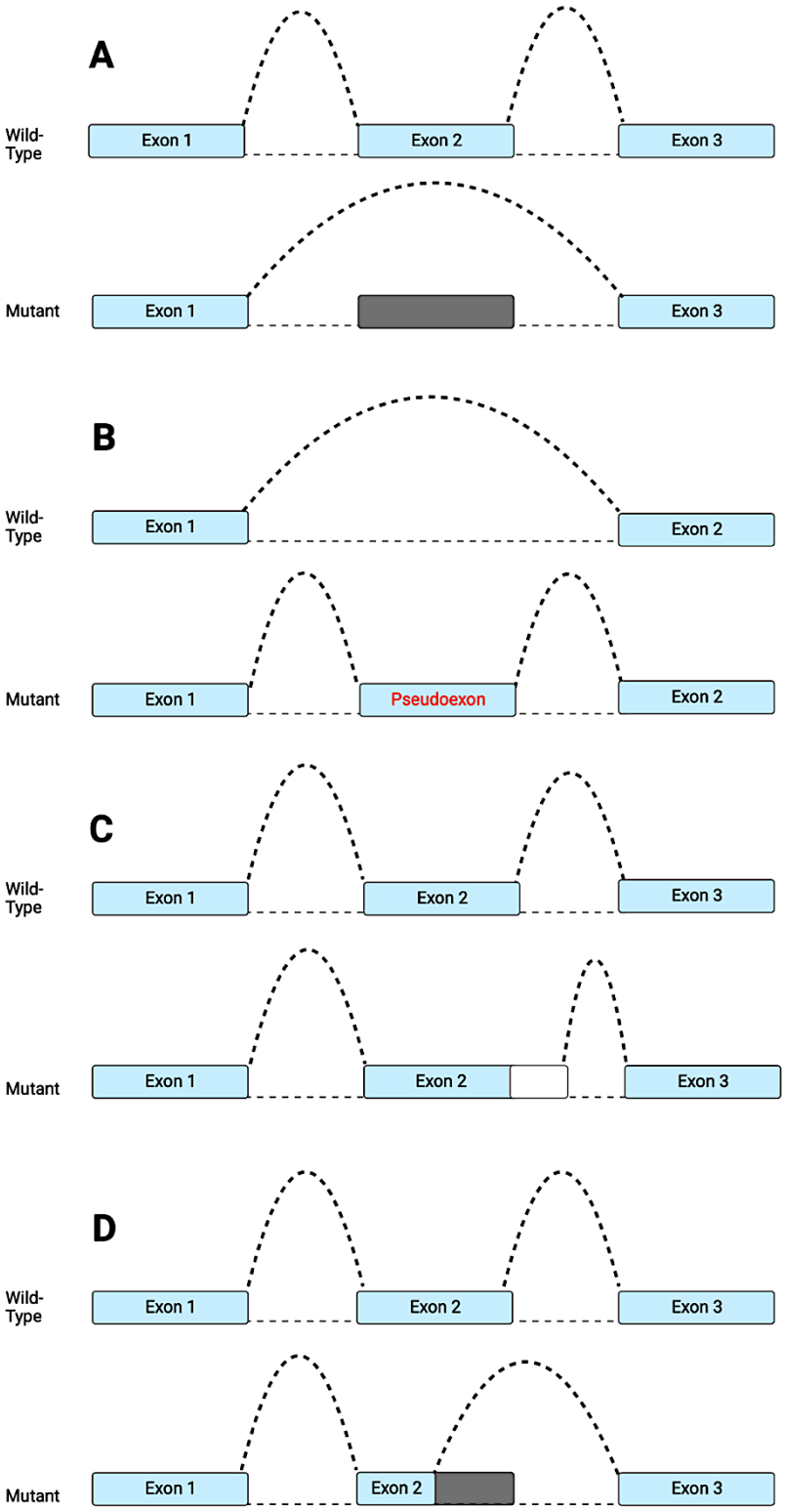


Figure 1.7 | Diagrams illustrating defects that may occur due to splice altering variants. A represents exon skipping, a defect that commonly occurs due to canonical and NCSS variants. B represents pseduoexon inclusion, most commonly a result of deep-intronic variation. C represents exon elongation, which can occur because of canonical and NCSS variants. D represents exon truncation, which can result from canonical or NCSS variants as well as through the creation/activation of cryptic splice donor or splice acceptor sites within the exon (Image created using BioRender[53]).

1.6 Target 5000

IRDs are a leading cause of visual dysfunction among those of working age, with huge implications on both quality of life and health economics[14]. Target 5000 is the national registry and care pathway for those affected by an IRD in Ireland. It is both a research and clinical study that endeavours to provide an accurate genetic diagnosis for those affected with a future goal of trial/treatment provision. Target 5000 was set in motion in 2009, following a meeting of Irish IRD specialists, scientists, and patient groups with a primary goal to develop a national register and individuals care plans where necessary. A first study was funded by the Irish Health Research Board. As described by Stephenson *et al.,* a survey was conducted in 2014 across 12 clinical ophthalmic sites in Ireland in order to assess infrastructure present at the time for IRD assessment and to identify any areas of concern. Results confirmed the absence of a standardised approach to assessment. An associated multidisciplinary study outlining work performed since then through a standardised all-Ireland pathway was published in 2021[116].

Target 5000 utilises target capture NGS, a time- and cost-effective screening method as its primary genetic testing procedure. Exonic regions of over 250 IRD-associated genes as well as a number of intronic regions harbouring previously identified pathogenic variants are captured, with a current diagnostic yield of ~70%[13], increased from previous reports[117,118]. This is described in detail in section 3. This doctoral thesis will demonstrate the integration of research grade WGS and whole gene sequencing in Target 5000 on a case dependent basis. Actively expanding the study's diagnostic reach is imperative in ensuring that every individual with an IRD has their underlying genetic cause investigated to the best of our current ability.

It will be illustrated herein how our study significantly highlights the benefits of target capture NGS as a first tier genetic test, before employing more comprehensive sequencing methods such as whole gene sequencing or WGS where needed. This doctoral thesis demonstrates the efficacy and advantages of this tiered diagnostic approach through the lens of Target 5000, serving as a template to other inherited rare diseases in Ireland and to other small countries embarking upon a diagnostic odyssey for individuals with IRD.

1.7 Aims and objectives of the doctoral thesis.

The overarching aim of this doctoral thesis was to shed light on genetically unexplained IRDs in Ireland. This was carried out using a variety of NGS methods and can be broken down into key objectives:

1. Establish the current diagnostic yield of panel sequencing as a first tier genetic test for IRDs in Ireland (Section 3).

2. Use WGS as a second tier genetic test in 100 unresolved cases of Irish, Dutch, and Israeli descent (Section 4).

3. Employ whole gene sequencing as a second tier genetic test and establish its efficacy as a first tier genetic test for *ABCA4* related disease in Ireland (Section 5).

4. Retrospectively apply the knowledge gained from objectives 2 and 3 to identify putative splice altering variants using panel sequencing data from those who remain unresolved following objective 1 (Section 6)

2. Materials and Methods

The methods outlined below are those that are employed in every study described within this thesis. Specific, separate methods for each study are detailed in full in their respective chapters.

2.1 Ethics

Prior to commencement, The Research and Medical Ethics committee of the Royal Victoria Eye and Ear Hospital (Dublin, Ireland), the Institutional Review Board of the Mater Misericordiae University Hospital (Dublin, Ireland) and Mater Private Hospital (Dublin, Ireland), The Belfast Health and Social Care Trust Hospitals (Belfast, Northern Ireland), the local ethics committee of Trinity College Dublin (Dublin, Ireland) and the local ethics committee of Radboud University Medical Center (Nijmegen, The Netherlands) awarded ethical approval for the studies detailed in this doctoral thesis. Written informed consent in accordance with the Declaration of Helsinki was given by all participants before commencement of these studies.

2.2 Patient sites and clinical assessment

Probands and other family members were recruited nationwide and primarily assessed at the Research Foundation of the Royal Victoria Eye and Ear Hospital (Dublin, Ireland), the Mater Misericordiae University Hospital (Dublin, Ireland), and the Belfast Health and Social Care Trust (Belfast, Northern Ireland, UK). This allows for the inclusion of participants from both The Republic of Ireland and Northern Ireland. informed consent, the applicable combination of tests were carried out to accurately assess the patients' phenotype (Table 2.1)

Table 2.1 | Table of typical tests, instruments, and manufacturers routinely used in clinical assessment as part of Target 5000.

Test	Method of testing	Manufacturer
Best corrected visual acuity	Revised 2000 early treatment diabetic retinopathy study charts	Precision Vision, La Salle, IL, USA
Color vision	Lanthony desaturated D-15 panel under standardized lighting conditions	Gulden Ophthalmics, Elkins Park, PA, USA
Peripheral visual fields	Goldmann perimeter 940 (uv4e,i4e and 04e targets)	Haag-Streit AG, Köniz, Switzerland
Full field electroretinogram	Roland Consult RET0-port retiscan	Brandenburg an der Havel, Germany
Fundus colour and autofluorescence photography	Topcon CRC50DX/Optos Daytona	Topcon Great Britain Ltd, Berkshire, England/Optos plc, Dunfermline, Scotland
Spectral domain optical coherence tomography	Cirrus HD-OCT	Carl Zeiss Meditec, Berlin, Germany

2.3 Sample acquisition and extraction

Following the above visual tests, if it was determined that the individual likely was affected by an IRD they were invited to participate in the Target 5000 study. Once informed consent was obtained a DNA sample was requested. DNA was provided by means of either a blood or saliva sample, depending on the patient's preference. 5ml of blood or 3ml of saliva was required to participate in the study. BD Vacutainer blood collection tubes were used to store blood samples at -20°C. Originally, each DNA isolation required 2ml of blood and was carried out using the Qiagen (Hilden, Germany) QIAamp DNA Blood Midi Kit. In addition to the manufacturer's protocol, samples were thoroughly vortexed before the lysis incubation step. 10 minutes into the lysis incubation at 70 °C samples were vortexed again and incubated for a further 5 minutes to increase yield of DNA. Saliva was collected and stored in DNA Genotek (Ottawa, Canada) ORAcollect DNA tubes. DNA was isolated using the DNA Genotek (Ottawa, Canada) prepIT-L2P kit or the Qiagen (Hilden, Germany) QIAamp DNA Blood Midi Kit with similar yields and DNA quality using both kits.

3. Target capture NGS of the Target 5000 cohort

3.1 Introduction

As described previously, Target 5000 aims to provide genetic testing and genetic diagnoses for individuals in Ireland who are affected by an IRD. Preceding this study, patients were diagnosed based on clinical criteria alone. This study is a collaborative effort between a multi-disciplinary team of academics, scientists, postgraduate researchers, clinicians and patient organisations to refine overall diagnoses through ascertaining a genetic diagnosis for each patient. This section presents the most updated findings from first-tier testing of over 1000 individuals across 710 pedigrees tested using TCS to date. 458 individuals were sequenced for variants in 210 genes, 689 individuals were sequenced for variants in 254 genes in addition to confirmatory sequencing of other affected and unaffected family members. TCS identified positive candidate variants in almost 70% (495/710) of pedigrees, including both previously reported pathogenic and novel potentially pathogenic variants. This includes novel SVs, single nucleotide variants (SNVs), and splice variants. Furthermore, 9% (64/710) of pedigrees were found to carry a pathogenic/likely pathogenic variant on one allele where the genotype and phenotype were associated with a recessive inheritance pattern. It is possible that these variants are indeed candidate disease causing variants but it is also plausible that these variants are carried by chance based on estimated carrier frequency of IRD variants globally (~2.7 billion individuals)[119].

3.2 Methods

3.2.1 DNA isolation

DNA was isolated from 2ml of blood or saliva as described in section 2.3. In order to perform targeted sequencing, fragments 200-250 base pairs in length were generated from 8µg of DNA via sonification using Diagenode Bioruptor Plus Microtubes (Denville, United Sates).

3.2.2 Target capture sequencing and variant confirmation

Sequencing libraries were created and the coding exonic regions of 254 IRD associated genes were captured using the Nimblegen SeqCap EZ kit (Roche Ireland Ltd., Dublin Ireland) as described by Dockery *et al,*[118]. This was performed in accordance with the manufacturers instructions. A small number of intronic regions harbouring known pathogenic non-coding variants were also captured in *CEP290, ABCA4* and *USH2A*.

Earlier studies involved 24-sample pool multiplexing using NimbleGen adapters on captured DNA (Roche Ireland Ltd., Dublin Ireland). Sequencing was performed on an Illumina MiSeq (Illumina Inc., San Diego, CA, USA). In later runs, 96 patient samples were multiplexed using dual-indexed adaptors from Integrated DNA Technologies (Coralville, Iowa, USA). To validate variants identified by NGS, relevant genomic loci containing the mutations were amplified by polymerase chain reaction (PCR) and investigated by direct sequencing. Primers were procured from Sigma-Aldrich (Gillingham, England, UK). DNA products were standardly amplified using Q5 High-Fidelity 2× Master Mix (New England Biolabs Inc., Ipswich, MA, USA). Samples from additional family members for segregation analyses were amplified similarly or alternatively, directly from blood (Phusion Blood Direct PCR Kit, Thermo Scientific, MA, USA) or saliva (Phusion Human Specimen Direct PCR Kit, Thermo Scientific) where applicable. The annealing temperatures for reactions were optimised for each variant; all other details were executed as per the supplier's recommendations. Sanger sequencing was performed by Eurofins Genomics (Ebersberg, Germany).

3.2.3 Data analysis and variant interpretation of target capture NGS data

Raw sequencing data was processed as previously described[120]. For completeness, NGS data was demultiplexed and mapped to the human genome (GRCh38) using BWA version 0.7.15[121]. Duplicate reads were flagged using Picard version 2.5.0[122] and downstream analysis and variant calling performed using Freebayes version 1.1.0[123]. Coding variants were annotated using SnpEff[124], dbNSFP[125], MetaLR[126] and M-CAP[127]. Additionally, structural variants were called by LUMPY[128] and Manta[129], which were used to identify structural variants via split reads or unusually paired-reads, and CoNIFER[130], which was used to identify structural variants based on read depth. The American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology (AMP) criteria for classifying pathogenic variants was utilised for interpretation[131]. Other proposed modifications to the ACMG-AMP guidelines to further quantitate certain lines of evidence were also employed during the study. These modifications affect three lines of evidence that can be applied. These adjustments consider using segregation data in a quantitative manner (code: PP1)[132], quantifying evidence that considers the rare incidence of variants (code: PS4) and the value of accurately phenotyping conditions with single gene aetiologies (code: PP4)[133]. To further implement these recommendations, REVEL and dbscSNV RF scores (a scale of 0-1 representing likelihood that splicing will be altered) were added to the existing suite of ensemble predictors to more stringently detect

agreement between predictors of pathogenicity[134,135]. Similarly, Manta was added to the SV pipeline to assist in the detection of SVs[129].

3.2.4 *CHM* deletion confirmation

As part of this study, large structural variants were idenitified in the *CHM* gene. Primers were designed to probe for exonic regions of *CHM* and neighbouring genes, *POF1B* and *DACH2*, as well as the intergenic space between these genes. A full list of the primers utilised can be found in Table 3.1.

Table 3.1. | Primers used to probe the intergenic space between *CHM, POF1B* and *DACH2.* Names are indicative of the genomic location, for example, 1K *POF1B* is 1kb in the direction of *POF1B* from *CHM.*

Primer Name	Forward Primer	Reverse Primer
1K POF1B	TGCTCGGCTAAGTTGGTACA	TTTGGTCTATCGTTGAAGGGG
5K POF1B	TCTTTCTTCCTCTGGCCTGG	ACCCTCCACTATTTCTGAAAAGG
10K POF1B	ACCTGCTTCACCAATTTTGCT	AACCAGCGTATTCCACAGAG
30K POF1B	CCAGGCCCCGTATAGCATAA	ACAAGCTATGAGAACAAACCCA
35K POF1B	GTGGGTTTCATATCAGGGATGC	TGGTGAGAGTGGGCATCTTT
39K POF1B	AGACAACCACCCGTACCAG	TGGTTTGGATCCATTGCTTGT
40K POF1B	TCTGCCCCTGAACATACACC	CACACCGGATTCATGCCTTC
41K POF1B	GCTCCTAAACTCCTCGTGCA	TCCATTGGTCTATGTGCCTGT
45K POF1B	TGGGGATCTGTCTTTTGCTT	ACCACCTATTTGCAGTTTGACA
50K POF1B	TGTTTCAGCTCCCCAATGTTC	ATGGATGAATGGGAGAAGCC
60K POF1B	AGACTCTGTTTCTTGGCGGT	GGTCCTGCCTTCCCTACTAG
70K POF1B	ATGCAGATTGAAGGTGGACC	AACGAGTTAATGGGTGCAGC
79.5K POF1B	GCAGAAATACCATTTGACTCAGC	ATGCGGTGTTTGGTTGTCTG
80.5K POF1B	TCAGAGATCACTAGCCCTGC	CTGTTCATGTCCTTTGCCCA
80K POF1B	ATGTCCCCACCGAAATCTCA	GTGAGGTGGTGTATGAGTCCA
90K POF1B	TCTGCCAAGAGATCAGCTGT	TGGGACTATGTGAAGAGAACAAA
100K POF1B	TCCACACTCGACTTTGTCCA	CCTCAAGCCTAAGTCTCCCA
110K POF1B	TTGTTGGCTGCATGAATGTCT	AGGGATGGGAAAAGACTTCATG
120K POF1B	TCAATTTCTTCCATGTCACCCA	AGGCCAGGAATTCAAGACCA
133K POF1B	TATCTGGGTGCTGTGGTGTT	ATGGCAGGCATAATCCCTCA
140K POF1B	TTGTGGTGTCTTTGTCTGGC	CCCTCTCCCTCTCCCTCTC
150K POF1B	AAGCCAGGCAGAGAAAGACA	GTATCCATCCCCTCAAGGCT
200K POF1B	AGACCAGGCACAGAAAGACA	TCAGGTTGGGTGAGCTATCC
20K POF1B	CACAGGTTGAGGCCCTTCTA	GCATGGGAGGTCTAATCAAAGG
250K POF1B	TCACATAGCATGACTGGAGC	GCCCTGTACAAGCCAATTCT
300K POF1B	CAGTGTGTTGAGCTGTGGAG	ACTAGCGATCCTATGCCACC
1K DACH2	ATGAGGGGCTTTAGGTCGTC	CCAGCATCCTTCCCCTTTTG
10K DACH2	CAAAGGCAGGGCATGTAGAC	GTTGAGCTCCTGAGTATGTGC
20K DACH2	TCAGGGAGAGCTGTAGGAGA	AACTGTTCCTAGAGCCCAGC

30K DACH2	GAGAGGAAGGTTGGGCTGAA	CCACCGCTGCTCCTTTTAG
40K DACH2	TCTCCCACTAGATCCCTCCC	TGCACTGACACAAAGCTTTCA
40K DACH2	CTGATTCCAACAGCGTATCCA	TCTCACAGGGGATGCTTTCA
45K DACH2	AGGCACATAGACCAATGGAAC	TGGTCATGAACTCTCTGCCT
50K DACH2	TTCATTCCATAGCACATAGGAGT	ACACTAAGAAGCCCCAGTTCA
50K DACH2	TCTTGGTGACTTTTATTGCCAGA	GAGGGCTGTTGTTTGCTTGA
55K DACH2	AGAGACAGTGGATTGGTGGG	AATGCTTTCAGGACGCTGTC
5K DACH2	CAGCCATGCCAAAAGGAGAG	TGGCTGCTCTACCCTACTTT
60K DACH2	TTCCCTTCTGCATTGCCCTA	GCGTAGAGGCTTAGGAAGGA
62K DACH2	ACAGACACATAGGCCAATGGA	TTTCTGCTTGGTTGCCTGTG
64K DACH2	TGTCACACAAATTCTTCCAGAGG	GAGCACTCATGAAACTAACGCT

3.2.5 Sequencing of *RPGR* ORF15

A previously reported sequencing strategy of *RPGR* was employed in order to cover the highly repetitive sequences of ORF15, a major cause of X-linked RP[136]. High-fidelity PCR was performed with Hotstart Takara LA Taq DNA polymerase using 50 ng of genomic DNA in a final volume of 25 µL. The PCR program used was 95°C for 15 minutes, 35 cycles of 95°C for 45 seconds, and 72°C for 3.5 minutes, followed by 72°C for 15 minutes, with a final hold at 4°C.

3.2.6 Single-molecule molecular inversion probe-based sequencing of *ABCA4*

Single-molecule molecular inversion probe (smMIP) based whole gene sequencing of *ABCA4* was carried out by Dutch colleagues at RadboudUMC using a NextSeq 500 as part of a large-scale study by Khan *et al*.,[23] which is outlined in detail in section 5 of this thesis (Figure 5.1). 3866 smMIPs were designed as described previously and employed to capture 110 nucleotide increments of both the sense and antisense strands of the *ABCA4* gene, as well as 40 kb of flanking sequences[137]. Sample processing, sequencing, data analysis, and variant interpretation was carried out as described in section 4.

3.3 Results

3.3.1 Clinical presentation and positive candidate detection rates

Thus far, 1004 pedigrees have been clinically recruited and 710 pedigrees have been genetically analysed as part of this study by TCS. To date, 458 individuals have been sequenced for variants in 210 genes, 689 individuals have been sequenced for variants in 254 genes, and confirmation sequencing is ongoing for other affected family members. A spectrum of IRDs were observed in this study, as can be seen in Figure 3.1.

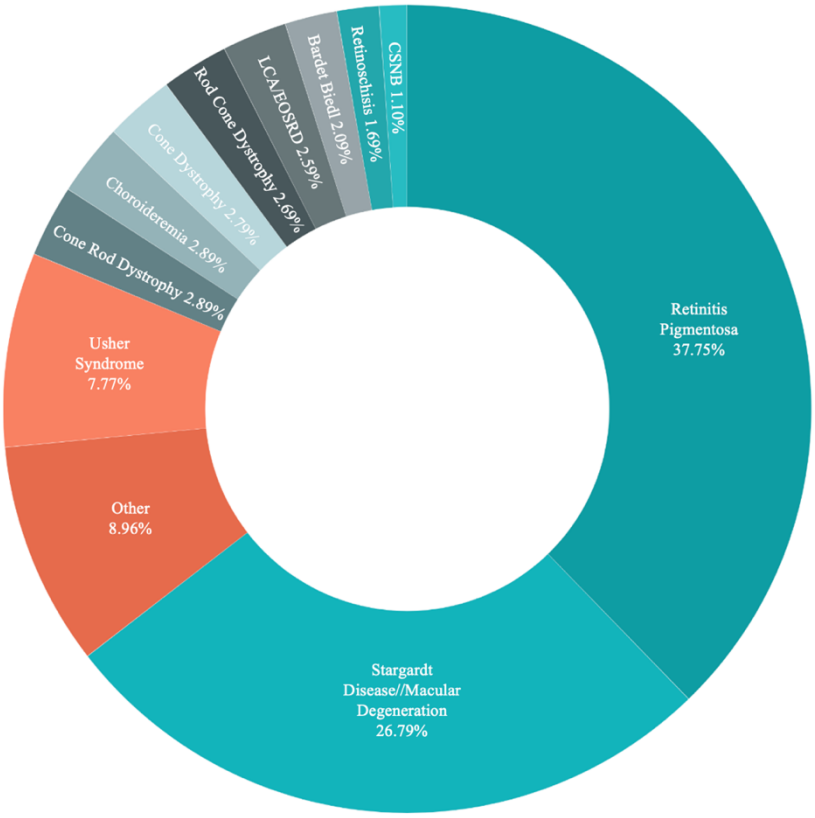


Figure 3.1 | Clinical presentation of 1004 recruited Target 5000 pedigrees. LCA: Leber Congenital Amaurosis, EOSRD: Early-Onset Retinal Degeneration, CSNB: Congenital Stationary Night Blindness.

The positive candidate detection rate has improved from previous reports[117,120], now reaching approximately 70% (495/710). Furthermore, a single candidate variant only has been identified in an additional 9% (64/710) of pedigrees diagnosed with an IRD that typically exhibits a recessive inheritance pattern (Figure 3.2). In addition, a number of novel missense variants outlined in section 3.3.6 have been detected during the course of this study.

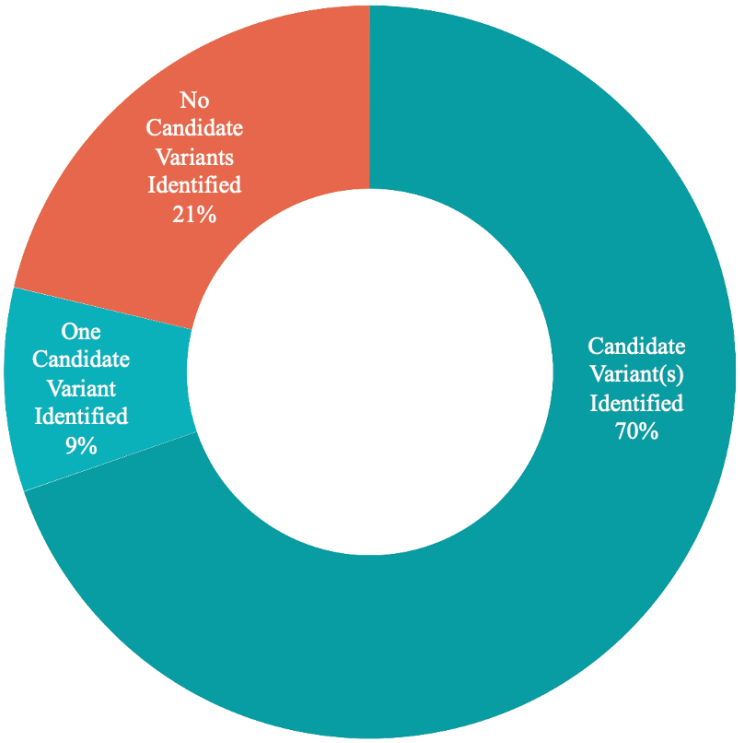


Figure 3.2 | Diagnostic yield rates for 710 Target 5000 pedigrees utilizing target capture NGS of the exonic regions of over 250 genes and previously identified pathogenic intronic variants. Positive candidate variants were identified in 69.72% of cases. One candidate was identified in a further 9.01% of cases where the inheritance pattern was thought to be autosomal recessive. No candidate variants were detected in 21.27% of cases.

3.3.2 Retinitis pigmentosa

The most prevalent clinical presentation was RP (MIM: #268000), accounting for 37.75% (379/1004) of all recruited pedigrees (Table 3.2). These figures incorporate various phenotypic presentations, including atypical, inverse, and paravenous RP. The total number of pedigrees analysed and clinically reported as having a family history of RP was 184. Of these, 131 have been genetically resolved. The total number of pedigrees analysed and clinically reported as having simplex RP was 79. 48 of these pedigrees have been genetically resolved. This gives a genetic diagnosis rate of 71% for RP with a suspected inheritance pattern and 61% for simplex RP. Of the resolved RP cases, 81 carried a variant/variants in genes associated with an autosomal dominant inheritance pattern, 71 with an autosomal recessive inheritance pattern, and 27 with an X-linked inheritance pattern (Table 3.2). This figure on dominant inheritance is larger than reported in other studies[138]. Variants in *RHO* were the most common candidate variants for RP in this cohort, accounting for over 14% (26/178) of all sequenced RP pedigrees, and over 32% (26/81) of dominant pedigrees alone. It is important to note that there is likely an even higher prevalence of *RHO*-linked RP in the Irish IRD cohort, as many pedigrees involved in single-gene studies were excluded from this study due to the discovery of a

causative variant previously[117,120]. Of these variants, namely c.533A>G (p.Tyr178Cys), c.541G>A (p.Glu181Lys), and c.620T>G (p.Met207Arg) collectively account for candidate *RHO* mutations across 13 different pedigrees to date. Novel variants identified in RP-associated genes can be found in section 3.3.6.

The most commonly observed candidate gene for autosomal recessive RP was *USH2A.* Two variants in this gene were found in over 21% (15/71) of pedigrees with this autosomal recessive RP phenotype (Table 3.2). As variants in *USH2A* are also the primary cause of type II Usher Syndrome (MIM: #276901), it is important to note that participants included in this section presented with recessive RP without syndromic disease manifestation. Non-syndromic *USH2A*-linked RP has been consistently reported in other IRD cohorts[139]. A proportion of clinically diagnosed recessive RP pedigrees (>9%) (7/71) carry variants in *ABCA4*, typically associated with STGD1 (MIM: #248200), a form of macular dystrophy. In the Target 5000 patient cohort, it has become apparent that an atypical RP-like phenotype appears to be caused by some variants in *ABCA4*, as previously described in other cohorts (Figures 3.3 and 3.4)[20,140].

Pathogenic variants in *RPGR* were the most frequent cause of X-Linked RP, explaining over 88% (23/26) of sequenced pedigrees (Table 3.2). Employment of a bespoke sequencing strategy of ORF15 has enhanced the success of sequencing this low complexity region, which typically presents a diagnostic challenge due to low coverage and poor variant detection[136].

Table 3.2 | A summary of the genetic architecture of retinitis pigmentosa (RP) in Target 5000 participants. Transcript IDs are as follows: *RPGR NM_000328.2, RP2 NM_006915.2, PRPF31 NM_015629.3, PRPH2 NM_000322.4, RHO NM_000539.3, RP1 NM_006269.1, NR2E3 NM_014249.3, RPE65 NM_000329.2, CRX NM_000554.4, PRPF8 NM_006445.3, SNRNP200 NM_014014.4, KLHL7 NM_001031710.2, USH2A NM_206933.2, KIZ NM_018474.4, NR2E3 NM_014249.3, EYS NM_001292009.1, IFT140 NM_014714.3, RP1 NM_006269.1, CERKL NM_001030311.2, IFT140 NM_014714.3, PROM1 NM_006017.2, PDE6B NM_000283.3, ABCA4 NM_000350.2, CNGB1 NM_001297.4, FLVCR1 NM_014053.3, MFRP NM_031433.3.*

Sub-Category	No. of Pedigrees	% of all RP cases	Gene	No. of Pedigrees	% of cases within sub-category
RP (X-linked)	27	14.61	*RPGR*	24	88.88
			RP2	3	11.54
RP (Dominant)	81	45.51	*PRPF31*	6	7.41
			PRPH2	6	7.41
			RHO	26	32.10
			RP1	15	18.52
			NR2E3	3	3.70
			RPE65	6	7.41
			CRX	2	2.47
			PRPF8	3	3.70
			SNRNP200	2	2.47
			KLHL7	2	2.47
			Other: One Observation	10	12.35
RP (Recessive)	71	39.89	*USH2A*	15	21.13
			KIZ	2	2.82
			NR2E3	4	5.63
			EYS	5	7.04
			IFT140	2	2.82
			RP1	3	4.23
			CERKL	3	4.23
			IFT140	2	2.82
			PROM1	2	2.82
			PDE6B	4	5.63
			ABCA4	7	9.86
			CNGB1	2	2.82
			FLVCR1	4	5.63
			MFRP	3	4.23
			Other: One Observation	13	18.31
Total	179				

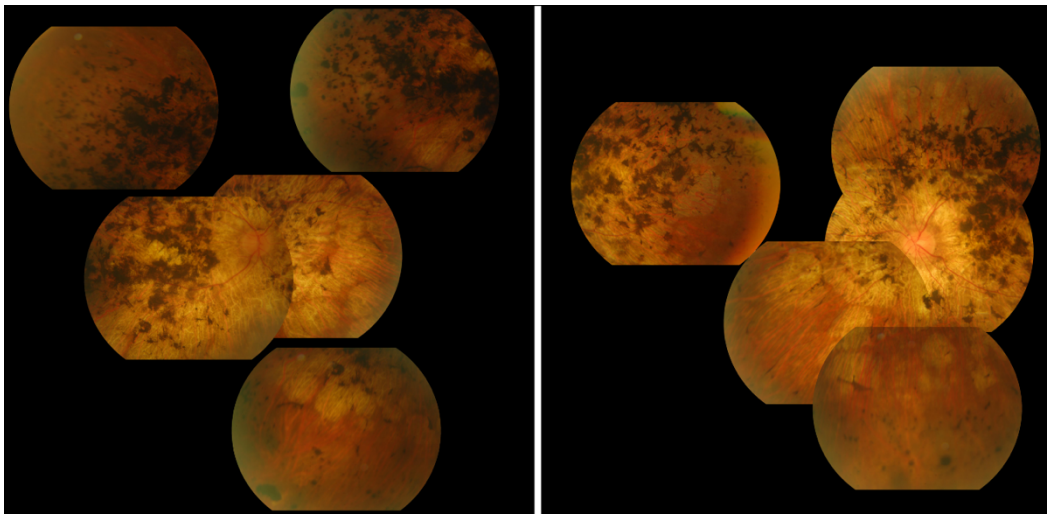


Figure 3.3 | Collage of fundus images of a patient in the Target 5000 cohort initially presenting with a Stargardt phenotype that progressed to phenotype suggestive of RP. Extensive pigmentation can be seen at both maculae and in the periphery of both eyes (left—right eye, right—left eye). There are also distinct punched-out areas in the pigment epithelium bilaterally.

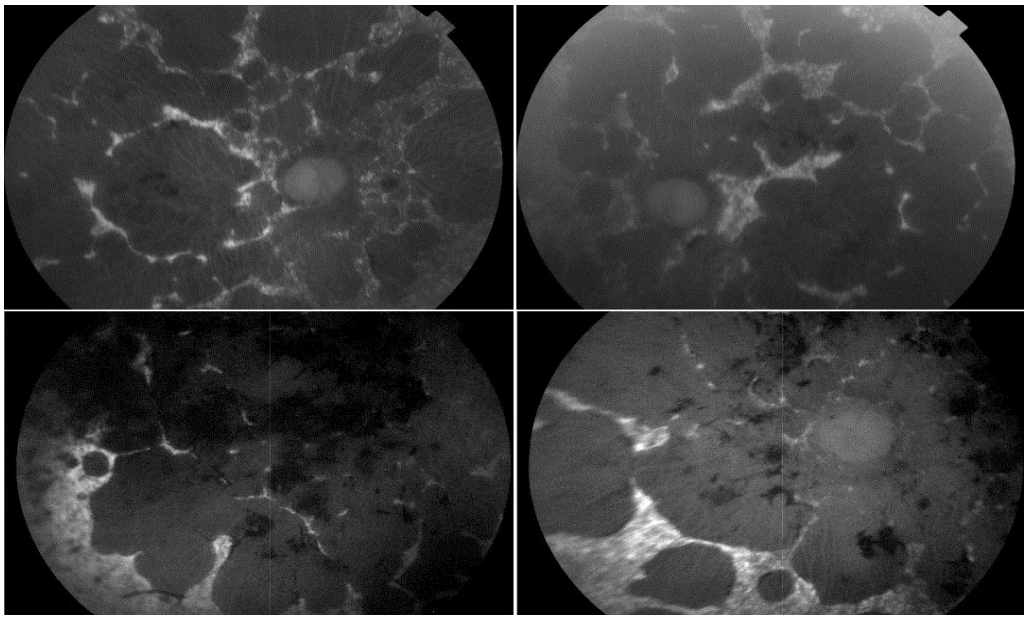


Figure 3.4 | Collage of autofluorescence fundus images of the patients initially presenting with a Stargardt phenotype that progressed to phenotype suggestive of RP. The right eyes are shown in the left column; the left eyes are shown in the right column. Patient example 1 is the first row and patient example 2 is the second row. All eyes demonstrate a dark choroid and large nummular region of hypoautofluorescence throughout the retinae.

3.3.3 Stargardt disease and other macular dystrophies

STGD1 is the second largest phenotypic presentation in this study (Figure 3.1). As mentioned previously, variants in *ABCA4*[141] were the most common cause, accounting for over 97% (109/112) of resolved STGD1 pedigrees in this cohort (Table 3.3) Six potentially novel coding variants in *ABCA4* were detected as part of this study, NM_000350.2: c.1865G>A (p.Ser622Asn), c.223T>C (p.Cys75Arg), c.4468T>C (p.Cys1490Arg), c.5329A>G (p.Met1777Val), c.5351T>C (p.Leu1784Pro), and c.6743T>C (Phe2248Ser). All of these variants, with the exception of c.5351T>C (p.Leu1784Pro), were observed in individual pedigrees where another known pathogenic *ABCA4* variant was also observed. Further details of all novel variants can be found in section 3.3.6. Variant *ABCA4* c.5351T>C (p.Leu1784Pro) was observed in three individuals of two not knowingly related families also carrying a known pathogenic variant. The probands of these pedigrees do not appear to share any other rare variants in common in the regions captured. This may be additional evidence suggestive of this novel variant's pathogenicity.

It is becoming increasingly clear that intronic variants play a major role in the causation of STGD1. Nine pathogenic intronic variants were included in the capture panel of this study and found in 26 cases in 20 pedigrees with another *ABCA4* variant. In recent years, single-molecule molecular inversion probe (smMIP) based sequencing of the whole *ABCA4* gene including intronic sequences has dramatically increased positive candidate detection rates for STGD1[142,143]. Thirty-six unresolved probands from Target 5000 presenting as STGD1 or CRD included in this cohort were analysed as part of a landmark study on the genetic landscape of *ABCA4* in over 1000 probands, with a positive candidate detection rate of 44% (16/36) [23]. This has increased the rate at which *ABCA4* is identified as the causal gene of STGD1 in the Target 5000 cohort and prompted sequencing of additional samples as outlined in section 5. Of note from the 36 Irish participants in this study, 5 individuals were found to carry c.4539+2028C>T (p.[=,Arg1514Leufs*36]) *in trans* with another pathogenic *ABCA4* variant. Upon retrospective analysis, 8 further incidences of this variant were detected in previously unresolved individuals, totalling 14 participants recruited by Target 5000 carrying this variant and another pathogenic *ABCA4* variant at the time of this study. To date, we have identified 25 individuals who harbour this variant including two homozygotes. This is discussed in detail in section 6.1. This suggests significant enrichment for the c.4539+2028C>T variant in the STGD1 population recruited in Ireland.

In agreement with previous reports from Target 5000 using a significantly smaller patient cohort[117,120], *ABCA4* c.5603A>T (p.Asn1868Ile) is a significant causal hypomorphic variant of a milder form of STGD1 in this expanded IRD cohort. Of 109 STGD1 pedigrees with 2 positive candidate variants in *ABCA4* identified, this variant was observed 14 times. *ELOVL4* and *PROM1* were determined to be the cause of clinically diagnosed STGD1 in 2.7% (3/112) of sequenced pedigrees. Other maculopathies that presented prominently include Best disease accounting for 9.4% (15/159) of macular dystrophies in this study, CRD, and a general macular dystrophy phenotype, each accounting for almost 7% (11/159) of all maculopathy pedigrees (Table 3.3). However, it is clear from Table 3.3 that STGD1 is the predominant cause of macular dystrophy in Target 5000 participants.

Table 3.3 | A summary of the genetic architecture of macular dystrophies in Target 5000 participants. CRD = cone-rod dystrophy. Transcript IDs are as follows: *ABCA4 NM_000350.2, PROM1 NM_006017.2, ELOVL4 NM_022726.3, BEST1 NM_001139443.1, RDH5 NM_001199771.1, PRPH2 NM_000322.4, GUCY2D NM_000180.3, CNGB3 NM_019098.4, CNGA3 NM_001298.2, GUCA1A NM_000409.4.*

Sub-Category	No. of Pedigrees	% of all macular disease cases	Gene	No. of Pedigrees	% of cases within sub-category
Stargardt Disease	112	70.44	*ABCA4*	109	97.32
			PROM1	2	1.79
			ELOVL4	1	0.89
Best Disease	15	9.43	*BEST1*	15	100.00
Fundus Albipunctatus	3	1.89	*RDH5*	3	100.00
Macular Dystrophy	11	6.92	*PROM1*	3	27.27
			PRPH2	3	27.27
			Other: One observation	5	45.45
CRD	11	6.92	*GUCY2D*	2	18.18
			CNGB3	3	27.27
			Other: One observation	6	54.55
Cone Dystrophy	7	4.40	*CNGA3*	2	28.57
			GUCA1A	2.00	28.57
			Other: One observation	3	42.86
Total	159				

3.3.4 Usher syndrome

Usher syndrome is the most common manifestation of syndromic-IRD in this cohort, accounting for over 7% (78/1004) of all clinical presentations in Target 5000 (Table 3.4). Usher syndrome is a form of ciliopathy characterised by RP and sensorineural hearing loss. It is generally divided into sub-groups based on severity of symptoms, ranging from most severe in Type 1 (MIM: #276900) to least severe in Type 3 (MIM: #276902). In total 57 pedigrees were genetically diagnosed with Usher syndrome, where 2 candidate variants were detected, with Type 2 (MIM: #276901) being the most common presentation. 78 pedigrees were clinically diagnosed with Usher syndrome, yielding a positive candidate detection rate of 73% (57/78). Usher type 2 is by far the predominant sub-group in the genetically characterised Usher syndrome pedigrees in the Target 5000 cohort (41/57), most frequently caused by variants in *USH2A* (28/41) (Table 3.4). Variants in *MYO7A* were the most frequent cause of Usher type 1 (10/14). Variants in *CLRN1* and *MTTS2* (1/2) contributed equally to cases of Usher Type 3 in this study, with one individual carrying a novel candidate variant in *CLRN1* as detailed in section 3.3.6.

Table 3.4 | A summary of the genetic architecture of Usher Syndrome (USH) in Target 5000 participants. Transcript IDs are as follows: *MYO7A NM_000260.3, CDH23 NM_022124.5, USH1C NM_153676.3, USH2A NM_206933.2, ADGRV1 NM_032119.3, CLRN1 NM_052995.2.*

Sub-Category	No. of Pedigrees	% of all USH cases	Gene	No. of Pedigrees	% of cases within sub-category
USH I	14	24.5614	*MYO7A*	10	71.43
			CDH23	3	21.43
			USH1C	1	7.14
USH II	41	71.9298	*USH2A*	28	68.29
			MYO7A	3	7.32
			ADGRV1	9	21.95
			CDH23	1	2.44
USH III	2	2.50877	*CLRN1*	1	50.00
			MTTS2	1	50.00

3.3.5 Other IRDs encompassed by Target 5000

The second most frequent syndromic IRD clinically diagnosed in this study was BBS (MIM: #209900), typically characterised by polydactyly, intellectual disability, and obesity. Over 2% of pedigrees (21/1004) in the entire cohort present with BBS at the clinic (Figure 3.1), however, this is thought to be an under-representation, as some of those diagnosed with simplex RP are later found to have had polydactyly removed as children. These individuals are re-diagnosed with BBS following detection of two candidate variants in a BBS-associated gene and confirmation with Target 5000 affiliated ophthalmologists. In total, 14 pedigrees received a genetic diagnosis of BBS having been initially clinically diagnosed with RP. Variants in *BBS1* were the most frequent cause of this syndrome accounting for over 71% (23/32) of sequenced pedigrees. The most common pathogenic variant identified was c.1169G>T (p.Met390Arg)[120]. This variant has been detected 28 times homozygously in 22 different pedigrees in this cohort. The second most frequent cause of BBS were variants in *BBS10* (4/32), one of which is believed to be a novel variant (Table 3.6). BBS causality is followed by *SDCCAG8* (n = 2), *BBS9, BBS4*, and *TTC8* (n = 1 each) in the Target 5000 cohort (Table 3.5).

Many other rare forms of IRD are also included in the Target 5000 study. Sixteen pedigrees with a diagnosis of Leber congenital amaurosis (LCA) (Figure 3.5) have been genotyped with a spectrum of genes implicated as positive candidates (Table 3.5). Other rare IRDs examined in this study include retinoschisis (MIM: #312700) (n = 10), achromatopsia (Figure 3.6) (MIM: #262300) (n = 5), Stickler syndrome (MIM: #108300) (n = 4), Leber hereditary optic neuropathy (LHON) (n = 3) (MIM: #535000), and optic atrophy (MIM: #165500) (n = 2), among others. In total, 90 pedigrees with positive candidate variants are encompassed by Table 3.5, illustrating the wide spectrum of disease presentation included in Target 5000.

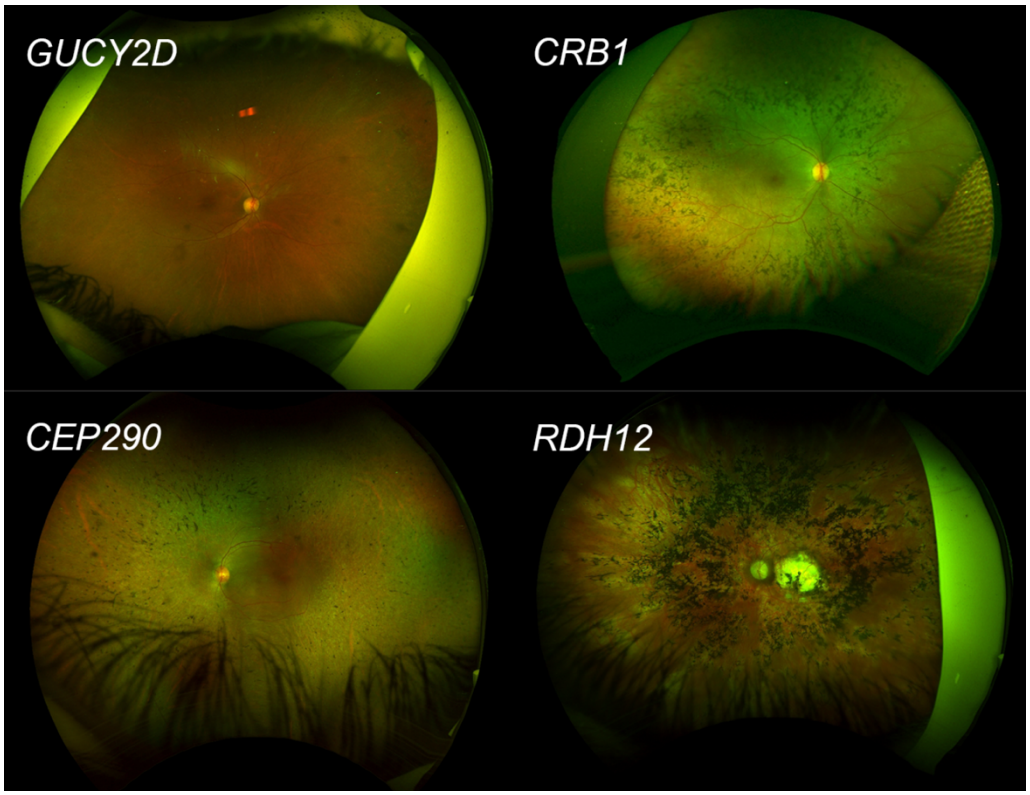


Figure 3.5 | The different fundal phenotypes of LCA/EOSRD. The fundus looks unremarkable in the patient with *GUCY2D* mutation. Nummular pigmentation and para-arteriolar sparing of the RPE are observed in *CRB1*-retinopathy. Bone-spicule hyperpigmentation and retinal atrophy in the mid-periphery are seen in the case with *CEP290* mutation. In autosomal recessive *RDH12* retinopathy, dense intra-retinal pigmentation and macular atrophy are seen (Photo from gene.vision[144]**).**

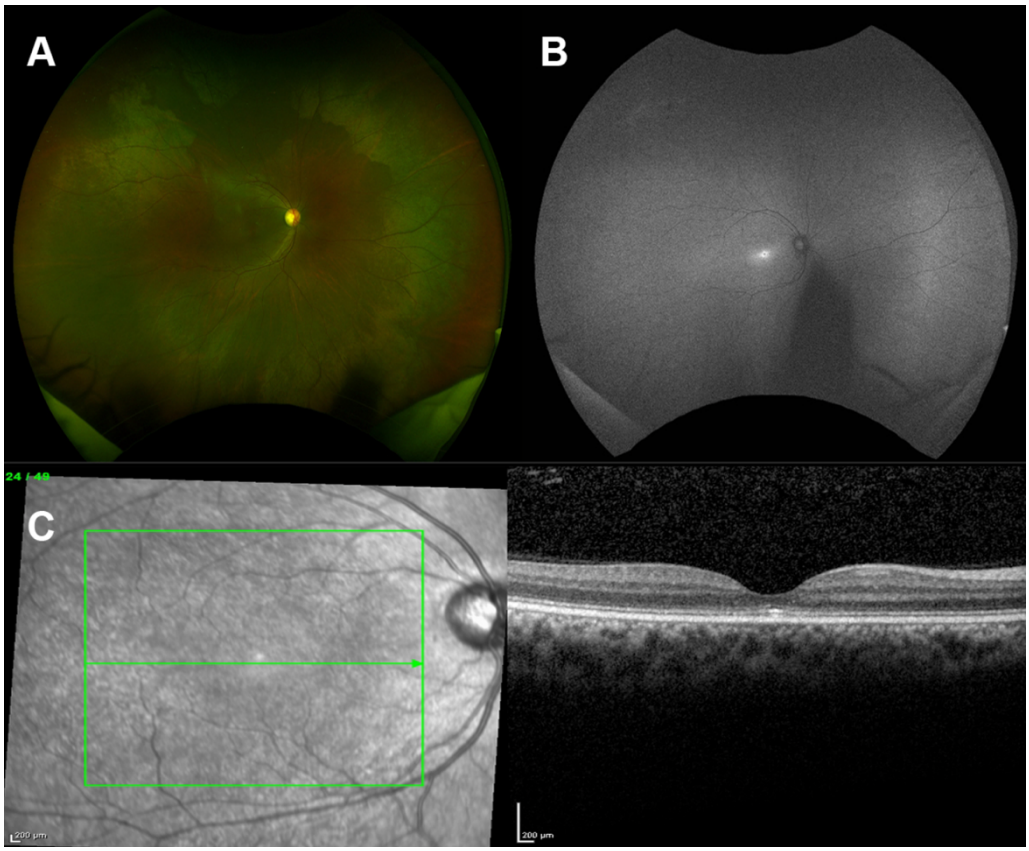


Figure 3.6 | Multimodal imaging of a patient with biallelic CNGA3 mutations. The wide field colour fundus photograph (A) is unremarkable. Wide field FAF imaging (B) shows hyperautofluorescence in the central macula. OCT scan through the macula (C) demonstrates some disturbance at the foveal ellipsoid zone. The outer retinal layers are relatively preserved throughout (Photo from gene.vision[144]**).**

Table 3.5 | A summary of the genetic architecture of less common phenotypes encountered in the Target 5000 study. LCA = Leber congenital amaurosis, LHON = Leber Hereditary Optic Neuropathy. Transcript IDs are as follows: *RS1 NM_000330.3, BBS10 NM_024685.3, SDCCAG8 NM_006642.3, BBS1 NM_024649.4, BBS9 NM_198428.2, BBS4 NM_033028.4, TTC8 NM_001288781.1, CNGB3 NM_019098.4, CNGA3 NM_001298.2, OPA1 NM_130837.2, COL2A1 NM_001844.4, COL11A1 NM_080629.2, CHM NM_000390.3, RPGRIP1 NM_020366.3, GUCY2D NM_000180.3, RDH12 NM_152443.2, CRB1 NM_201253.2, IQCB1 NM_001023570.3, TULP1 NM_003322.4, CRX NM_000554.4, LCA5 NM_001122769.2, RPE65 NM_000329.2, AIPL1 NM_014336.4, NYX NM_022567.2.*

Sub-Category	No. of Pedigrees	% of all fewer common phenotypes	Gene	No. of Pedigrees	% of cases within sub-category
Retinoschisis	10	11.11	*RS1*	10	100.00
Bardet-Biedl	32	35.56	*BBS10*	4	12.50
			SDCCAG8	2	6.25
			BBS1	23	71.88
			BBS9	1	3.13
			BBS4	1	3.13
			TTC8	1	3.13
Achromatopsia	5	5.56	*CNGB3*	3	60.00
			CNGA3	2	40.00
LHON	3	3.33	*ND4*	3	100.00
Optic Atrophy	2	2.22	*OPA1*	2	100.00
Stickler Syndrome	4	4.44	*COL2A1*	2	50.00
			COL11A1	2	50.00
Choroideremia	16	17.78	*CHM*	16	100.00
LCA	16	17.78	*RPGRIP1*	1	6.25
			GUCY2D	2	12.50
			RDH12	4	25.00
			CRB1	1	6.25
			IQCB1	2	12.50
			TULP1	1	6.25
			CRX	1	6.25
			LCA5	1	6.25
			OXT2	1	6.25
			RPE65	1	6.25
			AIPL1	1	6.25
Other	2	2.22	*NYX*	1	6.25
			FBN1	1	6.25
Total	90				

3.3.6 Novel variants

Here we present 19 novel missense variants that, to the best of our knowledge, have not yet been associated with an IRD (Table 3.6). *In silico* tools including MetaLR, M-CAP and REVEL were utilised to predict the pathogenicity of these variants. Segregation analyses were carried out where possible, results of which were used as evidence according to ACMG-AMP guidelines where appropriate[131]. Two of the variants listed here, *BBS10* c.155G>A (p.Gly52Asp) and *PRPH2* c.464C>T (p.Thr155Ile), had been allocated dbSNP IDs; however, this is likely due to their detection in population sequencing studies, such as gnomAD[145], where they have not previously been associated with an IRD. Ten of these variants have been classified as variants of unknown significance although having strong *in silico* predicted pathogenicity. This highlights the importance of functional analysis *in vitro* and *in vivo* in order to confidently call such variants likely pathogenic.

Table 3.6 | List of 19 novel missense variants discovered in Target 5000 as part of this analysis. American College of Medical Genetics and Genomics (ACMG) guidelines were applied[131], with modifications by Romanet *et al.*[133]. (S): reclassified as strong evidence. VUS = Variant of unknown significance. Oberserved with refers to other variants identified in the same patient that are candidate disease causing variants. However, phasing has not been performed.

Gene	Transcript ID	Chr	Genomic Location (GRCh38)	Nucleotide Change	Protein Change	ACMG-AMP Criteria	ACMG-AMP Classification	Observed With
ABCA4	NM_000350.2	chr1	94062649	c.1865G>A	p.Ser622Asn	PM2, PP3, PP4(S)	Likely Pathogenic	c.161G>A
ABCA4	NM_000350.2	chr1	g.94111517	c.223T>C	p.Cys75Arg	PM2, PM5, PP3	VUS	c.4253+43G>A
ABCA4	NM_000350.2	chr1	g.94029516	c.4468T>C	p.Cys1490Arg	PM2, PP3, PP4(S)	Likely Pathogenic	c.3305A>T
ABCA4	NM_000350.2	chr1	g.94014674	c.5329A>G	p.Met1777Val	PM2, PP3	VUS	c.1037A>C
ABCA4	NM_000350.2	chr1	g.93996182	c.6743T>C	p.Phe2248Ser	PM2, PP3, PP4(S)	Likely Pathogenic	c.302+68C>T
AHI1	NM_001134830.1	chr6	g.135428670	c.2582G>A	p.Gly861Glu	PM2, PM3, PP1, PP3	Likely Pathogenic	c.1267C>T
AIPL1	NM_014336.4	chr17	g.6425759	c.856G>C	p.Asp286His	PM2, PM3, PP1, PP3	Likely Pathogenic	c.834G>A
BBS10	NM_024685.3	chr12	g.76348204	c.155G>A	p.Gly52Asp	PM1, PM2, PP3	VUS	c.2119_2120delGT
BEST1	NM_001139443.1	chr11	g.61955783	c.133C>G	p.Arg45Gly	PM2, PP3, PP4(S)	Likely Pathogenic	/
BEST1	NM_001139443.1	chr11	g.61955157	c.23A>G	p.Tyr8Cys	PM2, PP3, PP4(S)	Likely Pathogenic	/
CLRN1	NM_001195794.1	chr3	g.150928039	c.635A>G	p.His212Arg	PM2, PM3, PP1, PP3	Likely Pathogenic	c.118T>G
CNGA3	NM_001298.2	chr2	g.98391904	c.607T>C	p.Trp203Arg	PM2, PP3	VUS	c.1694C>T
MYO7A	NM_000260.3	chr11	g.77192227	c.4101C>G	p.Ile1367Met	PM2, PP3	VUS	c.2904G>T
PEX7	NM_000288.3	chr6	g.136822742	c.77C>T	p.Pro26Leu	PM2, PP3, PP4	VUS	c.875T>A

PRPF31	NM_015629.3	chr19	g.54124558	c.757G>A	p.Gly253Arg	PM2, PP3	VUS	/
PRPH2	NM_000322.4	chr6	g.42721878	c.457A>G	p.Lys153Glu	PM2, PP3	VUS	/
PRPH2	NM_000322.4	chr6	g.42721871	c.464C>T	p.Thr155Ile	PM2, PP3	VUS	/
RDH5	NM_001199771.1	chr12	g.55721276	c.92T>C	p.Val31Ala	PM2, PP3	VUS	c.712G>T
USH2A	NM_206933.2	chr1	g.216247097	c.2297G>A	p.Cys766Tyr	PM2, PM3, PM5, PP3	Likely Pathogenic	c.2299delG

3.3.7 *RPE65*

In Ireland, the prevailing *RPE65* phenotype has been associated with a specific variant, c.1430A>G (p.Asp477Gly). The disease phenotype is comparatively much milder than biallelic *RPE65*-LCA or *RPE65*-RP and more closely resembles the clinical manifestations of choroideremia[146]. This amino acid position has been shown to be highly conserved across multiple species[147]. This variant is one that continues to be detected in additional cases in the ongoing Target 5000 study. Initial reports and phenotype characterisation of this mutation were based on several pedigrees originating from Ireland[147,148]. The dominant c.1430A>G (p.Asp477Gly) variant is probed for, yet is undetected, in large *RPE65* screening studies (n > 2000) of non-Irish ethnicity[149]. Currently in the Target 5000 cohort, there are 23 genetically confirmed affected individuals, with many patients from these pedigrees that are pending recruitment. These 23 affected patients span 7 pedigrees although likely originate from a single source as observed in previous studies[150].

3.3.8 *FLVCR1*

Typically, variants in *FLVCR1* have been associated with a neurological syndrome, posterior column ataxia with RP (PCARP) (MIM: #609033)[151–153], and more recently a specific splice variant (c.1092+5G>A) has been reported multiple times to be associated with non-syndromic RP [154–156]. Through Target 5000, substantial evidence has been obtained that suggests the first incidence of a protein coding *FLVCR1* variant c.1022A>G (p.Tyr341Cys) implicated in non-syndromic RP[157]. Patients with this variant present with typical RP, without any extraocular features (Figure 3.7). Since its initial detection, this variant has been observed and deemed to segregate with the condition in three additional not knowingly related families in the Target 5000 cohort. Each of the genotyped affected individuals in these three pedigrees are homozygous for this variant.

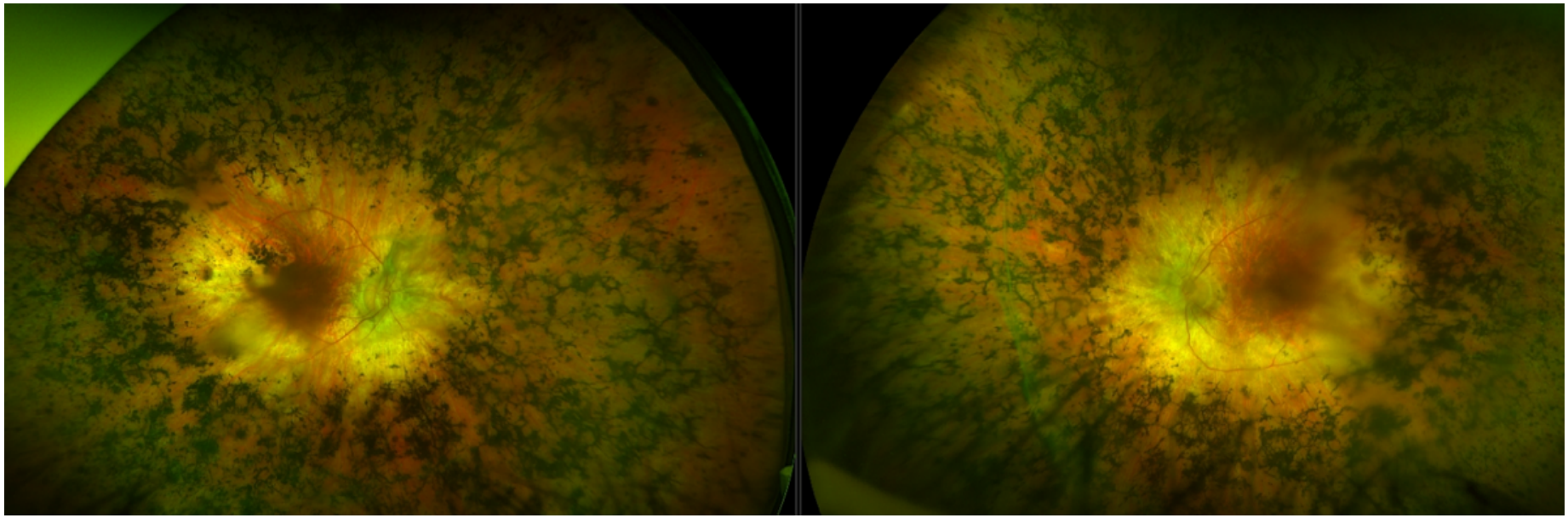


Figure 3.7 | Wide-field fundus images from a retinitis pigmentosa (RP) patient with a *FLVCR1* genotype. The *FLVCR1* genotype results in a phenotype (left- right eye, right- left eye) that appears to bear a resemblance to classical RP, with features such as masses of bony spicules in the periphery, attenuated blood vessels, and waxy disc pallor. There is also a relative preservation of the maculae

3.3.9 *CHM*

Choroideremia (MIM #303100) is an X-Linked recessive chorioretinal degenerative condition with progressive atrophy of various retinal cell types and the surrounding blood retinal barrier. Here we describe a novel deletion in the *CHM* gene found in two Irish pedigrees. This nearly 500 kb deletion represents the largest as yet detected IRD-associated gene deletion in Ireland (Figure 3.8). Two members of a large X-linked RP pedigree (Figure 3.9) clinically presented with choroideremia and tested negative for the segregating *RPGR* variant found in other affected members of this pedigree. Both males were analysed with TCS and found to possess large deletions spanning the *CHM* gene, approximating 500 kb. The observation of two different IRDs segregating in this one pedigree highlights the significant value of NGS-based diagnostics for IRDs.

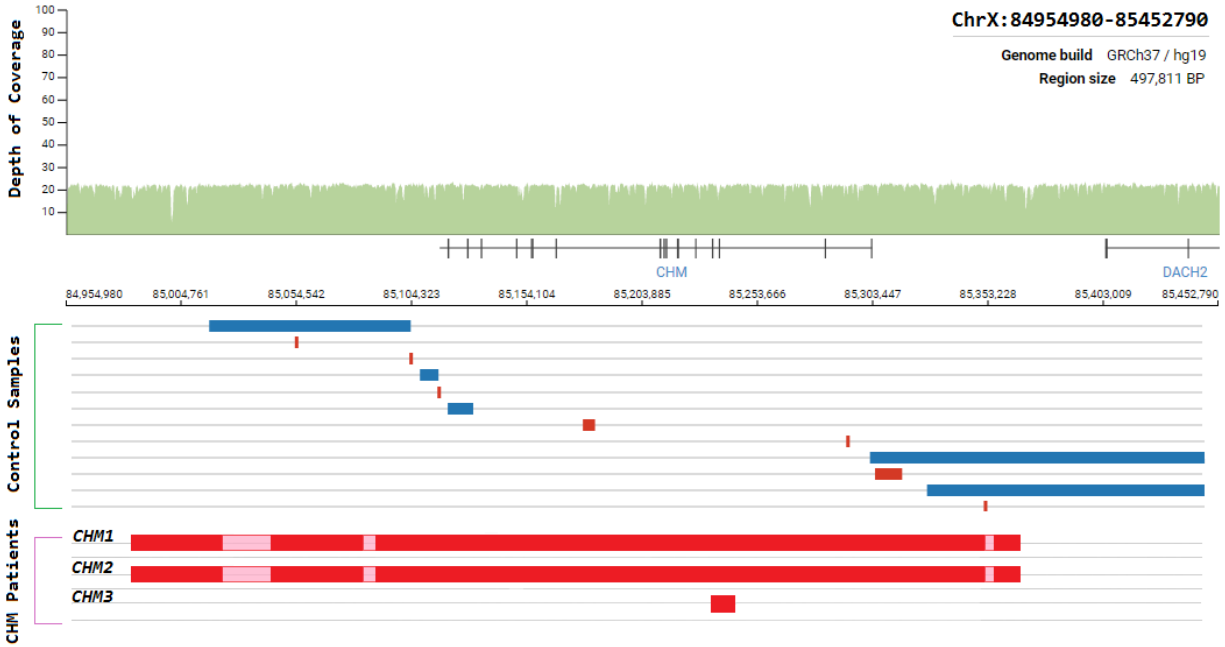


Figure 3.8 | Schematic representation of the genomic region surrounding the *CHM* gene. Whole-genome coverage data from a control population database (https://gnomad.broadinstitute.org/). SVs detected from controls samples and Target 5000 patients are aligned to this region to illustrate the instability of this genomic region. Given the presence of structural variation in gnomAD, it is possible that this area is prone to instability perhaps due to sequence complementarity. Blue = duplication, Red = deletion, Pink = template positive results likely due to highly similar sequence elsewhere in the genome i.e., these regions were successfully amplified via a PCR-based approach to determine breakpoints. However, they are highly homologous to other regions of the genome therefore it is not clear if they are truly present in this region or have been amplified from elsewhere. Genomic positions were lifted over from hg39 to hg19 as gnomadSV is aligned only to hg19.

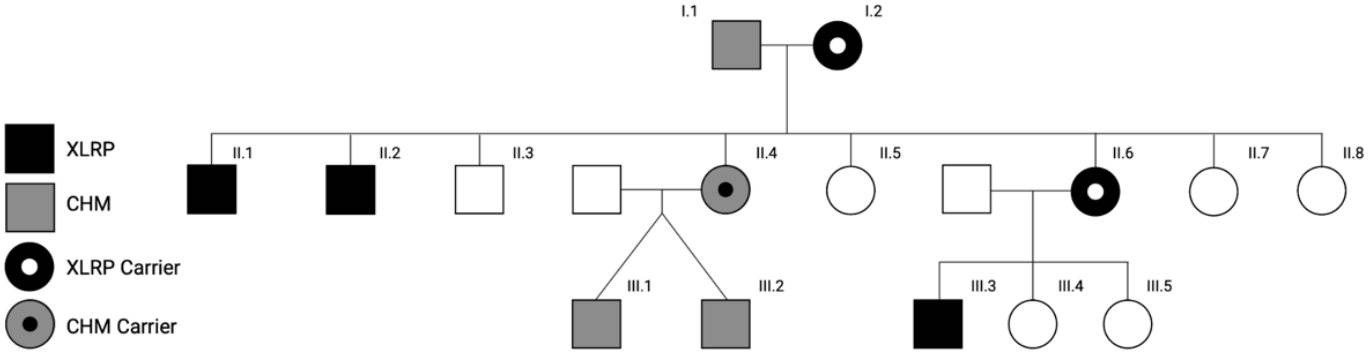


Figure 3.9 | Pedigree tree illustrating the segregation of two X-linked retinal conditions within one family

The same *CHM* deletion has also been detected in a second Irish pedigree since its initial discovery. Two additional males and two carrier females from this second pedigree were all found to be affected with progressive choroideremia. In this pedigree, the incidence of choroideremia could be traced back 5 generations, 4 of which have been assessed by the clinical team. A third Irish pedigree with a large *CHM* deletion was also detected previously as part of Target 5000[117]. This mutation spanned approximately 6.5 kb and encompassed exons 3 and 4 of *CHM* (Figure 3.8). In each instance, target capture NGS detected the presence of the deletion and breakpoints were approximated by tiled PCR analyses. Interestingly, some probing PCRs in the intergenic regions surrounding *CHM* suggested that genomic DNA was present between deleted regions. We hypothesise that these PCR products were potentially the result of large regions of homology that exist within this genomic area and other regions of the genome, having used BLAT to align the sequence obtained. BLAT illustrated that the sequence obtained via PCR and Sanger sequencing aligned to multiple regions of the genome. Multiple primer sets of this region achieved the same results. Several SVs in this *CHM*-proximal intergenic region have also been observed in control samples (Figure 3.8).

3.4 Discussion

To date, over 1000 individuals have been genotyped as part of Target 5000, accounting for over 20% of the estimated Irish IRD cohort. The results of this study thus far highlight the unique genetic architecture of IRDs in Ireland, cumulatively resulting in 89 novel variants identified to date[117,120], 19 of which were missense mutations in addition to 1 novel SV identified in the most recent analysis (Table 3.6, Figure 3.8). In this study, a positive candidate detection rate of almost 70% (495/710 pedigrees) was achieved.

This study also highlights the importance of an accurate genetic diagnosis in addition to a clinical diagnosis. Genetic diagnoses continue to become more relevant as gene-based medicines move towards the forefront of IRD treatment given a first approved gene therapy for an IRD and an array of gene therapies for other forms of IRD in clinical and preclinical development. Furthermore, genetic diagnoses facilitate a better understanding of disease progression and manifestation for both patients and clinicians.

STGD1 is an autosomal recessive disorder caused almost exclusively by variants in the *ABCA4* gene[158]. It is the most frequent form of macular dystrophy with an estimated prevalence of approximately 1 in 10,000 individuals[159]. STGD1 is characterised by progressive bilateral central vision loss, colour vision defects, delayed dark adaptation,

and flecks in the retinal pigmentary epithelium[160]. A spectrum of disease severity underlies STGD1, with age of onset and progression governed by the specific combination of variants an individual carries, highlighting again the value of a specific genetic diagnosis for disease progression and risk assessment. An extensive list of 5962 likely pathogenic *ABCA4* variants in 3928 cases was published in 2017[161], providing a valuable resource to aid in the interpretation of variants discovered as part of this study. In addition to coding variants captured, it has become increasingly clear that many of the unresolved one allele only (Figure 3.1) cases may likely carry a second pathogenic intronic variant where the disease is typically associated with a recessive inheritance pattern. This has been elegantly demonstrated by DIVs in *ABCA4* associated disease, such as c.4539+2028C>T (p.[=,Arg1514Leufs*36]). At the time of this study this variant had been observed as a candidate in 10 Target 5000 individuals, of which 5 were discovered as part of a landmark study of *ABCA4*-linked STGD1 employing smMIPs based whole gene sequencing[23] and 5 of which were discovered through TCS runs. To date, 25 individuals harbouring this variant as well as two homozygous individuals have been identified. This is discussed in detail in section 6.1. This highlights a significant enrichment of this variant in the Irish IRD patient cohort. C.4539+2028C>T (p.[=,Arg1514Leufs*36]) was previously functionally analysed in photoreceptor precursor cells (PPCs), and was shown to result in a 345 nucleotide pseudoexon inclusion due to strengthening of exonic splice enhancers. It is notable that such variants are an attractive target for antisense oligonucleotide (AON)-based splice correction therapy[112].

Identified variants proximal to the *RHO* p.Met207 region appear predominant in our sub-cohort of autosomal dominant RP patients, as reported in Section 3.3.2. Amino acids in this region of the protein form part of the fifth transmembrane domain of rhodopsin. This domain has been shown to be vital to retinal binding in bovine rhodopsin[162]. It is possible that undesirable changes in the protein folding of this region result in steric hindrance that prevents the apoprotein from successfully binding retinal. This impaired binding capacity may possibly underpin the reason for the prevalence of pathogenic variants found in the coding region of this protein domain.

FLVCR1 is a transmembrane protein involved in erythropoiesis and heme transport[163]. *FLVCR1* was first discovered for its role in aplastic anaemia in domestic cats and subsequently erythroblast destruction *in vivo*[164]. This was also the first study that theorised that *FLVCR1* was a receptor for an organic anion and identified its component domains as strikingly similar to other ancient major facilitator superfamily members. In a later study it was shown this receptor protein provided the first description of a mammalian

heme transporter[165]. The same group proceeded to investigate its effect in knockout mouse models, which resulted in embryonic lethality[163]. *FLVCR1* was first associated with PCARP by Higgins *et al.,* in 1997[166]. All individuals had syndromic features of this disease. Discovery of this link has prompted inclusion of *FLVCR1* in IRD sequencing panels. Indeed, candidate variants in *FLVCR1* continue to be identified in Target 5000.

It is notable the c.1022A>G (p.Tyr341Cys) variant is only the second disease-associated variant to be found in *FLVCR1* associated with RP without posterior column degeneration[155]. Moreover, it is the first protein coding variant found in this gene to be affiliated with non-syndromic RP, a finding discovered by the Target 5000 team[157]. There were no extraocular features associated with this variant (Figure 3.7). It is possible that the distinct families in this cohort with this variant share common ancestry. However, they do not share any rare variants in the regions captured and originate from distinct geographical locations in Ireland. The age of onset for the symptoms of ataxia in PCARP has been reported as typically in the third decade of life. Of note in the Target 5000 cohort, all patients carrying this variant have reached this age, with the oldest patient currently in their seventh decade of life. It is unclear yet as to whether the disease pathology associated with this variant will remain completely non-syndromic throughout the entirety of a patient's lifespan or will result in a milder, later onset of additional symptoms compared to other pathogenic variants found in this gene.

The RPE65 enzyme is active in the retinoid cycle that recycles retinoids in the retinal pigment epithelium (RPE), which are utilised by photoreceptor cells. Mutations in the *RPE65* gene were initially found to be causative of some cases of autosomal recessive LCA, a severe and clinically distinct IRD[167–169]. This was subsequently expanded to include a milder, dominant form of retinopathy and more recently, to include less severe autosomal recessive retinopathies [147,149]. Luxturna™, the first ocular gene therapy approved by the FDA, is for biallelic *RPE65* retinal disease[170]. *RPE65* c.1430A>G (p.Asp477Gly) remains the predominant *RPE65* genotype observed in the Irish population. This variant has since been detected in other population studies alongside the clinical presentation of similar phenotypes, largely described as a choroideremia phenocopy[148,150]. However, until recently the exact disease mechanism caused by this variant has evaded researchers. It was initially believed that this variant produced an abnormal protein structure that may affect enzymatic function[147]. More recently, it has been shown in a mouse model that although this variant is not located in a critical functional domain, the variant induces sufficient change to alter the physiochemical properties of the p.D477 loop to produce an aggregation-prone surface. This surface gains

the aberrant function of enabling abnormal protein-protein interactions[171]. One hypothesised type of interaction is with ubiquitin ligases, which would likely result in proteasomal degradation as seen in several other *RPE65* mutations[172]. This theory may support the variability observed phenotypically as severity might then become dependent on several other factors such as ubiquitination rates and proteasome response. It is of note that aberrant RNA splicing has been shown to have a major pathogenic effect in a knock-in mouse model with this variant[114]. Given the success of the gene therapy treatment of biallelic *RPE65* retinopathies with Luxturna™[170], it is interesting to speculate whether the therapy might have any utility for monoallelic dominant *RPE65* cases. Oral administration of synthetic retinoids may also be a promising treatment option. QLT091001, when taken daily for 7 days, has shown benefit in the majority of patients within six months as part of an open-label, Phase 1b, proof-of- concept study. In the study, 80% of patients responded positively to this treatment with 60% showing an increase in visual fields and 60% with improved visual acuity. There were no serious adverse effects reported as part of the study. This trial showed promising data that treatment of patients with an *RPE65* p.D477G genotype are likely to respond well to the safe treatment of orally-administered synthetic retinoids. All of the clinical studies referred to were undertaken by Mr. Paul F. Kenna[173]. These data highlight the significant outcomes that can be achieved as a result of a genotyping study such as Target 5000.

The *CHM* gene hosts the largest collection of pathogenic SVs detected in any IRD-related gene as part of the Target 5000 study to date. The presence of several SVs in the general population (Figure 3.8) may also indicate a natural instability in this genomic region[174]. Additionally, female carriers of *CHM* mutations typically show milder non-progressive signs with no symptoms, while males are severely affected. In this instance, some females were more severely affected than expected with advanced signs of degeneration and progressive visual decline. One female carrier showed radial pigmentation at the RPE level and scattered choroidal atrophy, whereas another showed more marked reactive pigmentation and extensive symmetrical atrophy of the choroid and outer retina. These manifestations were apparent both anatomically and functionally. The molecular pathogenesis of the more severe phenotype that usually observed in these *CHM* variant carrier female patients remains elusive. One possibility may be the presence of non-random X-chromosome inactivation, where a greater proportion of the X chromosome carrying the normal allele is inactivated; such a phenomenon has been observed previously in a number of X-linked conditions[175]. Alternatively, a variant in the promoter sequence of the normal allele such that the allele and encoded protein is

expressed at a lower level may underlie the observed phenotype in carriers. A more severe phenotype was also associated with whole *CHM* gene deletion in a previous genotype-phenotype choroideremia study[176]. However, in this study, the phenotype was described as choroideremia plus, due to additional syndromic features. These additional symptoms, such as deafness, were attributed to the extent of the deletion encompassing several surrounding genes, which is not the case for patients with whole gene deletions described here.

Thus far, the genetic analysis of IRD patients described herein has provided an overview of the genetic architecture of IRDs in Ireland and has helped to resolve ambiguous phenotypes and to identify causative mutations in nearly 70% (495/710) of sequenced pedigrees. The continuous expansion of our IRD cohort has enabled us to better interrogate the sequencing data and interpret the potential pathogenicity of novel variants when detected. Furthermore, whole gene and WGS analyses are highlighting the significant role of non-coding variants as causative of some IRDs. This study has established the efficacy of targeted sequencing as a first tier genetic test for IRDs in Ireland, but also highlights the need for additional testing of those that remain unresolved. Given the rapid development of the field of ocular therapeutics, it is clear that genetically characterising those affected by an IRD has become a diagnostic imperative of the utmost importance.

4. Whole genome sequencing and *in vitro* splice assays reveal genetic causes for inherited retinal diseases.

4.1 Introduction

High-throughput DNA sequencing technologies have been widely used to investigate IRD pathogenesis. Given a high degree of heterogeneity, NGS facilitates the analysis of genetic variation in multiple regions of the genome in a single sequencing procedure[177]. For many years, the application of NGS in IRD molecular diagnostics has been limited to TCS and WES. TCS has proven to be an effective preliminary diagnostic technique, boasting cost effectiveness, rapid data generation, and lower volumes of data interpretation as major advantages[13,57,58,118] Sequence analysis of only previously IRD-associated genes is one of the major disadvantages of this technique[178]. On the contrary, WES enables the analysis of genes not yet associated with IRD. It is estimated that 85% of pathogenic variants are located within protein-coding regions, the primary focus of WES[179,180]. However, this may vary depending on the disease entity. It has become evident that sequencing non-coding regions of the genome is required to also identify causal intronic variants and to improve the identification of SVs, CNVs and their breakpoints[54,181–183]

Using a TCS or WES approach, the analysis is restricted to genetic variations in protein-coding and flanking splice site regions of the genome. Furthermore, these techniques are limited in their ability to detect SVs such as deletions, duplications, or inversions when breakpoints lie in intronic or intergenic regions[52]. These limitations are less evident using WGS, which enables the identification of most types of variation across the entire genome[63,184]. Due to decreasing costs and increasing data interpretability, WGS is becoming an attractive alternative albeit predominantly in research settings, as diagnostic laboratories can interpret variants only in previously associated IRD genes with confidence using the ACMG-AMP guidelines[185]. In addition, IRD cases are ideal candidates for WGS analysis as they are largely presumed to be primarily monogenic diseases albeit with a high degree of genetic heterogeneity.

In this study, we aimed to investigate the sensitivity and accuracy of WGS in a cohort of 100 individuals with different IRDs. All individuals were previously tested by TCS or WES, but a genetic cause of disease remained undetermined. The study provides insights into the increased resolution of IRD cases that can be achieved by employing WGS.

4.2 Methods

4.2.1 Patient cohort

A total of 100 IRD probands were subjected to WGS, of which 43 were recruited from the Radboud University Medical Center, Nijmegen and The Rotterdam Eye Hospital (both in The Netherlands), 19 from the Technion-Israel Institute of Technology (Haifa, Israel), and 38 from the Research Foundation of the Royal Victoria Eye and Ear Hospital (Dublin, Ireland). Further details on ethnicity were not available. A suspected recessive inheritance pattern and unresolved status following prior preliminary genetic testing using TCS or WES were prerequisites for participation in this study.

4.2.2 DNA acquisition and preliminary genetic analysis

Participants provided a blood sample for analysis. DNA from Irish participants was extracted as described in section 2.3. TCS and subsequent data analysis of SNVs, SVs, and CNVs were carried out as described previously in the Irish cohort[13,117,118] and eight Israeli cases[50,186] All Dutch (43 cases) and remaining Israeli (11 cases) participants had undergone WES and the subsequent data were analyzed based on the identification of known pathogenic variants as well as likely pathogenic variants or variants of unknown significance, i.e., assessment of all type of variation including SNVs, SVs, and CNVs. Variant prioritization in WES data was based on a strong correlation between genotype and phenotype in individuals, i.e., a first pathogenic allele was identified in a gene previously found to be mutated in the IRD cases with a similar phenotype. WES was performed at BGI-Europe (Copenhagen, Denmark). In short, by application of Agilent's SureSelectXT Human all Exon V5 (Agilent Technologies, Santa Clara, CA), the exome was enriched initially followed by next-generation sequencing using an Illumina HiSeq 4000 sequencer (Illumina, Inc. San Diego, CA) to a mean sequence depth of at least 75-fold. The reads were aligned to the human reference genome (hg19) using Burrows-Wheeler Aligner. Subsequently, nucleotide and CNV calling were performed with the Genome Analysis Toolkit and CoNIFER 0.2.0, respectively[187]. Ultimately, variants were annotated using a bespoke in-house annotation strategy[188] For the Israeli cases, the targeted capture sequencing was performed by molecular inversion probes with 5-bp molecular tags incorporating 108 genes associated with IRDs. Sequencing was performed by the Illumina NextSeq 500 system and the reads were aligned to a human reference sequence (hg19) via Burrows-Wheeler Aligner (BWA) v.0.6.2. The variant calling tools and annotation pipeline were similar to the WES method[186]. Targeted capture sequencing of Irish cases was performed with the Nimblegen SeqCap EZ kit (Roche Ireland Ltd.,

Dublin, Ireland), incorporating the exonic regions of 254 genes associated with retinopathies followed by sequencing the enriched region using an Illumina MiSeq (Illumina Inc., San Diego, CA, USA) with the average read coverage of 125× per captured region[2,5]. The reads were aligned to the human reference genome (hg38) using Burrows-Wheeler Aligner 0.7.15. Subsequently, nucleotide, copy number and structural variant calling were performed using Freebayes 1.1.0 (arXiv:1207.3907), CoNIFER[187] and Manta[129] respectively. Thirty-three out of 43 Dutch, 14 out of 38 Irish, and nine out of 19 Israeli cases were monoallelic, carrying one reported pathogenic or candidate allele in IRD-associated genes prior to WGS.

WGS was performed by BGI (Hong Kong, China) on a BGISeq500 using either 2×100 or 2×150 bp paired-end reads, with a 30-fold minimal median coverage per genome. Burrows–Wheeler Aligner was utilized to map the data to the human genome (hg19). The quality of the WGS data was based on the insert size, percentage mapped reads, percentage duplicated mapped reads, coverage, bases with >20× coverage and error rate, which were evaluated using Qualimap V.2.2.1[189]. Variant calling was performed by xAtlas V.0.1[190] and variants were annotated using the Variant Effect Predictor (V.91)[191] and Gencode V.34lift37 basic gene annotations. It consists of various features such as chromosomal and nucleotide position, gene name and component (e.g., intron, exon, splice site, 5′-untranslated region (5′-UTR), 3′-UTR, intergenic). In addition, we used an in-house developed pipeline that provides further information from different population frequency databases (e.g., gnomAD (https://gnomad.broadinstitute.org), GoNL[192], Wellderly[193], 1000genomes[194], in-house variant frequencies database containing WES data of 15,576 individuals), various *in silico* prediction scores (e.g., Grantham, PhyloP, CADD_PHRED, SpliceAI)[195–198], predicted protein effect, gene and disease OMIM description, and gene regulation and expression data. Runs of homozygosity were detected using plink V.1.07[190] with the following parameters: homozyg-window-het=3, homozyg-snp=50 and homozyg-kb=300.

SVs were detected based on paired-end and split-read evidence using Manta SV Caller V.1.1.0 (Illumina) with default parameters[129]. CNVs were called using Control-FREEC, which detects copy number changes and allelic imbalances based on read depth[199]. SVs and CNVs were annotated using an in-house developed pipeline[188], i.e., annotations for chromosomal location and position, gene(s) and their component (e.g., intronic, exonic), type of SV (gain/loss), percentage overlap and frequency of various population frequency databases (e.g., gnomAD-SV[200], GoNL[192], Decipher[201], Wellderly[193], 1000genomes[194]), and disease OMIM description. Short tandem repeats for

33 known pathogenic sites were detected using Expansion Hunter V.3.1.2. using default settings[202].

4.2.3 Variant prioritization approach

Sequencing data from a customized panel of 446 IRD and ocular defect-associated genes and 163 ciliopathy associated genes were extracted from the WGS data obtained from each patient. To facilitate the identification of pathogenic variants, a two-step protocol was designed for variant prioritization, in which the first step was the utilization of an automated prioritization pipeline composed in R studio[203]. The details of this pipeline created by Dr Zeinab Fadaie and Mr B. van Gestel can be found in Appendix 9. The second step was a manual review of the remaining variants identified as follows. The manual variant prioritization approach differed depending on the findings of previous genetic studies. In cases where no first candidate variant was identified during previous genetic testing efforts, CNVs and SVs were evaluated first followed by assessment of SNVs. Subsequently, SNVs were manually prioritized according to their predicted pathogenic effect (i.e., nonsense, frameshift, CSS variants, NCSS variants, in-frame deletions/insertions, missense variants, and synonymous variants). Coding or noncoding SNVs with a minor allele frequency (AF) of >1% in the gnomAD database were not considered causative[200]. Once a candidate pathogenic variant was identified, the predicted effects of rare intronic variants (with AF <1%) in the gene of interest were interrogated *in silico.*

For cases where a first candidate variant in an IRD or ciliopathy associated gene was previously identified, the assessment of CNVs, SVs, and subsequently all rare intronic SNVs in the gene of interest was performed to identify a second causal variant. If a second causal variant remained elusive, an analysis of all CNVs, SVs, and SNVs was implemented in the same manner as for cases in which no first candidate variant was previously established. In addition, all IRD cases were analyzed for putative pathogenic short tandem repeat expansions in IRD-associated genes.

4.2.4 Variant frequency, pathogenicity prediction and midigene splice assays

After completing the data analysis of 100 cases, 10 deep-intronic and 3 NCSS candidate variants in 14 individuals were identified, which adhered to the stringent criteria defined above for noncoding variants. Subsequently, a midigene or minigene splice assay was employed to assess potential splicing defects in the presence of these variants as described previously[188,204]. The regions of interest of a genomic DNA sample were

amplified by primers that contain attB1 and attB2 tags at their 5′ end to facilitate Gateway cloning. After generating the wild-type constructs, they served as templates to generate mutant constructs by PCR-based site-directed mutagenesis. Constructs that could not be cloned via site-directed mutagenesis were cloned as follows; The *ABCA4* variant c.6148-89G>A was introduced into a vector containing the wild-type fragment of *ABCA4* from exons 43 to 47, as described previously[54]. For *PDE6B* c. 469-776C>G in Pt-61 and *RLBP1* c.525+425_525+433delinsATA in Pt-65, mutant and wild-type constructs were both generated from the patient's DNA as the mutagenesis PCR technique failed due to the complexity of the regions surrounding these variants. Subsequently, the wild-type and mutant constructs were separately incorporated into the pCI-*NEO*-*RHO* or pcDNA3native/DEST Gateway-adapted vector to generate wild-type and mutant midi(mini)genes. The pcDNA3native/DEST vector was utilized when the region of interest contained the first exon of the gene. Generated wild-type and mutant midigenes/minigenes were independently transfected into HEK293T cells. After 48 h of incubation, mRNA was isolated and utilized for transcript analysis by RT-PCR with primers in flanking exons for midigenes or primers in *RHO* exon 3 and 5 for minigenes. All primers designed for splice assays are listed in Table 4.1. Fiji software was utilized for quantification analysis in the presence of multiple mRNA fragments after gel electrophoresis. The variants with <25% of the wild-type fragments present in mutant constructs were classified as severe alleles. This variant classification refers only to the severity of the mRNA defect observed in HEK293T cells.

Table 4.1 | Details of primers synthesized for midigene splice assays of candidate variants. The sequence for all designed primers for midigene assays with GC content and melting temperature and size of PCR products. Bp, base pairs; n.a., not applicable.

Name	Sequence (5' to 3')	Tm (°C)	GC%	Size (bp)
ABCA4_ c.6148-89G>A_mut	gggctcccagatctcatgcctccagaatgaaat	80	52	10131
ABCA4_ c.6148-89G>A_Mut	atttcattctggaggcatgagatctgggagccc	80	52	
ABCA4 c.6148-89G>A validation R	gagagtttccgcttgttgc	63	53	n.a.
ABCA4_ c.6148-89G>A_RT-PCR	aagatgctcactggggacac	65	55	518
ABCA4_ c.6148-89G>A_RT-PCR	gaatggtgcccatacatcg	64	53	
ABCA4_BAC_Ex37_Ex40_F	aggtgtctgctgagcttcttc	59	50	7441
ABCA4_BAC_Ex37_Ex40_R	tgctctctcccacacaaaag	59	50	
Mut_ABCA4_c.5460+1315-17del F	tgttcttatccccccttccagagtctgactctg	77	50	7441
Mut_ABCA4_c.5460+1315-17del R	cagagtcagactctggaaaggggggataagaaca	77	50	
Mut_ABCA4_c.5460+1315insTA_F	tgttcttatcccctactttccagagtctgactctg	75	47	7441
Mut_ABCA4_c.5460+1315insTA_R	cagagtcagactctggaaagtagggggataagaaca	75	47	
ABCA4_Intron 38_F	ctggggaacacaggacaaat	60	50	392
ABCA4_Intron 38_R	gggtgcttagcaaatgttgg	61	50	
CACNA2D4_Exon19-23_tag_F	attcagttccttggccacac	60	50	6768
CACNA2D4_Exon19-23_tag_R	caacccagagaaaggagctg	60	55	
Mut_CACNA2D4_c.2009-3C>T_F	tcatgcaggcctaaaaccagagtccacattcaggg	81	51	6768
Mut_CACNA2D4_c.2009-3C>T_R	ccctgaatgtggactctggttttaggcctgcatga	81	51	
CACNA2D4_Intron20_F	acagaccccagtccctttct	60	55	400
CACNA2D4_Intron20_R	ctcctggcaaagtcctgaac	60	55	
CACN2D4_Exon 19_F	gcgagttcttttcctgacca	60	50	345
CACN2D4_Exon 23_R	gctgtccagtaggcttccat	59	55	
CYP4V2_Ex1-Ex3_tag_F	gtagagcaacctcgcagcac	61	60	4848
CYP4V2_Ex1-Ex3_tag_R	ccttcctttccagcctttct	60	50	
Mut_CYP4V2_ c.214+879del_F	cgttttacaggctgagcatctaactccaaagcttgaag	77	45	4848
Mut_CYP4V2_ c.214+879del_R	cttcaagctttggagttagatgctcagcctgtaaaacg	77	45	
Mut_CYP2V4_TAAC>G_F	cgttttacaggctgagcatgtaactccaaagcttgaag	78	45	4848
Mut_CYP2V4_TAAC>G_R	cttcaagctttggagttacatgctcagcctgtaaaacg	78	45	
CYP4V2_Intron 1_F	cctggtatttacaactccctgtc	59	48	365
CYP4V2_Intron 1_R	tgatgatatttggttggcaaag	60	36	
CYP4V2_RT-PCR_Exon 1_F	gccagtctggtcctgagc	60	67	320
CYP4V2_RT-PCR_Exon 3_R	aggccaagccatggttcta	61	53	
EYS_Intron 26_tag_F	ctccccagtttggtttgcta	60	50	982
EYS_Intron 26_tag_R	gtgtttgcttccctttttgc	60	45	
Mut_EYS_c.5644+70912A>G_F	tggtgctgggagtttcctgaggaaaaaaaggacagg	81	50	982
Mut_EYS_c.5644+70912A>G_R	cctgtccttttttttcctcaggaaactcccagcacca	81	50	
EYS_Intron 26_F	aagcccactaggatttctttcc	60	45	381
EYS_Intron 26_R	gggtaggtgtttgagtgagca	60	52	

GRM6_Ex5-Ex8_tag_F	gcagcatcctaacaggaaaga	60	48	3961
GRM6_Ex5-Ex8_tag_R	tctgacacccaagctggatt	61	50	
Mut_GRM6_c.1355-587dup_F	gtggcatcagacggtaatttgcgcatgtgtgac	81	52	3961
Mut_GRM6_c.1355-587dup_R	gtcacacatgcgcaaattaccgtctgatgccac	81	52	
GRM6_Intron 6_F	ttatgcaaccatcaccacca	61	45	451
GRM6_Intron 6_R	agttccactcccgggtagata	60	52	
GRM6_RT-PCR_Exon 5_F	agaacaaccgcaggaacatc	59	60	492
GRM6_RT-PCR_Exon 8_R	agcagggcagagtagctgag	61	55	
HGSNAT_Ex3-Ex5_tag _F	acaattccatagtgacacagtga	57	39	3959
HGSNAT_Ex3-Ex5_tag_R	ccagtcccttctccacattg	61	55	
Mut_HGSNAT_c.493+5G>A_F	gaaaatcagcagagaacatatatgtacgaaggttactatcaactg	74	36	3959
Mut_HGSNAT_c.493+5G>A_R	cagttgatagtaaccttcgtacatatatgttctctgctgattttc	74	36	
HGSNAT_Exon 4_F	tccgtacccaggaacatgtatt	61	45	517
HGSNAT_Exon 4_R	attacaagcatgagccaccat	60	43	
HGSNAT_RT-PCR_Exon 3_F	tcaggttctggtaaacgttcct	60	45	293
HGSNAT_RT-PCR_Exon 5_R	caatgatgacagcaagaccaat	60	41	
USH2A_Intron 4_tag_F	agcattaagcctgtgggtgt	60	50	935
USH2A_Intron 4_tag_R	gaggaggtgccacacacttt	60	55	
Mut_USH2A_c.784+14389G>T_F	ccatgcaagtctgaaacacaaaagggcagtcatca	81	46	935
Mut_USH2A_c.784+14389G>T_R	tgatgactgccctttgtgtttcagacttgcatgg	81	46	
USH2A_Intron 4_F	ctcactcctcccaccaatgt	60	55	318
USH2A_Intron 4_R	ttcccaacagtctcccaaag	60	50	
RHO_Exon 3_F	cggaggtcaacaacgagtct	65	55	-
RHO_Exon 5_F	atctgctgcggcaagaac	65	56	140
RHO_Exon 5_R	aggtgtaggggatgggagac	65	60	

4.3 Results

4.3.1 Patient characterization

We performed WGS on 100 IRD cases with a suspected autosomal recessive inheritance recruited in Ireland, Israel, and The Netherlands. Among these, TCS was previously performed on 38 Irish and eight Israeli cases, which did not result in a complete genetic diagnosis. The remaining 54 cases (43 Dutch and 11 Israeli) were previously examined using WES and remained genetically unresolved. In 56/100 cases, a first plausible pathogenic variant was selected from prior WES or TCS data based on a minor AF of <1% in the ExAC database[205] and 5% in dbSNP as well as <1% in an in-house database of 15,576 mostly Caucasian individuals and <2% in an in-house database of 454 mostly Asian individuals[200]. In the prior prioritization of monoallelic cases, all coding and splice site variants (nucleotides −17 to −1 and +1 to +6) were manually classified according to a five-class system based on the ACMG-AMP classification system[185], followed by evaluating the prioritized variants using practice guidelines by Wallis *et al*.[206]. Subsequent to this classification, these were included as monoallelic cases in this study (Table 4.2).

Table 4.2 | Overview of candidate causative variants as first pathogenic allele identified prior to this study. ACHM, achromatopsia; AS, Alström syndrome; CACD, central areolar choroidal dystrophy; CD, cone dystrophy; CRD, cone-rod dystrophy; CSNB, congenital stationary night blindness; FAP, fundus albipunctatus; LCA, Leber congenital amaurosis; MD, macular dystrophy; RCD, rod cone dystrophy; RP, retinitis pigmentosa; SLS, Senior-Løken syndrome; STGD1, Stargardt disease; TCS, Target capture sequencing; USH-III, Usher syndrome type 3; WES, whole exome sequencing. N.r. indicates no first candidate variant/gene identified via pre-screening methods.

ID	Phenotype	Gene	DNA allele 1	Protein allele 1	Pre-screening	Causative gene in this study
			Solved			
Pt-1	STGD1	*ABCA4*	c.6079C>T	p.(Leu2027Phe)	TCS	*ABCA4*
Pt-2	STGD1	*ABCA4*	c.6743T>C	p.(Phe2248Ser)	TCS	*ABCA4*
Pt-3	STGD1	*ABCA4*	c.5603A>T	p.(Asn1868Ile)	TCS	*ABCA4*
Pt-4	CRD*	*n.r.*	n.r.	n.r.	WES	*ATXN7*
Pt-5	RP	*n.r.*	n.r.	n.r.	TCS	*C21ORF2*
Pt-6	RP	*n.r.*	n.r.	n.r.	TCS	*C21ORF2*
Pt-7	RP	*n.r.*	n.r.	n.r.	TCS	*CDHR1*
Pt-8	AS	*ALMS1*	c.11997C>T	p.(Arg4003Trp)	TCS	*CEP78*
Pt-9	RP	*EYS*	c.[3906C>A;9468T>A]	p.[His1302Gln,Tyr3156*]	WES	*EYS*
Pt-10	RP	*EYS*	c.403_423delinsCTTTT	p.(Thr135Argfs*26)	TCS	*EYS*
Pt-11	CRD	*n.r.*	n.r.	n.r.	WES	*EYS*
Pt-12	RCD	*EYS*	c.1660T>A	p.(Cys554Ser)	WES	*EYS*
Pt-13	RP	*n.r.*	n.r.	n.r.	TCS	*FAM161A*
Pt-14	Oguchi disease	*GRM6*	c.1732C>T	p.(Arg578Cys)	WES	*GRM6*
Pt-15	RP	*n.r.*	n.r.	n.r.	WES	*HGSNAT*
Pt-16	RP	*n.r.*	n.r.	n.r.	WES	*HGSNAT*
Pt-17	RP	*HGSNAT*	c.1843G>A	p.Ala615Thr	WES	*HGSNAT*
Pt-18	RP	*n.r.*	n.r.	n.r.	WES	*HGSNAT*

Pt-19	RP	*n.r.*	n.r.	n.r.	TCS	*HGSNAT*
Pt-20	RP	*PCARE*	c.3604C>T	p.(Arg1202*)	TCS	*PCARE*
Pt-21	CRD	*n.r.*	n.r.	n.r.	WES	*RGS9BP*
Pt-22	LCA	*CEP290*	c.1079G>A	p.(Arg360Gln)	WES	*RPGRIP1L*
		PDE6A	c.2053G>A	p.(Val685Met)		
		RPE65	c.676G>A	p.(Val226Ile)		
Pt-23	RP	*n.r.*	n.r.	n.r.	TCS	*PRPF31*
Pt-24	RP	*PCARE*	c.2756_2768del	p.(Lys919Thrfs*2)	TCS	*USH2A*
			Unsolved			
Pt-25	STGD1	*ABCA4*	c.6416G>C	p.(Arg2139Pro)	TCS	
Pt-26	MD	*ABCA4*	c.5882G>A	p.(Gly1961Glu)	WES	
Pt-27	STGD1	*ABCA4*	c.161G>A	p.[Cys54Serfs*14,Cys54Tyr]	TCS	
Pt-28	RP	*ABCA4*	c.[3758C>T;5882G>A]	p.[Thr1253Met,Gly1961Glu]	WES	
		EYS	c.2137+1G>A	p.(?)		
Pt-29	RCD	*ABCA4*	c.2701A>G	p.(Thr901Ala)	WES	
		IMPG2	c.3038C>T	p.(Pro1013Leu)		
Pt-30	CACD	*ABCA4*	c.5603A>T	p.(Asn1868Ile)	WES	
		IMPG2	c.745C>T	p.(Leu249Phe)		
Pt-31	RP	*ABCA4*	c.6148G>C	p.(Val2050Leu)	WES	
		PDE6B	c.655T>C	p.(Tyr219His)		
Pt-32	Nystagmus	*ALMS1*	c.8002C>T	p.(Arg2668*)	WES	
Pt-33	RP	*BBS1*	c.1169T>G	p.(Met390Arg)	TCS	
Pt-34	RP	*BBS2*	c.1033A>C	p.(Lys345Gln)	WES	
		CLUAP1	c.1093-2A>G	p.(?)		
Pt-35	RP	*BBS4*	c.1388G>C	p.(Gly463Ala)	TCS	

Pt-36	Clumped pigmentary retinal	*BEST1*	c.404C>T	p.(Ala135Val)	WES	
	dystrophy					
Pt-37	MD	*BEST1*	c.1403C>T	p.(Pro468Leu)	WES	
Pt-38	RP	*CACNA2D4*	c.[1882C>T;2009-3C>T] #	p.[Arg628*,=]	TCS	
Pt-39	CD	*C21orf2*	c.862C>T	p.(Gln288*)	WES	
Pt-40	CRD	*CCDC138*	c.[1115C>T;1376G>A]	p.[Ser372Leu,Gly459Asp]	WES	
Pt-41	MD	*CEP290*	c.1079G>A	p.(Arg360Gln)	WES	
Pt-42	RP	*CEP290*	c.3176T>A	p.(Ile1059Lys)	WES	
Pt-43	CRD	*CEP290*	c.3812G>A	p.(Arg1271Gln)	WES	
		CLRN1	c.434-2A>T	p.(?)		
Pt-44	RP	*CNGA1*	c.959T>C	p.(Leu320Pro)	WES	
Pt-45	RP	*CNGB1*	c.2539G>A	p.(Gly847Arg)	TCS	
Pt-46	RCD	*CNGB1*	c.2747G>A	p.(Arg916His)	WES	
Pt-47	RP	*CNGB3*	c.1148del	p.(Thr383Ilefs*13)	WES	
Pt-48	MD	*CNGB3*	c.1208G>A	p.(Arg403Gln)	WES	
Pt-49	MD	*CNNM4*	c.502T>C	p.(Ser168Pro)	TCS	
Pt-50	RP	*CRB1*	c.135C>G	p.(Cys45Trp)	TCS	
Pt-51	USH-III	*CRB1*	c.664G>A	p.(Glu222Lys)	WES	
		USH1G	c.83C>T	p.(Pro28Leu)		
Pt-52	Bietti crystalline corneoretinal	*CYP4V2*	c.610G>A	p.(Ala204Thr)	WES	
	dystrophy					
Pt-53	MD	*FAM161A*	c.1355_6delCA	p.(Thr452Serfs*3)	WES	
Pt-54	BBS	*FGD1*	c.935C>T	p.(Pro312Leu)	WES	
Pt-55	RP	*HGSNAT*	c.370A>T	p.(Arg124Trp)	TCS	

Pt-56	RP	*HSPA9*	c.1042C>T	p.(Arg348Cys)	WES	
Pt-57	RP	*IFT172*	c.406C>T	p.(Arg136Cys)	WES	
Pt-58	SLS	*IQCB1*	c.424_425del	p.(Phe142Profs*5)	WES	
Pt-59	LCA	*NMNAT1*	c.-57+13798_440-343dup	p.(Ala147Aspfs*8)	WES	
Pt-60	CRD	*PDE6B*	c.2363G>A	p.(Arg788His)	WES	
Pt-61	RP	*PDE6B*	c.621+97_*785del	p.(Val208_855del)	WES	
Pt-62	RP$	*PRPF31*	c.632G>A	p.(Arg211Gln)	WES	
Pt-63	CRD	*PRPH2*	c.133C>T	p.(Leu45Phe)	WES	
Pt-64	RP	*RBP3*	c.2153delC	p.(Pro718fs)	TCS	
Pt-65	RP	*RLBP1*	c.545T>G	p.(Phe182Cys)	WES	
Pt-66	RP	*SNRNP200*	c.4835C>T	p.(Thr1612Met)	WES	
Pt-67	CRD	*SNRNP200*	c.5644C>A	p.(Pro1882Thr)	WES	
Pt-68	CSNB	*TRPM1*	c.3121A>T	p.(Ile1041Phe)	TCS	
Pt-69	RP	*TULP1*	c.821delA	p.(Lys274Argfs*36)	TCS	
Pt-70	CD	n.r.	n.r.	n.r.	WES	
Pt-71	RP	n.r.	n.r.	n.r.	TCS	
Pt-72	CRD	n.r.	n.r.	n.r.	TCS	
Pt-73	RP	n.r.	n.r.	n.r.	TCS	
Pt-74	RP	n.r.	n.r.	n.r.	TCS	
Pt-75	RP	n.r.	n.r.	n.r.	TCS	
Pt-76	RP	n.r.	n.r.	n.r.	TCS	
Pt-77	RP	n.r.	n.r.	n.r.	TCS	
Pt-78	ACHM	n.r.	n.r.	n.r.	TCS	
Pt-79	CD	n.r.	n.r.	n.r.	TCS	
Pt-80	RP	n.r.	n.r.	n.r.	TCS	

Pt-81	RP	n.r.	n.r.	n.r.	TCS	
Pt-82	RP	n.r.	n.r.	n.r.	TCS	
Pt-83	STGD1	n.r.	n.r.	n.r.	TCS	
Pt-84	STGD1	n.r.	n.r.	n.r.	TCS	
Pt-85	RP	n.r.	n.r.	n.r.	TCS	
Pt-86	RP	n.r.	n.r.	n.r.	TCS	
Pt-87	RP	n.r.	n.r.	n.r.	TCS	
Pt-88	CRD	n.r.	n.r.	n.r.	TCS	
Pt-89	RP	n.r.	n.r.	n.r.	TCS	
Pt-90	RP	n.r.	n.r.	n.r.	WES	
Pt-91	RP	n.r.	n.r.	n.r.	WES	
Pt-92	FAP	n.r.	n.r.	n.r.	WES	
Pt-93	RP	n.r.	n.r.	n.r.	WES	
Pt-94	MD	n.r.	n.r.	n.r.	WES	
Pt-95	MD	n.r.	n.r.	n.r.	WES	
Pt-96	RP	n.r.	n.r.	n.r.	TCS	
Pt-97	RP	n.r.	n.r.	n.r.	TCS	
Pt-98	RP	n.r.	n.r.	n.r.	WES	
Pt-99	RP	n.r.	n.r.	n.r.	WES	
Pt-100	MD	n.r.	n.r.	n.r.	WES	

A broad range of IRDs were included in this cohort (Table 4.3 and Figure 4.1). RP, (MIM: #268000) is the most frequently listed phenotype (53%), followed by CRD (9%, MIM: #601777), macular dystrophy (8%, MIM: #616152), and STGD1 (6%, MIM: #248200). Rare forms of IRD such as LCA (MIM: #204000), BBS (MIM: #209900), and Senior–Løken syndrome (MIM: #266900) were also included.

Table 4.3 | Phenotypic diversity of patients and the number of cases referred for TCS, WES, and WGS. WES whole-exome sequencing, TCS target-capture sequencing, WGS whole-genome sequencing.

Clinical diagnosis	No. of cases previously screened by WES	No. of cases previously screened by TCS	No. of cases screened by WGS
Retinitis pigmentosa	19	38	57
Cone dystrophy	3	2	5
Rod dystrophy	0	1	1
Cone-rod dystrophy	4	4	8
Alström syndrome	0	1	1
Bardet–Biedl syndrome	1	0	1
Leber congenital amaurosis	2	0	2
Macular dystrophy	7	1	8
Oguchi disease	1	0	1
Central areolar choroidal dystrophy	1	0	1
Fundus albipunctatus	0	1	1
Usher syndrome type III	1	0	1
Senior-Løken syndrome	1	0	1
Nystagmus	0	1	1
Stargardt disease	0	6	6
Clumped pigmentary retinal dystrophy	1	0	1
Retinal dystrophy with progressive neurodegeneration	1	0	1
Retinal dystrophy with hypogonadism	1	0	1
Retinal dystrophy	0	2	2
Total number	43	57	100

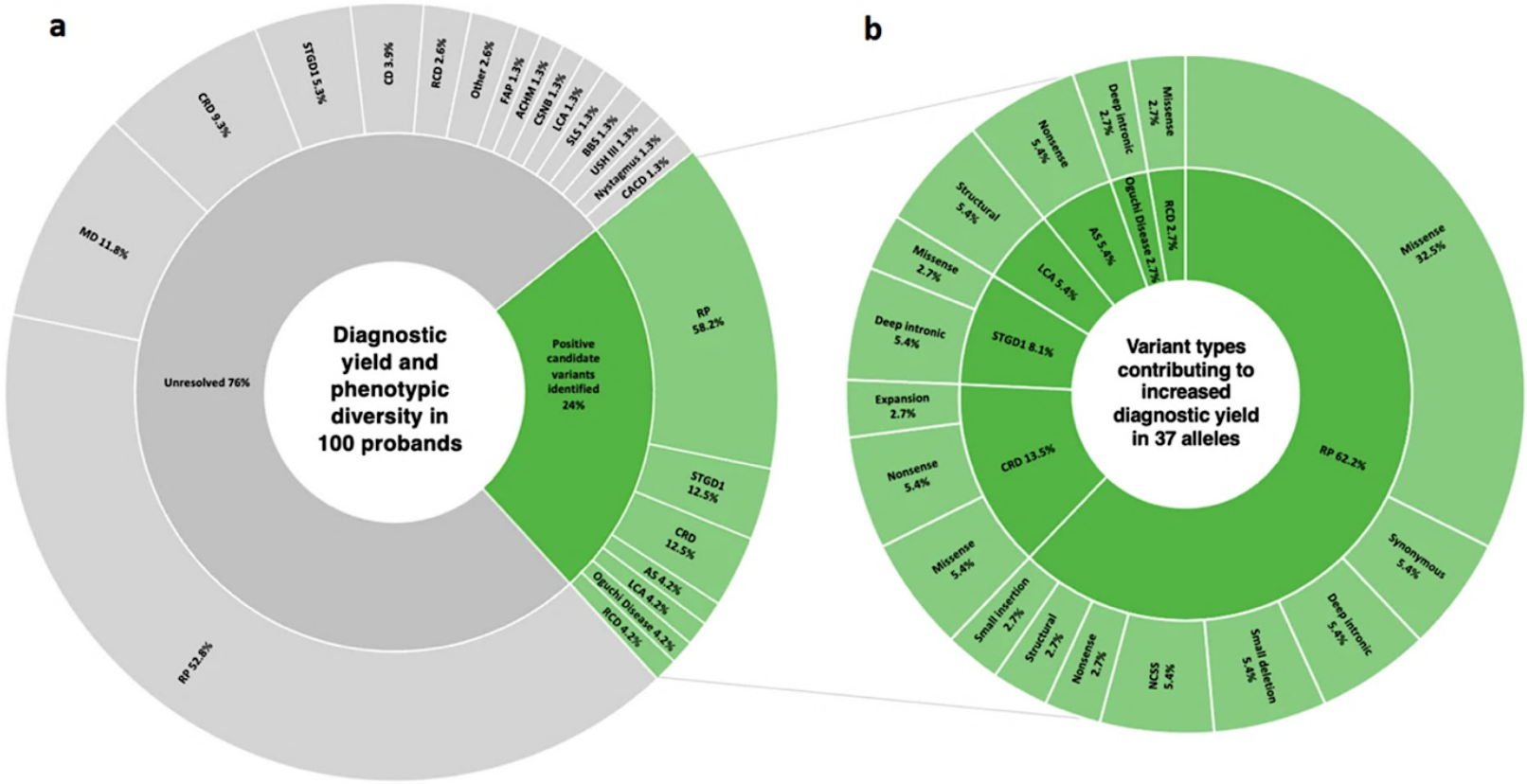


Figure 4.1 | Sunburst diagrams illustrating diagnostic yield, phenotypic diversity and variants contributing to increased diagnostic yield in this study. A Diagnostic yield and phenotypic diversity. The inner rings represent solved (green) and unresolved (gray) cases following whole-genome sequencing. The outer ring encompasses the phenotypic presentation of patients in these groups. B Variant types contributing to increased diagnostic yield. The types of variants contributing to increased diagnostic yield are illustrated on the outer ring, with their corresponding clinical phenotypes on the inner ring. ACHM achromatopsia, AS Alström syndrome, BBS Bardet–Biedl syndrome, CACD central areolar choroidal dystrophy, CD cone dystrophy, CRD cone–rod dystrophy, CSNB congenital stationary night blindness, FAP fundus albipunctatus, LCA Leber congenital amaurosis, MD macular dystrophy, NCSS non-CSS, RCD rod–cone dystrophy, RP retinitis pigmentosa, SLS Senior–Loken syndrome, STGD1 Stargardt disease type 1, USH III Usher syndrome type III. Small insertions include insertions <4-bp in length. Small deletions include deletions <4-bp in length.

4.3.2 Identification of coding and noncoding candidate variants after WGS data analysis

Following comprehensive WGS analysis and *in vitro* functional assays, we identified a genetic cause of disease in 24 probands (24%). Pathogenic variants were confirmed by Sanger sequencing and segregation analysis when family members were available (9/24) (Table 4.4 and Table 4.5). Among these, 15 probands (55.5%) carried previously undetected pathogenic variants in coding regions and three individuals (11.1%) carried SVs in *RPGRIP1* and *PRPF31*, or a triplet tandem repeat expansion in *ATXN7*. In proband Pt-4, presenting with CRD and progressive neurodegeneration (MIM: #164500), we identified a pathogenic CAG repeat expansion in exon 3 of *ATXN7* (NG_008227.1, p.(Gln30[70])) (Table 4.4 and Figure 4.2). While the wild-type allele carries 3–19 CAG repeats in *ATXN7*, the pathogenic allele can be between 37 and 460 repeats[207].

Table 4.4 | DNA notations for each variant found in 24 solved cases in this study. Transcript IDs are as follows: *ABCA4: NG_009073.1; ATXN7: NG_008227.1; C21ORF2: NG_032952.1; CDHR1: NG_028034.1; CEP 78: NG_053171.1; EYS: NG_023443.2; FAM161A: NG_028125.1; GRM6: NG_008105.1; HGSNAT: NG_009552.1; PCARE: NG_021427.1; RGS9BP: NG_016751.1; RLBP1: NG_008116.1; RPGRIP1: NG_008933.1; PRPF31: NG_009759.1; USH2A: NG_009497.2*. AS, Alström syndrome; CD, cone dystrophy; CRD, cone-rod dystrophy; LCA, Leber congenital amaurosis; MD, macular dystrophy; RCD, rod cone dystrophy; RP, retinitis pigmentosa; TCS, target capture sequencing; STGD1, Stargardt disease; WES, whole exome sequencing. – no second allele in case of dominant disease. *With neurodegeneration. gnomAD allele frequency (AF, accessed on April 27, 2021) is the total AF across all populations. An AF of zero indicates the variant is not identified among the tested alleles in gnomAD.

ID	Phenotype	Gene	g. position allele 1	DNA allele 1	gnomAD-AF	g. position allele 2	DNA allele 2	gnomAD-AF
Pt-1	STGD1	*ABCA4*	g.94471065G>A	c.6079C>T	0.0001980	g.94484001C>T	c.5196+1137G>A	0.00009558
Pt-2	STGD1	*ABCA4*	g.94461738A>G	c.6743T>c	0[#]	g.94492973G>A	c.4539+2028C> T	0.00006371
Pt-3	STGD1	*ABCA4*	g.94476467T>A	c.5603A>T	0.04219	g. 94577010T>G	c.286A>C	0.000003983
Pt-4	CRD*	*ATXN7*	g..63898362_63898391CAG[70]	c.88_117CAG[43]	0	-	-	n.a.
Pt-5	RP	*C21ORF2*	g.45753071C>G	c.218G>C	0.0003255	g.45759002A>G	c.76T>C	0
Pt-6	RP	*C21ORF2*	g.45753071C>G	c.218G>C	0.0003255	g. 45753071C>G	c.218G>C	0.0003255
Pt.7	RP	*CDHR1*	g.85962879G>A	c.783G>A	0.003052	g.85962879G>A	c.783G>A	0.003052
Pt-8	AS	*CEP78*	g.80863753G>T	c.1033G>T	0	g. 80863753G>T	c.1033G>T	0
Pt-9	RP	*EYS*	g.[64430522A>T, 65301854G>T]	e.[3906C>A;9468T>A]	0.00002564, 0.00002574	g.65229204T>C	c.5644+70912A>G	0
Pt-10	RP	*EYS*	g.66204881_66204901delinsAAAAG	c.403_423delinsCTTTT	0	g.6204765G>T	c.539C>A	0
Pt-11	CRD	*EYS*	g.64430522A>T	c.9468T>A	0.00002564	g.4430522A>T	c.9468T>A	0.00002564
Pt-12	RCD	*EYS*	g.66044979A>T	c.1660T>A	0.0003055	g.65300716C>A	c.5044G>T	0.003442
Pt-13	RP	*FAM161A*	g..62063240_62063243del	c.1753_1756del	0	g.62063240_6206_3243del	c.1753_1756del	0
Pt-14	Oguchi disease	*GRM6*	g.178413523G>A	e.1732C>T	0.001955	g.178414571dup	c.1355-587dup	0
Pt-15	RP	*HGSNAT*	g.43054647G>A	c.1843G>A	0.004030	g.43054647G>A	c.1843G>A	0.004030
Pt-16	RP	*HGSNAT*	g.43037305C>T	c.1030C>T	0.00001203	g.43054647G>A	c.1843G>A	0.004030
Pt-17	RP	*HGSNAT*	g.43054647G>A	c.1843G>A	0.004030	g..493+5G>A	c.493+5G>A	0.00002049

Pt-18	RP	*HGSNAT*	g.43054647G>A	c.1843G>A	0.004030	g..493+5G>A	c.1843G>A	0.004030
Pt-19	RP	*HGSNAT*	g.43052991C-T	c.1622C>T	0.0001319	g..43054647G>A	c.1843G>A	0.004030
Pt-20	RP	*PCARE*	g..29293524G>A	c.3604C>T	0.00000802 2	g..29294028_292 4029insCCTG	c.3099_3100ins CAGG	0.00001370
Pt-21	CRD	*RGS9BP*	g.33167752T>G	c.583T>G	0.0001314	g.33167752T>G	c.583T>G	0.0001314
Pt-22	LCA	*RPGRIP1*	g.21799664_2180 3187del	c.3238+1118_333 9+323del	0	g.21799664_2180 3187del	c.3238+1118_3 339+323del	0
Pt-23	RP	*PRPF31*	g.54106454- 54133135del	-	0	-	-	n.a.
Pt-24	RP	*USH2A*	.216270422T>C	c.4758+3A>G	0.001500	g..216523906C>A	c.784+14389G> T	0.0004781

Table 4.5 | Overview of pathogenic variants in genetically solved cases. Transcript IDs are as follows: *ABCA4: NG_009073.1; ATXN7: NG_008227.1; C21ORF2: NG_032952.1; CDHR1: NG_028034.1; CEP78: NG_053171.1; EYS: NG_023443.2; FAM161A: NG_028125.1; GRM6: NG_008105.1; HGSNAT: NG_009552.1; PCARE: NG_021427.1; RGS9BP: NG_016751.1; RPGRIP1: NG_008933.1; PRPF31: NG_009759.1; USH2A: NG_009497.2.* AS Alström syndrome, CD cone dystrophy, CRD cone–rod dystrophy, LCA Leber congenital amaurosis, RCD rod–cone dystrophy, RP retinitis pigmentosa, TCS target-capture sequencing, WES whole-exome sequencing. A = With neurodegeneration. B = A first pathogenic allele was identified in previous genetic testing. C = This patient is assigned as likely solved.

ID	Country where recruited	Phenotype	Gene	DNA allele 1	Protein allele 1	DNA allele 2	Protein allele 2	Previous screening
Pt-1	Ireland	STGD1	*ABCA4*	c.6079C>T[b]	p.(Leu2027Phe)	c.5196+1137G>A	p.[=,Met1733Glufs*78]	TCS
Pt-2	Ireland	STGD1	*ABCA4*	c.6743T>C[b]	p.(Phe2248Ser)	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	TCS
Pt-3	Ireland	STGD1	*ABCA4*	c.5603A>T[b]	p.(Asn1868Ile)	c.286A>C	p.(Asn96His)	TCS
Pt-4	Netherlands	CRD[a]	*ATXN7*	c.88_117CAG[70]	p.(Gln30[70])	–	–	WES
Pt-5	Ireland	RP	*C21ORF2*	c.218G>C	p.(Arg73Pro)	c.76T>C	p.(Trp26Arg)	TCS
Pt-6	Ireland	RP	*C21ORF2*	c.218G>C	p.(Arg73Pro)	c.218G>C	p.(Arg73Pro)	TCS
Pt-7	Ireland	RP	*CDHR1*	c.783G>A	p.Asp214_Pro261del	c.783G>A	p.Asp214_Pro261del	TCS
Pt-8	Israel	AS	*CEP78*	c.1033G>T	p.(Gly345*)	c.1033G>T	p.(Gly345*)	TCS
Pt-9	Netherlands	RP	*EYS*	c.[3906C>A;9468T>A]b	p.[Tyr3156*;His1302Gln]	c.5644+70912A>G	p.Asp1882Glyfs*18	WES
Pt-10	Israel	RP	*EYS*	c.403_423delinsCTTTT[b]	p.(Thr135Argfs*26)	c.539C>A	p.(Ser180*)	TCS
Pt-11	Israel	CRD	*EYS*	c.9468T>A	p.(Tyr3156*)	c.9468T>A	p.(Tyr3156*)	WES
Pt-12[c]	Netherlands	RCD	*EYS*	c.1660T>A[b]	p.(Cys554Ser)	c.5044G>T	p.(Asp1682Tyr)	WES
Pt-13	Israel	RP	*FAM161A*	c.1753_1756del	p.(Lys585Alafs*6)	c.1753_1756del	p.(Lys585Alafs*6)	TCS

Pt-14	Netherlands	Oguchi disease	*GRM6*	c.1732C>T[b]	p.(Arg578Cys)	c.1355-587dup	p.[Asn451_Ala452ins25, Gly385Alafs*42,=]	WES
Pt-15	Netherlands	RP	*HGSNAT*	c.1843G>A	p.Ala615Thr	c.1843G>A	p.Ala615Thr	WES
Pt-16	Netherlands	RP	*HGSNAT*	c.1030C>T	p.(Arg344Cys)	c.1843G>A	p.Ala615Thr	WES
Pt-17	Netherlands	RP	*HGSNAT*	c.1843G>A[b]	p.Ala615Thr	c.493+5G>A	p.Arg124Serfs*25	WES
Pt-18	Netherlands	RP	*HGSNAT*	c.1843G>A	p.Ala615Thr	c.1843G>A	p.Ala615Thr	WES
Pt-19	Ireland	RP	*HGSNAT*	c.1622C>T	p.(Ser541Leu)	c.1843G>A	p.Ala615Thr	TCS
Pt-20	Ireland	RP	*PCARE*	c.3604C>T[b]	p.(Arg1202*)	c.3099_3100insCAGG	p.(Val1034Glnfs*74)	TCS
Pt-21	Israel	CRD	*RGS9BP*	c.583T>G	p.(Ser195Ala)	c.583T>G	p.(Ser195Ala)	WES
Pt-22	Netherlands	LCA	*RPGRIP1*	c.3238+1118_3339+323del	p.(Asp1080Glyfs*6)	c.3238+1118_3339+323del	p.Asp1080Glyfs*6	WES
Pt-23	Ireland	RP	*PRPF31*	c.-396_*287(0) (g.54106454-54133135del)	p.(0)	–	–	TCS
Pt-24	Israel	RP	USH2A	c.4758+3A>G	p.Gln1586_Gly1587ins*5	c.784+14389G>T	p.Gly262Aspfs*26	TCS

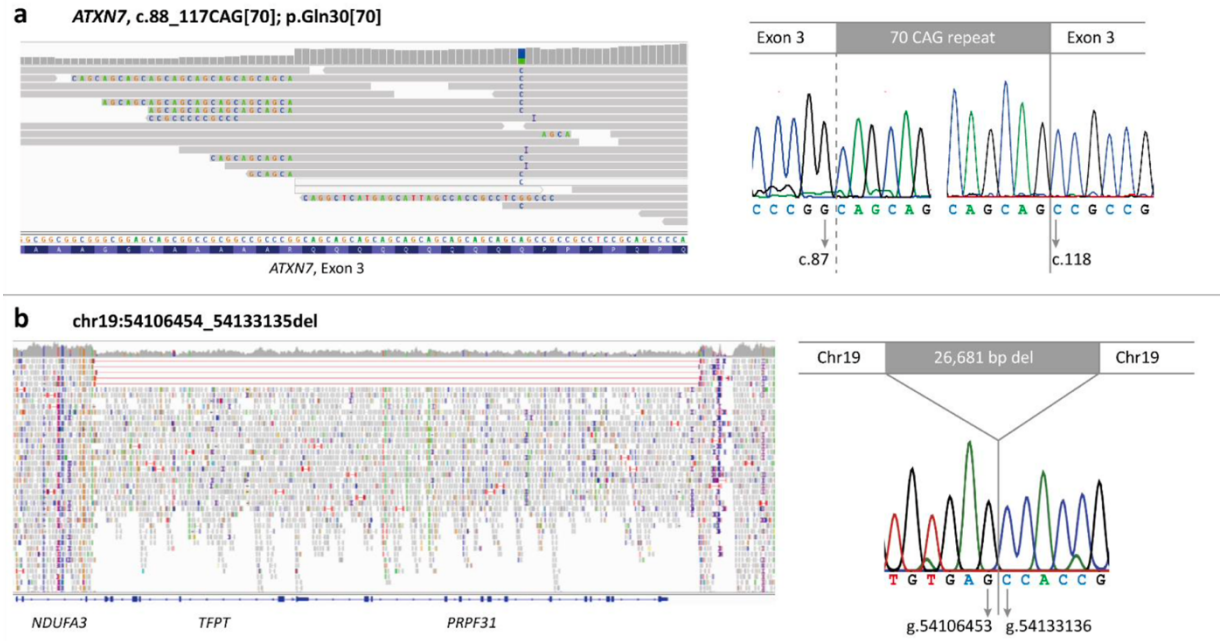


Figure 4.2 | Schematic representation of IGV software for pathogenic structural variants in this study. A) The CAG expansion repeat in *ATXN7* exon 3 is depicted in soft-clipped reads. We observed 70 repeats in the mutant allele, while the normal allele does not exceed 19 repeats. B) An ~26-kb pathogenic deletion on chromosome 19 detected by Control-FREEC, found in one proband is shown. Red lines across the region represent reads where the patients insert size is smaller than that of the reference, indicative of a deletion. Similarly, the depth of coverage above the reads appears to decrease by approximately half by comparison with the surrounding sequences, pointing towards a heterozygous deletion. A cross-deletion PCR was performed that yielded a 700-bp fragment, which was subsequently Sanger sequenced (right) to confirm the breakpoints at the nucleotide level. The deletion encompasses *PRPF31, TFPT*, and the promoter of *NDUFA3*.

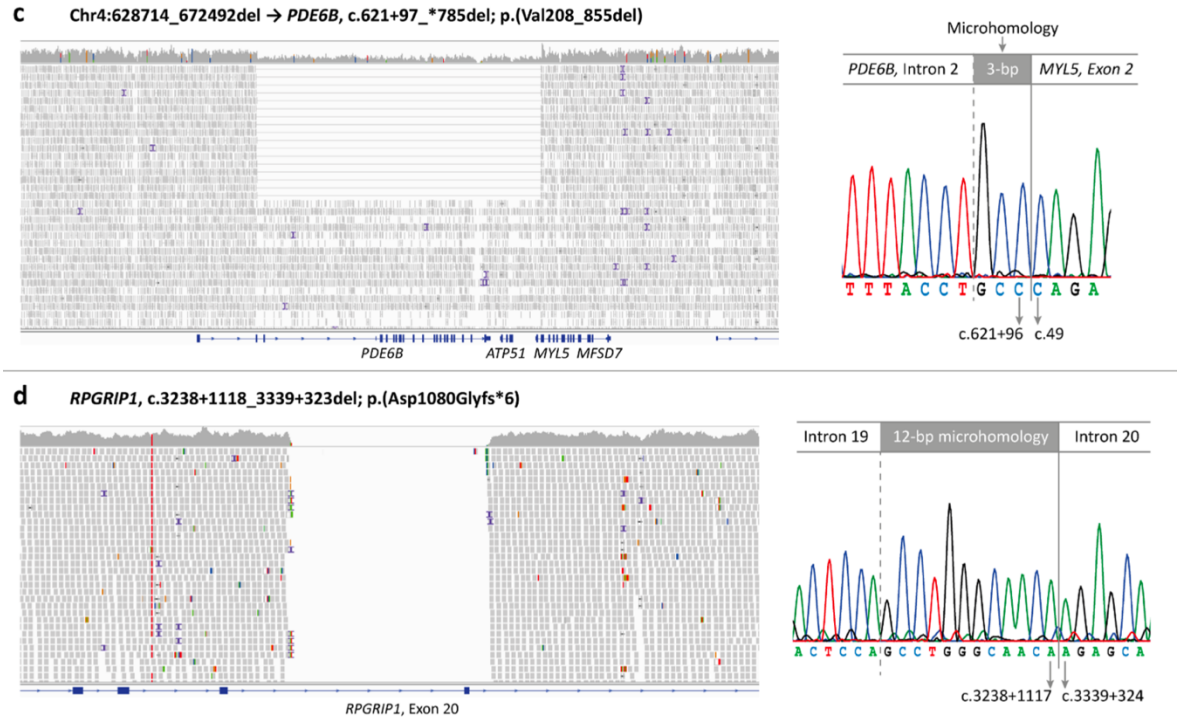


Figure 4.2. | continued: c) The heterozygous deletion in *PDE6B* was detected by Control-FREEC tool in this affected individual. Gray lines across the region represent reads where the patient insert size is smaller than that the reference, indicative of a deletion. The Sanger sequence analysis confirmed the deletion at the nucleotide level. D) We identified the homozygous deletion in the proband which led to the deletion of exon 20 of *RPGRIP1*. The SV breakpoints were confirmed by Sanger sequencing shown in the right part. In *PDE6B* and *RPGRIP1* SVs, the 3-bp and 12-bp microhomologies at the breakpoints were indicated in the grey boxes.

One proband (Pt-22) clearly illustrates the strength of WGS for SV detection compared to WES. The patient was diagnosed with LCA (MIM: #605446), with one pathogenic allele in *CEP290* (p.(Arg360Gln)), *PDE6A* (p.(Val685Met)), and *RPE65* (p.(Val226Ile)) (Table 4.2). However, WGS data revealed—and segregation analysis confirmed—a homozygous *RPGRIP1* exon 20 deletion (p.(Asp1080Glyfs*6)), missed previously due to incapability of WES analysis tools to detect CNVs using the SureSelect XT Human All Exon V4 (SOLiD 5500xl sequencer in 2012). While we cannot rule out that a more recent WES strategy would have detected this single exon deletion, this approach would not have revealed the breakpoints. A 12-bp microhomology region was identified in the breakpoint boundaries (Table 4.5 and Figure 4.2D). Completing the variant analysis, the noncoding regions of *CEP290, RPE65*, and *PDE6A* did not contain additional potential pathogenic variants. Another proband (Pt-23) had a 26.68-kb heterozygous deletion on chromosome 19, encompassing *PRPF31, TFPT*, and the promoter region of *NDUFA3*. This large deletion was not detected previously using TCS assessment (Table 4.4 and Figure 4.2).

Furthermore, we prioritized ten DIVs and three NCSS variants in 14 individuals using criteria in Alamut and SpliceAI (Tables 4.6 and 4.7) outlined in the variant frequency and pathogenicity prediction parameters section of the methods. For the indel variants detected in *ABCA4* (in Pt-28, Pt-29, and Pt-30), *CYP4V2* (in Pt-52), and RLBP1 (in Pt-65), no SpliceAI scores were provided due to the nature of these variants. Therefore, inclusion of these three variants for *in vitro* splice assays was based on Alamut prediction scores only. The pathogenicity of noncoding variants was determined in seven cases (46.6%) after *in vitro* splice assays (see below) (Table 4.4 and Figure 4.1). While employing WGS, significant progress was made in defining the genetic pathogenesis of IRDs in this prescreened IRD cohort; 76% of cases remained genetically unexplained and clearly for such cases additional levels of interrogation of sequence data will be required in the future.

Table 4.6 | Wild-type and mutant *in silico* prediction scores embedded in Alamut for candidate variants. SpliceSiteFinder-Like (0-100); MaxEntScan (0-12 for SDSs_0-16 for SASs); NNSPLICE (0-1); GeneSplicer (0-24 for SDSs_0-21 for SASs); HSF, Human Splicing Finder (0-100).

Gene	DNA-variant	c. position splice site	SSFL			MaxEntScan			NN SPLICE			Gene Splicer			HSF		
			WT	Mut	Delta %	WT	Mut	Delta %	WT	Mut	Delta %	WT	Mut	Delta %	WT	Mut	Delta %
ABCA4	c.6148-89G>A	c.6148-90	68	79	-12	-	-	-	-	-	-	-	-	-	-	-	-
ABCA4	c.5460+1315_5460+1317delinsTA	c.5460+1315	71	73	-2	-	3	-23	-	0	-10	-	-	-	-	78	-78
CACNA2D4	c.2009-3C>T	c.2009-1	83	76	9	6	7	-3	0	0	0	2	1	4	91	83	8
CYP4V2	c.214+879_214+882delinsG	c.214+881	-	75	-75	-	5	-45	-	0	-30	-	-	-	-	77	-77
EYS	c.5644+70912A>G	c.5644+70914	86	96	-11	7	12	-33	1	1	-10	6	8	-6	86	97	-11
GRM6	c.1355-587dup	c.1355-589	-	78	-78	6	7	-15	-	1	-80	3	4	-3	73	82	-9
HGSNAT	c.493+5G>A	c.493+1	63	51	12	6	-	53	1	0	80	1	-	3	-	-	-
PDE6B	c.469-776C>G	c.469-775	29	88	-59	-	8	-70	-	1	-100	-	8	-32	-	95	-95
RLBP1	c.525+425_525+433delinsATA	c.525+436	58	67	-9	1	1	-7	0	0	-10	-	-	-	71	76	-5
USH2A	c.784+14389G>T	c.784+14397	88	93	-5	9	11	-14	1	1	-30	2	5	-15	87	88	-2

Table 4.7. | SpliceAI scores of candidate variants. The positions are based on the distance from the variant. DS, delta score; AG, acceptor gain; AL, acceptor loss; DG, donor gain; DL, donor loss; DP, delta position. 500 bp up- and downstream of a variant were included as an input sequence.

Gene	DNA allele	DS_AG	DS_AL	DS_DG	DS_DL	DP_AG	DP_AL	DP_DG	DP_DL
ABCA4	c.6148-89G>A	0	0	0.12	0.23	3	1	3	1
ABCA4	c.5460+1315_5460+1317delinsTA	-	-	-	-	-	-	-	-
CYP4V2	c.214+879_214+882delinsG	-	-	-	-	-	-	-	-
EYS	c.5644+70912A>G	0.6	0.05	0.66	0	-1	-2	-104	-2
GRM6	c.1355-587dup	0.05	0	0	0	-49	-66	3	-70
HGSNAT	c.493+5G>A	0	0.22	0.6	0.74	-474	-126	-14	-5
PDE6B	c.469-776C>G	0	0	0.05	0.01	-207	-405	-1	-97
RLBP1	c.525+425_525+433delinsATA	-	-	-	-	-	-	-	-
USH2A	c.784+14389G>T	0.03	0	0.12	0	-8	-132	-104	-200

4.3.3 Splice defects due to NCSS variants in *CDHR1, HGSNAT,* and *USH2A*

In Pt-7, we identified a homozygous synonymous variant in the last nucleotide of *CDHR1* exon 8, c.783G>A. In 2017, Stingl *et al*., reported that a patient was homozygous for the same variant. A midigene splice assay encompassing exons 7, 8, and 9 demonstrated that this variant causes in-frame skipping of exon 8 (p.Asp214_Pro261del)[208]. Likewise, a similar defect was identified in messenger RNA (mRNA) from the retina of an individual heterozygous for this variant. The authors suggest that this variant is associated with a relatively mild form of IRD, based on clinical examination of a patient homozygous for this variant. Similarly, our patient displays a mild phenotype and late age-at-onset as previously described[208].

In Pt-17, a pathogenic missense variant (c.1843G>A) was identified in *HGSNAT* as the first allele. Through WGS analysis, a NCSS variant, c.493+5G>A, was identified in the same gene, resulting in exon 4 skipping (p.Arg124Serfs*25) as evaluated using a midigene splice assay spanning exons 3, 4, and 5 (Figure 4.3a).

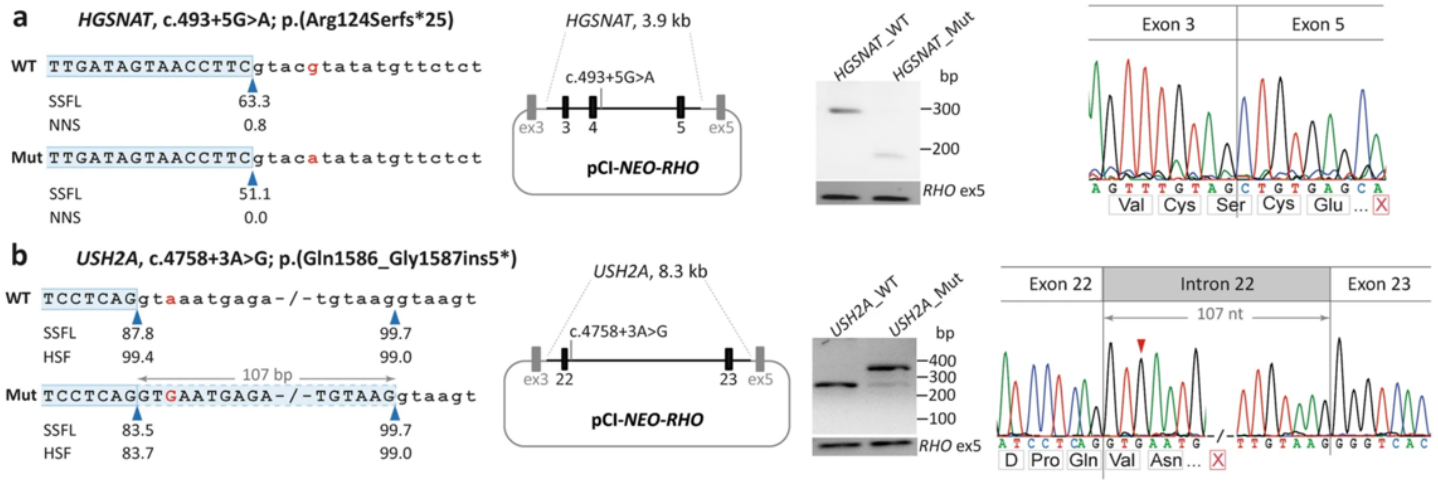


Figure 4.3 | Splice defects due to non-CSS variants. For *HGSNAT* c.493+5G > A (a) and *USH2A*, c.4758+3A>G (b), the left panels are schematic representations of the gene segments of interest containing the corresponding variants in the wild-type (WT) and mutant (Mut) sequences. The SpliceSiteFinder-like (SSFL, range 0–100), and Human Splicing Finder (HSF, range 0–100) or NNSPLICE (NNS, range 0–1) scores for the splice sites are indicated below each sequence. The blue triangles indicate the position of splice donor sites in both wild-type and mutant sequences. The red highlighted nucleotides indicate the variants in the mutant sequence. The second panels show the mutant pCI-*NEO-RHO* vectors. They contain the variants and flanking exons 3, 4, and 5 in *HGSNAT* and exons 22 and 23 in *USH2A* flanked by *RHO* exons 3 and 5. Wild-type and mutant constructs were used to transfect HEK293T cells. The next panels show the gel images of RT-PCR products in wild-type and mutant midigenes. RT-PCR analysis of *RHO* exon 5 was performed as a control for transfection efficiencies. In the last panels, the Sanger sequence analysis of the RT-PCR fragments confirmed the exon 4 skipping in *HGSNAT*, c.493+5G>A (a) and 107-nt exon elongation in *USH2A* c.4758+3A>G (b). The red "X" in a red box indicates the stop codon in each schematic chromatogram. For each variant, both WT and mutant variants were examined in parallel in the same experiment and processed in parallel.

An individual (Pt-24) with autosomal recessive RP carried a heterozygous variant, c.4758+3A>G in *USH2A* together with a DIV, c.784+14389G>T, in the same gene. The NCSS variant was found to be pathogenic after performing a midigene splice assay showing exon elongation (p.Gln1586_Gly1587ins5*) (Figure 4.3b). Two fragments were observed upon transfection of the mutant midigene in HEK293T cells. After quantification, we observed 78.9% of a larger fragment (332 nt) corresponding to 107-bp elongation of exon 22 and 21.1% of the wild-type fragment (225 nt) (Table 4.8).

4.3.4 Splice defects due to deep-intronic variants in *ABCA4, EYS, GRM6*, and *USH2A*

In aforementioned Pt-24, after confirming the pathogenicity of the c.4758+3A>G NCSS variant in *USH2A*, an extensive analysis of the intronic regions of this gene was performed and a DIV, c.784+14389G>T, was detected *in trans* in *USH2A* that led to a pseudoexon insertion (p.Gly262Aspfs*26) (Figure 4.4c). Using the mutant construct for midigene analysis, 77.1% of the mutant fragment (371 nt) corresponding to the presence of a pseudoexon in mature mRNA was observed and 22.9% of the wild-type fragment (274 nt) (Table 4.8). This patient also carried a frameshift variant in *PCARE* (p.(Lys919Thrfs*2)), which was detected previously by TCS. However, WGS analysis did not reveal any plausible second candidate variant in intronic regions of the *PCARE* gene (Table 4.2).

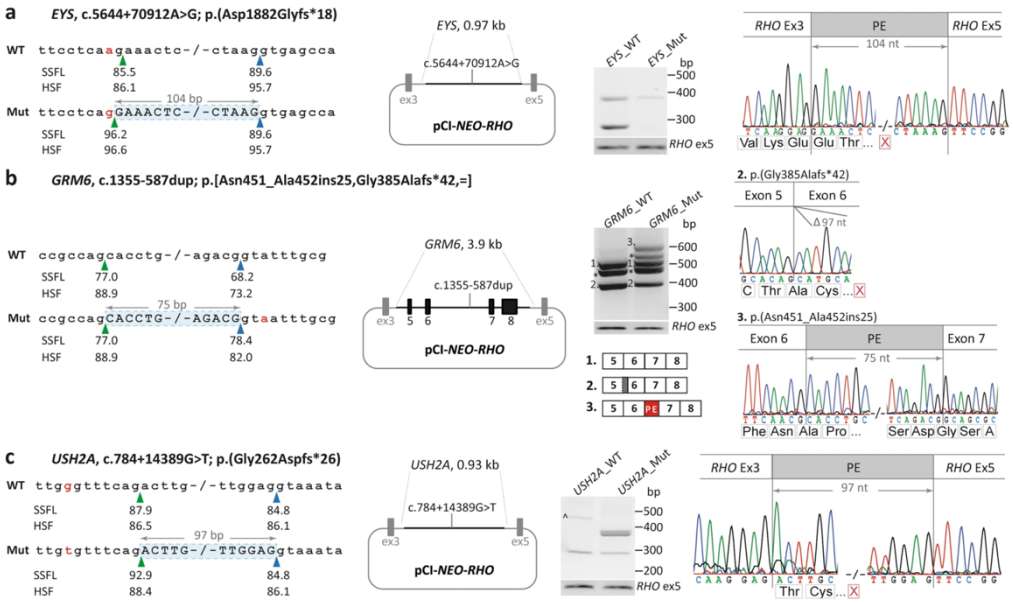


Figure 4.4 | Splice defects due to deep-intronic variants. For each variant, i.e., a *EYS*, c.5644+70912A>G, b *GRM6*, c.1355-587dup, and c *USH2A*, c.784+14389G>T, the left panels are schematic representations of the gene segments of interest containing the corresponding variants in the wild-type (WT) and mutant (Mut) sequences. The SpliceSiteFinder-like (SSFL, range 0–100), and Human Splicing Finder (HSF, range 0–100) scores for the splice sites are indicated below each sequence. The green and blue triangles indicate the position of splice acceptor sites and splice donor sites in both wild-type and mutant sequences. The red highlighted nucleotides indicate the variants in the mutant sequence. The second panels show the mutant pCI-*NEO-RHO* vectors. They contain the variants and flanking exons (except for *EYS*, as c.5644+70912A>G is in a minigene construct) flanked by *RHO* exons 3 and 5. Wild-type and mutant constructs were used to transfect HEK293T cells. The next panels show the gel images of RT-PCR products in wild-type and mutant midi- or minigenes. RT-PCR analysis of *RHO* exon 5 was performed as a control for transfection efficiencies. For the *GRM6* c.1355-587dup variant, the asterisks in the gel image indicate the heteroduplex fragments. For the *USH2A* c.784+14389G>T variant, the "^" in the wild-type construct indicates an artifact product. In the last panels, the Sanger sequence analysis of the RT-PCR fragments confirmed the predicted splice defects for each variant. For *GRM6* c.1355-587dup, the sequence of fragment 2 in both midigenes showed the 97-bp deletion of exon 6, which led to the frameshift variant. The third fragment was only observed in the mutant construct and corresponded to the insertion of an in-frame pseudoexon in intron 6. The red "X" in a red box indicates the stop codon in each schematic chromatogram. For each variant, both WT and mutant variants were examined in parallel in the same experiment and processed in parallel.

Two STGD1 cases (Pt-1 and Pt-2) carried c.6079C>T and c.674T>C missense variants, individually, as pathogenic alleles (Table 4.2). Upon WGS analysis, we identified previously published pathogenic deep-intronic *ABCA4* variants in these probands, c.4539+2028C>T and c.5196+1137G>A, respectively[209]. The c.4539+2028C>T variant did not show a splice defect in fibroblast cells, but revealed a retina-specific 345-nt pseudoexon insertion in a proportion of the mRNA (p.[=,Arg1514Leufs*36]) extracted from patient-derived PPCs. The variant c.5196+1137G>A showed a 73-nt pseudoexon insertion in a small fraction of mRNA in a HEK293T cell–midigene-based splice assay. This effect was stronger in patient-derived PPCs (p.[=,Met1733Glufs*78])[210] Based on genotype–phenotype correlations, variants c.4539+2028C>T and c.5196+1137G>A were deemed moderately severe[209,210].

EYS c.5644+70912A>G (NG_023443.2) was identified in Pt-9, who carried c.[3906C>A;9468T>A] (p.[His1302Gln;Tyr3156*]) in *EYS* as the first pathogenic allele for which the truncating variant is mostly reported as c.9405T>A (p.(Tyr3135*) (Table 4.2). Due to the large size of intron 26 (150.8 kb), c.5644+70912A>G was tested in a minigene with a small genomic DNA insert (0.97 kb; Figure 4.4a). Reverse transcription-polymerase chain reaction (RT-PCR) analysis of the mRNA resulting from the wild-type *EYS* minigene revealed the expected wild-type fragment of 274 nt (76.0% of the total product) and a larger fragment of 378 nt (24.0% of the total product) corresponding to the presence of a 104-nt pseudoexon, which leads to frameshift variant after 18 amino acids (p.Asp1882Glyfs*18). mRNA products from the mutant *EYS* c.5644+70912A>G minigene only contained the larger fragment (378 nt), with no remaining wild-type mRNA (Figure 4.4a and Table 4.8).

Table 4.8 | The percentage of each fragment in wild-type and mutant midigenes for candidate variants with multiple mRNA molecules in RT-PCR analysis. # The size of the wild-type fragment in each variant was utilized as the default for the correction factor. To do that, we divided each fragment size to the corresponding wild-type fragment. WT, wild-type; Mut, mutant; bp, base pair.

	Raw intensity	Fragment size (bp)	Correction factor[#]	Corrected intensity	Percentage
***EYS*_c.5644+70912A>G; p.(Asp1882Glyfs*18)**					
EYS_WT_ larger fragment	21989	378	1.38	15939	24%
EYS_WT_ smaller fragment	51722	274	1	51722	76%
EYS Mut larger fragment	22284	378	1	16153	100%
***GRM6*_c.1355-587dup; p.[Asn451_Ala452ins25,Gly385Alafs*42,=]**					
GRM6_ WT_ larger fragment	60863	492	1	60863	46%
GRM6_WT_ smaller fragment	53111	359	1	72787	54%
GRM6_Mut_ larger fragment	33375	567	1	28960	21%
GRM6_Mut_ middle fragment	59069	492	1	59069	42%
GRM6_Mut_ smaller fragment	41692	395	1	51930	37%
***USH2A*_c.784+14389G>T; p.(Gly262Aspfs*26)**					
USH2A_WT fragment	3501	274	1	3501	100%
USH2A_Mut_ larger fragment	11286	371	1	8335	77%
USH2A Mut smaller fragment	2483	274	1	2483	23%
***USH2A*_c.4758+3A>G; p.(Gln1586_Gly1587ins5*)**					
USH2A_WT fragment	26777	225	1	26777	100%
USH2A_Mut_ larger fragment	28628	332	1	19402	79%
USH2A_Mut_ smaller fragment	5172	225	1	5172	21%

In Pt-14, who was monoallelic for *GRM6*, c.1732C>T (Table 4.2), *GRM6* c.1355-587du, p.[Asn451_Ala452ins25,Gly385Alafs*42,=] was identified, which resulted in the insertion of an in-frame pseudoexon encoding 25 amino acids (p.Asn451_Ala452ins25) (Figure 4.4b). Surprisingly, a 97-nt partial deletion of exon 6 (359 nt) in both wild-type and mutant midigene assays was observed, which was calculated to be 54.4% and 37.0% of total RNA molecules, respectively. In the *GRM6* mutant midigene assay, the 567-nt fragment corresponding to the presence of a pseudoexon was quantified as 21.0% of total mRNA molecules. The percentage of full-length wild-type fragments (492 nt) for wild-type and mutant *GRM6* midigene assays were 45.6% and 42.0%, respectively. Therefore, we normalized the percentage of wild-type fragment in the assays using the mutant construct (42.0%) to the wild-type construct (46.0%). The mutant trascripts produced as a result of *GRM6* c.1355-587dup was observed alongside 91.0% of the wild-type fragment (42/46 × 100) (Table 4.8).

Finally, five noncoding variants in seven cases did not show any splice defects after mini- or midigene *in vitro* splice assays (Figure 4.5A–E): *ABCA4*, c.6148-89G>A in Pt-27; *ABCA4*, c.5460+1315_5460+1317delinsTA in three cases (Pt-28–30); *CYP4V2*, c.214+879_214+882delinsG in Pt-52; *PDE6B* c. 469-776C>G in Pt-61; *RLBP1* c.525+425_525+433delinsATA in Pt-65. These patients were included in this study as monoallelic cases having the first pathogenic allele in *ABCA4*, *CYP4V2, PDE6B*, and *RLBP1*, respectively (Table 4.2). However, all seven probands remained genetically unresolved given that a potential pathogenic second allele was not identified. Raw gel electrophoresis images of all midigene assays can be found in Figure 4.6. Of note, midigene assays were solely undertaken in HEK293T cells and therefore it may be that splice defects would be observed if a more relevant cell type was utilised.

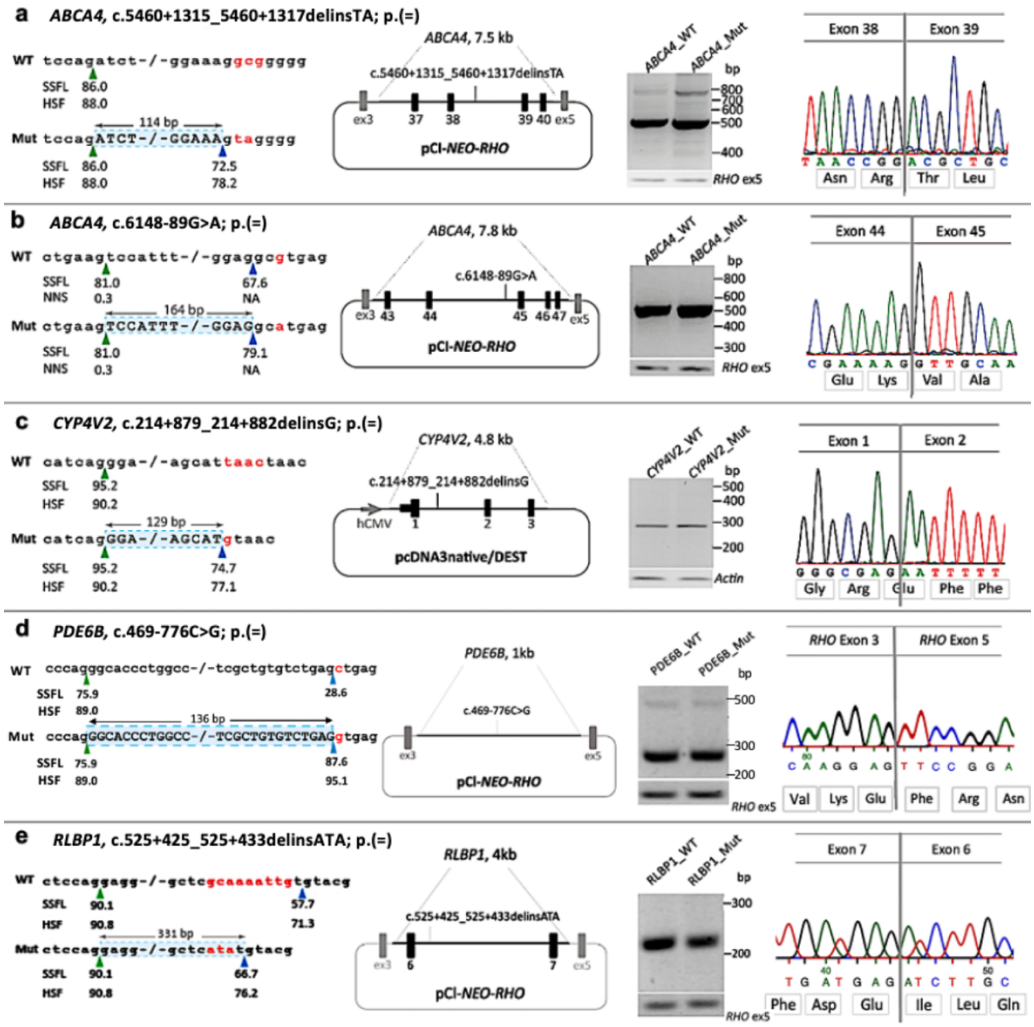


Figure 4.5 | Molecular genetic characterization of non-coding variants with no splice defects. For each variant, i.e. *ABCA4* 5460+1315_5460+1317delinsTA (a), *ABCA4*, c.6148-89G>A (b) *CYP4V2* c.214+879_214+882delinsG (c) *PDE6B* c.469-776C>G (d) and *RLBP1* c.525+425_525+433delinsATA (e) the left panels show the schematic representation of the gene of interest in the position of the corresponding variant in wild-type (WT) and in the presence of the variant (Mut). The SpliceSiteFinder-like (SSFL, range 0–100), NNSPLICE (NNS, range 0-1), and Human Splicing Finder (HSF, range 0–100) scores for the splice sites are indicated below each sequence. The green and blue triangles indicate the position of splice acceptor sites and splice donor sites in both wild-type and mutant sequences. The red highlighted nucleotides indicate the variants in the mutant sequence. The second panels showed the schematic representation of the mutant pCI-*NEO-RHO*.

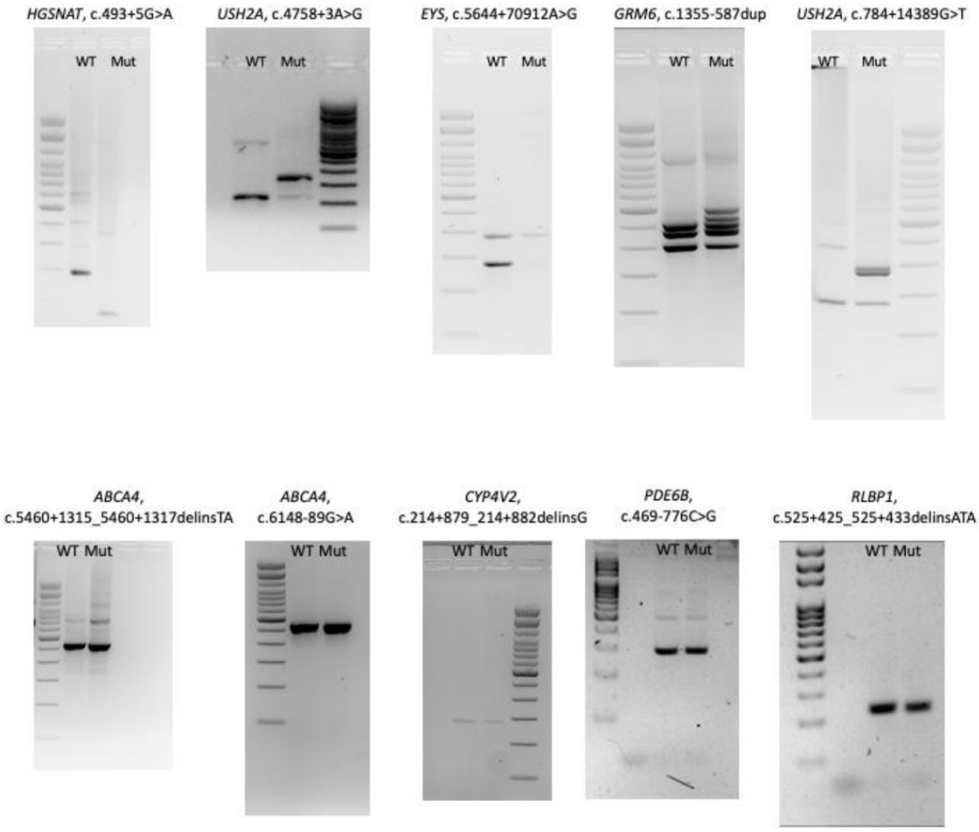


Figure 4.6 | Raw electrophoresis gel images obtained in this study. These are included for transparency purposes as the images are cropped in the above graphics. All gels, including those that illustrate defects and those that do not are included. The ladder for *HGSNAT – CYP4V2* corresponds to a 100bp ladder starting at 100bp and increasing in 100 nt increments to 1000bp. The ladder for *PDE6B* and *RLBP1* corresponds to a 100bp ladder starting at 100bp and increasing in 100 nt increments to 1000bp as well as 1500bp, 2000bp and 3000bp bands.

Table 4.9 | The allele frequency of the European (non-Finnish) population, the allele count and number of homozygous individuals in the pathogenic variants identified in 24 solved cases. AF, allele frequency. Allele frequency of zero indicates the variant is not reported in the database.

ID	Gene	DNA allele 1	gnomAD-AF (European (non-Finnish))	Total allele count	Total number of homozygotes	DNA allele 2	gnomAD-AF (European (non-Finnish))	Total allele count	Total number of homozygotes
Pt-1	*ABCA4*	c.6079C>T	0	56	0	c.5196+1137G>A	0	3	0
Pt-2	*ABCA4*	c.6743T>C	0	0	0	c.4539+2028C>T	0	2	0
Pt-3	*ABCA4*	c.5603A>T	0	11928	364	c.286A>C	0	1	0
Pt-4	*ATXN7*	c.88 117CAG(43]	0	0	0		n.a.	n.a.	n.a.
Pt-5	*C21ORF2*	c.218G>C	0	90	0	c.76T>C	0	0	0
Pt-6	*C21ORF2*	c.218G>C	0	90	0	c.218G>C	0	90	0
Pt-7	*CDHR1*	c.783G>A	0	863	4	c.783G>A	0	863	4
Pt-8	*CEP78*	c.1033G>T	0	0	0	c.1033G>T	0	0	0
Pt-9	*EYS*	c.(3906C>A;9468T >A]	0.00003077, 0.00006172	4 4	0 0	c.5644+70912A>G	0	0	0
Pt-10	*EYS*	c.403 423delinsC TTTT	0	0	0	c.539C>A	0	0	0
Pt-11	*EYS*	c.9468T>A	0	4	0	c.9468T>A	0	4	0
Pt-12	*EYS*	c.1660T>A	0	86	1	c.5044G>T	0	631	9
Pt-13	*FAM161A*	c.1753_1756de	0	0	0	c.1753_1756del	0	0	0
Pt-14	*GRM6*	c.1732C>T	0	532	1	c.1355-587dup	0	0	0
Pt-15	*HGSNAT*	c.1843G>A	0	1119	4	c.1843G>A	0	1119	4
Pt-16	*HGSNAT*	c.1030C>T	0	3	0	c.1843G>A	0	1119	4
Pt-17	*HGSNAT*	c.1843G>A	0	1119	4	c.493+5G>A	0	5	0
Pt-18	*HGSNAT*	c.1843G>A	0	1119	4	c.1843G>A	0	1119	4
Pt-19	*HGSNAT*	c.1622C>T	0	37	0	c.1843G>A	0	1119	4

Pt-20	*PCARE*	c.3604C>T	0	2	0	c.3099_3100InsCAGG	0	3	0
Pt-21	*RGS9BP*	c.583T>G	0	31	0	c.583T>G	0	31	0
Pt-22	*RPGRIP1*	e.3238+1118_3339+323del	0	0	0	c.3238+1118_3339+323del	0	0	0
Pt-23	*PRPF31*	_	0	0	0	_	n.a.	n.a.	n.a.
Pt-24	*USH2A*	c.4758+3A>G	0	422	5	c.784+14389G>T	0	15	0

4.4 Discussion

The sophistication of sequencing methods employed for the detection of disease-causing variants in the human genome has increased rapidly over the past decade, from targeted approaches such as panel sequencing to more comprehensive diagnostic methods such as WES and WGS. TCS and WES have undoubtedly played a valuable role in the genetic diagnosis of IRDs. For genetically undiagnosed cases following TCS and WES, it has become clear that more comprehensive sequencing procedures are required. Importantly, WGS facilitates the identification of noncoding variants, now widely considered a prominent cause of a wide spectrum of IRDs[20,52,142,211–214]. In order to investigate these possible advantages, we performed WGS on 100 probands with clinical indications of IRD who had received inconclusive results from TCS and WES previously.

This study cohort consisted of 56 monoallelic individuals who carried one candidate pathogenic variant and 44 individuals with no candidate pathogenic variants previously identified in an IRD-associated gene via WES or TCS (Table 4.2). Following extensive WGS data analysis and employment of *in vitro* splice assays, we identified the genetic cause of disease in 24 cases. These findings demonstrate that WGS is a more comprehensive sequencing method than TCS and WES for IRDs. As previously described in the literature, we have also observed a higher detection rate of pathogenic variants in WGS compared to WES and TCS in the coding and NCSS regions[63,184]. For instance, in seven of 24 solved cases (29.2%), one or both pathogenic alleles in the coding regions of *EYS* and *HGSNAT* were not detected or remained unrecognized as pathogenic in the previous genetic testing methods. This is due to incomplete gene panels (previous absence of *HGSNAT*), flaws in the annotation pipeline, or a lack of evidence or knowledge regarding the implication of these variants as IRD-associated at the time of WES or TCS data analysis. In one case (Pt-12), we have identified a missense variant in *EYS*, c.5044G>T, which was assigned previously as a "variant with unknown significance"[215]. However, due the strong *in silico* predictions for this missense variant and the highly conserved amino acid and nucleotide, we categorized this patient as likely solved. We have identified the hypomorphic allele p.Ala615Thr in *HGSNAT* in two homozygous (Pt-15 and Pt-18) and three heterozygous (Pt-16, Pt-17, and Pt-19) patients. *HGSNAT* mutations are cauastive of Mucopolysaccharidosis type IIIC, however, mutations in this gene can also cause non-syndromic forms of retinal degeneration[216]. After extensive analysis of IRD and ciliopathies-associated genes as well as the complete sequence of *HGSNAT*, no other candidate variant was identified in these probands. Recently, Schiff *et al*., investigated the enzymatic activity of the *HGSNAT* protein and showed the decreased

enzyme activity and urinary GAG/creatinine ratio in patients homozygous for p.Ala615Thr. They suggested that this allele is associated with non-syndromic retinal disease with the influence of *trans*-acting genetic and/or environmental modifiers in the retina[217]. Currently, our cases have non-syndromic RP; however, we suggest monitoring as a protracted syndromic phenotype may arise.

In addition, WGS facilitated the discovery of pathogenic noncoding variants that were not previously captured by TCS and WES. Two NCSS variants, i.e. *HGSNAT*, c.493+5G>A and *USH2A*, c.4758+3A>G, were detected using WGS. These noncoding variants were overlooked in earlier testing methods due to the lack of evidence for pathogenicity at the time of sequencing. Following *in vitro* splice assay analysis of these variants, less than 25% of the wild-type fragment in the total mRNA remained[211] (Figure 4.3 and Table 4.8).

In Pt-7, we identified *CDHR1* c.783G>A in a homozygous state, which previously was determined to lead to in-frame skipping of exon 8 (p.Asp214_Pro261del)[208]. No other potentially causative variants were identified as part of our analysis. The AF of this variant (total AF of 0.003052 and AF of 0.004903 in non-Finnish Europeans; Table 4.9) indicates that this variant may not be fully penetrant or acts as a hypomorphic allele. The splice defect is in-frame and may confer a subtle change in protein function. Additionally, Stingl *et al*., showed reduced expression of the mutant allele by comparison with the wild type. In line with this, Pt-7 displays a mild phenotype consistent with previously described *CDHR1*-related disease and late age at disease onset. It is also possible that other genetic factors affect the penetrance of this variant, potentially explaining the relatively high AF[217].

Importantly, WGS identified pathogenic coding variants in IRD-associated genes that did not reside in the TCS panels. Genes such as *FAM161A*, *RGS9BP*, *C21ORF2*, and *CDHR1* had not been included in the TCS panels at the time of sequencing (Table 4.4). These findings prompt redesign of TCS panels to include these genes where WES and WGS remain too costly to apply on a wider scale.

In seven out of 24 (29.2%) resolved cases, we identified five incidents of DIVs that could not be detected with previous sequencing techniques. The tested DIVs in this study led to frameshift variants, except for *GRM6,* c.1355-587dup. We observed the insertion of a 75-nt pseudoexon in *GRM6* midigene assays using the mutant construct (21%), which encodes 25 amino acids. In addition, a partial deletion of exon 6 in wild-type and mutant midigene assays was identified, which was more abundant in products from the wild-type midigene compared to those from the mutant midigene (54% compared

to 37% in the mutant) (Figure 4.4b). Of note, the partial exon 6 deletion in *GRM6* was not reported previously, and further research needs to be undertaken to investigate this phenomenon in a retina-specific cell model. In addition, we classified *GRM6* c.1355-587dup as a mild allele in this study, although a stronger effect on splicing in retina-specific cells cannot be discarded as a similar effect was previously observed for the *ABCA4* c.4539+2028C>T and c.5196+1137G>A variants. These findings highlight the importance of functional analyses in proving intronic variant pathogenicity, the contribution of such variants to IRD pathogenesis, and the diagnostic imperative imposed by them.

No splice defects were observed as a result of *ABCA4*, c.6148-89G>A (Pt-27); *ABCA4*, c.5460+1315_5460+1317delinsTA (Pt-28-30); *CYP4V2*, c.214+879_214+882delinsG (Pt-52); *PDE6B* c. 469-776C>G (Pt-61); and *RLBP1* c.525+425_525+433delinsATA (Pt-65). For the *ABCA4, CYP4V2*, and *RLBP1* indel variants, only Alamut prediction scores were used for their inclusion in the *in vitro* functional analysis as the SpliceAI algorithm cannot provide prediction scores for indel variants. All three variants create a cryptic donor site and have not been reported previously in the gnomAD database (Figure 4.5A, C, and E). The *ABCA4*, c.6148-89G>A variant strengthens the cryptic splice donor site in intron 44, from 67.6 to 79.1 in SpliceSiteFinder-like (SSFL) (range 0–100), and also activates an exonic splice enhancer (SF2/ASF) motif in the presence of c.6148-89G>A (Figure 4.5B). Similarly, the SpliceAI algorithm predicts a gain of a nearby splice donor site with a delta score of 0.12 (Table 4.7). Surprisingly, we did not observe a splice defect in the presence of variant c.6148-89G>A in *ABCA4*. Possibly, the functional assay in HEK293T cells is not a representative model for this variant due to differences in the presence of splice enhancer and silencer proteins in kidney and retina-specific cells. These five variants and in particular the *ABCA4*, c.6148-89G>A variant, may still be proven pathogenic when investigated in more relvant cell types, i.e. retinal cells. Such findings have been reported previously, e.g., splice defects resulting from *ABCA4* c.4539+2001G>A and c.4539+2028C>T were only detected in PPCs derived from patient fibroblasts[112]. Therefore, for variants that do not display a splice defect, exploring variant pathogenicity in models that better recapitulate the retinal cell identity should be considered. Likewise, we have established that a minority of wild-type mRNA remains for a number of variant alleles, while analysis of variants in PPCs may determine an alternative allele severity compared to the result from *in vitro* splice assays in HEK293T cells. Furthermore, Pt-52 was included as a suspected Bietti crystalline corneoretinal dystrophy (MIM: #210370) case, which has a strong

genotype–phenotype association with pathogenic variants in *CYP4V2*. Given that the phenotype supports this gene as a candidate gene and a good first candidate variant was identified via WES, this may support our hypothesis that providing pathogenic evidence for the DIV detected in *CYP4V2* in this patient possibly failed due to the disadvantages of the HEK293T midigene system. It is possible that retina specific splice factors, such as silencers or enhancers play a vital role in coordintating splicing in the retina itself. Alternatively, the disadvantage lies within the flaws of short-read sequencing to detect the actual pathogenic variants, or the initial pathogenic allele may represent a chance-finding.

WGS is the optimal sequencing method for the detection of SVs by comparison with WES and TCS, as indicated by previous studies[52,63,218]. In line with this, we have detected a homozygous 3.5 kb SV in *RPGRIP1* in an individual with LCA (Pt-22), which remained undetected by WES. The SV led to exon 20 deletion in the mRNA, altering the open-reading frame and resulting in a stop codon after six amino acids (p.Asp1080Glyfs*6) (Figure 4.2D). The assessment of WES was set to only detect deletions when spanning 2 exons or more and therefore could not detect this SV. Previously, heterozygous pathogenic variants in *CEP290, PDE6A*, and *RPE65* were reported as the potential first allele in Pt-22, with all three genes strongly associated with the given phenotype, which may have distracted from assessing other clinically relevant genes such as *RPGRIP1* (Table 4.2). However, a comprehensive analysis of the entirety of these genes did not identify additional pathogenic variants. This finding emphasizes the importance of analysis of all IRD-associated genes to identify the true genetic cause of disease.

In addition, a 26.86 kb heterozygous deletion on chromosome 19 was detected in proband Pt-23 (Figure 4.5B). This deletion encompasses the entire *PRPF31* and *TFPT* genes, and the promoter of *NDUFA3*. This large deletion was not detected using TCS as the breakpoints of the variant lay within noncoding regions not covered by the TCS panel. In general, it is difficult to detect read depth reduction in heterozygous SVs when there is no read coverage in the surrounding region for comparison. In addition, this specific individual was thought to have an autosomal recessive disease given the lack of family history. Segregation analysis confirmed maternal inheritance of the SV. However, the proband's mother was phenotypically unaffected, due to incomplete penetrance associated with variants in *PRPF31*, noted in other studies[219–221]. TFPT is a molecular partner of TCF3 and is associated with childhood acute lymphoblastic leukemia (MIM #613065)[222] NDUFA3 is required for the formation of the extramembrane arm of human mitochondrial complex 1[223]. No additional disease features could be attributed to the disruption of these

genes at this time. In addition, there are no reported pathogenic variants in either *TFPT* or *NDUFA3* in the ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/). A similar deletion encompassing *PRPF31, TFPT*, and *NDUFA3* on chromosome 19 has been reported previously[224].

In this study, 76 (76%) cases remain genetically unresolved following WGS analysis, indicating the presence of pathogenic variants that eluded our data analysis approach. By comparison, the 100,000 genomes project performed a pilot study that used WGS to investigate the genetic cause of rare disease in 2183 families. This study had a diagnostic yield of 25%, similar to that of our study. However it should be notes that 100,000 genomes examined a broad range of disease and the diagnostic yield was higher for monogenic disease (35%)[225]. The pipeline employed in our study was designed to be stringent, comprehensive, and time effective, given the immense volume of data created by WGS. Unresolved cases may harbor pathogenic variants that were called, but not individually interpreted because they did not pass the prioritization criteria employed. Our data analysis approach relied on the presence of one coding or NCSS pathogenic variant before interrogation of intronic regions of the same gene. Therefore, it is possible that some unresolved probands carry homozygous or compound heterozygous pathogenic intronic variants. Alternatively, some may carry pathogenic variants in genes not yet associated with IRD and therefore were not included in the WGS data analysis.

Furthermore, it is likely that some individuals carry pathogenic variants located within regulatory regions of IRD genes, as reported in other studies. However, the identification and interpretation of such variants remains complex. These variants are possibly not prioritized by our current strategy as they likely have their own interpretation criteria that are yet to be established. In addition, it is important to note that the line between Mendelian and complex disease may be blurred by modifier loci and environmental influences[226]. Thus, we should consider that some of the remaining unresolved IRD cases in this study may have a more complex aetiology than previously thought as shown by the presence of hypomorphic alleles in cases of our cohort.

Lastly, it is possible that some variants may not have been called by variant calling software as they may be located in repetitive or low complexity regions with poor coverage, e.g., *RPGR* open-reading frame 15 (ORF15). Pathogenic variants in *RPGR* ORF15 are responsible for 50–90% of X-linked RP (MIM: 312610)[227]. While in many instances male cases are analyzed for pathogenic variants in *RPGR* ORF15 prior to WES, in our WGS cohort, 35/76 (46%) of unresolved cases are male individuals with an RP phenotype and it is possible that a proportion carries pathogenic variants in *RPGR* ORF15.

As data analysis tools for WGS and knowledge surrounding variant interpretation improve, it is anticipated that further cases will receive a genetic diagnosis in the future.

We performed WGS on a cohort of 100 IRD probands that remained unresolved following the employment of TCS and WES as preliminary sequencing measures. A genetic cause of the disease was determined in 24 (24%) cases through detection of variants in genes that were not included in TCS panels or by identification and functional analyses of pathogenic intronic variants as well as detection of both small and large SVs not previously identified or recognized as causative. While a 24% yield may be modest, the number of "obvious" genetic causes detected in the coding regions of established IRD-associated genes is surprising, especially, considering the extensively previously examined cohort we assessed. Reasons for this may include an incomplete gene panel, poor read quality in previous testing methods, or lack of evidence surrounding the implication of some IRD genes at the time of initial sequencing. While the coverage and quality of reads in WGS are higher than prescreening techniques enabling better variant detection in this cohort, our findings emphasize the necessity of improvements in variant interpretation and bioinformatics pipelines as well as performing functional assays to obtain a higher diagnostic yield not only by WGS but also by less comprehensive sequencing techniques such as WES and TCS.

The findings of this study demonstrate that WGS can be an optimal sequencing approach for individuals diagnosed with an IRD. In future studies, a combined approach of WGS and optical genome mapping or long-read sequencing can be beneficial to overcome the difficulties of SV detection in short-read sequencing and provide a higher genetic diagnostic yield. It is a diagnostic imperative to continue to analyze individuals who have received an inconclusive genetic testing result following less comprehensive sequencing methods. Identification of the underlying genetic pathogenesis of a disease is becoming increasingly important in the advent of gene-based therapeutics, where a known genetic cause is essential to access clinical trials and approved treatments. With decreasing costs, increasing data interpretability, and continued functional analysis, we foresee that WGS may become the sequencing methodology of choice to provide IRD patients with an accurate genetic diagnosis.

5. Entire *ABCA4* gene sequencing of 162 Irish Stargardt disease/cone-rod dystrophy probands.

5.1 Introduction

High-throughput WGS is considered to be the most comprehensive genetic screening procedure to identify the underlying genetic cause of inherited human disease, as illustrated in section 4 of this doctoral thesis. It allows for detection of a multitude of variation across the entire genome, including both protein coding and non-coding SNVs, large and small SVs as well as CNVs. However, the costs, data handling and variant interpretation involved are significant challenges to employ WGS as a routine diagnostic measure for all IRDs. These limitations highlight the value of a screening procedure that can comprehensively and economically sequence a small number of genes that are associated with a very clinically distinct condition.

Autosomal recessive STGD1, primarily caused by variants in *ABCA4*, is a prime example of such a condition. STGD1 is the most commonly inherited form of macular dystrophy (Figure 5.1), with an estimated prevalence of 1 in 10,000[228], however, an exact number is difficult to determine due to its high level of clinical heterogeneity. STGD1 encompasses a wide spectrum of phenotypic severity, as well as displaying a large degree of allelic heterogeneity. The combination of variants carried by the affected individual determines the severity of disease[142,229,230]. This ranges from early-onset pan-retinal dystrophy[231] to a more typical STGD1 presentation of onset in adolescence or late onset STGD1 due to a combination of mild or hypomorphic variants, including the relatively common c.5603T>C (p.Asn1868Ile) variant (AF of ~6%). A large study published in 2017 found that at this time, 1180 unique *ABCA4* variants had been reported across 8777 alleles in 6684 cases[161]. In addition to coding variants, many of these variants are NCSS variants as well as a smaller number of DIVs[23,65,142,143,210,211,229,232,233]. These variants have varying effects on the splicing process, some weakening the strength of canonical splice sites (CSSs), activating cryptic splice acceptor or donor sites, or incorporating a pseudoexon into the resulting mature mRNA. SVs in *ABCA4* appear to be rare[65,232–234], although analysis of CNVs has not been routinely employed in many STGD1 cases.

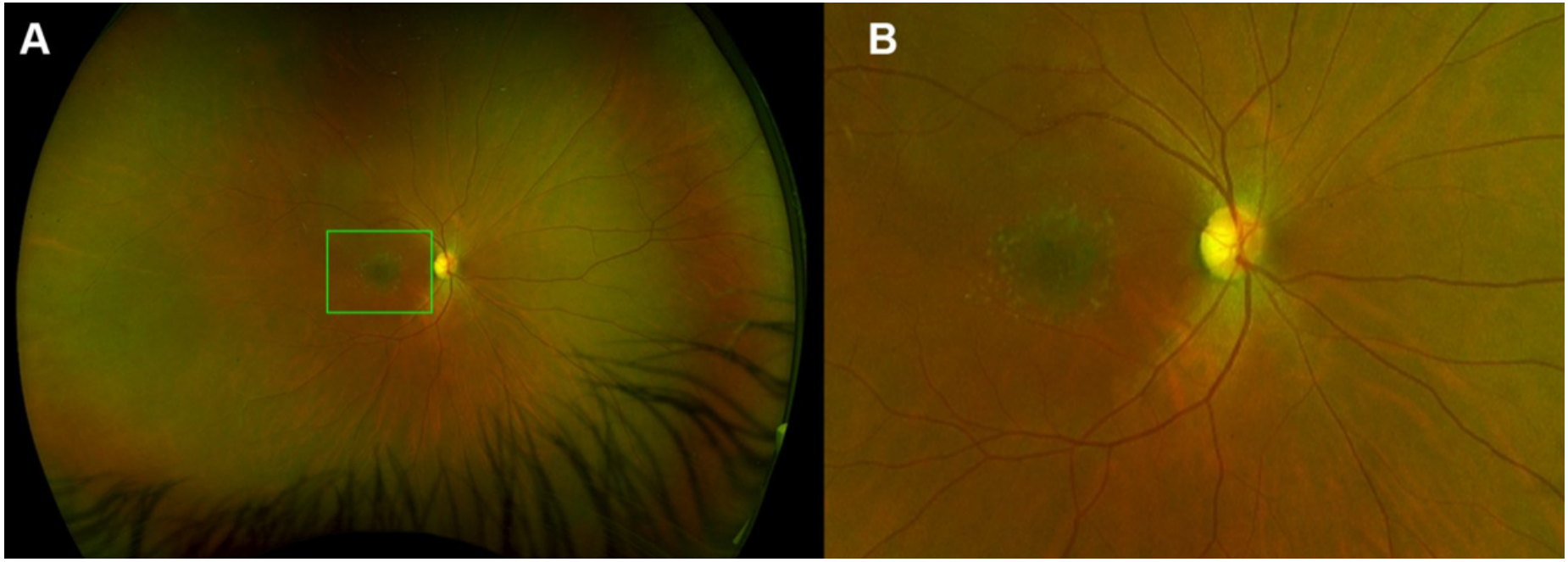


Figure 5.1 | **Retinal photographs of a patient with Stargardt disease (**A). Fleck deposits are visible surrounding an area of central photoreceptor degeneration in the macula, a defining characteristic of STGD1 (Photo from gene.vision[144]).

As *ABCA4* is a relatively large gene, comprising 128,313 base pairs, traditional screening methods have focused their attention on the protein coding proportion of the gene (50 exons) and flanking splice sites. STGD1 is predominantly caused by variants in *ABCA4*. Therefore, sequencing methods such as panel TCS or WES tend to be relatively more expensive and generate a surplus of data from genes that do not correlate with the STGD1 phenotype. Furthermore, these broader capture methods, including that employed as a preliminary diagnostic measure in Target 5000, typically only detect intronic variants that have been previously associated with disease pathology. This highlights the demand for a sequencing strategy that can generate data from one or a few genes in a cost-effective manner.

A recent study by Khan *et al*., utilised fully automated, high throughput, cost-effective and comprehensive sequencing of the entire *ABCA4* gene using smMIPs on a large cohort of unresolved STGD1 and CRD probands[23]. 3866 smMIPs were employed to capture 110 nt increments of both the sense and antisense strands of the entire *ABCA4* gene, as well as 40kb of flanking sequences. 36 Irish probands were included in this initial study, with an additional 126 patients from Ireland sequenced in subsequent runs. The candidate variant detection rate was 61% and 59% for prescreened and unscreened cases respectively. These findings highlight the effectiveness of smMIPs-based sequencing for unresolved *ABCA4* related IRDs, emphasising the potential use of smMIPs as an extremely cost effective first-pass measure when determining the underlying genetic cause of those affected by STGD1 and STDG1-like phenotypes.

5.2 Methods

5.2.1 smMIPs design

For the targeted sequencing of the *ABCA4* gene, as well as 23.9 kb up and 14.9 kb downstream sequences, 3,866 smMIPs were designed using the MIPgen pipeline, as described previously[137]. 496 of the 3,866 smMIPs spanned known single nucleotide polymorphisms (SNPs) at the extension or ligation arm. 3,866 smMIPs effectively capture 110-nt regions. All smMIPs were designed in a way that each nucleotide position in principle is targeted by two different smMIPs to increase sensitivity and avoid dropout of alleles due to variants not annealing to the smMIP probes. A schematic of the workflow can be found in Figure 5.2.

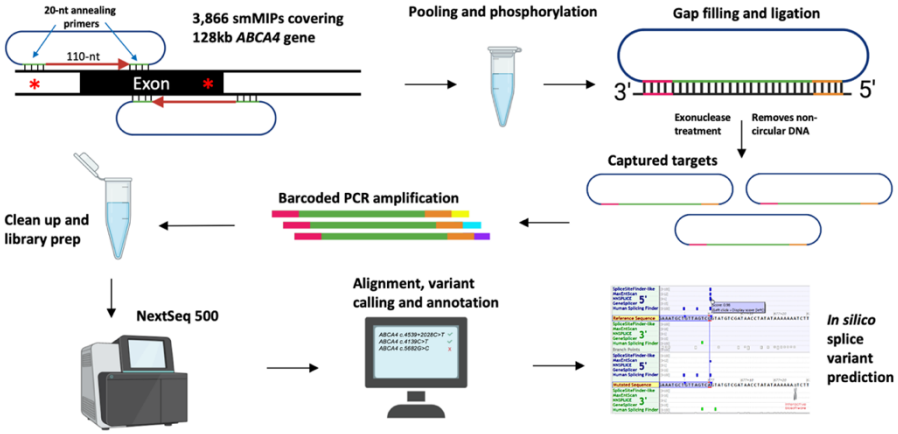


Figure 5.2 | single-molecule molecular inversion probe (smMIP) sequencing workflow. smMIPs contain two targeting arms each 20 nucleotides in length each. These probes are complementary to the targeted region. In the capture reaction, the pooled smMIPs hybridise to the genomic DNA. This is followed by gap filling extension and ligation to circularise the smMIPs. PCR is carried out using primers that recognise sequences in the smMIPs backbone. These include indexing barcodes to allow for multiplexing. Following this library preparation is performed and the samples are sequenced on a NextSeq500. The sequencing data is aligned, variants are called and annotated for manual inspection. All variants are assessed for their potential impact on the splicing process.

5.2.2 Data analysis and variant classification

To select the causative variants, several filters were applied. Firstly, variants with AFs <1% in the gnomAD[145] population database and the Genome of the Netherlands (GoNL)[235] were retained. A filter based on a 'quality score' of >200, was applied based on the average number of reads targeting each genomic position in a single run and other standard filters based on the position of SNVs (e.g. exons, CSSs and known NCSS variants). A number of previously listed causal variants with relatively high AFs such as c.2588G>C, c.5461-10T>C, c.5603A>T and c.5882G>A (AFs in gnomAD: 0.00430, 0.00022, 0.04219 and 0.00456 respectively) were selected separately[161,230,236,237]. To assess the variants, *in silico* analysis was performed using pathogenicity assessment tools such as Sorting Intolerant From Tolerant (SIFT; https://sift.bii.a-star.edu.sg)[238], Polymorphism Phenotyping v2 (PolyPhen-2)[239], and Mutation Taster[240]. Among others, Grantham[197] and PhyloP[195] scores were also assessed to predict the physiochemical effect of nucleotide changes on *ABCA4* protein and conservation of nucleotide positions, respectively.

For the selection of known coding and non-coding causal variants, *ABCA4* variant databases, such as the Leiden Open Variation Database[161,241] and the Human Genome Mutation Database[242] were used. We supplemented the list of SNVs and small indels with additional variants reported in the literature and employed an excel-based script developed by Stephanie Cornelis to identify these in an automated manner. Candidate variants were selected based on AFs and *in silico* predictions. Thereafter, variants were classified

according to the guidelines of the ACMG-AMP[185]. Similarly, we detected previously reported causal splice variants (NCSS variants and DI variants) by listing all 165 variants previously tested at this time using *in vitro* splice assays[65,142,143,210,211,229,243,244] or RT-PCR analysis of PPCs[112,243].

5.2.3 Data processing

Raw sequencing data (FASTQ) files were used for mapping and variant annotation by an in-house bioinformatics pipeline. Raw reads were aligned to the reference genome (GRCh37) using BWA mem (v.0.7.12). The extension and ligation arms as well as overlap between read pairs was trimmed after alignment. Duplicated reads were also removed. Variant calling and annotation were performed as described by Khan *et al*[143]. Briefly, variants were called using both Unified Genotyper[245] and Haplotype Caller[246] from GATK (v3.4-46), with both sets then merged using GATK Combine Variants. Variants were annotated with predicted effects, gene component information as well as frequency information from various population databases.

5.2.4 CNV analysis

An excel script was employed to detect CNVs using smMIP read depth. Only large (>400 bp) deletions and duplications in principle could be identified, but not insertions and inversions. Therefore the term CNV instead of SV was used in the following section. The read coverage of a single target was first normalized in comparison to all 3,866 targets by dividing the number of reads for a single target by the total number of reads for a given DNA sample. The coverage of each target of a patient was subsequently compared with the coverage of each target in all probands of that series using the following ratio: R = average coverage of each target of a proband divided by average coverage of each target of all probands in the run. In order to identify large deletions and duplications, the following ranges were defined: homozygous deletions, $R < 0.3$; heterozygous deletions, $0.3<R<0.7$; no CNV, $0.7<R<1.2$; heterozygous duplications, $1.2<R<1.7$; homozygous duplications, $R>1.7$. These parameters were defined based on the assumption that a relative drop or increase in read depth by comparison with the average for that run was indicative of a deletion or duplication event. To discriminate actual CNVs from artifacts, ratios per target were compared within each patient and among all the patients. Candidate CNVs were considered if at least 6 consecutive targets were deleted or duplicated. Originally candidates with less than 6 consecutive targets were analysed but these were highly numerous and when interrogated were artifacts, so the threshold was set at 6.

5.3 Results

5.3.1 Patient cohort

A total of 162 (36 from Khan *et al*.,[54] and 126 in subsequent runs) probands were referred for smMIPs based sequencing of the entire *ABCA4* gene from The Royal Victoria Eye and Ear Hospital, The Mater Misericordiae University Hospital and the Belfast Health and Social Care Trust. Patients were included given that they had a clinical presentation of either STGD1 or CRD. 114 patients presented with STGD1 and 48 presented with CRD.

5.3.2 Candidate variant detection rate

Target 5000 target-capture NGS detects positive candidate variants in ~80% of STGD1/CRD cases. A further 10% of cases carry one likely pathogenic variant (Figure 5.3). smMIPs-based whole gene sequencing of *ABCA4* provided a refined genetic diagnosis in 61% (40/65) of previously screened cases where TCS did not and 59% of unscreened cases (57/97), giving an overall diagnostic yield of ~60% (Figure 5.4). It is important to note here that 20 pre-screened cases carried a mild *ABCA4* variant that was detected using TCS. These cases were included in this study to determine the possibility of additional undetected variants influencing the phenotype. These 20 probands are included in the positive diagnostic yield for pre-screened cases as smMIPs sequencing provided a more definitive and comprehensive diagnosis than TCS. However, the possibility remains that additional undiscovered variants do influence phenotypic outcome in these patients. For the other 20 (31%) pre-screened cases, the increase in candidate variant detection for previously screened cases was due to the discovery of both exonic and intronic variants not detected previously by target capture NGS as part of Target 5000. The positive candidate variant detection rate varied greatly between those clinically diagnosed with STGD1 and CRD, at 75% (86/114) and 23% (11/48) respectively (Figure 5.5). 65 cases remained unresolved following this study, of which 43% (28/65) were clinically diagnosed with STGD1 and 57% (37/65) with CRD. Of these 65 cases, 34% (22/65) were found to be monoallelic for variants in *ABCA4.* Of the 22 monoallelic cases, ~64% (14/22) were clinically diagnosed as CRDs and ~36% (8/22) were STGD1. Of the 43 zero *ABCA4* allele cases, 51% (22/43) were clinically diagnosed with CRD and 49% (21/43) with STGD1. A summary of all pathogenic variants detected can be found in Table 5.1.

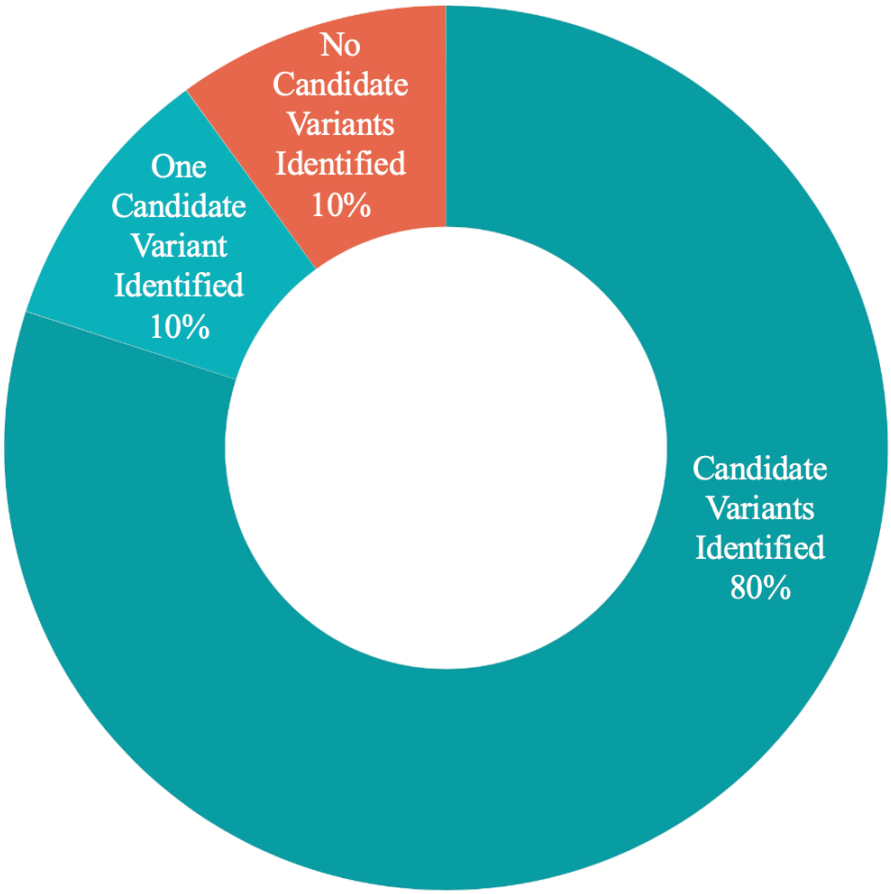


Figure 5.3 | Positive candidate variant detection rate of Target 5000 panel based NGS for STGD1 and CRD combined. 80% of STGD1/CRD cases receive a positive genetic diagnosis following Target 5000 target capture NGS. 10% has a single *ABCA4* allele and 10% remain unresolved.

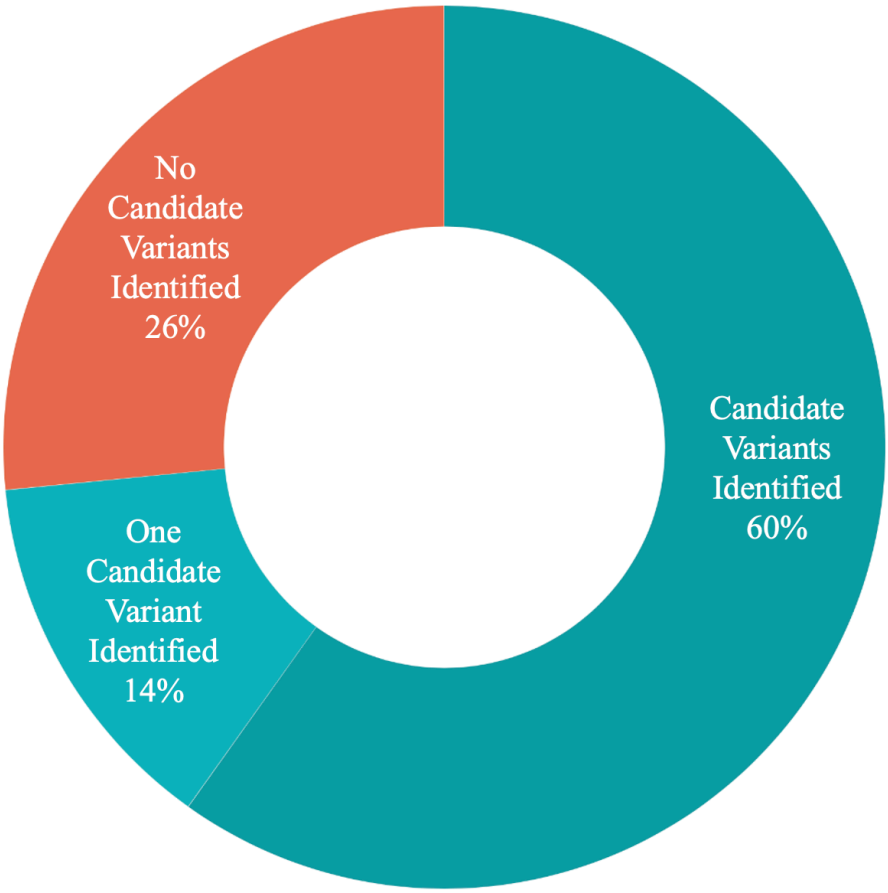


Figure 5.4 | Positive candidate variant detection rate of single-molecule molecular inversion probe (smMIP)s-based sequencing for previously unresolved STGD1 and CRD combined. 60% of STGD1/CRD cases receive a positive genetic diagnosis following smMIPs-based sequencing.

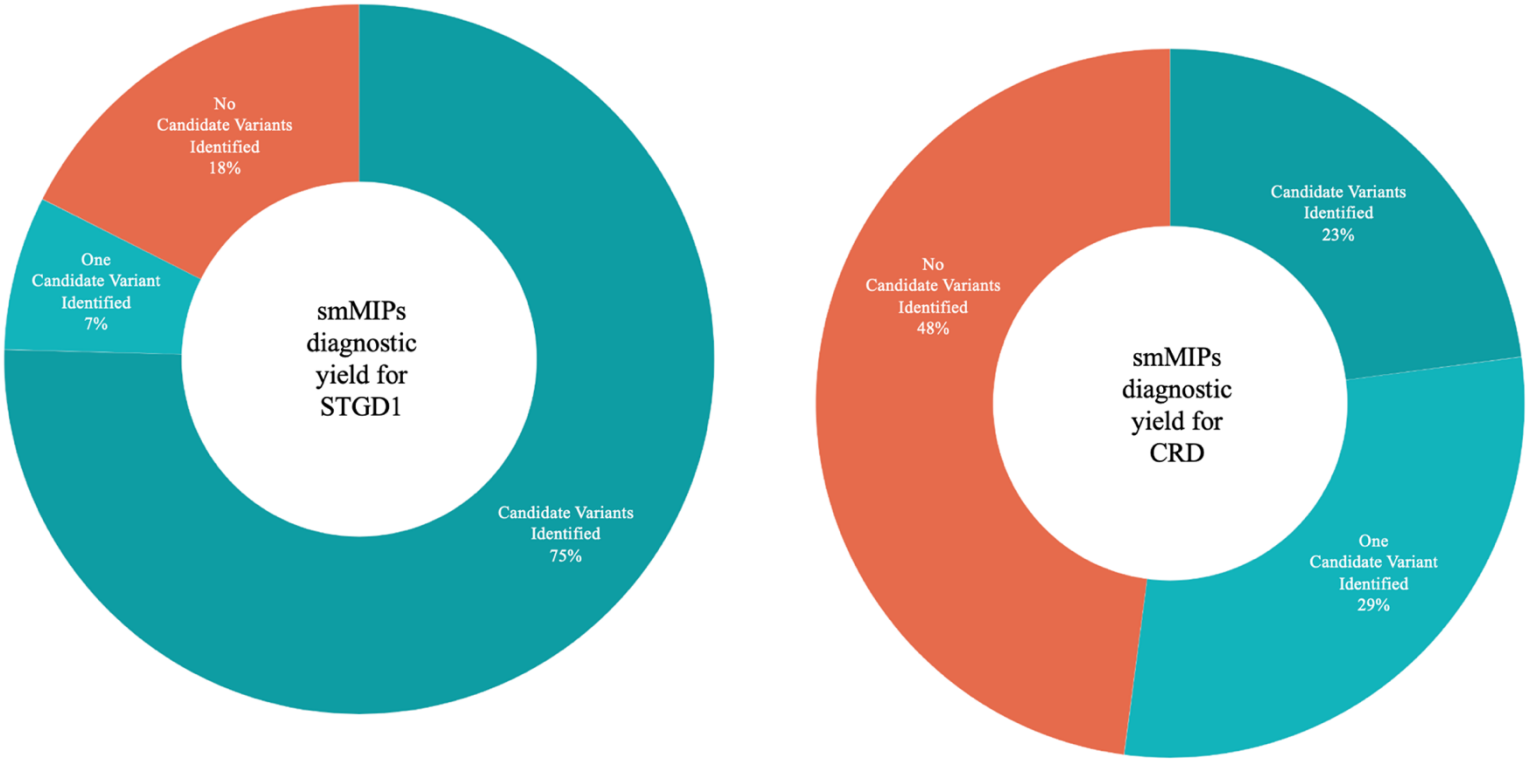


Figure 5.5 | Positive candidate variant detection rate of single-molecule molecular inversion probe (smMIP)s-based sequencing for Stargardt Disease (STGD1) and cone-rod dystrophy (CRD) separately. 75% of clinically defined STGD1 cases received a positive genetic diagnosis after this study, by comparison with only 29% of CRD cases.

Table 5.1 | A summary of the pathogenic variants identified in 97 resolved probands that underwent smMIPs sequencing of *ABCA4*. STDG1 = Stargardt Disease, CRD = Cone-rod dystrophy.

Patient ID	Clinical Diagnosis	Allele 1 Nucleotide Change	Allele 1 Protein Change	Allele 2 Nucleotide Change	Allele 2 Protein Change
Pt-1	CRD	c.3322C>T	p.(Arg1108Cys)	c.4539+2028C>T	p.[=,Arg1514Leufs*36]
Pt-2	CRD	c.4253+43G>A	p.[=,Ile1377Hisfs*3]	c.4363T>C	p.(Cys1455Arg)
Pt-3	CRD	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.6472A>G	p.(Lys2158Glu)
Pt-4	STGD1	c.161G>A	p.[Cys54Serfs*14,Cys54Tyr,=]	c.5196+1137G>A	p.[=,Met1733Glufs*78]
Pt-5	STGD1	c.4139C>T	p.(Pro1380Leu)	c.4539+2028C>T	p.[=,Arg1514Leufs*36]
Pt-6	STGD1	c.5196+1137G>A	p.[=,Met1733Glufs*78]	c.6079C>T	p.(Leu2027Phe)
Pt-7	STGD1	c.4222T>C	p.(Trp1408Arg)	c.4918C>T	p.(Arg1640Trp)
Pt-8	STGD1	c.735T>G	p.(Tyr245*)	c.5603A>T	p.(Asn1868Ile)
Pt-9	STGD1	c.286A>C	p.(Asn96His)	c.5603A>T	p.(Asn1868Ile)
Pt-10	STGD1	c.4577C>T	p.(Thr1526Met)	c.5603A>T	p.(Asn1868Ile)
Pt-11	STGD1	c.[5461-10T>C;5603A>T]	p.[Thr1821Aspfs*6,Thr1821Valfs*13;(Asn1868Ile)]	c.5929G>A	p.(Gly1977Ser)
Pt-12	STGD1	c.1222C>T	p.(Arg408*)	c.5603A>T	p.(Asn1868Ile)
Pt-13	STGD1	c.1804C>T	p.(Arg602Trp)	c.5603A>T	p.(Asn1868Ile)
Pt-14	STGD1	c.[4469G>A;5603A>T]	p.[(Cys1490Tyr);(Asn1868Ile)]	c.5603A>T	p.(Asn1868Ile)
Pt-15	STGD1	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.4918C>T(;)c.4222T>C	p.(Arg1640Trp)(;)(Trp1408Arg)
Pt-16	STGD1	c.4139C>T	p.(Pro1380Leu)	c.4539+2028C>T	p.[=,Arg1514Leufs*36]
Pt-17	CRD	c.768G>T	p.(Leu257Valfs*17)	c.6079C>T	p.(Leu2027Phe)
Pt-18	CRD	c.[4469G>A;5603A>T]	p.[Cys1490Tyr;Asn1868Ile]	c.5196+1137G>A	p.[=,Met1733Glufs*78]
Pt-19	CRD	c.[2588G>C;5603A>T]	p.[Gly863Ala,Gly863del;(Asn1868Ile)]	c.4139C>T	p.(Pro1380Leu)
Pt-20	CRD	c.1715G>A(;)4577C>T	p.(Arg572Gln)(;)(Thr1526Met)	c.[2588G>C;5603A>T]	p.[Gly863Ala,Gly863del;(Asn1868Ile)]

Pt-21	CRD	c.1715G>A	p.(Arg572Gln)	c.[2588G>C;5603A>T]	p.[Gly863Ala,Gly863del;(Asn1868Ile)]
Pt-22	CRD	c.[2588G>C;5603A>T]	p.[Gly863Ala,Gly863del;(Asn1868Ile)]	c.3323G>A	p.(Arg1108His)
Pt-23	CRD	c.[2588G>C;5603A>T]	p.[Gly863Ala,Gly863del;(Asn1868Ile)]	c.3323G>A	p.(Arg1108His)
Pt-24	CRD	c.[1622T>C;3113C>T]	p.[Leu541Pro;Ala1038Val]	c.[3758C>T;5882G>A]	p.[Thr1253Met;Gly1961Glu]
Pt-25	STGD1	c.[3481C>A(;)5603A>T]	p.[Arg1161Ser(;)Asn1868Ile]	c.5882G>A	p.(Gly1961Glu)
Pt-26	STGD1	c.5312+1G>A	p.(?)	c.5882G>A	p.(Gly1961Glu)
Pt-27	STGD1	c.1532G>A(;)5882G>A	p.(Arg511His)(;)(Gly1961Glu)	c.[2852T>C(;)5603A>T]	p.[Ile951Thr(;)Asn1868Ile]
Pt-28	STGD1	c.4577del	p.(Thr1526Argfs*10)	c.5882G>A	p.(Gly1961Glu)
Pt-29	STGD1	c.5882G>A	p.(Gly1961Glu)	c.6221G>A	p.(Gly2074Asp)
Pt-30	STGD1	c.4234C>T	p.(Gln1412*)	c.5882G>A	p.(Gly1961Glu)
Pt-31	STGD1	c.4139C>T	p.(Pro1380Leu)	c.[5461-10T>C;5603A>T]	p.[Thr1821Valfs*13,Thr1821Aspfs*6;(Asn1868Ile)]
Pt-32	STGD1	c.4139C>T	p.(Pro1380Leu)	c.5882G>A	p.(Gly1961Glu)
Pt-33	STGD1	c.2549A>G(;)5882G>A	p.(Tyr850Cys)(;)(Gly1961Glu)	c.6098T>G	p.(Leu2033Arg)
Pt-34	STGD1	c.4793C>A	p.(Ala1598Asp)	c.3322C>T	p.(Arg1108Cys)
Pt-35	STGD1	c.4139C>T	p.(Pro1380Leu)	c.6079C>T	p.(Leu2027Phe)
Pt-36	STGD1	c.3322C>T	p.(Arg1108Cys)	c.4139C>T	p.(Pro1380Leu)
Pt-37	STGD1	c.6658C>T	p.(Gln2220*)	c.[550T>C;551C>G]	p.[Ser184Pro;Ser184Cys]
Pt-38	STGD1	c.3754G>T	p.(Glu1252*)	c.4539+2028C>T(;)5603A>T	p.[=,Arg1514Leufs*36](;)(Asn1868Ile)
Pt-39	STGD1	c.1519G>T	p.(Asp507Tyr)	c.4139C>T	p.(Pro1380Leu)
Pt-40	STGD1	c.4253+43G>A	p.[=,Ile1377Hisfs*3]	c.5917del	p.(Val1973*)
Pt-41	STGD1	c.4363T>C	p.(Cys1455Arg)	c.6079C>T	p.(Leu2027Phe)
Pt-42	STGD1	c.4139C>T	p.(Pro1380Leu)	c.[4469G>A;5603A>T]	p.[(Cys1490Tyr;Asn1868Ile)]
Pt-43	STGD1	c.3056C>T	p.(Thr1019Met)	c.6472A>G	p.(Lys2158Glu)
Pt-44	STGD1	c.161G>A	p.(Cys54Tyr)	c.[5461-10T>C;5603A>T]	p.[Thr1821Valfs*13,Thr1821Aspfs*6;(Asn1868Ile)]

Pt-45	STGD1	c.4469G>A(;)5603A>T	p.(Cys1490Tyr)(;)(Asn1868Ile)	c.466A>G(;)4771G>A	p.(Ile156Val)(;)(Gly1591Arg)
Pt-46	STGD1	c.4139C>T	p.(Pro1380Leu)	c.4577C>T	p.(Thr1526Met)
Pt-47	STGD1	c.2041C>T	p.(Arg681*)	c.5882G>A	p.(Gly1961Glu)
Pt-48	STGD1	c.5603A>T	p.(Asn1868Ile)	c.185C>T(;)6098T>G	p.(Pro62Leu)(;)(Leu2033Arg)
Pt-49	STGD1	c.5113C>T	p.(Arg1705Trp)	c.5929G>A	p.(Gly1977Ser)
Pt-50	STGD1	c.3056C>T	p.(Thr1019Met)	c.5714+5G>A	p.[=,Glu1863Leufs*33]
Pt-51	STGD1	c.3113C>T	p.(Ala1038Val)	c.[4222T>C;4918C>T]	p.[Trp1408Arg;Arg1640Trp]
Pt-52	STGD1	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.6449G>A	p.(Cys2150Tyr)
Pt-53	STGD1	c.4253+43G>A	p.[=,Ile1377Hisfs*3]	c.4861A>T	p.(Asn1621Tyr)
Pt-54	STGD1	c.3056C>T	p.(Thr1019Met)	c.3322C>T	p.(Arg1108Cys)
Pt-55	STGD1	c.3322C>T	p.(Arg1108Cys)	c.4139C>T	p.(Pro1380Leu)
Pt-56	STGD1	c.161G>A	p.(Cys54Tyr)	c.2483C>T	p.(Pro828Leu)
Pt-57	STGD1	c.2382+1dup	p.(?)	c.5603A>T	p.(Asn1868Ile)
Pt-58	STGD1	c.5603A>T	p.(Asn1868Ile)	+	+
Pt-59	STGD1	c.4957G>A	p.(Gly1653Arg)	c.4957G>A	p.(Gly1653Arg)
Pt-60	STGD1	c.3137T>G	p.(Leu1046Trp)	c.3322C>T	p.(Arg1108Cys)
Pt-61	STGD1	c.[1622T>C;3113C>T]	p.[(Leu541Pro;Ala1038Val)]	c.5882G>A	p.(Gly1961Glu)
Pt-62	STGD1	c.4577C>T	p.(Thr1526Met)	c.869G>A	p.(Arg290Gln)
Pt-63	STGD1	c.[5113C>T;5882G>A]	p.[Arg1705Trp;Gly1961Glu]	c.5882G>A	p.(Gly1961Glu)
Pt-64	STGD1	c.[4253+43G>A;5603A>T]	p.[=,Ile1377Hisfs*3;(Asn1868Ile)]	c.5603A>T	p.(Asn1868Ile)
Pt-65	STGD1	c.4363T>C	p.(Cys1455Arg)	c.5882G>A	p.(Gly1961Glu)
Pt-66	STGD1	c.2701A>G	p.(Thr901Ala)	c.5603A>T	p.(Asn1868Ile)
Pt-67	STGD1	c.4320del(;)5603A>T	p.(Phe1440Leufs*6)(;)(Asn1868Ile)	c.455G>A	p.(Arg152Gln)
Pt-68	STGD1	c.[1622T>C;3113C>T]	p.[Leu541Pro;Ala1038Val]	c.5603A>T	p.(Asn1868Ile)
Pt-69	STGD1	c.3322C>T	p.(Arg1108Cys)	c.4577C>T	p.(Thr1526Met)

Pt-70	STGD1	c.1906C>T	p.(Gln636*)	c.[5461-10T>C;5603A>T]	p.[Thr1821Valfs*13,Thr1821Aspfs*6;(Asn1868Ile)]
Pt-71	STGD1	c.5882G>A	p.(Gly1961Glu)	c.5714+5G>A(;)587C>T	p.[=,Glu1863Leufs*33](;)(Pro196Leu)
Pt-72	STGD1	c.1317G>A(;)3352C>G	p.(Trp439*)(;)(His1118Asp)	c.5603A>T	p.(Asn1868Ile)
Pt-73	STGD1	c.4577C>T	p.(Thr1526Met)	c.6089G>A	p.(Arg2030Gln)
Pt-74	STGD1	c.[4222T>C;4918C>T]	p.[Trp1408Arg;Arg1640Trp]	c.6089G>A	p.(Arg2030Gln)
Pt-75	STGD1	c.1804C>T(;)2894A>G	p.(Arg602Trp)(;)(Asn965Ser)	c.2311A>T(;)4283C>T	p.(Thr771Ser)(;)(Thr1428Met)
Pt-76	STGD1	c.1317G>A	p.(Trp439*)	c.[2588G>C;5603A>T]	p.[Gly863Ala,Gly863del;(Asn1868Ile)]
Pt-77	STGD1	c.1317G>A(;)4577C>T	p.(Trp439*)(;)(Thr1526Met)	c.5908C>T	p.(Leu1970Phe)
Pt-78	STGD1	c.6079C>T	p.(Leu2027Phe)	c.6079C>T	p.(Leu2027Phe)
Pt-79	STGD1	c.655A>T	p.(Arg219*)	c.5603A>T	p.(Asn1868Ile)
Pt-80	STGD1	c.1715G>A(;)5312+3A>T	p.(Arg572Gln)(;)(Asn1734Glyfs*14)	c.[2588G>C;5603A>T]	p.[Gly863Ala,Gly863del;(Asn1868Ile)]
Pt-81	STGD1	c.161G>A(;)1715G>A	p.(Cys54Tyr)(;)Arg572Gln	c.[2588G>C;5603A>T]	p.[Gly863Ala,Gly863del;(Asn1868Ile)]
Pt-82	STGD1	c.4139C>T	p.(Pro1380Leu)	c.5381C>A	p.(Ala1794Asp)
Pt-83	STGD1	c.1519G>T	p.(Asp507Tyr)	c.5312+3A>T	p.(Asn1734Glyfs*14)
Pt-84	STGD1	c.[2588G>C;5603A>T]	p.[Gly863Ala,Gly863del;(Asn1868Ile)]	c.6089G>A	p.(Arg2030Gln)
Pt-85	STGD1	c.[2588G>C;5603A>T]	p.[Gly863Ala,Gly863del;(Asn1868Ile)]	c.5461-10T>C(;)5603A>T	p.[Thr1821Aspfs*6,Thr1821Valfs*13](;)(Asn1868Ile)
Pt-86	STGD1	c.3323G>A	p.(Arg1108His)	c.6416G>C	p.(Arg2139Pro)
Pt-87	STGD1	c.2041C>T	p.(Arg681*)	c.[2588G>C;5603A>T]	p.[Gly863Ala,Gly863del;(Asn1868Ile)]
Pt-88	STGD1	c.4577C>T	p.(Thr1526Met)	c.5603A>T	p.(Asn1868Ile)
Pt-89	STGD1	c.[2588G>C;5603A>T]	p.[Gly863Ala,Gly863del;(Asn1868Ile)]	c.4469G>A	p.(Cys1490Tyr)
Pt-90	STGD1	c.1792G>A	p.(Val598Met)	+	+
Pt-91	STGD1	+	+	+	+
Pt-92	STGD1	c.6148G>C	p.(Val2050Leu)	+	+
Pt-93	STGD1	c.161G>A	p.(Cys54Tyr)	c.4254-5T>A	p.(?)

Pt-94	STGD1	c.161G>A	p.(Cys54Tyr)	c.4254-5T>A	p.(?)
Pt-95	STGD1	c.161G>A	p.(Cys54Tyr)	c.161G>A	p.(Cys54Tyr)
Pt-96	STGD1	c.161G>A	p.(Cys54Tyr)	c.4685T>C	p.(Ile1562Thr)
Pt-97	STGD1	c.1817G>A	p.(Gly606Asp)	c.4577C>T	p.(Thr1526Met)

5.3.3 Intronic variants are a significant cause of STGD1 and CRD in Ireland

Of 97 cases with positive candidate variants identified, 26 carried a likely pathogenic intronic variant. These variants and their incident rates are illustrated in Figure 5.6. Notably, 27% of cases that have now received a likely genetic diagnosis via smMIPs sequencing did so due to the detection of a likely pathogenic splice altering variant. Clearly for such cases additional functional studies are required to provide empirical evidence of the pathogenicity of the candidate intronic splice altering variant.

Figure 5.6 | Pathogenic intronic variants detected in *ABCA4* as part of this study and their prevalence. There is a distinct enrichment of c.4539C>T.



5.3.4 Enrichment of a deep-intronic *ABCA4* variant in the Irish population

As noted, 7 probands were found to carry *ABCA4* c.4539+2028C>T (p.[=,Arg1514Leufs*36]) with another pathogenic variant. Upon retrospective and ongoing analysis of unresolved STGD1 cases in the Target 5000 cohort, a total of 27 incidents of this variant in affected individuals have been detected, including the only two known homozygous patients globally. Indeed, such homozygotes can provide valuable insights as to the phenotypic consequences of a specific *ABCA4* mutation in STGD1 patients. The frequency of *ABCA4* c.4539+2028C>T (p.[=,Arg1514Leufs*36]) indicates enrichment of this variant in the Irish population with a possible founder effect. Detailed investigation of the clinical outcome and severity of this variant is presented in section 6.1.

5.4 Discussion

3866 smMIPs were employed to sequence 97.4% of the 128kb *ABCA4* gene in 162 STGD1/CRD cases from the Target 5000 cohort. A comprehensive genetic diagnosis was ascertained for 60% (97/162) of cases (Figure 5.4, Table 5.1). Furthermore, 26 of 97 resolved cases were found to carry a pathogenic intronic variant in *ABCA4,* in conjunction with a pathogenic coding variant (Figure 5.6). This highlights the significant role of intronic variants in STGD1/CRD pathology in Ireland. These results emphasise the importance of sequencing the entirety of the *ABCA4* gene in order to fully elucidate the spectrum of variants that may be involved in the manifestation of STGD1 and CRD. In addition, this highlights the efficacy of a smMIPs-based approach to sequence one or a small number of genes to obtain a genetic diagnosis in a clinically distinct phenotype.

Figure 5.6 illustrates the spectrum of non-canonical and DIVs identified in the probands of this study. Due to the nature of intronic variants, their pathogenicity cannot be determined in the same way as coding variants. Functional analysis is imperative to prove the pathogenicity of intronic variants and can be done through a number a means, including *in vitro* splice assays, using mammalian cells such as HEK293T cells, PPCs and/or in an even more sophisticated manner, employing retinal organoid technology to better recapitulate actual retinal context. Most of the intronic variants identified in this study have been previously functionally analysed and have been shown to produce aberrant mature mRNA transcripts, as detailed below.

ABCA4 c.5461-10T>C (p.[Thr1821Aspfs*6,Thr1821Valfs*13]), a NCSS variant, was detected in four probands, and is the third most frequent variant reported in STGD1 patients, with ~7% of affected individuals carrying this variant. First identified by Maugeri *et al*.,[29] and functionally analysed by Rivera *et al*.,[247] using an exon trapping technique, this variant did not appear to result in any functional defect. However, functional analysis carried out by Sangermano *et al*.,[243] using both a midigene *in vitro* splice assay and RNA extracted from patient derived PPCs found skipping of exon 39 of *ABCA4* and in some incidences skipping of both exon 39 and 40. Another study by Aukrust *et al*.,[108] verified this finding in patient derived fibroblast cells.

ABCA4 c.4253+43G>A (p.[=,Ile1377Hisfs*3]) was identified in four probands in this study. Functional analysis carried out by Sangermano *et al*.,[142] via employment of a midigene-based splice assay showed that this variant results in partial skipping of exon 28. Interestingly, this variant does not weaken the canonical splice donor site of exon 28, nor does it affect any other cryptic splice sites as predicted by *in silico* analysis. It does, however, affect the strength of several splice enhancer and silencer motifs, facilitating this skipping phenomenon. Although some natural skipping of exon 28 has been observed, the level of aberrantly spliced transcript is much higher in the presence of this variant and likely leads to the production of inefficient levels of wild-type transcript.

ABCA4 c.5196+1137G>A (p.[=,Met1733Glufs*78]) was detected in 3 probands in this analysis and was first reported on and functionally analysed by Braun *et al*.,[210] using patient derived keratinocytes. This variant was found to activate a cryptic 3' acceptor site located within intron 36, facilitating a 76 base pair pseudoexon inclusion, ultimately affecting the reading frame and resulting in the incorporation of a premature stop codon.

ABCA4 c.4539+2028C>T (p.[=,Arg1514Leufs*36]) was also first reported on by Braun *et al*.,[210] with functional analysis carried out by Albert *et al*.,[112]. Interestingly, when functionally analysed in patient derived fibroblasts, this variant did not appear to impact splicing or the levels of wild-type transcript in any way. However, when RNA extracted from patient derived PPCs was examined, it was determined that exonic splice enhancers attracted to the region in the presence of the variant led to the inclusion of a 345 nucleotide pseudoexon. The discrepancy between models indicated that retinal specific splice factors play a significant role in disease manifestation and that conclusions regarding pathogenicity cannot be drawn based on one model alone. Rather, variants such as this should be tested in the most accessible and time effective model first, midigene *in vitro* splice assays for example, followed by more sophisticated models such as PPCs and organoids should an effect not be found. A significant number of individuals recruited to

Target 5000 have been found to carry this variant with another pathogenic *ABCA4* variant, including two homozygous patients. This is discussed in detail in section 6.1.

65 cases remained unresolved following this study, of which ~43% (28/65) were clinically diagnosed with STGD1 and ~56% (37/65) with CRD. Of these 65 cases, 34% (22/65) were found to be monoallelic for variants in *ABCA4.* Of the 22 monoallelic cases, 64% (14/22) were clinically diagnosed as CRDs and 36% (8/22) were STGD1. Of the 43 zero *ABCA4* allele cases, 51% (22/43) were clinically diagnosed with CRD and 49% (21/43) with STGD1. These distinctions prove important because the potential causes of missing variability may vary between STGD1 cases and CRD cases. In the monoalleic cases, one explanation may be the relatively high carrier frequency of *ABCA4* variants in the general population, estimated to be approximately 5%[161]. It may be the case that variants in *ABCA4* are not causative of disease in these individuals and that they happen to carry a variant in *ABCA4* by chance. Bias towards this may indeed be high as unresolved individuals who were found to carry one pathogenic *ABCA4* variant by TCS were intentionally recruited to this study.

Another possibility is that some causal variants were not identified, for example, those that affect transcriptional regulation. Some regulatory *ABCA4* variants that appear to a have a quantitative effect on mRNA have been identified[65], however, these variants were not found in this study. It is possible that as-of-yet unidentified regulatory variants contribute to disease in the remaining unresolved cases. Furthermore, *in silico* tools employed to identify potentially pathogenic intronic variants may not predict the effect of retina specific splice defects, and thus, potentially pathogenic variants may have been overlooked. Alamut visual allows for the examination of retinal specific transcripts where such transcripts have been identified. While smMIPS represent an excellent sequencing method with robust coverage and depth, it is possible that this method may miss small heterozygous deletions (<400 base pairs) and cannot detect inversions that are larger than ~40 base pairs.

Another source of missing heritability is potentially due to the overlapping phenotypes observed in IRDs. Phenocopies of conditions can be particularly misleading. Therefore, causative variants in other IRD genes may be the actual cause of disease in these probands. This is particularly relevant for the unresolved CRD cases, as over 80 genes have been associated with central vision loss[23]. This is also evident from the significantly different positive candidate detection rate for STGD1 and CRDs in this study, at 73% and 29% respectively (Figure 5.5). The diagnostic yield for both STGD1 and CRDs

could be improved by employing a similar smMIPs based approach on a wider panel of maculopathy associated genes.

It should also be considered that the aetiology of disease in some of these cases may be more complex than previously thought. While IRDs are traditionally viewed as Mendelian diseases, the field is expanding to investigate the potential effect of complex inheritance, with a number of modifier loci already elucidated which will be discussed further in section 7.7. Indeed, a significant overlap between Mendelian macular disease and age-related macular disease (AMD) (MIM #603075) may exist, with some forms of AMD actually caused by very mild inherited macular disease associated variants. Conversely, it is possible that some presumed Mendelian macular diseases that remain unresolved are caused by variants in AMD associated genes. In a more complex scenario, it is possible that these two types of diseases may co-manifest or that variants in some macular genes may affect the function or expression of others in a polygenic manner.

The advantage of whole gene sequencing versus target capture or WES for diseases like STGD1 has been demonstrated by the above study. Namely, the detection of pathogenic deep-intronic, structural and CNVs is better facilitated by sequencing the entire gene. In addition, resources better channelled elsewhere are not used in sequencing genes not associated with such diseases. The main advantages of smMIPs-based whole gene sequencing as a technique by comparison with WGS are that firstly, smMIPs-based sequencing of a singular gene is far more cost effective than WGS, at approximately ~€28 per sample, which is crucial for cohorts in developing countries. Secondly, the volume of data produced by smMIPs is far lower than that produced by WGS, easing pressure on data storage and analysis demands. Thirdly, incidental or secondary findings that do not pertain to a genetic IRD diagnosis are avoided. In turn, simplifying the reporting process. Nonetheless, disadvantages of the smMIPs-based approach exist. Firstly, if including the non-coding proportion of the gene to be sequenced, as in this study, this technology is restricted to one or a few genes. Secondly, large scale employment of this technology is needed to keep costs low. Thirdly, this type of analysis does not readily identify inversions and insertions. Lastly, a highly specialised sequencing procedure and set up is required to employ smMIPs successfully.

A total of 10 probands in this study were found to carry DIVs (c.5196+1137G>A (n=3) and c.4539+2028C>T (n=7)) in *ABCA4* that lead to the inclusion of pseudoexons in the final mRNA transcript. An additional 18 individuals carrying c.4539+2028C>T have been identified in Target 5000. These variants are ideal targets for AON based therapy. AONs are short synthetic RNA and/or DNA molecules with chemistries to optimise

longevity that can interfere with the processing of pre-mRNA and thus can modulate splicing. The therapeutic potential of AONs is being readily explored for pseudoexon exclusion for *CEP290*[248–252] , *OPA1*[253], *CHM*[254] *USH2A*[255], and *ABCA4*[65,112,142,209,256,257] at a preclinical and clinical trial level. Specifically in relation to *ABCA4* and the variants observed in the current study, promising results have stemmed from a study on modulating the splice defect caused by *ABCA4* c.4539+2028C>T in PPCs[142], restoring correct *ABCA4* splicing in a mutation dependent manner. In addition, AONs designed to facilitate exon inclusion where the variant results in exon skipping are being studied. One such AON targets *ABCA4* c.4253+43G>A, identified in three probands in this study, resulting in 10% extra inclusion of exon 28 where this variant has resulted in its being skipped[142]. Indeed, AONs are becoming realistic therapeutic options for some IRDs given recent FDA/EMA approvals for splice modifying *AONs* for disorders including Duchenne muscular dystrophy and spinal muscular atrophy, among others[258].

Increasing the number of individuals who undergo this method of sequencing will provide better insights into the prevalence of intronic variants in the Irish STDG1/CRD cohort and will possibly facilitate a genetic diagnosis where one could not have been provided through Target 5000 TCS alone. Employing this smMIPs sequencing approach on unscreened cases is more cost effective than TCS and could be used as a first pass measure before widening the net of IRD genes sequenced to determine the genetic diagnosis. Furthermore, as additional non-coding variants in *ABCA4* are discovered, and an estimate made as to their prevalence in the Irish population, the target capture panel can be amended to facilitate detection by this means also. Adding these regions to the target capture panel is important to establish the prevalence of these variants not only in the STGD1/CRD population, but also in the Irish IRD population as a whole in order to work towards establishing the prevalence of *ABCA4* related STGD1/CRD in Ireland.

Future work will place an emphasis on the identification and functional analysis of putative cis-regulatory variants in *ABCA4.* Identification of such variants relies heavily on large sequencing datasets, in order to identify potentially enriched and rare variants in monoallelic cases and cases with no *ABCA4* variants detected. Analysis of data from any remaining unresolved cases will be carried out to determine if any rare or shared variants are present upstream of *ABCA4.* Once a variant is identified, assessment of ChIP-Seq data[259], Circularised Chromosome Conformation Capture[260], and RNA-seq data (ENCODE)[261] will be undertaken to determine if the region in which this variant is located may contribute significantly to levels of gene expression. If so, DNA segments containing the putative cis regulatory variants and corresponding wild-type sequence will be

amplified, cloned into expression vectors and analysed via a Luciferase assay, to determine their effect on gene expression.

Another avenue to potentially identify non-coding variants that may have been overlooked in the analyses thus far (as discussed in section 4.5.2) is that of retinal organoids. These are three dimensional structures derived from human pluripotent stem cells that better recapitulate many aspects the human retina in terms of both structure and function than other commonly used cell models. This form of analysis facilitates the discovery of variants that impact splicing due to changes in splice factors that are specific to the retina itself, that may not appear as viable candidates through *in silico* analyses alone. It would be of interest to survey remaining unresolved cases, obtain relevant tissue samples from willing participants and to collaborate with other laboratories to form a combined sequencing and retinal organoid based approach to provide these patients with a genetic diagnosis. This type of approach was elegantly demonstrated by Bronstein *et al.,*[262] wherein they discovered a novel pathogenic DIV in *CNGB3*. Using a combined WGS and RNA-sequencing approach of the retinal organoid transcriptome, they determined the cause of retinal disease in previously unresolved family, facilitating access to an ongoing clinical trial.

smMIPs-based analysis of the complete sequence(s) of one or a few genes implicated in a clinically well-defined phenotype allows for cost and time effective candidate variant detection. Of 162 STGD1/CRD probands sequenced in this study a genetic diagnosis was ascertained in 97 cases using smMIPs to target the entire *ABCA4* gene. 27% these now resolved cases carried an intronic *ABCA4* variant, highlighting the need to sequence both intronic and coding regions to fully define the genetic architecture underlying STGD1/CRDs in Ireland. Such non-coding variants highlight the need for continued functional studies to be able to provide a more definitive categorisation in terms of pathogenicity of these novel variants. Future efforts will focus on utilising sequencing approaches similar to smMIPs to retrospectively analyse more unresolved and unscreened cases while also determining the sources of missing heritability in this cohort. As AON-based therapies continues to show promising results, identifying such variants may potentially facilitate access to a future therapeutic option where appropriate. More generally, dual AAV-vector *ABCA4* gene replacement therapies are in development at a number of sites and progressing towards human clinical trial for STGD1 again highlighting the importance of a definitive genetic diagnosis for STGD1 patients[263].

6. Case studies further illustrating the disease-causing potential of traditionally overlooked splice altering variants.

6.1 Enrichment of ABCA4 c.4539+2028C>T in Stargardt disease patients in Ireland - a detailed genotype-phenotype analysis

6.1.1 Introduction

As detailed in the previous section, autosomal recessive STGD1 is primarily caused by biallelic variants in the *ABCA4* gene[141]. *ABCA4* encompasses 50 exons, encoding a protein 2,273 amino acids in length. It is expressed in rod, cone and RPE cells [264]. It functions as a flippase for 11-cis and all-trans isomers of N-retinylidene-phosphatidylethanolamine at the rim of rod and cone outer segment disks, allowing for the removal of potentially toxic retinoids[265–267]. If ABCA4 function is impaired, these by-products and other bisretinoids cumulatively termed lipofuscin, accumulate in the RPE, causing dysfunction and cell death followed by impairment and eventually loss of photoreceptors[268–271].

As *ABCA4* is a large gene, comprising 128,315 nucleotides, traditional methods of screening have primarily utilized an exon focused sequencing approach similar to that outlined in section 3. The aim of such studies is to identify disease causing protein coding variants. However, these methods do not ascertain the cause of disease in all cases[234,247,272,273]. This highlights the need to broaden the sequencing approach to include non-coding regions of *ABCA4* in order to fully understand the landscape of *ABCA4* related disease[54]. Studies that employ this methodology have identified several novel non-coding pathogenic *ABCA4* variants[65,112,142,229,233,274,275]. Variants detected in the NCSS regions can lead to exon skipping or perturbation of the natural length of the exon[244]. In addition, many non-coding variants are located deep within introns and can result in the incorporation of a pseudoexon into the mature mRNA transcript due to the recruitment of splicing machinery to cryptic splice sites[54,210]. These pseudoexons often disrupt the reading frame or contain premature stop codons, potentially causing the transcript to undergo nonsense mediated decay[276].

The aberrant splicing patterns identified as a result of intronic variants often do not occur in all transcripts, with a degree of wild-type mature mRNA produced in some incidents[54]. This can depend on the level of similarity of the model utilized for functional analyses to the retina[112,209,277]. Thus, considerable attention should be paid to both the model and the ratio of wild type to mutant transcript observed in order to understand

disease manifestation and variant severity. However, it is imperative that follow up clinical studies are carried out to ascertain the true contribution of these variants to disease. Examples include *ABCA4* c.5461-10T>C which has been classified as severe based functional studies using patient derived fibroblasts[108] and PPCs as well as the clinical presentation in four homozygous individuals[278]. Additionally, *ABCA4* c.5196+1137G>A was examined in keratinocytes derived from patients heterozygous for this variant resulting in a 73 nucleotide pseudoexon inclusion[210]. It has since been examined in patient-derived PPCs and has been categorized as a variant of intermediate severity following detailed genotype-phenotype analysis in a substantial cohort of patients[209].

A landmark study on *ABCA4* to which we contributed recently provided the most comprehensive analysis of genetic variation in the entire *ABCA4* gene to date[54]. In this study, seven patients harboured *ABCA4* c.4539+2028C>T, p.[=,Arg1514Leufs*36] of which, five were patients recruited from Ireland. In the current study, through retrospective and continued analyses of the Irish STGD1 cohort, a total of 25 individuals have been identified to date, including two individuals who carry this variant homozygously. This, in conjunction with an evaluation of population controls from Ireland, suggests significant enrichment of this variant in the Target 5000 STGD1 cohort. *ABCA4* c.4539+2028C>T was first reported by Braun *et al.,*[210] with functional analyses carried out by Albert *et al*[112]. RNA extracted from patient-derived PPCs was examined with a 345-nucleotide pseudoexon inclusion observed in ~15% of all transcripts and ~30% of transcripts from this allele. A comprehensive clinical characterization of a group of patients with *ABCA4* c.4539+2028C>T is provided in this section. Impact of the variant on disease severity is evaluated through assessment of the phenotype of homozygous patients versus those who carry the variant *in trans* with a null or null-like variant. In addition, we phenotypically compare these individuals to those who carry *ABCA4* variants of known severity *in trans* with a null or null-like variant[279] in order to determine clinical severity of *ABCA4* c.4539+2028C>T .

6.1.2. Materials and methods

6.1.2.1 Participants

Participants were recruited through The Royal Victoria Eye and Ear Hospital and The Mater Misericordiae University Hospital, Dublin as part of the Target 5000 study[116]. The participant cohort consisted of 25 individuals across 18 pedigrees clinically diagnosed with STGD1 and found to possess at least one allele harboring the *ABCA4* c.4539+2028C>T variant. 13 unaffected relatives were also recruited in order to determine

the phase of variants identified. In addition, 15 affected participants carrying two *ABCA4* variants of previously categorised severity and null/null-like variants[279] were recruited from Target 5000 and their phenotypic data used for comparative purposes to determine the severity the *ABCA4* c.4539+2028C>T variant. Previously published data on age at onset from 56 affected individuals carrying variants of known severity was also included in this study. Of the 25 participants with *ABCA4* c.4539+2028C>T, 13 were female while 12 were male, with a median age at last examination of 38 years (range:13-70). All patients were of Irish ethnicity.

6.1.2.2 Clinical assessment

Clinical assessment was performed by Dr Kirk A.J. Stephenson, Dr Julia Zhu, Ms Emma Duignan, Mr David J. Keegan and Mr Paul F. Kenna. An excel template was provided to all clinicians to provide clinical data for the study. All study participants had a clinical diagnosis of either STGD1 or CRD. Informed consent was obtained from each participant and the appropriate assessments were performed in order to determine the patient's phenotype. Clinical data collected and analysed included sex, ethnicity, age at last examination, age at onset of disease, visual acuity (Revised 2000 early treatment diabetic retinopathy study (ETDRS) charts, Precision Vision, La Salle, IL, USA), visual fields (Goldmann perimeter 940 (iv4e, i4e and 04e targets), Haag-Streit AG, Köniz, Switzerland), colour vision (Lanthony desaturated D-15 panel under standardised lighting conditions, Gulden Ophthalmics, Elkins Park, PA, USA), dilated slit lamp biomicroscopy, colour fundus photography and fundus autofluorescence (FAF) (Topcon CRC50DX/Optos Daytona, Topcon Great Britain Ltd., Berkshire, England/Optos plc, Dunfermline, Scotland), optical coherence tomography (OCT) (Cirrus 5000, Carl Zeiss Meditec, Berlin, Germany), and electroretinography (ERG) Roland Consult RETI-port retiscan, Brandenburg an der Havel, Germany), where available. The least affected reading from the most recent patient assessment was used for data illustration purposes in all cases. Age at examination is also provided.

6.1.2.3 Sequencing procedures

13 affected individuals with *ABCA4* c.4539+2028C>T underwent target capture NGS of the exons and known pathogenic intronic regions of *ABCA4* as described previously[13,118,280], 9 affected individuals with *ABCA4* c.4539+2028C>T underwent whole-gene single molecule molecular inversion probe (smMIP) based sequencing of *ABCA4* as well as 40 kb of flanking sequence as described previously[54], including both homozygotes. Two individuals with *ABCA4* c.4539+2028C>T underwent direct Sanger sequencing. (Eurofins Genomics, Germany). The remaining individual with c.4539+2028C>T underwent WGS[51]. WGS was performed by BGI (Hong Kong, China) on a BGISeq500 using 2x100bp paired end reads, with a 30-fold minimal median coverage per genome. BWA was utilized to map data to the human genome (GRCh37). The quality of the WGS data was based on insert size, percentage mapped reads, percentage duplicated mapped reads, coverage, bases with >20X coverage and error rate, which were evaluated using Qualimap V.2.2.1[189]. Variant calling was performed by xAtlas V.0.1 and annotated in-house with data such as variant type, *in silico* pathogenicity prediction scores and

population frequency. Variant validation was carried out by amplifying the surrounding region via PCR This was followed by Sanger sequencing (Eurofins Genomics, Germany). Primers for use in this procedure were manufactured by Sigma Aldrich (Gillingham, England, UK). New England Biolabs Inc.'s (Ipswich, MA, USA) Q5 High-Fidelity 2× Master Mix was used to amplify the regions surrounding the variant. Where available, relatives samples were processed similarly using either the Phusion Human Specimen Direct PCR Kit or the Phusion Blood Direct PCR Kit (Thermo Scientific, MA, USA) where appropriate. The relevant accession numbers for *ABCA4* in this study are NM_000350.2 and NC_000001.11.

6.1.2.4 Variant categorisation

Variants designated "mild", "intermediate" or "null-like" were based on a previous comprehensive study by Fakin *et al*[279]. Variants categorised as intermediate: *ABCA4* c.6079C>T,p.(Leu2027Phe), c.4139C>T,p.(Pro1380Leu), c.3113C>T,p.(Ala1038Val) and c.3322C>T,p.(Arg1108Cys). Variants categorised as mild: *ABCA4* c.5882G>A,p.(Gly1961Glu) and c.5603A>T,p.(Asn1868Ile). Variants categorised as null or null-like: c.6449G>A,(p.Cys2150Tyr), c.2564G>A,p.(Trp855*), c.4469G>A,p.(Cys1490Tyr), c.1906C>T,(p.Gln636*), c.5312+1G>A,p.(?), c.2160+1G>C,p.(?), c.4577delC,p.(Thr1526Argfs*10), c.1222C>T,(p.Arg408*), c.4234C>T,(p.Gln1412*), c.4320delT,p.(Phe1440Leufs*6), c.2041C>T,p.(Arg681*), c.735T>G,p.(Tyr245*), and c.3329-1G>A,p.(?) Variant categorisation of c.4539+2028C>T was carried out by comparison of participants homozygous for c.4539+2028C>T to participants harbouring c.4539+2028C>T *in trans* with a null or null-like variant to determine the effect of the variant in isolation. In addition, participants harbouring c.4539+2028C>T *in trans* with a null or null-like variant were phenotypically compared to those who carry intermediate or mild variants *in trans* with a null or null-like variant to gauge severity.

6.1.2.5 Statistical analyses and data visualisation.

Statistical analyses and data visualisation were performed using R version 4.0.3 (R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.) The median age at onset of those harboring *ABCA4* c.4539+2028C>T were compared with the age at onset of individuals who carried two null/null-like variants, those who carried mild variants *in trans* with a null or null-like variants and those who carried intermediate variants *in trans* with a null or null-like variants previously reported[279] using a Kruskal-

Wallis ranked sum test with post hoc analysis carried out using a pairwise Wilcoxon rank sum test, correcting for multiple testing using the Holm-Bonferroni method. The association between changes in vision tests and age as well as symmetry between eyes were assessed using Spearman's rank correlation coefficient.

6.1.3 Results

6.1.3.1 ABCA4 c.4539+2028C>T is enriched in the Irish STGD1 population.

27 alleles of *ABCA4* c.4539+2028C>T have now been identified in the *ABCA4* sequenced Irish STGD1 population. Sex and genotype information for these patients is provided in Table 6.1.1. This includes two homozygous individuals, resulting in a total of 25 patients carrying *ABCA4* c.4539+2028C>T in Ireland to date. A total of 15 other incidents of the variant have been identified across 5 STGD1 studies in 3,940 alleles tested globally[54,143,210,232,274] (Figure 6.1.1). *ABCA4* c.4539+2028C>T has a population frequency of 0.00003943 in gnomAD (v3.1)[145], totaling 6/152,182 alleles. A population frequency of 0.000044 was observed in a recent study of 5579 bi-allelic STGD1 patients[281]. Of the 27 alleles of *ABCA4* c.4539+2028C>T, 24 of these had the variant ABCA4 c.302+68C>T detected concurrently. *ABCA4* c.302+68C>T has an AF of 0.00004598 in gnomAD (v3.1)[145]. The complex allele consisting of these two variants has been previously reported several times[210,274,282]. The numbers of carriers of the complex allele and the variant of interest only in this study were not sufficient to determine if the presence of the second variant had an additional effect on the severity of the phenotype. In addition, *ABCA4* c.302+68C>T does not have a qualitative effect on the resulting mRNA transcript in PPCs[112]. Furthermore, to explore the frequency of the *ABCA4* c.4539+2028C>T and *ABCA4* c.302+68C>T allele in the Irish population an unaffected control cohort previously subjected to WGS has been interrogated, and a single allele harboring both variants has been observed in the 408 healthy individuals evaluated to date. Of note, the two individuals that are homozygous for *ABCA4* c.4539+2028C>T are the only homozygotes observed to date globally. Homozygotes for this or any *ABCA4* variant, in principle provide an opportunity to explore the functional consequences of variants in patients

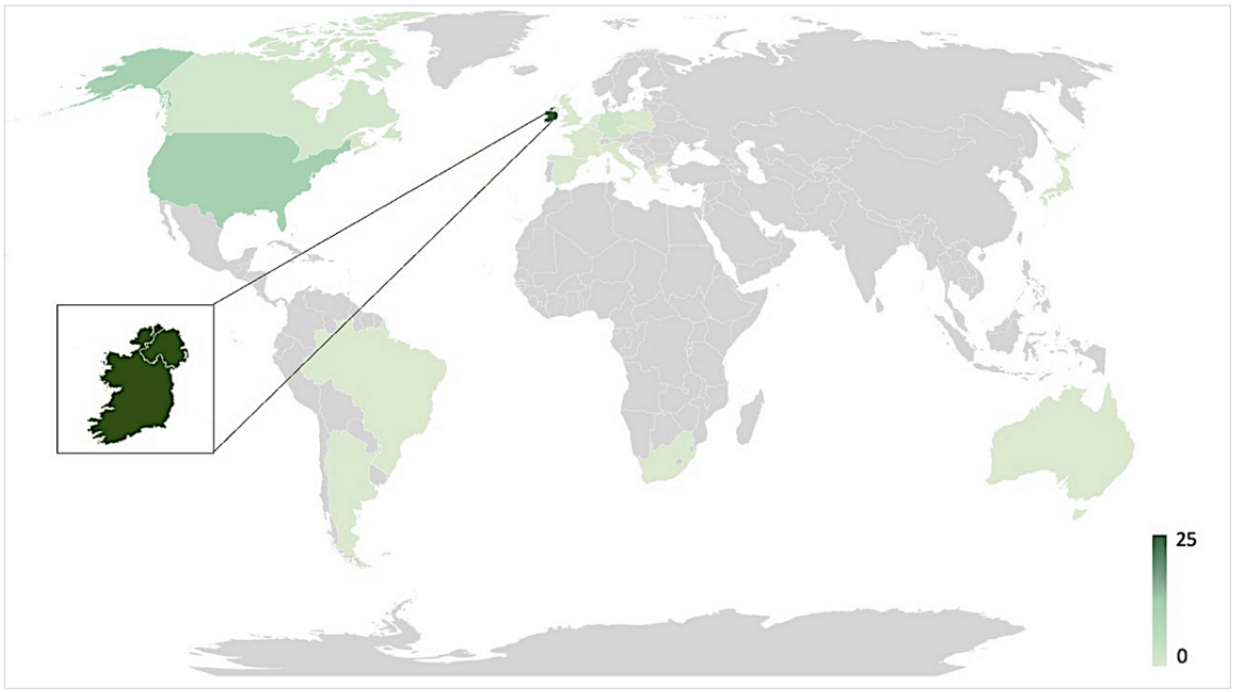


Figure 6.1.1 | Distribution of *ABCA4* c.4539+2028C>T cases worldwide from publications and the Target 5000 database. The lightest shade of green represents countries where the region harboring the variant has been sequenced, but no cases were identified. 1 case was detected in Canada, 3 cases were detected in German cohorts, 11 in cohorts in the United States and 25 cases in Ireland including two homozygous individuals[54,143,210,232,274]. Data has not been published for countries shaded in grey.

Table 6.1.1 | Patient ID, Pedigree, sex, and genotype information of all individuals harboring *ABCA4* c.4539+2028C>T, p.[=,Arg1514Leufs*36]. †= Homozygous individuals; ‡ = those who carry a null or null-like variant; M=Male; F=Female. The relevant accession numbers for *ABCA4* are NM_000350.2 and NC_000001.11.

Patient ID	Pedigree	Sex	Allele 1		Allele 2	
			cDNA	Protein	cDNA	Protein
Pt-1‡	1	M	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.2160+1G>C	p.(?)
Pt-2	2	F	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.[4222T>C;4918C>T]	p.[Trp1408Arg;Arg1640Trp]
Pt-3	2	M	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.[4222T>C;4918C>T]	p.[Trp1408Arg;Arg1640Trp]
Pt-4	2	M	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.[4222T>C;4918C>T]	p.[Trp1408Arg;Arg1640Trp]
Pt-5	3	M	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.3322C>T	p.(Arg1108Cys)
Pt-6	4	F	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.6472A>G	p.(Lys2158Glu)
Pt-7	5	M	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.2453G>A	p.(Gly818Glu)
Pt-8	6	M	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.3322C>T	p.(Arg1108Cys)
Pt-9‡	7	F	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.3754G>T	p.(Glu1252*)
Pt-10†	8	M	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.4539+2028C>T	p.[=,Arg1514Leufs*36]
Pt-11	9	F	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.4139C>T	p.(Pro1380Leu)
Pt-12‡	10	M	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.6449G>A	p.(Cys2150Tyr)
Pt-13	11	F	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.161G>A	p.(Cys54Tyr)
Pt-14	12	M	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.[4222T>C;4918C>T]	p.[Trp1408Arg;Arg1640Trp]
Pt-15	12	F	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.[4222T>C;4918C>T]	p.[Trp1408Arg;Arg1640Trp]
Pt-16	12	F	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.[4222T>C;4918C>T]	p.[Trp1408Arg;Arg1640Trp]
Pt-17‡	13	F	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.3212C>A	p.(Ser1071*)
Pt-18	14	F	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.4139C>T	p.(Pro1380Leu)
Pt-19	15	M	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.223T>G	p.(Cys75Gly)
Pt-20	15	M	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.223T>G	p.(Cys75Gly)
Pt-21	16	M	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.4139C>T	p.(Pro1380Leu)

Pt-22	17	F	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.6743T>C	p.(Phe2248Ser)
Pt-23[†]	18	F	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.4539+2028C>T	p.[=,Arg1514Leufs*36]
Pt-24	18	F	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.[4222T>C;4918C>T]	p.[Trp1408Arg;Arg1640Trp]
Pt-25	18	F	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	c.[4222T>C;4918C>T]	p.[Trp1408Arg;Arg1640Trp]

6.1.3.2 Clinical characterisation of individuals carrying ABCA4 c.4539+2028C>T

Firstly, it is important to note that when using patient data, data sets are often imperfect, however, this study was undertaken given the presence of the only two homozygotes globally with this *ABCA4* splice altering variant to try to shed some light on the severity or otherwise of this mutation compative to other *ABCA4* mutations. Detailed phenotypic data including visual acuity, FAF, OCT, ERG and age at onset was collected from all 25 patients harboring *ABCA4* c.4539+2028C>T, where available, including two homozygous individuals. The median age at last examination was 38 (range = 13-70) years. In order to determine the severity of *ABCA4* c.4539+2028C>T, two indivduals who carry this variant in a homozygous manner were compared to four individuals who carry this variant *in trans* with a null or null-like variant. The phenotype of these four individuals allows surveyance of the effect of *ABCA4* c.4539+2028C>T without residual function from the other allele, while the clinical presentation of the homozygous individuals highlights the direct phenotypic outcome of *ABCA4* c.4539+2028C>T without the presence of any additional pathogenic *ABCA4* variants. The individuals who carry this variant *in trans* with a null or null-like variant were also compared to individuals who carry variants of known severity *in trans* with a null or null-like variant as described by Fakin *et al*[279]. Details of these variants are included in the methods section. These comparisons were performed in order to gauge the pathogenicity of *ABCA4* c.4539+2028C>T on the STGD1 severity spectrum.

6.1.3.2.1 Visual acuity

The visual acuity of all individuals harbouring *ABCA4* c.4539+2028C>T, where available, can be found in Table 6.1.2. The median visual acuity of the homozygotes was 6/33.5 at a median age of 47.5, with the female having visual acuity of 6/48 at age 58 and the male having a visual acuity of 6/19 at age 37. By comparison, the median visual acuity of those who harboured *ABCA4* c.4539+2028C>T *in trans* with a null or null-like variant was 6/48 (range = 6/36-60, n=4) at a median age of 36.5 (range = 16-41, n=4). In addition, the median visual acuity of those who carried mild variants *in trans* with a null or null-like variant was 6/60 (range = 6/7.5-60, n=9) at a median age of 33 (range = 17-50, n=9). The median visual of those who carried intermediate variants *in trans* with a null or null-like variant was 6/60 (range = 6/48-75. n=4) at a median age of 22.5 (range = 13-54, n=4). Figure 6.1.2 illustrates best corrected visual acuity (BVCA) in the right eye vs age (A) and the left eye vs age (B). Spearman's correlation showed a signifiant correlation between

BVCA and age in the left eye only. (Spearman's r_s: right eye $r_s = 0.39572$, p (2-tailed) = 0.0556, left eye $r_s = 0.45256$, p (2-tailed) = 0.02638). A significant correlation between the BVCA of both eyes was found ($r_s = 0.8501$, p (2-tailed) = 0) indicating a high degree of symmetry.

Table 6.1.2 | Pedigree numbers and visual acuity of all individuals harbouring *ABCA4* c.4539+2028C>T. †= Homozygous individuals; ‡ = those who carry a null or null-like variant; CF = counting fingers; BVCA = best corrected visual acuity; OD = right eye; OS = left eye, /// = not available.

Patient ID	Pedigree	BCVA - OD	BCVA – OS
Pt-1‡	1	6/36	6/36
Pt-2	2	6/90	6/75
Pt-3	2	6/48	6/48
Pt-4	2	6/120	6/600
Pt-5	3	6/6	6/4.5
Pt-6	4	6/36	6/36
Pt-7	5	6/36	6/120
Pt-8	6	6/90	6/75
Pt-9‡	7	6/60	6/36
Pt-10†	8	6/19	6/19
Pt-11	9	6/48	6/48
Pt-12‡	10	6/60	6/60
Pt-13	11	CF	6/60
Pt-14	12	6/48	6/30
Pt-15	12	6/48	6/48
Pt-16	12	6/48	6/48
Pt-17‡	13	6/60	6/60
Pt-18	14	6/60	6/60
Pt-19	15	6/60	6/48
Pt-20	15	6/60	6/60
Pt-21	16	6/38	6/38
Pt-22	17	///	///
Pt-23†	18	6/60	6/48
Pt-24	18	6/38	6/38
Pt-25	18	6/7.5	6/7.5

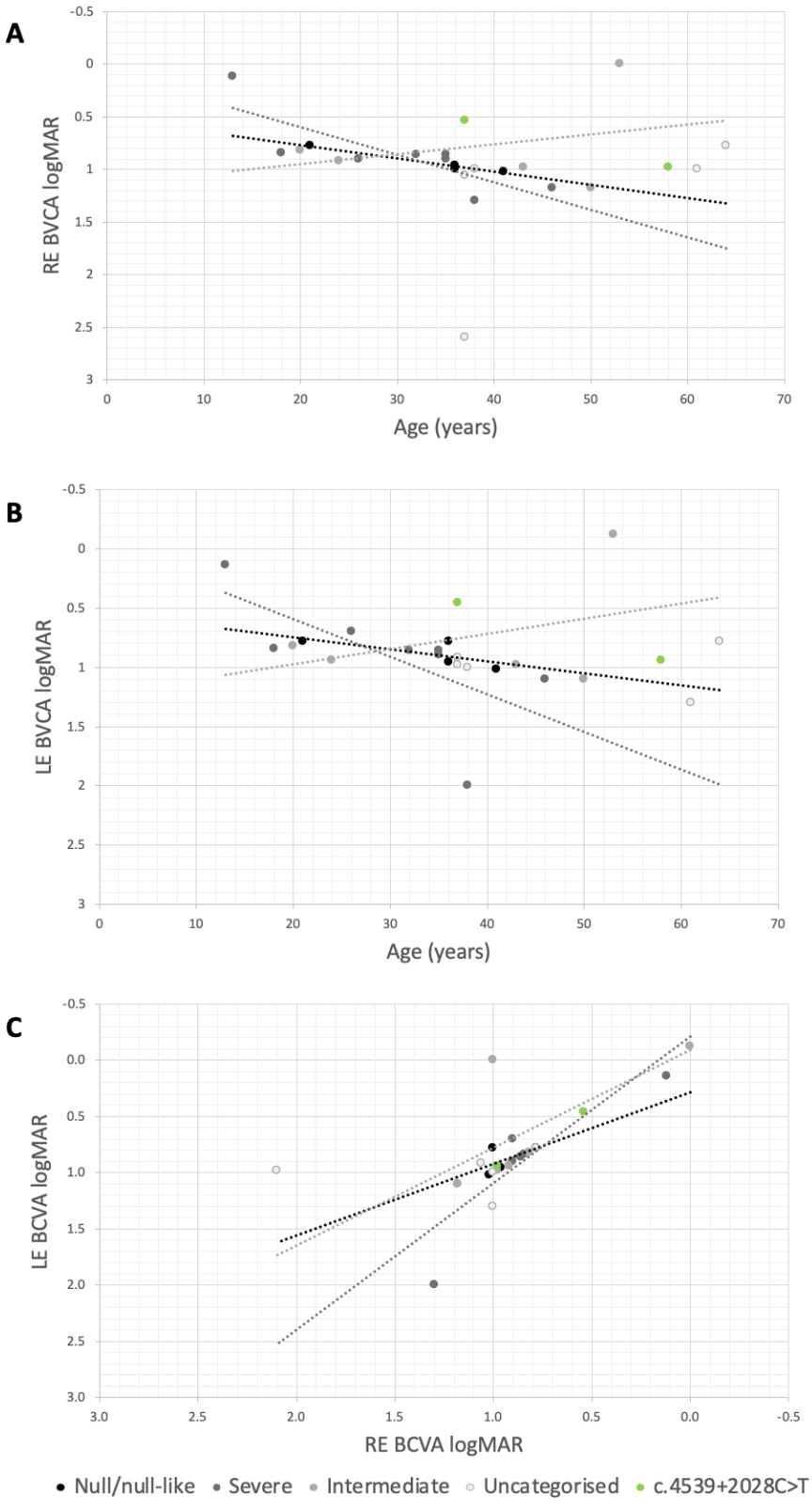


Figure 6.1.2 | Scatterplots illustrating (A) best corrected visual acuity (BCVA) of the right eye (RE) vs age, (B) BCVA in the left eye (LE) vs age and (C) BCVA in the RE vs the LE. C illustrates the degree of symmetry between the eyes. All patient's carry c.4539+2028C>T and the color legend represents the severity categorisation of the disease causing variant *in trans*. Green circles represent patients homozygous for c.4539+2028C>T.

6.1.3.2.2 Fundus autofluorescence and optical coherence tomography.

All individuals exhibited FAF abnormalities (Figure 6.1.3). In 13/25 individuals, this abnormality extended beyond the vascular arcades. 11/25 individuals were designated a group 1 Fishmann classification[283], including both homozygous individuals. 9/25 individuals were classified as group 2, four individuals were classified as group 3 and one individual was classified as group 4. A distinctive well-circumscribed circular area of profound atrophy on auto-fluorescent imaging was observed in the majority of the 25 individuals harboring *ABCA4* c.4539+2028C>T, including both homozygous individuals. 20/25 had peripapillary sparing. Peripheral retinal yellow flecks were noted in 9/25 individuals, while central flecks were seen in 23/25 individuals. Of those without flecks, one case (EEJ13, age 56, Figure 6.1.3) had advanced, diffuse retinal atrophy without flecks and the second (EECC41, age 14, Figure 6.1.3) had early foveal RPE disturbance only. In addition, 20/25 individuals had a beaten-bronze appearance and 19/25 exhibited a bull's eye maculopathy pattern. Symmetry was noted in 21/25 individuals. Details of FAF analysis are provided in Table 6.1.3.

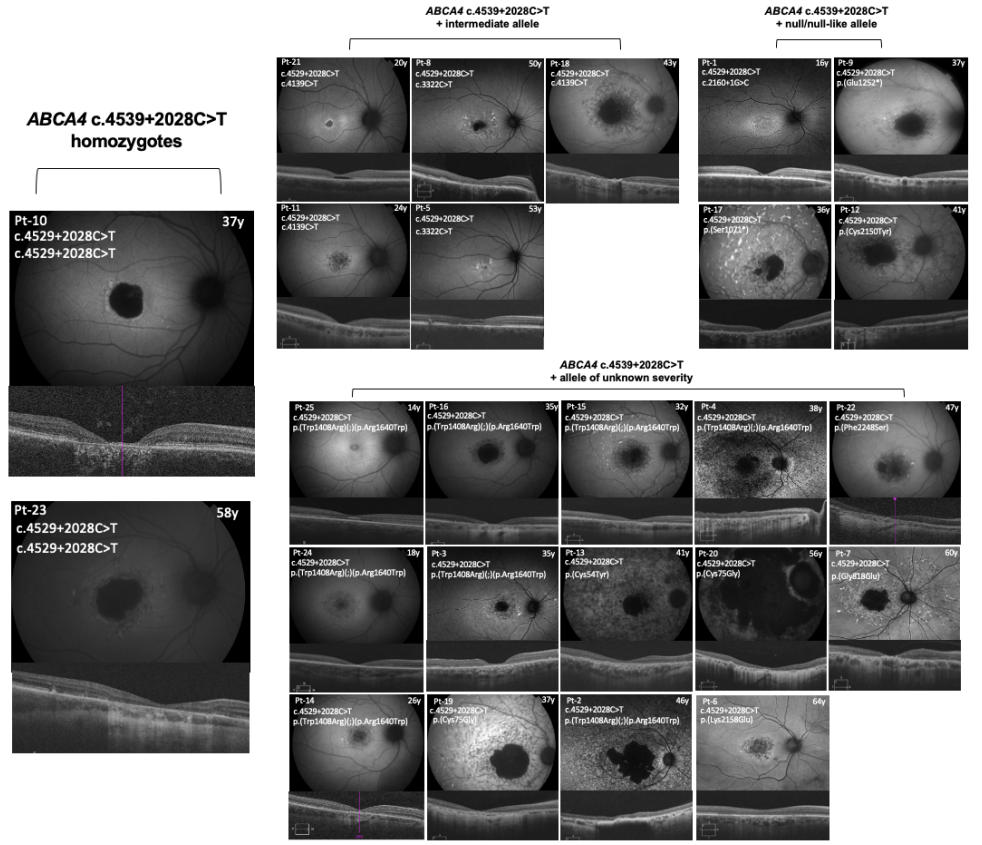


Figure 6.1.3 | Fundus autofluorescence (FAF) and optical coherence tomography (OCT) analysis of all individuals harboring *ABCA4* c.4539+2028C>T. FAF and OCT is shown for the right eye of each patient, with variant details and age at time of image capture overlaid. c.4539+2028C>T homozygotes show focal, well-circumscribed macular hypoautofluorescence surrounded by minimal retinal flecks (mean horizonal width of atrophy on OCT 2596µm). c.4539+2028C>T with an *ABCA4* allele of intermediate severity manifested retinal changes confined mainly to the perifoveal area, whereas c.4539+2028C>T in conjunction with a null/null-like allele showed more widespread disease, extending beyond the arcades in older probands. c.4539+2028C>T with alleles of unknown pathogenic potential show a range of distribution of retinal pathology, however, the majority have central profound hypoautofluorescence as seen in the c.4539+2028C>T homozygotes.

Table 6.1.3 | Detailed fundus examination of all individuals harboring *ABCA4* c.4539+2028C>T. †= Homozygous individuals; ‡ = those who carry a null or null-like variant; WRT = with respect to. Y=yes, N=no.

Patient ID	Pedigree	FAF WRT Vascular Arcades	Bull's Eye Pattern	Beaten Bronze Appearance	Yellow Flecks Centrally	Flecks Peripherally	Peripapillary Sparing
Pt-1‡	1	within	Y	Y	Y	N	Y
Pt-2	2	beyond	N	Y	Y	Y	Y
Pt-3	2	beyond	Y	Y	Y	N	Y
Pt-4	2	beyond	N	N	Y	Y	N
Pt-5	3	within	Y	Y	Y	N	Y
Pt-6	4	within	Y	Y	Y	N	Y
Pt-7	5	beyond	N	Y	Y	N	Y
Pt-8	6	within	Y	Y	Y	N	Y
Pt-9‡	7	beyond	N	Y	Y	Y	Y
Pt-10†	8	within	Y	N	Y	N	Y
Pt-11	9	within	Y	Y	Y	N	Y
Pt-12‡	10	beyond	Y	Y	Y	Y	Y
Pt-13	11	beyond	Y	Y	Y	Y	N
Pt-14	12	beyond	Y	Y	Y	N	Y
Pt-15	12	within	Y	Y	Y	Y	N
Pt-16	12	beyond	Y	Y	Y	N	Y
Pt-17‡	13	beyond	Y	Y	Y	Y	N
Pt-18	14	beyond	Y	Y	Y	N	Y
Pt-19	15	beyond	N	N	Y	Y	Y
Pt-20	15	beyond	N	N	N	Y	N
Pt-21	16	within	Y	Y	Y	N	Y

Pt-22	17	within	Y	Y	Y	N	Y
Pt-23[†]	18	within	Y	Y	Y	N	Y
Pt-24	18	within	Y	N	N	N	Y

The foveal photoreceptors were affected on OCT in all individuals, with none exhibiting foveal sparing (Figure 6.1.3). One participant displayed only mild foveal deterioration, however this individual was 13 years old at the time of multimodal imaging and the severe manifestations of the phenotype may not yet be fully apparent at this age (Figure 6.1.3, Pt-25). It is of note that 16/25 patients exhibited profound foveal atrophy (including both homozygotes), 8/25 presented with an intermediate macular phenotype and 1/25 (Pt-25) with a mild macular phenotype when OCT and FAF phenotype were surveyed. In contrast, 1/25 (Pt-20) 6/25, 3/25 and 14/25 (including both homozygotes) exhibited severe, intermediate, mild and normal peripheral phenotypes respectively. The median central retinal thickness (CRT) of all individuals harboring ABCA4 c.4539+2028C>T was 121 microns (range = 73-256, n=25). The median horizontal OCT atrophy area was 3123 microns (range = 860-6000, n=25). Data for each individual can be found in Table 6.1.4. The median CRT, OCT atrophy area, age and sample size for the c.4539+2028C>T homozygotes, patients who carried *ABCA4* c.4539+2028C>T *in trans* with a null or null-like variant, patients who carried mild variants *in trans* with a null or null-like variant and patients who carried intermediate *in trans* with a null or null-like variant can be found in Table 6.1.5. The homozygous individuals have a less severely affected median CRT and OCT atrophy area at a later median age than all three other groups. The patients who carried *ABCA4* c.4539+2028C>T *in trans* with a null or null-like variant had median CRT and OCT atrophy areas comparative to that of patients who carried the mild and intermediate variants *in trans* with a null or null-like variant. Figure 6.1.4 illustrates CRT in the right eye vs age (A) and the left eye vs age (B). A significant correlation between either eye and age was not detected (Spearman's r_s: right eye r_s = -0.17544, p (2-tailed) = 0.40157, Left eye -0.16194, p (2-tailed) = 0.44964.) A significant correlation between the CRT of both eyes was found (Spearman's r_s: r_s = 0.64216, p (2-tailed) = 0.00072) indicating a high degree of symmetry. Figure 6.1.5 illustrates OCT atrophy area in the right eye vs age (A) and the left eye vs age (B). A significant correlation was detected between OCT atrophy area in both eyes and increasing age (Spearman's r_s: right eye r_s = 0.58994, p (2-tailed) = 0.00241, left r_s = 0.60415, p (2-tailed) = 0.00177). Additionally a significant correlation between the OCT atrophy are of both eyes was detected (r_s = 0.93129, p (2-tailed) = 0.

Table 6.1.5 | Median central retinal thickness (CRT), optical coherence tomography (OCT) atrophy area (AA) and age for patients homozygous for *ABCA4* c.4539+2028C>T and who carry *ABCA4* c.4539+2028C>T *in trans* with a null or null-like variant, mild variants *in trans* with a null or null-like variant and intermediate *in trans* with a null or null-like variant. R = range.

Genotype	Median CRT (microns)	Median OCT AA Horizontal Width (microns)	Median Age (years)	Sample
c.4539+2028C>T Homozygotes	164.5 (R:73-256)	2401.6 (R:2182-2621)	47.5 (R:37-58)	2
c.4539+2028C>T + null/null-like	122 (R:72-164)	3908 (R:1064-4919)	36.5 (R:21-60)	4
Mild + null/null-like	124 (R:87-195)	2480 (R:1470-6000)	33 (R:13-54)	5
Intermediate + null/null-like	116 (R:83-127)	5444 (R:1933-6000)	38.5 (R:17-50)	9

Table 6.1.4 | Examination of central retinal thickness (CRT) and optical coherence tomography (OCT) atrophy area in all individuals harboring *ABCA4* c.4539+2028C>T. 6000 microns was the maximum horizontal width reading measurable by OCT and implies a larger field of outer retinal atrophy. †=Homozygous individuals. ‡ = those who carry a null or null-like variant; /// = not available. OD = oculus dextrus, right eye. OS = oculus sinister, left eye.

Patient	Pedigree	CRT - OD	CRT - OS	OCT Atrophy Horizontal Width OD	OCT Atrophy Horizontal Width OS
Pt-1‡	1	162	164	1732	1064
Pt-2	2	140	90	6000	6000
Pt-3	2	109	///	2824	3467
Pt-4	2	176	87	6000	6000
Pt-5	3	196	210	2322	2615
Pt-6	4	94	98	3604	3257
Pt-7	5	100	117	6000	5446
Pt-8	6	117	100	2990	3735
Pt-9‡	7	120	123	4608	4473
Pt-10†	8	73	66	2189	2182
Pt-11	9	121	119	2234	2162
Pt-12‡	10	99	121	4996	4919
Pt-13	11	109	106	6000	6000
Pt-14	12	168	179	2436	2623
Pt-15	12	104	111	3360	2748
Pt-16	12	106	118	3739	3585
Pt-17‡	13	68	72	4630	3344
Pt-18	14	65	91	5099	4965
Pt-19	15	184	74	4536	3828
Pt-20	15	99	106	6000	6000
Pt-21	16	189	182	1055	1157
Pt-22	17	197	249	4290	4999
Pt-23†	18	225	256	2621	3392
Pt-24	18	115	121	2072	2805
Pt-25	18	200	198	860	1106

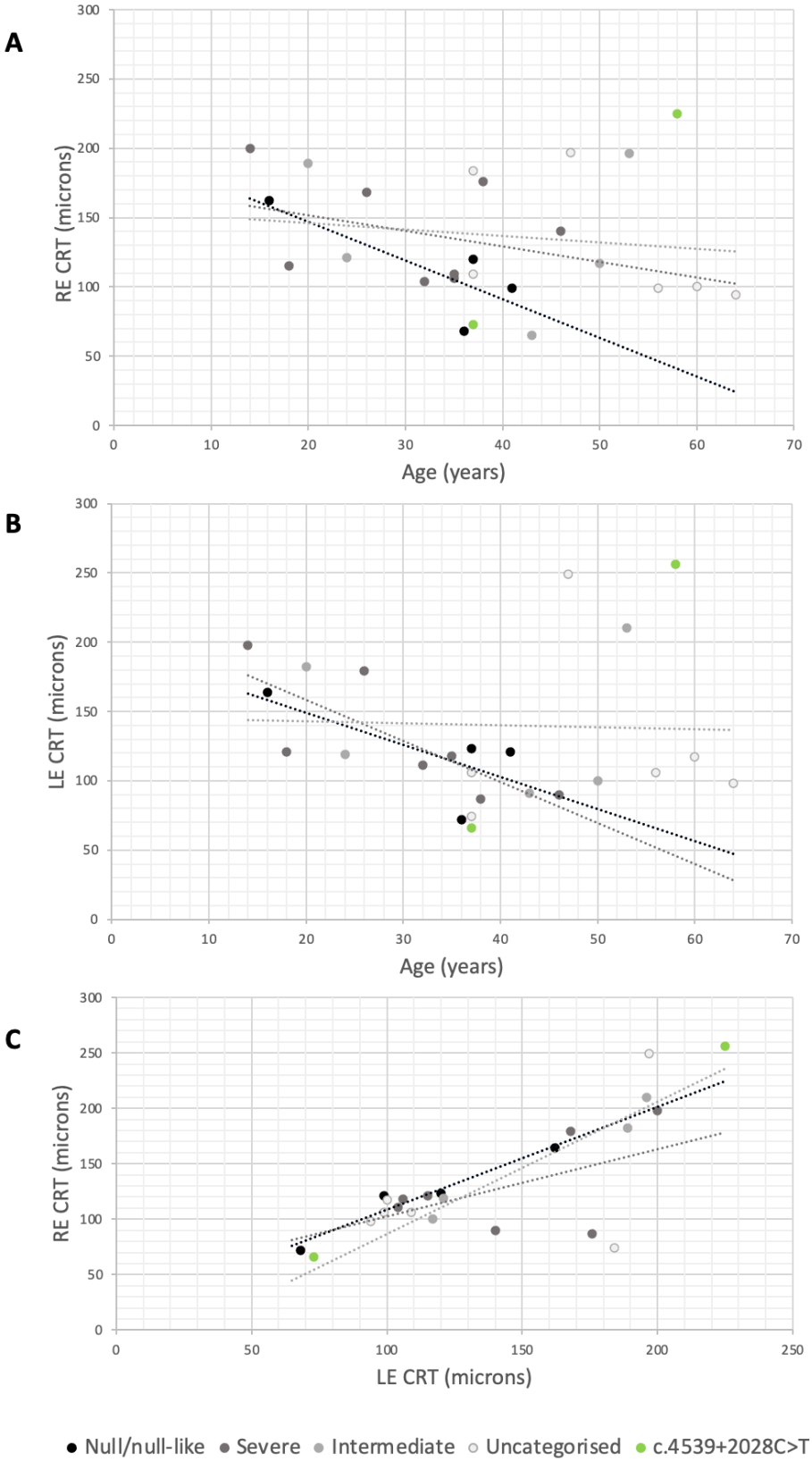


Figure 6.1.4 | Scatterplots illustrating (A) central retinal thickness (CRT) of the right eye (RE) vs age, (B) CRT in the left eye (LE) vs age and (C) CRT in the RE vs the LE. C illustrates the degree of symmetry between the eyes. All patient's carry c.4539+2028C>T and the color legend represents the severity categorisation of the disease causing variant *in trans*. Green circles represent patients homozygous for c.4539+2028C>T.

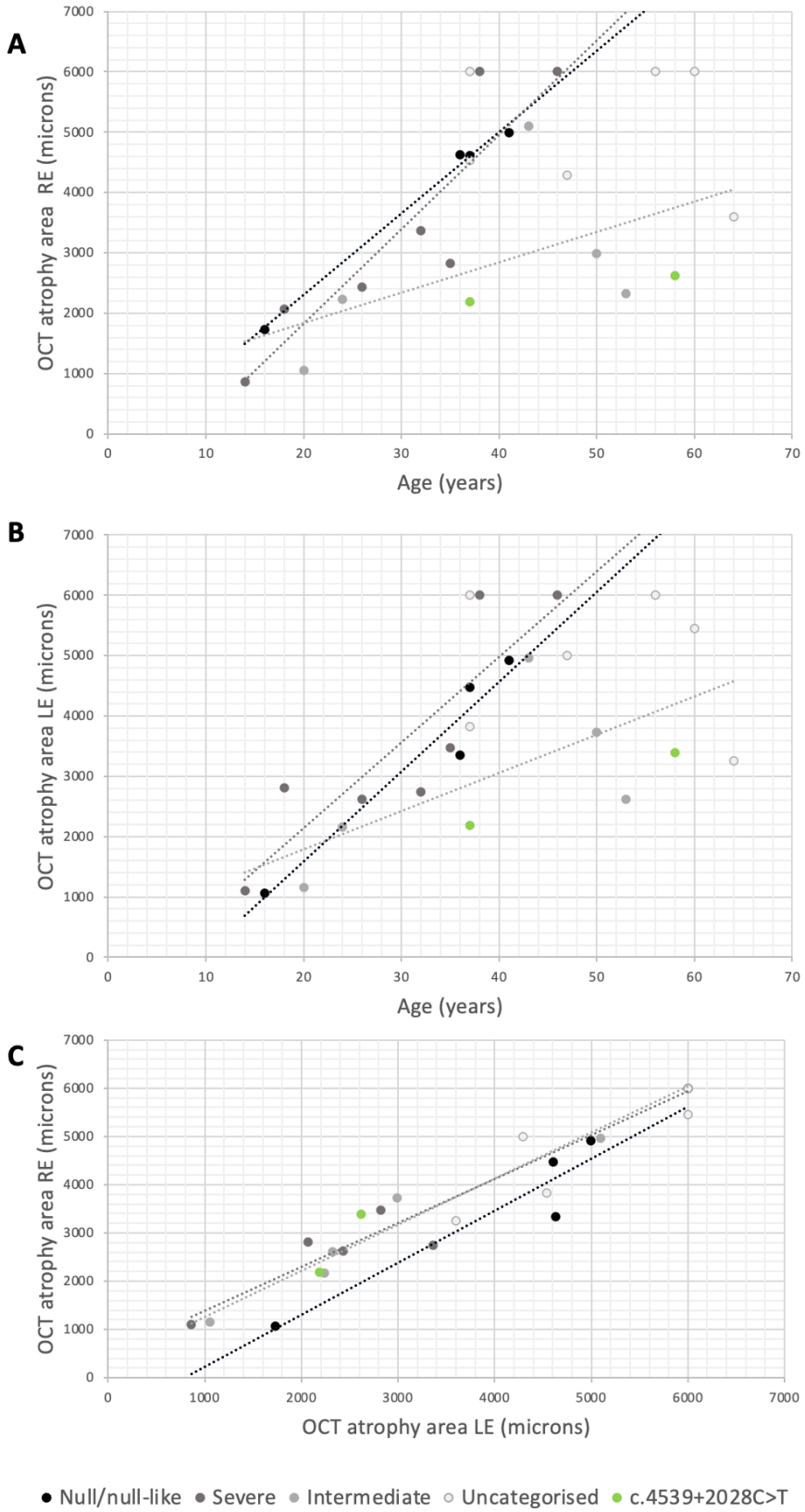


Figure 6.1.5 | Scatterplots illustrating (A) optical coherance tomography (OCT) atrophy area of the right eye (RE) vs age, (B) OCT in the left eye (LE) vs age and (C) OCT atrophy area in the RE vs the LE. C illustrates the degree of symmetry between the eyes. All patient's carry c.4539+2028C>T and the color legend represents the severity categorisation of the disease causing variant *in trans*. Green circles represent patients homozygous for c.4539+2028C>T**.**

6.1.3.2.3 Electroretinography

ERG data was available for 14/25 participants (Table 6), all recruited from one institution and examined by the same clinician. The median age at ERG recording was 37 (range = 18-56). In accordance with a published ERG classification system[284], 9/25 individuals were designated a group 1 status (normal full-field ERG [ffERG]; median age of 31 (range 18-56), including both homozygotes. 4/25 individuals were classified into group 2 (reduced photopic function; median age of 41 (range 18-43). One individual was classified as group 3 (reduced photopic and scotopic function; age 37). Figure 6.1.6 illustrates dark adapted a wave response in the right eye vs age (A) and the left eye vs age (B). Spearman's correlation did not find a signifiant correlation between dark adapted a wave response and age in either eye. (Spearman's r_s: right eye r_s = -0.3978, p (2-tailed) = 0.17827, left eye r_s = -0.41437, p (2-tailed) = 0.15921). A significant correlation between the dark adapted a wave response of both eyes was found r_s = 0.91209, p (2-tailed) = 0.00001) indicating a high degree of symmetry. Figure 6.1.7 illustrates low amplitude 30hz flicker response in the right eye vs age (A) and the left eye vs age (B). Spearman's correlation did not show signifiant correlation between low amplitude 30hz flicker response and age in either eye. (Spearman's r_s: right eye r_s = -0.24862, p (2-tailed) = 0.41274, left eye r_s = -0.12155, p (2-tailed) = 0.69243). A significant correlation between the dark adapted a wave rod response of both eyes was found r_s = 0.83516, p (2-tailed) = 0.00038. indicating a high degree of symmetry.

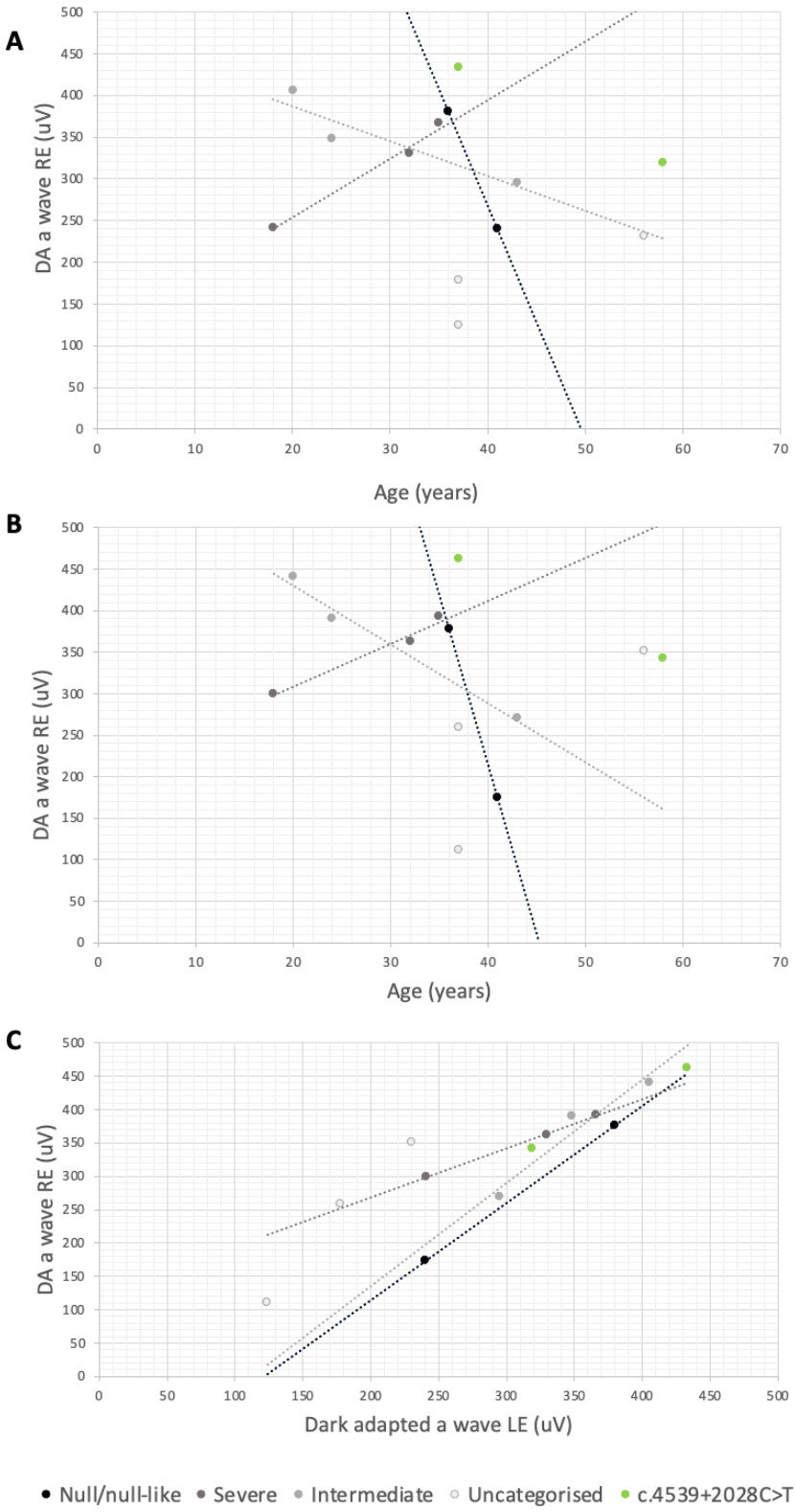


Figure 6.1.6 | Scatterplots illustrating (A) dark adapted (DA) a wave amplitude of the right eye (RE) vs age, (B) dark adapted a wave amplitude of the left eye (LE) vs age and (C) dark adapted a wave amplitude of the RE vs the LE. C illustrates the degree of symmetry between the eyes. All patient's carry c.4539+2028C>T and the color legend relects the severity categorisation of the disease causing variant *in trans*. Green circles represent patients homozygous for c.4539+2028C>T

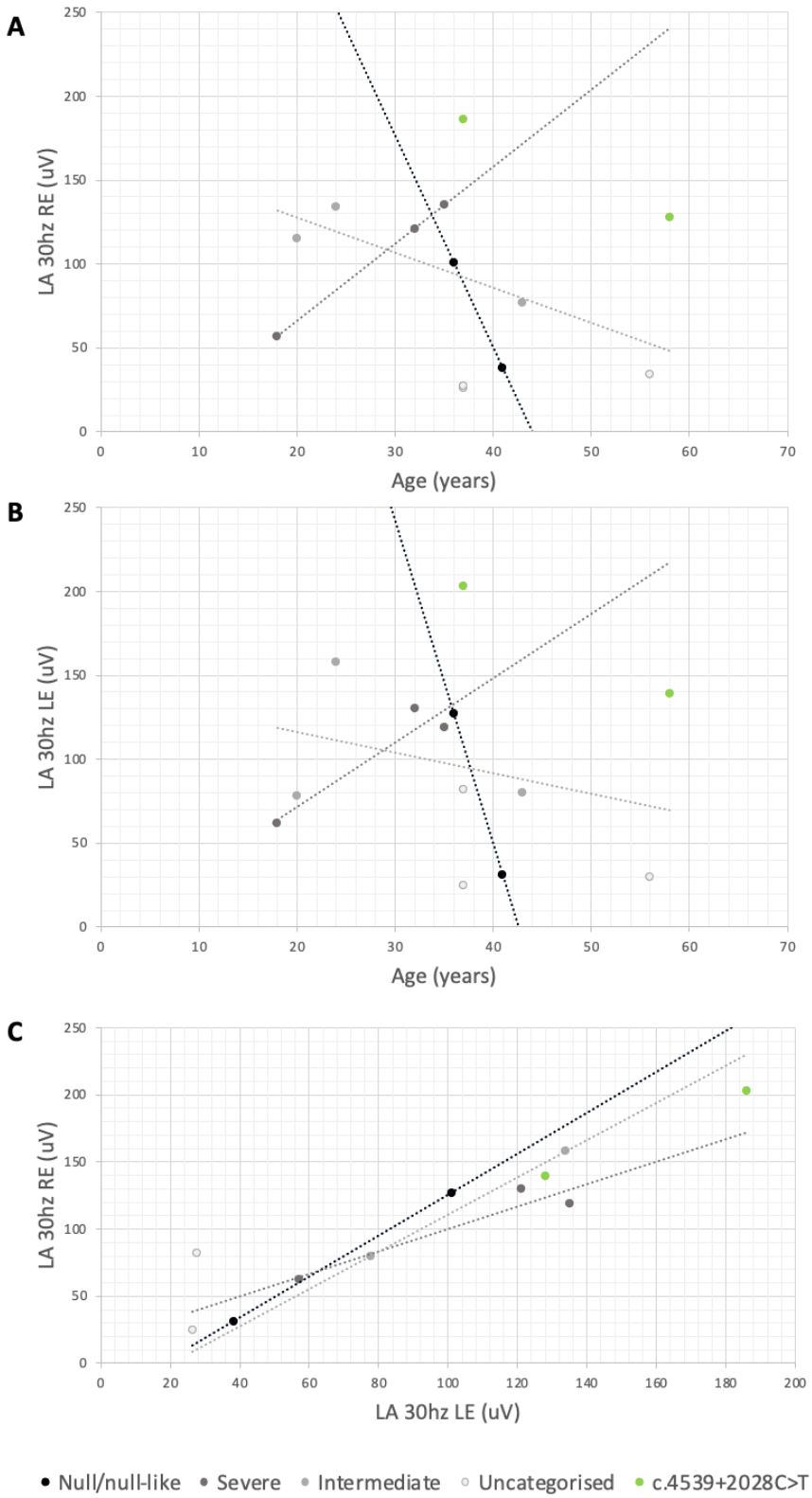


Figure 6.1.7 | Scatterplots illustrating (A) light adapated (LA) 30hz amplitude of the right eye (RE) vs age, (B) LA 30hz amplitude of the left eye (LE) vs age and (C) LA 30hz amplitude of the RE vs the LE. C illustrates the degree of symmetry between the eyes. All patient's carry c.4539+2028C>T and the color legend relects the severity categorisation of the disease causing variant *in trans*. Green circles represent patients homozygous for c.4539+2028C>T.

Table 6.1.6 | Electroretinography (ERG) data collected from a subset of individuals harboring *ABCA4* c.4539+2028C>T. ERG dark adapted 3 a wave amplitude lower limit of normal is 292 uV, ERG light adapted 30 Hz amplitude lower limit of normal is 125uV. †= Homozygous individuals; ‡ = those who carry a null or null-like variant; ffERG = full field electroretinography. PERG = pattern electroretinography. N = normal, A = Abnormal, /// = Not available, DA = dark adapted, LA = light adapted.

Patient ID	Pedigree	Age at ERG	ffERG group	PERG	ERG DA 3	ERG DA 3	ERG LA 30Hz (uV)	ERG LA 30Hz (uV)
Pt-10†	8	37	1	N	433	462	186	203
Pt-11	9	22	1	///	348	390	134	158
Pt-12‡	10	41	2	///	240	173	38	32
Pt-13	11	37	3	A	124	111	26	25
Pt-15	12	28	1	A	330	362	121	130
Pt-16	12	31	3	///	336	392	135	119
Pt-17‡	13	31	1	///	380	337	101	127
Pt-18	14	43	2	A	295	269	78	80
Pt-19	15	37	1	///	178	258	28	82
Pt-20	15	53	1	///	230	351	34	31
Pt-21	16	18	1	A	405	440	115	79
Pt-22	17	47	2	///	210	267	116	157
Pt-23†	18	56	1	A	319	341	128	139
Pt-24	18	18	2	///	241	299	57	63

6.1.3.2.4 Age at onset

The median age at onset for those with *ABCA4* c.4539+2028C>T was 15 (range = 6-49 years). As age at onset data on *ABCA4* variants of known severity was available through prior publications, these data were used in statistical analyses[279]. In order to assess the severity of *ABCA4* c.4539+2028C>T, based on age at disease onset, 4 patients who carry this variant *in trans* with a null or null-like variant were compared with 12 previously reported double null/null-like individuals[279], 28 individuals who carry mild variants *in trans* with a null or null-like variant (8 from the Target 5000 cohort, 20 previously reported[279]) and 29 individuals who carry intermediate variants *in trans* with a null or null-like variant (5 from the Target 5000 cohort, 24 previously reported[279]) (Figure 6.1.7). Median ages of onset were compared across all groups using a Kruskal-Wallis ranked sum test (p=0.0000008837), with *post hoc* analysis carried out using a pairwise Wilcoxon rank sum test, correcting for multiple testing using the Holm-Bonferroni method. Only one individual from each pedigree was included in the tests in order to control for factors such as shared environment, which may influence phenotypic outcome. A significant difference was observed between the median age at onset of those who carried *ABCA4* c.4539+2028C>T *in trans* with a null or null-like variant and those who were carried two null/null-like variants (p=0.0082). The median age at onset of those who carried *ABCA4* c.4539+2028C>T *in trans* with a null or null-like variant was 14 (range = 12-17, n=4), while the median age at onset of those who carried two null/null-like variants was 6 (range = 4-10, n=12). Those who carried a variant of intermediate severity *in trans* with a null or null-like variant had a median age of onset of 16 (range = 6-53, n=28). The median age at onset for those who carried a mild variant *in trans* with a null or null-like variant was 25 (range = 11-56, n=28). A significant difference was not observed between the median age at onset of those who carried *ABCA4* c.4539+2028C>T *in trans* with a null or null-like variant and those who carried a mild or intermediate variant *in trans* with a null or null-like variant (p=0.1042 and p=0.6581 respectively) (Figure 6.1.8). Comparison of a variant of interest found *in trans* with a null or null-like allele with individuals who carry two null/null-like variants has been used to interrogate variant severity previously[209,279]. However, as age at onset is subject to recall bias (the proband may not have attended the current clinical centre at first presentation) it cannot be used in isolation as a definitive marker of variant severity.

The ages at onset of two patients homozygous for *ABCA4* c.4539+2028C>T were then compared to the four individuals who carried this variant *in trans* with a null or null-like variant. The homozygotes had a median age at onset of 22.5 (19 and 26) versus a

median age at onset of 14 (range = 12-17, n=4) in those who carry *ABCA4* c.4539+2028C>T *in trans* with a null or null-like variant.

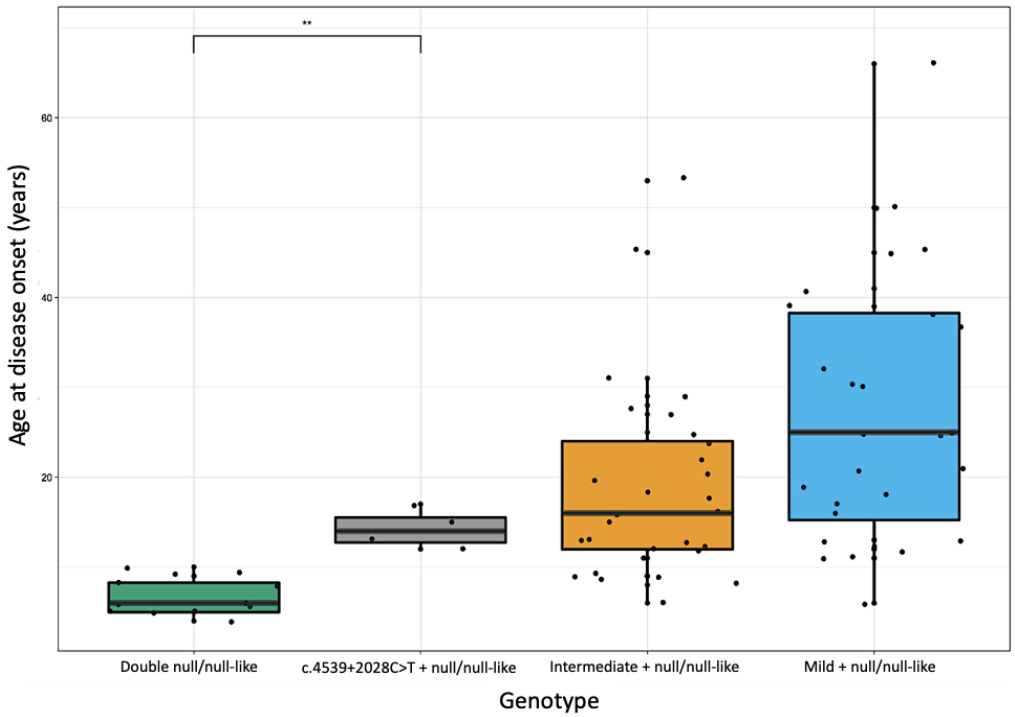


Figure 6.1.8 | Boxplot illustrating age at disease onset (years) vs genotype. Patients who carry *ABCA4* c.4539+2028C>T *in trans* with a null or null-like variant have a significantly later age at disease onset than double null/null-like individuals (p=0.0082). The median age at onset of those who carry c.4539+2028C>T *in trans* with a null or null-like variant appears to be earlier than that of the patients who carry mild and intermediate variants *in trans* with a null or null-like variant, although sample size is limited.

6.1.3.3 Variant Assessment in accordance with the ACMG-AMP guidelines

The publication of the ACMG-AMP guidelines on variant interpretation has helped steer the assessment of potentially pathogenic variants in the medical genetics community since 2015[185]. Here we present several lines of evidence based on these principles that have been implemented to interpret the pathogenicity of the *ABCA4* c.4539+2028C>T variant. **1) PS3**: Well-established *in vitro* or *in vivo* functional studies supportive of a damaging effect on the gene or gene product[112]. **2) PS4 (moderate):** The prevalence of the variant in affected individuals is significantly increased compared to the prevalence in controls. However, as this is a very rare variant, this criterion is being used as moderate evidence based on observation of the variant in multiple unrelated patients with the same phenotype and almost complete absence in controls (1/816 Irish control alleles). **(3) PM2:** This variant is found at an extremely low frequency in the gnomAD database[145] (0.00006371). **4) PM3:** STGD1 is a recessive disorder, and this variant has been detected *in trans* with a pathogenic *ABCA4* variant multiple times. Therefore, using

the ACMG-AMP guidelines this variant can be designated a pathogenic status based on the combination of one strong and 3 moderate lines of evidence.

6.1.4 Discussion

In this study, we collated data on Irish participants with STGD1 who possess a pathogenic deep-intronic *ABCA4* variant, c.4539+2028C>T representing the largest cohort of such patients identified globally to date. We clearly illustrate the enrichment of *ABCA4* c.4539+2028C>T in the Irish STGD1 population, through analysis of this group and examination of the literature and databases for additional cases globally. A detailed genotype-phenotype analysis was performed suggesting that this deep-intronic *ABCA4* variant results in a phenotype between mild and intermediate severity. Homozygotes present during adulthood and display well-circumscribed, localized, profound macular atrophy with minimal fleck lesions.

ABCA4 c.4539+2028C>T was first described by Braun *et al.,*[210] in a study that described some of the first DIVs in *ABCA4.* Functional analysis was performed by Albert *et al*[112]. RNA was first extracted from patient fibroblast cell lines, but aberrantly spliced product was not detected. To further investigate retina-specific splicing defects, RNA extracted from patient derived PPCs was examined and a 345 nucleotide pseudoexon inclusion was observed in ~7.5% of all transcripts and ~15% of transcripts derived from the *ABCA4* c.4539+2028C>T allele. This phenomenon, wherein a model that better recapitulates the patient displays a splice defect where one was not previously detected, has been noted previously[112,285]. Reasons for this may include genomic context that is more recapitualative of the patients disease or perhaps the existence of retina specific splice factors, such as enhancers or silencers. For example, the presence of *ABCA4* c.4539+2028C>T is predicted to create one SC35 and two SRp40 motifs by ESE finder 3.0[286]. *ABCA4* c.4539+2028C>T and *ABCA4* c.4539+2001G>A were the first reports of a pseudoexon inclusion in IRD due to the creation of novel exon splice enhancer motifs, as opposed to the direct creation of cryptic splice acceptor/donor sites[112]. The neighboring c.4539+2001G>A variant that also results in the 345-nt pseudo-exon insertion showed 75% wild-type transcript though genotype-phenotype correlations classified it as ‘severe’. Given the nature of c.4539+2028C>T and additionally its enrichment in the Irish STGD1 patient cohort, it was of interest to explore the phenotypic consequences of this variant in patients.

All patients harboring *ABCA4* c.4539+2028C>T and a known disease-causing variant presented with phenotypes that fall on the spectrum of STGD1. No other

pathogenic candidate variants were detected using a variety of sequencing procedures, such as target capture NGS[13,118,280], entire *ABCA4* gene sequencing[54], and WGS. All individuals exhibited FAF abnormalities, with this abnormality extending beyond the vascular arcades in 13/25 individuals. In general, a normal/mild peripheral phenotype was observed. However, severe atrophy of the central macula was detected in most cases. Four participants harboured *ABCA4* c.4539+2028C>T *in trans* with a null or null-like variant and two participants were found to carry the variant in a homozygous manner. These individuals were phenotypically compared with those harboring mild or intermediate variants *in trans* with null or null-like variants. Age at onset, ERG amplitudes as well as FAF and OCT patterns were used to assess variant severity, suggesting that the phenotype of the individuals who were homozygous for *ABCA4* c.4539+2028C>T was most similar to those that had been described as mild to intermediate previously[279]. Other lines of evidence support this hypothesis, for example, *ABCA4* c.4539+2028C>T was not detected *in trans* with a variant previously categorized as mild. This may be because when *in trans* with a mild variant, the level of functional ABCA4 protein produced is sufficient to prevent significant loss of vision and thus individuals did not present for examination. As mentioned previously, functional analyses in PPCs indicated that residual wild-type transcript is produced in the presence of *ABCA4* c.4539+2028C>T[112]. The resulting phenotype also greatly depends on the pathogenic variant found *in trans,* providing additional evidence of between mild and intermediate severity. This is supported by the phenotype of both homozygous individuals who were older at disease onset, have a more well-preserved peripheral phenotype, better ERG amplitudes (ERG group 1: normal ffERG) and better visual acuity over time than those who carry the variant *in trans* with a null or null-like variant. It is of note that the disease observed in the two homozygotes is not widespread within the retina, but the fovea and perifoveal area is profoundly impacted in both.

Population-specific studies have provided a wealth of data describing the genomic architecture of IRDs in particular countries, identifying variants present at distinctly higher frequencies in one cohort than other cohorts[50–53,54,55]. Similarly, it has become apparent that *ABCA4* c.4539+2028C>T is enriched in the Irish STGD1 population. A total of 27 incidents of this variant have been identified in our cohort to date, by comparison with 15 incidents of the variant identified across 5 studies in 3,940 alleles tested globally[54,143,210,232,274]. The Irish cohort includes, to the best of our knowledge, the only two patients homozygous for this variant identified to date. A complex allele containing *ABCA4* c.4539+2028C>T and *ABCA4* c.302+68C>T, is frequently observed, but not in all

incidents. Among the 27 *ABCA4* c.4539+2028C>T alleles detected in Irish STGD1 patients, 24 incidents of c.302+68C>T were also detected. The segregation of these two variants in a complex allele has been reported previously[210,274,282] and also accounts for the majority of Irish cases with c.4539+2028C>T. In our cohort, although *ABCA4* c.4539+2028C>T could be observed both in the presence of *ABCA4* c.302+68C>T or in isolation, *ABCA4* c.302+68C>T was only observed as part of the complex allele. Importantly, for those with sequencing datasets that are limited to the exonic regions of *ABCA4,* c.302+68C>T may be captured due to the nature of TCS by hybridization and its proximity to exon 3. As *ABCA4* c.302+68C>T has not been shown to have a qualitative effect on the resulting mRNA[112], directly sequencing for *ABCA4* c.4539+2028C>T, in cases where c.302+68C>T has been detected may aid in providing a more conclusive genetic diagnosis where one has not yet been determined. This approach adds value to existing exome or target panel data and may reduce the need for whole gene or WGS to resolve partial STGD1 diagnoses, particularly in cases where the patient is of Irish descent. This is internationally relevant as approximately 10% of the population in some western countries, including the UK, US, Canada and Australia claim Irish heritage[293,294,295]. Pedigrees harboring *ABCA4* c.4539+2028C>T account for 14% of resolved STGD1 pedigrees in the Irish cohort to date.

The identification of IRD patients homozygous for putative pathogenic variants is imperative, as these Mendelian conditions present with phenotypes that can be greatly attributed to that particular variant, as illustrated by other STGD1 studies[209,279]. This is particularly true of splice altering variants. Given their mostly intronic nature, pathogenicity can currently only be proven by *in vitro* and *ex vivo* models that recapitulate the patient context to a limited extent, as demonstrated by the lack of identifiable splice defect in patient derived fibroblasts harboring *ABCA4* c.4539+2028C>T and low percentage of mutant transcript in PPCs[112]. The participation of both homozygous patients in this study provides unique and invaluable insights into the phenotypic outcome of this variant, facilitating variant categorisation and interpretation. In addition, as increasing numbers of homozygous patients are identified and if their phenotypes vary despite non-detection of other pathogenic variants, they may provide unique insights into potential modifier variants. *ABCA4* c.4539+2028C>T appears to result in a small area of profound central macular atrophy without evidence of disease in the peripheral retina. Describing this genotype-phenotype correlation may enable more accurate clinical phenotyping, facilitating greater positive pre-test probability for diagnosis of this particular allele, particularly in unresolved mono-allelic cases where Irish ancestry is noted. However, as

this study examines the effect of a rare variant on a rare disease in a small number of individuals, it must be noted that caution should be taken when attributing severity to individual alleles. There is variance between the individuals in each group and the numbers of individuals in each group. While cumulatively these patient groups represent by far the largest patient cohort globally for the *ABCA4* c.4539+2028C>T variant, the numbers are still limited as expected for a rare disease. However, the study provides insights into the disease features associated with the *ABCA4* c.4539+2028C>T variant. In addition, interrogation of this variant in *ex vivo* retinal organoids, three dimensional structures derived from patient pluripotent stem cells, may more accurately illustrate the level of mutant versus wild-type transcript produced; this may more accurately recapitulate the splice defect. Of note, another consideration when considering the phenotypic consequences of a genetic variant is that IRDs may not be strictly Mendelian at times. Modifier variants, environmental influences and other factors are likely to affect phenotypic outcome, with a recent study indicating that sex may play a modifying role in STGD1[226,296].

A variety of therapies are under consideration for STGD1. Preclinical trials have employed lentiviral vectors, chosen based on their cargo capacity relative to the size of the *ABCA4* gene, resulting in reduced levels of lipofuscin[297]. A phase I/II clinical trial of SAR422459 is investigating the safety and efficacy of subretinal injection and lentiviral-mediated delivery of *ABCA4*[298]. Additionally, dual AAV therapies for STGD1 are being considered[263,299]. Interestingly, the possibility of subretinal transplantation of human embryonic stem cell-derived RPE cells (MA09-hRPE) is also under investigation for cases with more advanced macular/foveal atrophy. A phase I/II clinical trial has resulted in no adverse immune effects in 7/7 patients and 3 of these 7 reported increased visual acuity during a 12-month trial[300,301]. In addition, intronic variants such c.4539+2028C>T represent ideal candidates for AON-based therapy, employing short synthetic RNA molecules to restore correct splicing patterns. Promising pre-clinical results modulating the splice defect associated with *ABCA4* c.4539+2028C>T have been obtained, rescuing 80% of the observed splice defect[112]. AONs are increasingly explored as therapies to modulate splicing, and indeed a number have received approval for non-ocular conditions including Duchenne muscular dystrophy (Eteplirsen™, Golodirsen™, Casimersen™) and spinal muscalar atrophy (Spinraza™), among others[258,302].

In this study, we provide detailed genetic and clinical characterization of 25 individuals with STGD1 in Ireland who carry *ABCA4* c.4539+2028C>T, which we define as a having a macula-centred phenotype of between mild and intermediate severity. We

also illustrate the enrichment of this variant in Ireland and describe a specific clinical phenotype in homozygotes as a result of this variant. Notably these findings represent important considerations for clinicians when making an accurate phenotypic diagnosis, as well as for geneticists in ascertaining a genetic diagnosis in unresolved STGD1 cases, particularly in those of Irish descent. This study also highlights the necessity of intronic as well as exonic sequencing in previously unresolved STGD1 cases. Given the widespread availability of DNA sequencing in this era of genomic medicine, it is imperative that patients' interests are at the forefront, with access to novel therapies dependent on a genetic diagnosis. With this wealth of genetic information comes the responsibility to try to establish accurate associations between genotypes and phenotypes. The current study aids this important objective with respect to STGD1. The findings should serve to facilitate additional harmonized clinical and genetic diagnoses as well as improving both clinician and patient understanding of disease progression and management in STGD1 patients who carry *ABCA4* c.4539+2028C>T.

6.2 A synonymous variant in PROM1 causes recessive early onset retinal dystrophy through disrupted splicing

6.2.1 Introduction

Variants in the prominin 1 gene (*PROM1)* are associated with a spectrum of disease including macular dystrophy, RP, dominant STGD1 like macular dystrophy as well as recessive and dominant CRD[303]. *PROM1* is located on the short arm chromosome 4 at p15.32 and is comprised of 23 exons encoding a 865 amino acid 5-transmembrane domain glycoprotein[304]. It is ubiquitously expressed and was originally identified as a surface antigen of human hematopoietic stem and progenitor cells[305]. In the retina, *PROM1* is localised to the evaginations of the photoreceptor epithelium and functions in the maintenance of lipid levels within the plasma membrane, disc morphogenesis and photoreceptor autophagy in the RPE[306,307]. ClinVar currently lists 41 and 119 reports of likely pathogenic and pathogenic variants in *PROM1* respectively[308].

In recent years, a multitude of studies have demonstrated the contribution of pathogenic splice-altering variants to the causation of IRD[51,56,65,108–115], as described throughout this thesis. This case study focuses on a patient diagnosed with early onset retinal degeneration who carries a synonymous splice-disrupting variant in *PROM1,* NM_006017.2:c.1632G>T,p.Gly544Gly, in a homozygous manner. Given the clinically heterogenous nature of *PROM1*-associated disease with a broad spectrum of severity, detailed molecular assessment of patients is paramount in order to establish accurate genotype-phenotype associations.

6.2.2 Materials and Methods

6.2.2.1 Patient details

The pedigree recruited consisted of proband II.1, a 38-year old female and her unaffected parents (I.1 and I.2), who presented to RVEEH, Dublin, Ireland. This family is of south Asian descent. A detailed pedigree tree can be found in figure 6.2.1.

6.2.2.2 Clinical evaluation

A number of examinations were performed in order to assess the clinical phenotype. These included the use of revised 2000 early treatment diabetic retinopathy study charts (Precision Vision, La Salle, IL, USA) to survey best corrected visual acuity. Colour vision was assessed using a Lanthony desaturated D-15 panel (Gulden Ophthalmics, Elkins Park, PA, USA) under standardised lighting conditions. A Goldmann

perimeter 940 (iv4e, i4e and 04e targets – Haag Streit AG, Köniz, Switzerland) was used to measure visual fields. Full field electroretinograms were carried out using the Roland Consult RETI-port retiscan (Brandenburg an der Havel, Germany). The Topcon CRC50DX (Topcon Great Britain Ltd., Berkshire, England) and Optos Daytona (Optos plc, Dunfermline, Scotland) systems were used to capture detailed fundus colour and autofluorescence photography.

6.2.2.3 Genetic analyses

Participants provided a blood or saliva sample for genetic analysis and were processed as described previously[26–28]. Briefly, DNA was extracted from participant samples and underwent target capture NGS of the exons of 254 IRD-associated genes. Ascertaining the causative variant in this case was based on genotype-phenotype correlation, considering variants with a minor AF of less than 1% in the Irish in-house and gnomad databases[309]. In addition all candidate variants were assessed using computational prediction tools such as MetaLR[126], M-CAP[127] and REVEL[134]. *In silico* splice prediction tools employed included SSFL[310], MaxEntScan (MES)[311], NNSPLICE (NNS)[312] and GeneSplicer[313] accessed within Alamut Visual software version 2.1 (Interactive Biosoftware, Rouen, France). Candidate causative variants were confirmed by amplification of the region harbouring the variant via PCR and subsequent Sanger sequencing in all family members.

6.2.2.4 *In vitro* midigene splice assay

Midigene *in vitro* splice assays were employed in order to assess the effect of candidate variants on the splicing process. Regions encompassing the exons harbouring the candidate variants and at least one flanking exon on either side were amplified via PCR using the TaKaRa LA PCR Kit Version 2.1 (TaKaRa Bio, Kusatsu, Shiga, Japan) and primers with 5' attB1 and attB2 tags to facilitate Gateway cloning. Proband A-II.1 father's DNA was used as a template for this reaction in order to obtain both wild-type and mutant fragments simultaneously. Wild-type and mutant fragments were separately introduced into pDonor221 using a Gateway BP Clonase II enzyme kit (Thermofisher, Waltham, MA, USA) and subsequently incorporated into the pCI-*NEO-RHO* Gateway-adapted vector[278] using a Gateway LR Clonase II enzyme kit (Thermofisher, Waltham, MA, USA) as per the manufacturer's instructions. This facilitated the creation wild-type and mutant midigene plasmids. Midigenes were transfected into HEK293T cells in parallel using Fugene transfection reagent (Promega, Madison, WI, USA) and mRNA was isolated 48 hours later using the RNeasy RNA isolation kit (QIAGEN, Hilden, Germany). cDNA

was produced using the iScript cDNA synthesis kit (Bio-Rad, Hercules, CA, USA) and subsequent PCR was performed using Q solution and buffers from the QIAGEN Taq polymerase PCR kit (QIAGEN, Hilden, Germany) and *Taq* DNA polymerase from Merck (Merck Life Science, Kenilworth, NJ, USA). Primers in the first and last exons of the cloned *PROM1* fragments were used in this PCR. Amplified products were analysed via agarose gel electrophoresis and Sanger sequencing. For each assay, two independent experiments were performed. All primers used are detailed in Table 6.2.1.

Table 6.2.1 | Details of primers used in this study.

Primer name	Sequence
c.1632G>T_midigene_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCCA TCTCCATCATCCACCAAC
c.1632G>T_midigene_R	GGGGACCACTTTGTACAAGAAAGCTGGGTGTCTT TCTTCAGGCCAGTGTG
c.1632G>T_Sanger_F	CAAATGCAAAATTCTCATTCCA
c.1632G>T_Sanger_R	CTTCCTAATGTGTTCTAGGTTTTGG
c.1632G>T_RTPCR_F	GTGAAGAGTGCCGTAAGTGC
c.1632G>T_RTPCR_R	TGGAGTTGGATTAAGTTTCC
c.1946C>T_midigene_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCTA TGGTGGATGGCAGTTGGA
c.1946C>T_midigene_R	GGGGACCACTTTGTACAAGAAAGCTGGGTGGCA CTTAACCTCACCTTCACA
c.1946C>T_Sanger_F	CTCTCCAGCAGCTTCCCTAA
c.1946C>T_Sanger_R	AATGTTGCCACCTGTTTAAGAGT
c.1946C>T_RTPCR_F	TGGAAGCATAAGCAGTGAATTG
c.1946C>T_RTPCR_R	TTGTTCTATAGGAAGGACTCGTTG
*RHO*_exon5_F	ATCTGCTGCGGCAAGAAC
*RHO*_exon5_R	AGGTGTAGGGGATGGGAGAC

6.2.3 Results

6.2.3.1 Clinical findings

A pedigree tree describing the participating family can be found in Figure 6.2.1. A summary of the clinical findings can be found in Tables 6.2.2-3. Proband II-1 presented at 31 years of age with a history of central vision loss. She described central visual loss since the age of 5. Her central vision deteriorated over time, followed by her peripheral visual field. She also developed nystagmus. At presentation, her visual acuity was hand movements in both eyes. Fundus examination revealed bilateral marked disc pallor, vessel attenuation and macular retinal atrophy, with punched out chorioretinal atrophy peripherally and bone spicule pigmentary deposits nasally (Figure 6.2.2). Goldmann

visual fields were severely constricted in the left eye and undetectable centrally in the right eye (Figure 6.2.3). There was a small retained arc of visual field inferiorly in both eyes. OCT could not be carried out due to nystagmus. ERG revealed extinguished rod and cone responses (Table 6.2.3).

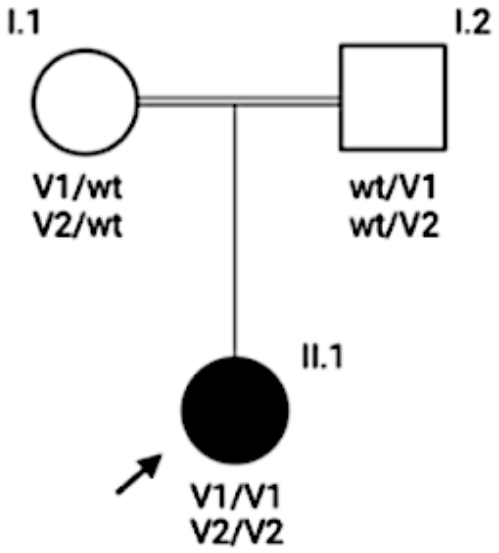


Figure 6.2.1 | Pedigree tree representing the family recruited to this study. V1 = *PROM1* c.1632G>T,p.Gly544Gly. V2 = *PROM1* c.1946C>T,p.Ser649Leu. The homozygous state of both variants segregates with disease. The relevant accession number for *PROM1* is NM_006017.2.

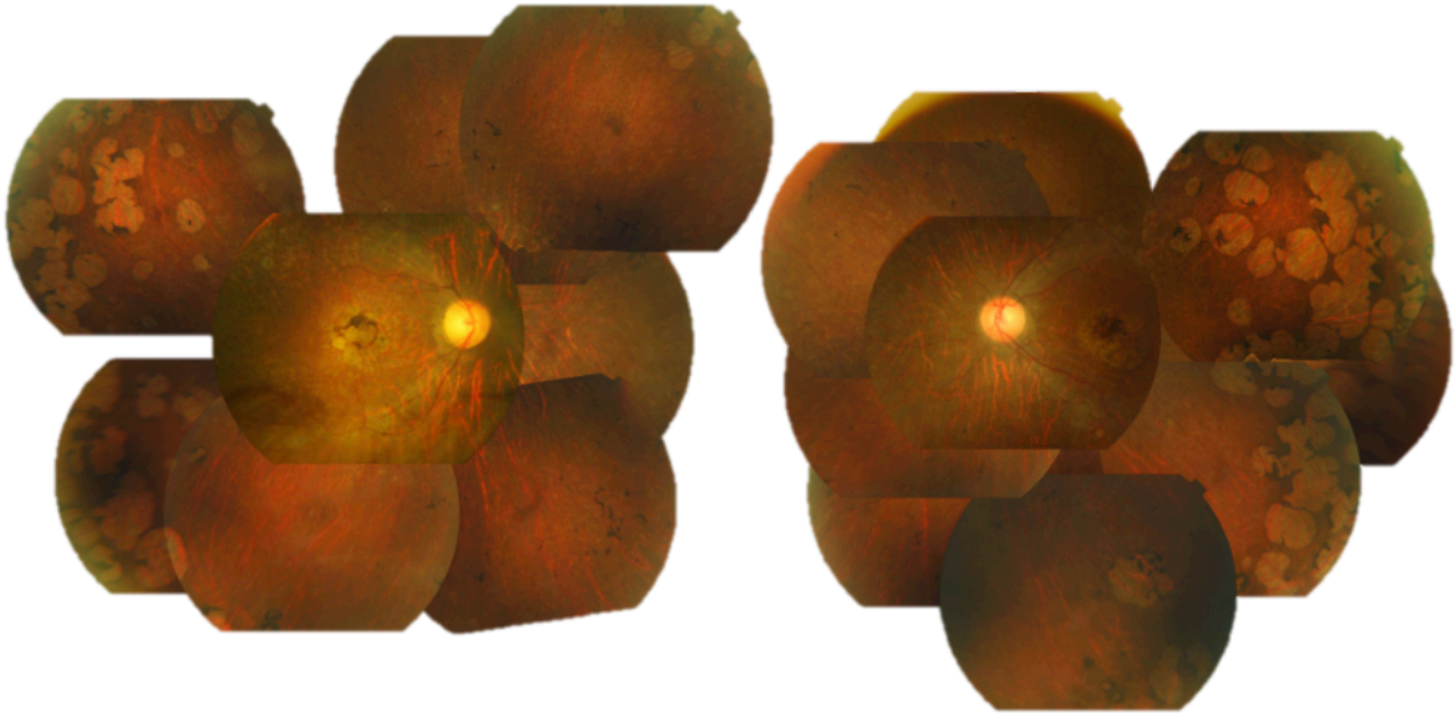


Figure 6.2.2 | Colour fundus images of patient II.1. The maculae reveal punched out atrophic areas with pigment migration. There are numerous circular areas of atrophy with pigmentary deposits in the peripheral temporal retina. Bone spicule pigmentary deposits are noted in the mid-peripheral retina.

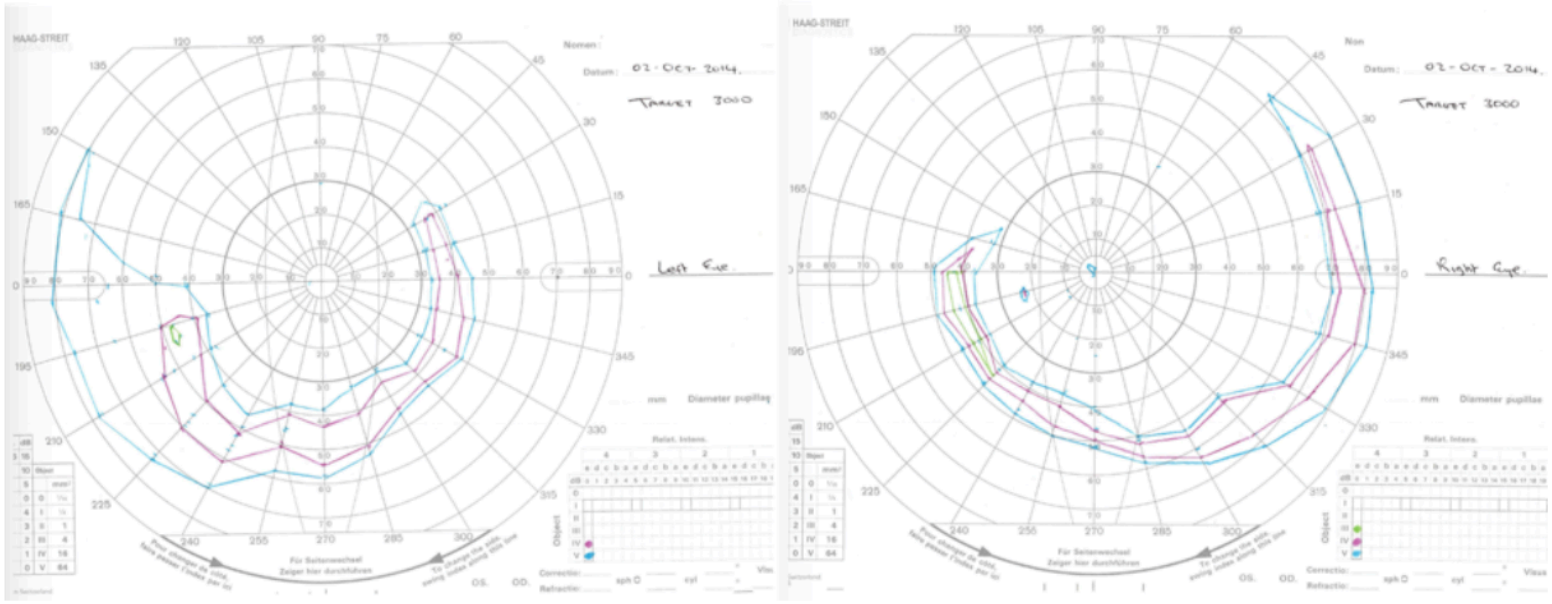


Figure 6.2.3 | Goldmann visual fields of patient II.1. Field testing revealed almost full loss of the central visual field with small remaining inferior arcuate islands

Table 6.2.2 | Detailed clinical examination of the affected individual in this study. CRD = Cone-rod dystrophy, OD = right eye, OS = left eye, HM = hand movements, CF = counting fingers, Y= yes N= No.

Patient ID	Clinical Diagnosis	Age at Examination	Vision OD/OS	Nystagmus	Pisciform Flecks	Symmetry	Peripapillary Atrophy	Foveal Atrophy	Peripheral Atrophy	Circular Chorioretinal Atrophy	Pigment Migration	Disc Pallor
II.1	CRD	29	HM/HM	Y	N	Y	Y	Y	Y	Y	Y	Y

Table 6.2.3 | Electroretinography (ERG) data obtained from the affected individual in this study, where available. OD = right eye, OS = left eye, PERG = pattern electroretinogram. ND = not detectable.

Patient ID	Age at Scan	ERG dark adapted max response (DA3) OD (µV)	ERG dark adapted max response (DA3) OS (µV)	ERG Light adapted flicker response (LA 30Hz) OD (µV)	ERG Light adapted flicker response (LA 30Hz) OS (µV)	ERG manufacturer	PERG P50 amplitude OD (µV)	PERG P50 amplitude OS (µV)
A-II.1	29	0	0	0	0	Roland Consult	0	0

6.2.3.2 Genetic findings and in silico analyses

The proband underwent target capture NGS of the exons of 254 IRD associated genes. Two candidate variants in *PROM1* were detected homozygously: c.1946C>T,p.(Ser649Leu) and c.1632G>T,p.(Gly544Gly), with AFs of 0.00002898 and 0.00003680 respectively in gnomad v2.1.1. These variants were identified *in cis* in each participant. No other potentially causative variants were identified. *In silico* pathogenicity predictor tools MetaLR, M-CAP and REVEL scored c.1946C>T,p.(Ser649Leu) 0.0896, 0.0085 and 0.055 respectively, suggesting that this variant is not disease causing. These scores are not available for the latter variant given its synonymous nature. *In silico* splice prediction tools SSFL, MES, NNS and GeneSplicer did not predict any mis-splicing events as a consequence of the c.1946C>T,p.(Ser649Leu) variant, while SpliceAI produced a low mis-splicing probability score (0.17). In contrast, SpliceAI indicated a high probability of donor site creation (0.89) resulting from the c.1632G>T,p.(Gly544Gly) variant. SSFL, MES, NNS and GS also unanimously predicted the creation of a novel donor site within exon 15 that would be stronger than the canonical donor site (Figure 6.2.4).

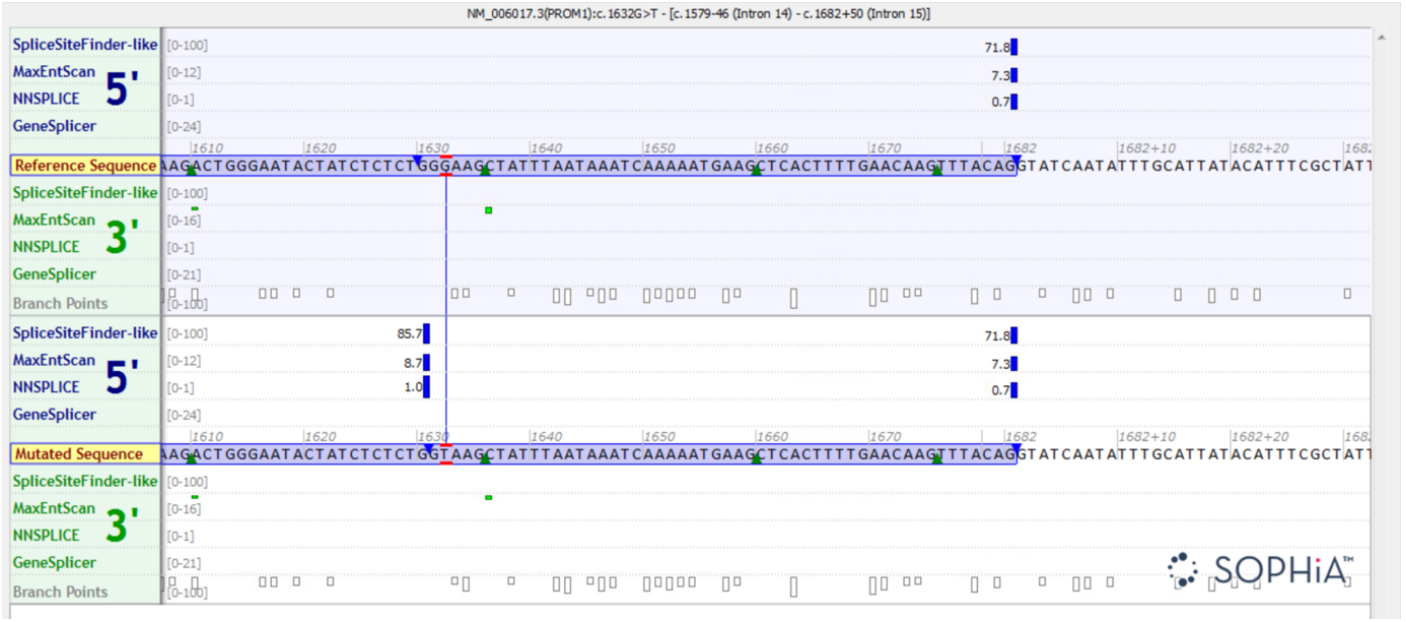


Figure 6.2.4 | *In silico* splice prediction analysis using Alamut Visual of the cryptic splice donor site created by c.1632G>T,p.Gly544Gly versus the canonical donor site. The various splice prediction tools employed can be seen on the far left. Green corresponds to splice acceptor predictions while blue corresponds to splice donor predictions. The wild-type state is displayed on the upper half, while the mutant state is displayed on the bottom half. The variant is highlighted by red lines, with a visible creation of a splice donor motif in the mutant half. The prediction scores for this cryptic donor site are higher than the prediction scores for the canonical splice donor site, indicating that this donor site has the potential to take precedence.

6.2.3.3 Functional analyses

To determine if c.1632G>T,p.(Gly544Gly) did affect the splicing process, we first employed a midigene-based splice assay approach. Wild-type and mutant midigenes encompassing exons 14-16 (c.1632G>T,p.(Gly544Gly)) of *PROM1* were transfected into HEK293T cells and the resulting mRNA transcripts were analysed via RT-PCR. The c.1632G>T,p.(Gly544Gly) wild-type midigene produced a transcript of expected size, the mutant midigene yielded a mRNA fragment containing a 52 base pair deletion in exon 15 (Figure 6.2.5). This partial deletion corresponds to use of the predicted novel donor site and results in a frameshift. No wild-type fragment appeared to be produced by the mutant midigene. These results indicate that *PROM1* c.1632G>T results in a protein defect corresponding to p.Gly544Valfs*34 and a lack of detectable wild-type product expression in HEK293T cells.

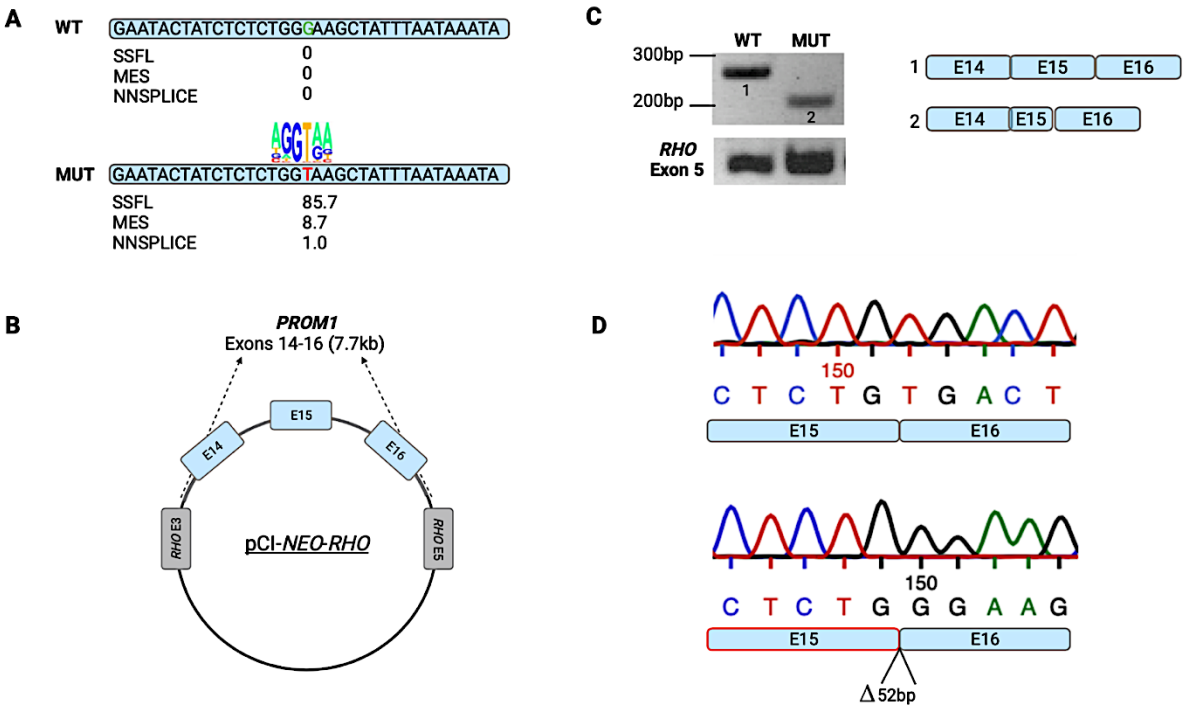


Figure 6.2.5 | Midigene *in vitro* splice assay of c.1632G>T,p.(Gly544Gly). A: Diagrammatic representation of the variant (red), including *in silico* splice prediction scores from SpliceSiteFinderLike (SSFL), MaxEntScan (MES) and NNSplice (NNS). All algorithms shown indicate that c.1632G>T,p.(Gly544Gly) catalyses the formation of a novel donor site within exon 15. **B:** Schematic of the *PROM1* c.1632G>T,p.(Gly544Gly) midigene splice vector. The plasmid includes a *PROM1* fragment encompassing exons 14-16, flanked by rhodopsin (*RHO)* exons 3 and 5. **C:** Agarose gel electrophoresis analysis of midigene RT-PCR products. HEK293 cells transfected with the wild-type midigene expressed a mRNA transcript of expected size (268bp). In cells that received the mutant midigene, a shorter fragment corresponding to exclusion of 52bp was observed and no wild-type product was detected. A region of *RHO* exon 5 was amplified as a transfection efficiency and loading control. **D:** Sanger sequencing of the mutant RT-PCR product confirmed the presence of a 52bp deletion in exon 15, which causes a frameshift and formation of a premature stop codon 34 amino acids downstream.

To investigate c.1946C>T,p.(Ser649Leu) thoroughly, we also investigated its potential to disrupt the splicing process using a midigene splice assay. We found no qualitative difference in the mRNA transcript produced from a midigene containing this variant (Figure 6.2.6). However, it is important to note that it is possible that c.1946C>T,p.(Ser649Leu) may influence the phenotypic outcome in the retina specifically in addition to c.1632G>T,p.(Gly544Gly). The availability of a larger cohort of patients to compare to patients who carry c.1632G>T homozygously in isolation may shed light on this, however we have not identified any patients with the latter genotype.

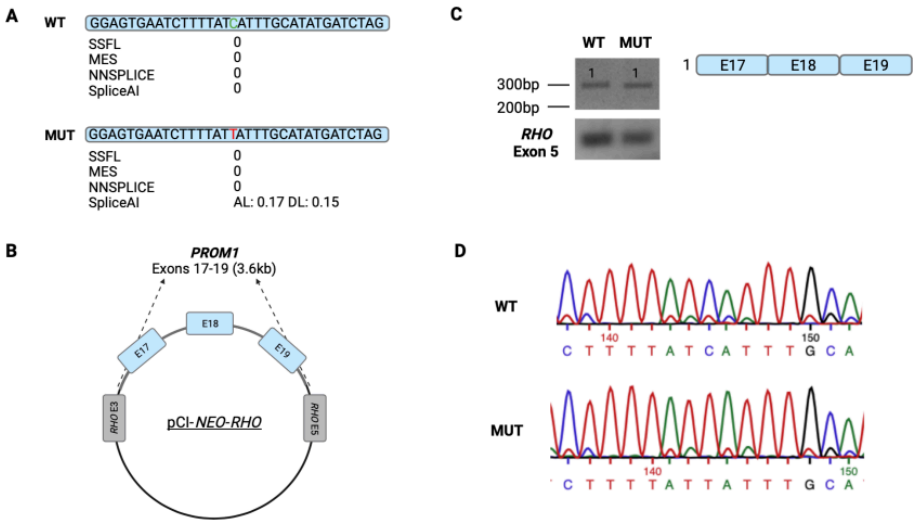


Figure 6.2.6 | Midigene *in vitro* splice assay of c.1946C>T,p.(Ser649Leu). A: Diagrammatic representation of the variant (red), including *in silico* splice prediction scores from SpliceSiteFinderLike (SSFL), MaxEntScan (MES), NNSplice (NNS) and SpliceAI. The only tool that predicts mis-splicing is SpliceAI, hence midigene-based investigation of this variant. B: Schematic of the c.1946C>T,p.(Ser649Leu) midigene splice vector. The plasmid includes a *PROM1* fragment encompassing exons 17-19, flanked by rhodopsin (*RHO*) exons 3 and 5. C: Agarose gel electrophoresis analysis of midigene RT-PCR products. Both HEK 293 cells transfected with the wild-type midigene only and the mutant midigene only expressed an mRNA transcript corresponding to the expected wild-type size. A region of *RHO* exon 5 was amplified as a transfection efficiency and loading control. D: Sanger sequencing of both RT-PCR products confirmed that both transcripts were spliced correctly.

6.2.4 Discussion

This section describes functional analysis of a synonymous variant in *PROM1,* c.1632G>T, which results in the protein defect p.(Gly544Valfs*34) as a result of aberrant splicing. Detailed clinical characterisation of patients harbouring this variant homozygously is provided. The functional consequences of *PROM1* c.1632G>T at the mRNA level in this study was examined using an *in vitro* midigene splice assay. Correspondence from collaborators has also revealed the same defect is observed when RNA is examined from patient-derived peripheral leucocytes (personal correspondence). This assay demonstrated that this synonymous variant results in a frameshift mutation (p.Gly544Valfs*34) due to a 52 base pair deletion induced by mis-splicing of exon 15 (Figure 6.2.5). No residual wild-type transcript was observed. Synonymous splice altering variants have previously been reported in IRD associated genes. These include *CDHR1*[208]*, CHM*[314]*, USH2A*[56]*, BEST1*[315] *and ABCA4*[210] among others. The defects associated with these variants vary in complexity and include exon skipping and partial exon deletion. However, to the best of our knowledge this is the first report of such a variant in *PROM1.*

Cases such as this, where two potentially pathogenic variants are identified in a gene associated with the phenotype, should be handled with caution. When presented with a case where one of the variants is nonsynonymous and one is synonymous, some may be inclined to lean towards the missense variant as disease-causing given its more apparent effect on the protein product. However, as splice altering variants are increasingly implicated in disease, it is imperative to investigate their pathogenic potential at the RNA level. In this particular case, the missense variant, c.1946C>T, displays poor *in silico* pathogenicity prediction scores concurrent across multiple tools (MetaLR: 0.0896, M-CAP: 0.0085 and REVEL: 0.055). The variant was also found homozygously in a control individual in a study by Litink *et al*[316]. Despite this, an alternative cause of disease had not been determined until now.

This study demonstrates that *PROM1* c.1632G>T results in the deletion of a 52 base pair segment of exon 15 *in vitro*. This is based on creation of a GT donor motif in the exon, where the reference sequence naturally resembles a consensus splice site sequence motif that is activated by the variant. This cryptic donor site takes precedence over the canonical donor site, resulting in truncation of the exon. This hypothesis is supported by functional data obtained in a midigene model and consensus among numerous splice prediction tools (Figure 6.2.4). Analysis of *PROM1* mRNA produced from patient-derived midigenes revealed that 100% of detectable transcripts contained this deletion. This suggests that the variant is severe in nature, which is supported by an early age at onset

and rapid decline in vision. However, as mentioned previously, it must be noted that assays such as these are limited in their ability to recapitulate the patient and retinal context. Patient-derived PPCs or retinal organoids may provide a clearer insight into the levels of wild-type and mutant transcripts present in the retina, due to the potential presence of retina-specific splice factors. This phenomenon, wherein splice defects are only visible or differ in a retina-like model has been demonstrated for other IRD genes including *CEP290* and *ABCA4*[112,285].

PROM1 variants result in a spectrum of disease encompassing both dominant and recessive inheritance patterns, with macular involvement representing a unifying phenotypic feature. Disease associations include recessive RP with macular degeneration, recessive and dominant CRD as well as STGD1-like macular disease and bull's eye maculopathy[317,318]. It has been postulated that homozygous and compound heterozygous null variants in *PROM1* cause more severe, recessive RP and CORD phenotypes[319–321] while pathogenic missense variants in the heterozygous state result in the dominant macular presentations of disease[322]. However, in a study that clinically and genetically characterised the largest cohort of patients affected by pathogenic variants in *PROM1* to date these hypotheses were not supported[323]. In addition, both variable intra- and interfamilial age at onset was observed in this study making accurate genotype-phenotype correlations difficult to establish[323]. The inheritance pattern in this pedigree appears to be autosomal recessive, given carrier relatives have not presented with any visual impairment even in their 7th decades of life.

Characterisation of the molecular effects of putative splice-altering variants is important for elucidating mechanisms of disease, ensuring accurate classification of pathogenicity and enabling access to emerging therapeutics. As of yet, no pre-clinical studies for therapeutics targeted specifically toward *PROM1*-related retinal dystrophy have been documented in the literature. Development of a *PROM1* gene replacement therapy may be somewhat challenging due to a limited window for therapeutic intervention in recessive cases, which typically display early-onset and extensive retinal degeneration. Conversely, while dominant *PROM1* mutations appear to give rise to a late-onset, macula-restricted disease[324], treatment of these cases may require suppression of the dominant allele in parallel to gene augmentation, as seen with therapeutic approaches targets towards dominant *RHO* variants[325]. Mutation-specific correction strategies may also be facilitated by gene editing or AON technologies. Notably, an AON strategy was successfully used to block an aberrant splice donor site created by an exonic variant in a mouse model of Usher syndrome, leading to rescue of hearing and vestibular

dysfunction[326]. In principle, it therefore may be feasible to use such an approach to correct *PROM1* pre-mRNA splicing in cases of c.1632G>T-associated disease.

In summary, a detailed clinical and genetic characterisation of a patient harbouring *PROM1* c.1632G>T,p.Gly544Gly homozygously is provided. Establishing true pathogenicity and indeed the severity of a given variant is challenging and is particularly true of synonymous variants. Traditionally overlooked, these variants have become increasingly implicated in IRD causation alongside their nonsynonymous counterparts. Identifying patients homozygous for any potential disease causing variant is diagnostically important, providing unique insights into their clinical consequences in isolation. However, given the nature of synonymous variants, evidence of pathogenicity provided by functional interrogation is imperative. This study demonstrates that this variant causes a 52 base pair deletion in exon 15 of *PROM1*, causing a frameshift followed by truncation via a premature stop codon encountered 34 amino acids downstream. Homozygous patients present with a severe early onset retinal degeneration phenotype. To the best of our knowledge this is the first report of a pathogenic synonymous variant in *PROM1*. This study underscores the role of splice-altering variants in disease manifestation. Moreover, it highlights the potential mis-categorisation of the pathogenicity of synonymous variants in *PROM1* and other genes. Thoroughly investigating the disease causing potential of these variants is central to a comprehensive and accurate genetic diagnosis.

6.3 A BBS1 branchpoint variant is associated with non-syndromic retinitis pigmentosa

6.3.1 Introduction

The *BBS1* (Bardet-Biedl syndrome 1) gene has been reported in cases of both the syndromic form of disease[327,328] and in the non-syndromic RP[329,330]. BBS is the second most common syndromic IRD in Ireland and globally[14,329]. In concordance with IRDs in general, BBS is highly variable dependent on the genes and variants driving the phenotype[330,331].

. Located on chromosome 11, *BBS1* comprises 17 exons and 593 amino acids. The gene encodes an intergral part of a large complex known as the BBSome which is associated with membrane trafficking to the outer segments of photoreceptors[332]. Studies in animal models indicate that disease in driven by the accumulation of abberant BBS1 protein product or loss of wild-type protein function[329]. To date, 191 pathogenic or likely pathogenic variants have been reported in *BBS1*[308].

As detailed in the introduction of this thesis, the splicing process is a highly complex phenomenon involving multiple proteins with and interaction between *cis* and *trans* elements[333]. In addition to the canonical and NCSS sequences, the consensus branchpoint sequence plays a key role in ensuring accurate execution of the splicing process. The branchpoint sequence is located 9-400 nucleotides upstream of the canonical acceptor site. The consensus sequence is yUnAy-in humans with the branchpoint conserved A underlined, the lowercase 'y' represents pyrimidines that are less conserved compared to the U and A bases[334].

This section of the thesis describes the pathogenicity of a near-exon aberrant RNA splice variant, c.592-21A>T in *BBS1* identified in four seemingly unrelated pedigrees by Fadaie *et al*[109]. This section discusses in detail the findings from one of these patients, recruited as part of the Target 5000 study. This variant alters the branchpoint sequence upstream of exon 8 at its most critical position. The functional effect of this variant was explored by collaborators using a midigene splice assay and highlights the relevance of interrogating branchpoint motif regions in IRDs.

6.3.2 Material and methods

6.3.2.1 Clinical evaluations

The proband was recruited from the Royal Victoria Eye and Ear Hospital, Dublin, Ireland. The clinical diagnoses were based on ophthal- mological examinations, including assessment of best-corrected visual acuity, Goldmann visual field testing, slit-lamp

biomicros- copy, detailed fundoscopic examination and fundus photog- raphy. Furthermore, electrophysiological evaluations, including full-field flash ERG and multifocal ERG and were performed according to the guidelines of the International Society of Clinical Electrophysiology of Vision[335].

Following the detection of *BBS1* variants, the patient and their records were re-examined for any characteritics indicative of a syndromic disease including polydactyly, obesity, cognitive impairment, hypogonadism and renal abnormalities as well as speech disorders, developmental delay, behavioural abnormalities, orodental abnormalities, brachydactyly/syndactyly, ataxia, diabetes mellitus, cardiovascular anom- alies, hepatic involvement, mild hypertonia and additional eye abnormalities such as strabismus, cataracts and astigmatism were also investigated. Unless the patient's health was indicative of any renal or hepatic symptoms these features were not examined.

6.3.2.2 Genetic analysis and variant prioritisation

Genomic DNA was isolated from peripheral lymphocytes as detailed in section 2. The patient was genetically undiagnosed after genetic assessment through Target 5000 panel sequencing (section 3). This individual was sequenced and analysed in parallel with the WGS study outlined in section 4, therefore the genetic analysis and variant prioritisation/selection was identifical to that detailed in the methods of that chapter.

6.3.2.3 In vitro splice assay

The *in vitro* splice assay was performed by Dr Zeinab Fadaie at RadboudUMC, Nijmegen, The Netherlands, however, for completeness is briefly outlined below. As described in Fadaie *et al*.,[109] The wild-type and mutant constructs, with a 7.3kb insert, were designed according to a previously described protocol[109,244]. In short, BBS1 exons 4 to 9 of a genomic control DNA sample were amplified using Gateway adapted primers. This wild-type construct served as a template to generate the mutant construct with c.592-21A>T by site-directed mutagenesis PCR. Subsequently, wild-type and mutant constructs were independently incorporated into the pCI-NEO-RHO Gateway-adapted vector to generate wild-type and mutant midigenes. Both midigenes were separately transfected into HEK293T cells and, after 48 hours of incubation, RNA was collected, and the transcripts were analysed by reverse transcription–polymerase chain reaction (RT-PCR) with primers in exons 6 and 9. In the presence of multiple fragments after gel electrophoresis, they were quantified using ImageJ Fiji software[336].

6.3.3 Results

6.3.2.1 Clinical findings

A pedigree tree describing the family can be seen in Figure 6.3.1. The proband presented with non-syndromic RP. Symptoms began with nightblindness in mid-teens followed by impaired mobility due to visual field loss in the probands early twenties. At last examination visual acuity was 3/36-2 in the right eye and hand movements in the left eye. The spherical equivalent refraction was -4.125 in the right eye (-1.75/-4.75cx@15°) and -4.125 in the left eye (-2.0/-4.25cx@150°). Early posterior sub-capsular opacities were detected in the right and left eye. Ophthalmoscopy findings included disc pallor, vascular attenuation, scattered nummular areas of RPE thinning, extensive bone-spicule intra-retinal pigmentary and deposits mid-peripherally. Goldmann perimetry showed marked concentric constriction to within 5° of fixation with preservation of a temporal island of field in each eye to the IV4e target (Figure 6.3.2). Electroretinogram demonstrated non-recordable rod-isolated responses and significantly delayed and reduced amplitude cone-isolated responses. There were no non-ocular features such as polydactyly, truncal obesity, intellectual, impairment, hypogonadism. The proband was of general good health.

The clinical presentation concurrent with findings from 2 of the other patients in Fadaie *et al*. One of the other patients did display a mild cognitive impairment. None of the four patients had complaints related to cardiac, hepatic or renal features of BBS.

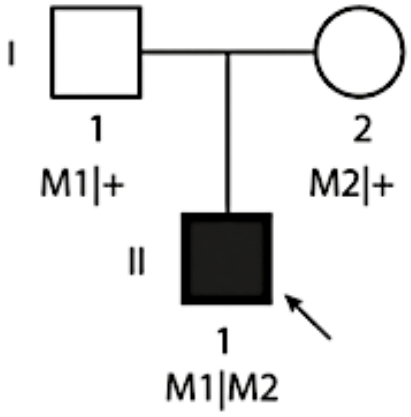


Figure 6.3.1 | Pedigree tree of one of the affected individuals in this study. The *BBS1* variants (M1) c.592-21A>T amd (M2) c.1169T>G segregate with the disease. The arrow indicates the proband in each family.
.

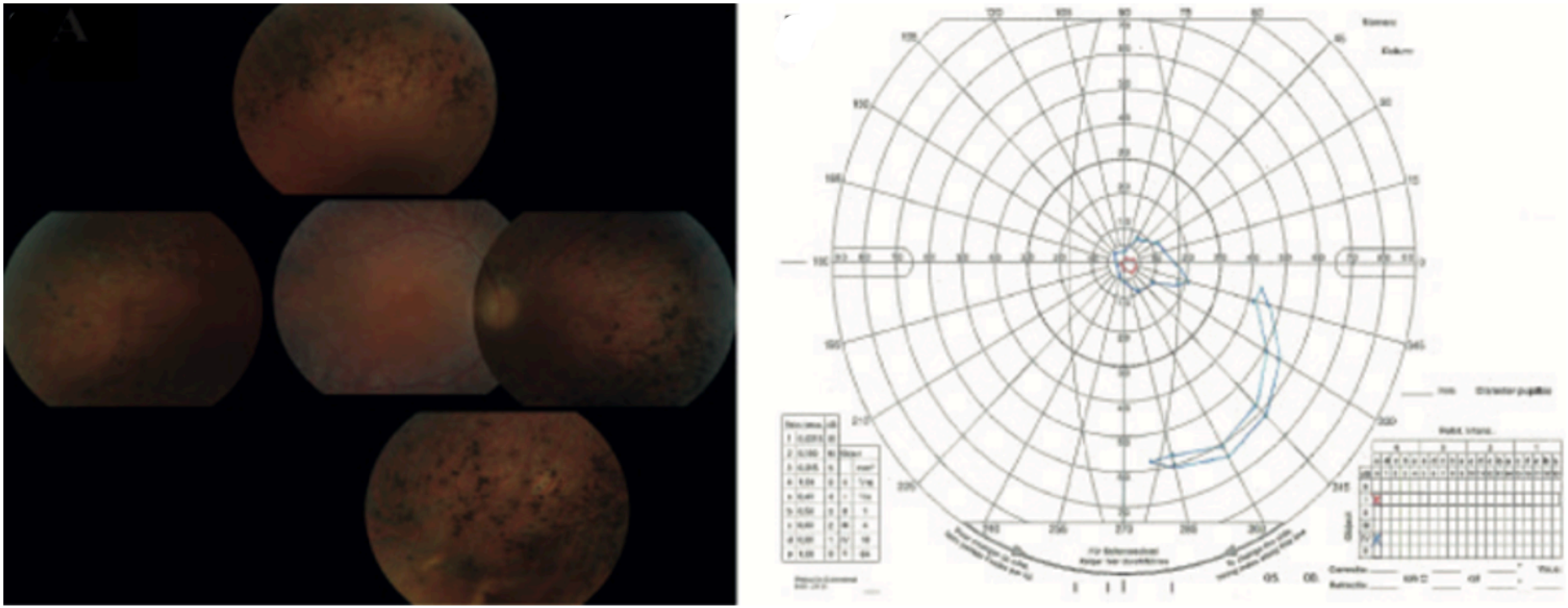


Figure 6.3.2 | Ophthalmoscopy and Goldmann perimetry images of the proband. Right: Ophthalmoscopy image displays disc pallor, extensive bone- spicule intra-retinal pigmentary deposits and vascular attenuation in the right eye. Left: Goldmann perimetry demonstrates very extensive visual field loss, with gross concentric constriction to the IV4e target and preservation of a narrow island of field temporally in the right eye.

6.3.2.1 Genetic findings

As the proband remained unresolved following TCS, WGS was performed as a second tier genetic test. p.(Met390Arg) in *BBS1* was detected as part of first tier testing, however a second causative variant remained elusive. This variant was also detected via WGS. *In silico* pathogenicity tools indicate that this variant is disease causing, with a MetaLR score of 0.7304, an M-CAP score of 0.124373324325 and a REVEL score of 0.662. p.(Met390Arg) is the most common pathogenic *BBS1* variant with a total AF of 0.001570 in gnomAD v2.1.1 (0.002773 in non-Finnish. European population).

No other pathogenic coding or canonoical/non-CSS variants were detcted in *BBS1* or other IRD-associated genes. An intronic variant, c.592-21A>T was detected and was also found in the 3 other probands examined as part of this study. While *in silico* analysis did not suggest that this variant has a damaging effect on the nearby splice site motif, the Branch Point algorithm with Alamut Visual predicted a damaging effect on the canonoical branh point motif. The recognition score for the wild-type sequence is 73.5 (0-100), however, in the presense of this variant a score of 0 is observed, completely destroying the canonoical branchpoint. In addition, SpliceAI predicts both acceptor and donor loss of exon 8 (0.44 and 0.04 repectively). Segregation analysis was performed on the family and the variants were found to be *in trans*. No other incidents of this variant have been detected in patients in Ireland.

The results of the midigene splice assay performed by Fadaie *et al*., are detailed in Figure 6.3.3 and were as follows: cDNA from the wild-type construct constituted a 300bp wild-type fragment and a 95bp fragment that corresponded with skipping of both exons 7 and 8. cDNA from the mutant contruct also contained these two fragments as well as a 270 base pair fragment corrresponding to a 30-nucleotide deletion of exon 8 ad a 168 base pair fragment corresponding to exon 8 skipping.

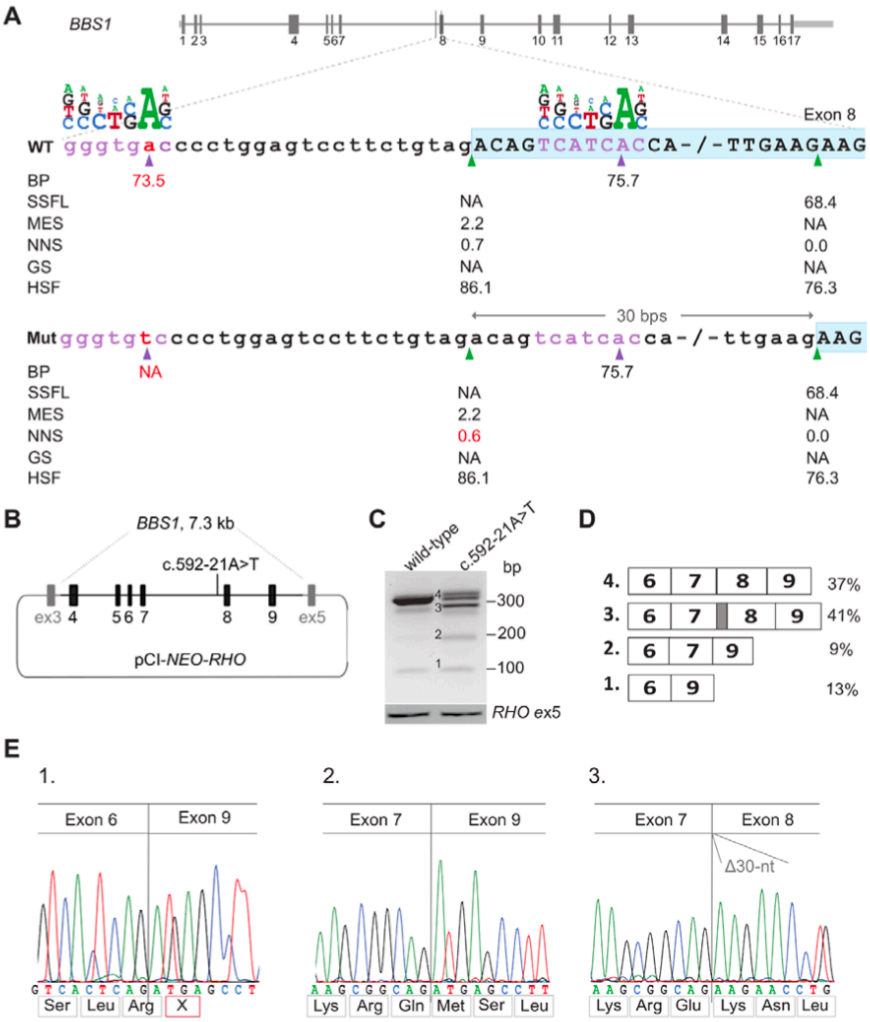


Figure 6.3.3 | Molecular genetic characterisation of variant c.592-21A>T in *BBS1*. (A) Schematic representation of the *BBS1* gene and enlargement of the exon–intron boundary of exon 8 in the wild-type and mutant sequences. The SpliceSiteFinder-like (SSFL, range 0–100), MaxEntScan (MES, range 0–12), GeneSplicer (GS, range 0–24) and Human Splicing Finder (HSF, range 0–100) scores for the splice acceptor site, and Branch Points (BP, 0–100) scores for the branchpoint sequence are indicated below each sequence. The green and purple triangles indicate the position of splice acceptor site and the branchpoint in both wild-type and mutant sequences. The purple sequences highlight the 7-mer branchpoint motif and the sequence above that shows the general branchpoint motif. The red "t" highlights the c.592-21A>T variant in the mutant sequence. The red numbers represent the altered scores compared to wild-type. (B) Schematic representation of the mutant pCI-*NEO-RHO* vector, containing exons 4–9 of *BBS1* flacked by *RHO* exons 3 and 5, which was used to transfect HEK293T cells with wild-type and mutant constructs, respectively. (C, D) The gel image of RT-PCR products of the wild-type midigene reveals the expected 300 bp wild-type fragment along with 95 bp fragment of natural exon 7/8 skipping. In the mutant midigene, two additional fragments were observed in which the larger one corresponds to a 270 bp fragment representing a 30 bp deletion of exon 8 and a smaller fragment of 168 bp corresponding to the complete deletion of exon 8. RT-PCR analysis of *RHO* exon 5 was performed as a control for efficient transfection. Fiji software was used for quantification of each fragment in mutant construct. (E) Sanger sequence analysis of the RT-PCR fragments confirmed the wild-type and exon 7/8 deletion in the wild-type midigene as well as exon 8 deletion, and the 30 bp deletion of exon 8 in the mutant. Mut, mutant; WT, wild- type (From Fadaie *et al.*[109]).

6.3.4 Discussion

As part of a study by Fadaie *et al.*, A branchpoint variant, c.592-21A>T, was detected in one proband recruited in Ireland as well as three other probands, two of Dutch and one of British descent. Pathogenic branchpoint variants have been detected in other genetic diseases, for example, in Ehlers-Danlos syndrome, Norum disease,

neurofibromatosis type II and Xeroderma pigmentosum[337–341]. To the best of our knowledge this is the first report of a pathogenic branchpoint variant causative of an IRD. A midigene splice assay was peformed and resulted in partial and complete deletion of exon 8 in the RNA (p.[Thr198_Lys207del,=]). In addition, natural exon 7/8 skipping was observed in both wild-type and mutant midigenes.

Only in recent years has the scale at which intronic variants cause disease begun to be illuminated. The more "clear cut" intronic variants, such as canonical and NCSS variants are now frequently interrogated[341]. However, less explored motifs such as the branchpoint sequence still remain elusive. Reports on defects caused by branchpoint variants have been published. For example, one case of ALPHA-LCAT deficiency where the recognition of an alternative branchpoint sequence with subsequent intron retention in the upstream region was reported[337,339]. Additionally, branchpoint variants can also lead to pseudoexon inclusion when in the presence of a cryptic splice acceptor site[339].

While actual phenotypic severity cannot be estimated from *in vitro* studies acurately, one can hypothesise that given the non-syndromic nature of the proband, and the other three affected individuals, the disease severity may be in part influenced by the degree of wild-type fragment remaining as splicing is carried out correctly in some instances. To more accurately recapitulate the patient context, a model derived from the affected individual would be more appropriate. As mentioned in section 5, studies on PPCs derived from STGD1 patients have illustrated that additional defects can be observed when using a more "retina-like" model[112].

BBS1 variants are associated with sytemic and ocular-only disease that lies on a phenotypic spectrum[330]. Studies suggest that, as with other IRDs, the type of variant driving disease influences phenotypic outcome[342]. On the BBS spectrum, BBS1 has been associated with a releatively milder disease presentation by comparison with pathogenic variants in other BBS genes[327,328,343]. The phenotype caused by p.(Met390Arg) is well established given its high frequency. It is associated with relatively milder disease symptoms such as eye abnormalities, non-syndromic RP or milder forms of BBS[330]. The proband involved in this study has been classified as having non-syndromic RP with macular involvement, which has been described in other studies[344]. As mentioned above, this clinical classification is concurrent with the categorisation of the other three probands described in Fadaie *et al*. Upon retrospective analysis of the sequencing data generated as part of Target 5000, the c.592-21A>T was indeed detected at the sequence level. However, at the time this variant was not included in a pathogenic variants shortlist given a lack of branchpoint variant identification software and distance from the exon. This highlights the

imperative of updating *in silico* tools as new types of IRD-associated variants come to light.

To conclude, *BBS1* c.592-21A>T was detected in four not knowingly related families with non-syndromic RP from three distinct genetic backgrounds. It results in a complex splice defect and was found *in trans* with the common *BBS1* p.(Met390Arg) variant in the proband discussed in this section of the thesis as well as the other three probands described by Fadaie *et al*. To the best of our knowledge, this collaborative effort is the first report of a pathogenic branchpoint variants causative of an IRD. This study highlights the role that non-coding variants, and more specifically, branchpoint variants play in the pathogenesis of IRD.

7. Discussion

7.1 The evolution and expansion of Target 5000

Over the past four decades, the process of DNA sequencing has undergone an extraordinary transformation from inception to diagnostic implementation. It is now possible to sequence the entirety of a patient's DNA and possibly provide them with an underlying molecular basis for their disease. The field of IRDs has significantly benefited from advances in this technology, with over 280 genes implicated as causative of disease by RetNet [15], where there was once thought to only be 3 genes, one autosomal recessive, one autosomal dominant and one x-linked. Within Target 5000, an increase in the scope of sequencing was undertaken throughout this doctoral degree from target-capture NGS to the employment of whole gene and WGS. It is the overall aim of Target 5000 to provide it's participants with a precise genetic diagnosis and to fully elucidate the vast spectrum of variants that underly IRDs in Ireland.

Section 3 outlines data obtained from 1000 Target 5000 participants who have been sequenced for the exons of up to 254 IRD genes as well as some known pathogenic intronic regions. It provides an overview of the current genetic landscape in Ireland across a phenotypic spectrum, including the discovery of 19 novel potentially pathogenic missense variants and a novel SV in *CHM*[13]. Another key point in this section is the discovery of additional cases of what is thought to be the first *FLVCR1* protein coding variant that causes non-syndromic RP[157], presented and discussed in detail in sections 3.3.8 and 3.4. In addition, the prevalence of an *RPE65* variant (c.1430A>G (p.Asp477Gly) in Ireland causing autosomal dominant RP is presented and discussed in sections 3.3.7 and 3.4.

Section 4 describes the results of a WGS study on 100 unresolved IRD patients, including 38 Target 5000 participants. A bespoke variant prioritisation strategy was employed to identify the underlying genetic cause of IRD in these patients. This was done by extracting sequence data from 446 ocular and 163 ciliopathy associated genes and manually interpreting any variants that arose in both the exons and introns of these genes. A detailed description of this process can be found in section 4.2.2. 24 of the 100 patients recruited to this study received a genetic diagnosis having undergone WGS, with positive candidate variants identified in genes that were not previously included in the target capture panel as well as non-coding variants located in intronic regions not previously captured. Functional analysis was employed where necessary. Furthermore, a CNV was detected in *PRPF31* and SVs were identified in *ATXN7, PRPF31, PDE6B and RPGRIP1*.

Hypotheses for why target capture did not detect these variants are discussed in Section 4.4.

Section 5 highlights the need for a cost-effective approach that is focused on sequencing one or a few genes of interest when attempting to ascertain the genetic cause in a clinically distinct phenotype. STGD1 serves as a prime example in this case, most typically associated with variants in *ABCA4.* smMIPs, a type of small circular probe, were designed and employed on 162 unresolved STGD1 and STGD1-like phenotypes[137]. The entirety of the *ABCA4* gene was sequenced in 110 nucleotide increments, including introns and 40kb of flanking sequences[23]. The candidate variant detection rate was 61% and 59% for prescreened and unscreened cases respectively. 27% of cases that have been genetically resolved via smMIPs sequencing were solved due to the detection of a likely pathogenic splice-altering variant. These results highlight that splice-altering variants represent a significant cause of disease in the Irish IRD cohort.

These smaller scale studies carried out as part of Target 5000 illustrate the importance of moving towards more comprehensive sequencing methods for unresolved cases. In addition the studies highlight the value of international collaboration within the IRD field and the benefit that this work can have for those affected by an IRD. These pilot studies into alternative detection methods have proved very successful. Moving forward, unscreened and unresolved STGD1 cases will undergo smMIPs based sequencing of *ABCA4.* Other IRD cases that remained unresolved following TCS will undergo WGS. Target 5000 target capture NGS achieves a relatively high diagnostic yield of almost 70%, indicating that it remains an effective preliminary diagnostic measure before WGS is employed on a case dependent basis.

The following sections will focus on broader areas pertinent to the genetics and IRD community in Ireland and internationally. Finally a perspective on the future of our field is provided, centred on what is important for the investigation of IRD patients who have yet to receive a comprehensive genetic diagnosis.

7.2 Challenges faced in the interpretation of IRD variants

A vitally important part of every genetics service is the accurate interpretation of genetic variants. A basis for this was provided by the publication of guidelines by the American College of Medical Genetics and the Association for Molecular Pathology (ACMG-AMP) in 2015[185] (Appendix 10). Parameters such as population, segregation, and functional data among others are used to gauge where on a spectrum of benign to pathogenic a variant lies. Over the course of this doctoral degree, it has become apparent

that current standard guidelines do not accurately assess all variants that are potentially causative of IRD. Challenges are not limited to those provided - only those directly encountered through Target 5000 will be discussed, in addition to challenges specific to genomics in Ireland as a whole.

This doctoral research, in line with the literature, demonstrates the contribution of splice-altering variants to the causation of IRD. Only two lines of evidence in the current ACMG-AMP guidelines reference splice-altering variants directly. That is *"PVS1 Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion) in a gene where loss of function (LOF) is a known mechanism of disease"* and "*BP7 A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site AND the nucleotide is not highly conserved"*. PVS1 also operates on the caveat "use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact".

Starting with PVS1, it could be argued that immediately awarding a very strong piece of evidence (given that it does not leave the rest of the protein intact) to a CSS variant because of its genomic location is conjectural. For example, the thirteenth exon of *USH2A* can be skipped, and is even being explored as a therapeutic option for individuals with variants within that exon[345]. Additionally, natural exon skipping is a recognised phenomenon and should be considered for any variants affecting canonical/CSSs. Splicing is not binary. It is not just a question of whether an exon is skipped or not, but also to what degree it is skipped and if the reading frame is maintained. Certain percentages of skipping may be tolerable and this will vary by variant, exon, gene and disease. The ratio of wild-type to mutant transcript produced is paramount in determining the pathogenicity of a variant that results in exon skipping. Recommendations published by the ClinGen Sequence Variant Interpretation Working Group in 2018 offered additional guidance on the implementation of PVS1[346]. Their decision tree includes considerations such as frame, nonsense medicated decay, transcript relevance, importance of the region to protein function and percentage of remaining wild-type protein. These recommendations do facilitate more accurate interpretation of PVS1-associated variants. BP7 accurately describes that synonymous variants may have an impact on splicing process, as demonstrated by section 6.2 of this doctoral thesis. A "pathogenic" counterpart of this line of evidence would be useful, and is somewhat covered by "*PP3 Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc*". However, the problems listed above in

relation to wild-type:mutant transcript ratio persist. These problems suggest that a line of evidence or caveat directly addressing this issue would be appropriate not just for IRDs but also genetic diseases as a whole.

In order to provide evidence for the pathogenicity of putative splice altering variants, functional analysis is required. This may take many forms, and could prove difficult in terms of interpretation in both clinical and research settings. Currently, this is addressed by PS3: *"Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product. Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well-established"* and by BS3 (no functional effect demonstrated). There are a number of problems associated with these descriptors. Firstly, functional effects may vary depending on what model system is used. The effect demonstrated by a midigene splice assay may not be the same as that from an organoid model, and neither completely recapitulate the actual tissue or cells type(s) perturbed in the patient. Should we consider that these lines of evidence be weighted based on model sophistication/similarity to the patient? Secondly, statements like PS3/BS3 are somewhat vague, resulting in different clinics and laboratories interpreting them in different ways. What one group may consider well-established, another will not. The Association for Clinical Genomic Science in UK automatically use PS3 if a splice assay demonstrates a defect for example[347], while the Australian Consortium for RNA diagnostics do not[348]. Recommendations published by the Clinical Genome Resource Sequence Variant Interpretation Working Group in 2019 offered additional guidance on the use of PS3/BS3 criteria in order to facilitate more accurate and consistent use[349]. In short they suggest that the method of functional analysis defines the disease mechanism in its results, that basic and variant controls are included, replicates performed and that sufficient statistical analyses are carried out. Adherence or nonadherence to these parameters place the use of PS3/BS3 criteria on a spectrum of don't use, use as supporting evidence (PP), use as moderate evidence (PM) or to use as strong evidence (PS). This detailed guidance is welcomed and should be considered in any future updates to the ACMG-AMP guidelines.

A number of groups have suggested criteria for the interpretation of splice altering variants[277,348,350–352]. One such study by Bournazos *et al*., on behalf of the Australian Consortium for RNA Diagnostics provides suggested standardised practices for RNA diagnostics. They demonstrated that the use of functional analyses in a cohort of 74 families with diverse monogenic conditions enabled variant reclassification for 75% of variants. Informed by this, they recommend that the PS3/BS3 criteria are currently the

most appropriate for interpreting splice altering variation. Moreover they suggest that PS3/BS3 is subdivided into very strong, strong and moderate evidence levels based on level of mis-splicing, technical approach design, variant zygosity, RNA-seq data, tissue examined/affected and alternative splicing among other parameters.

Another recently published study by Ellingford *et al*., recommends guidelines for the clinical interpretation of non-coding variants in general and not just those that affect the splicing process[350]. This includes variants that cause disruption of promoter/enhancer elements, the 5' untranslated region (UTR) and the 3' UTR/miRNAs. This study aims to apply existing guidelines to non-coding region variants. They also express caution in the application of PVS1 to CSS variants by making points regarding alternative splicing, dosage and the ability for splice-altering variants to have multiple impacts or partial effects. They also suggest changes to PS1, PM5, PM3, PS3/BS3 and PP3/BP4. They recommend updates to these criteria based on considerations such as non-coding sequence constraint, previous reports of pathogenic non-coding variation in conjunction with splice prediction software, non-coding region variants that are predicted to have exactly the same impact, on the same gene, as established pathogenic variants, stringent AF cut offs, genetic ancestry and types of functional assessment employed. Importantly these suggestions are conservative, given the breadth of knowledge that remains to be discovered in relation to non-coding region variants.

Another complication to variant interpretation is that of phenotypic variability, with particular reference to incomplete penetrance. This is a phenomenon where a known likely causative genotype is present however the associated phenotype is not observed. Patient 23 and their family in section 4 of this thesis are an epitomical example. This patient harbours a whole gene deletion of the *PRPF31* gene, as does their mother. However, their mother is phenotypically unaffected. Similar *PRPF31* cases have been noted in the literature and it is thought that *cis* or *trans* modifiers affecting the expression levels of this gene are the underlying cause of the incomplete penetrance[220,353]. Cases such as these are not readily addressed by current guidelines. When attempting to determine if a variant is pathogenic or not, a common procedure is to check segregation of disease with potentially causative genotypes within the family. If an unaffected family member has the variant you are investigating, it raises concerns about its pathogenicity. This is not as simple when genes with a known history of incomplete penetrance are involved. In the above case, given that the variant is a gross deletion, the individual had WGS and no other potential cause of disease was identified, it is likely that this is indeed the causative variant. Guideline stipulations that consider incomplete penetrance are important.

Over the course of this doctoral research it has also become apparent that additional data could be added to open resource databases in order to reflect the complexities of genetic disease. In the acknowledgements section of this thesis you will read an Irish seanfhocail (proverb) "*Ar scáth a chéile a mhaireann na daoine"* meaning that we live in each other's shadows. It speaks to interdependency, and genomics is no different. While our exact genomic make-up is individual to us, it exists as part of a genomic context nationally and globally that affects the way that changes to it are interpreted. We cannot simply view singular genetic variants in isolation, and this is where genomic databases are important. Examining if the variant has been seen in other populations and at what frequency or if it is highly frequent in a certain population is extremely informative. A significant limitation of the studies included in this thesis was the lack of available data on age, gender and ethnicity, which should not be overlooked. Future studies should and will endeavour to collect this data where possible. Later in this discussion the lack of an open source Irish genome project and why it would be extremely useful will be discussed. However, here it is important to point out that at least Irish individuals are in part represented by non-finnish European genomes, which greatly predominate large population databases like gnomAD. For example, v2.1.1 contains 126708 non-finish Europeans samples versus 24928 African/African American samples. It is imperative that we include more samples from under-represented populations in such databases otherwise we are doing patients from these backgrounds a great disservice. What may appear rare to us at this moment in time may actually be a relatively common SNP in an under-represented cohort, and misattribution of disease causation may result from this. This is a common occurrence due to population bottlenecks and founder effects.

In terms of IRD variants specifically but without doubt broader applications as well, a key improvement to databases such as ClinVar and LOVD would be the mandatory disclosure of allelic context. This refers specifically to genes where, depending on variant location, the disease can manifest in different inheritance patterns. One such example is the *BEST1* gene, pathogenic variants in which cause dominant and recessive forms of Best vitelliform macular dystrophy. For illustration purposes, imagine a patient presents with one singular *BEST1* variant. It is not immediately clear if there was a family history of disease and therefore the inheritance patten is in question. If this is a recessive form of the disease then only a partial genetic diagnosis has been ascertained. If this is a dominant form of disease then a more accurate potential diagnosis is now available. A geneticist decides it is best to consult a database like ClinVar for example. The variant has been reported multiple times as pathogenic, but the evidence supplied in the database does not

detail if this variant has been associated with dominant or recessive disease. Moreover, if the variant is reported as recessive, rarely is the variant on the other allele disclosed. This lack of information can make variant interpretation difficult.

An update that was highly welcomed was that of integrated variant co-occurrence in gnomAD in 2021[354]. For example, in section 6.2 of this thesis a patient with a complex allele in *PROM1* is described. gnomAD's variant phasing data provided valuable insight to this study, illustrating that every person who was a carrier of the synonymous splice altering variant was also a carrier of the missense variant on the same allele. This aided with the interpretation of both variants. As long read sequencing technologies such as PacBio or nanopore sequencing become more frequently employed, phasing data such as this will become more readily available.

While these points represent significant challenges in the interpretation of IRD and genetic variants in general, many groups are working towards improved and more comprehensive guidelines. Open data sharing and collaboration is imperative to this endeavour. The field of IRD genomics has the potential to massively contribute to the development of additional standardised guidelines given the wealth of sequencing and clinical data available and the breath of cases that represent challenges to some of the existing parameters. The Target 5000 team is actively involved in this work.

7.3 Genomic services in Ireland with a focus on IRD patients.

Currently, there is no designated genetics and genomics strategy functioning within the Irish Health Service Executive (HSE). The clinical application of genomics is highly dependent upon the genomics research community which can negatively affect both research endeavours and clinical service provision in terms of resources and wait times. On May 12th of 2022 a National genetics and genomics strategy was initiated by the HSE at the inaugural meeting of the HSE National Genetic and Genomic Steering Group[355]. A detailed strategy is set to be published by the end of 2022.

2021 saw the publication of an overview of Target 5000 as a standardised all-Ireland pathway for the diagnosis and management of IRDs[116]. The publication described how Target 5000 was set in motion in 2009, following a meeting of Irish IRD specialists, scientists, and patient groups with a primary goal to develop a national register and individuals care plans where necessary. This publication also descrived a survey that was conducted in 2014 across 12 clinical ophthalmic sites in Ireland. This was done in order to assess infrastructure present at the time for IRD assessment and to identify any areas of concern. Results of the survey confirmed the absence of a standardised approach to

assessment. Strikingly, survey response indicated that an estimated 52% of the Irish IRD population were not being reviewed and only 33% of visits were at IRD subspecialty clinics. As of 2021, 1482 IRD patients had been clinically assessed by the pathway with 1004 individuals genotyped as described in section 3. The general approach is continuous clinical characterisation of new IRD patients and genotyping using panel-based sequencing of 254 IRD genes as funding becomes available. Additional, research-oriented studies such whole gene, WES, WGS and functional analyses are also performed on a case dependent basis as outlined in this doctoral thesis.

Genetic testing prior to Target 5000 was on an individual basis with funding from public departmental resources or private patient funds. Predominant funding sources now include charity and research grants. In recent years some IRD patients have undergone DNA sequencing through the HSE on a case-dependent basis. Prior to this, all patients sequenced in Target 5000 have undergone testing in a research grade setting. Confirmation of variants identified is performed by a commercial accredited clinical laboratory. The Target 5000 team has pioneered the integration of genomics in standard healthcare practice in Ireland and the value that a collaborative effort between research and the clinic can provide. However, it is imperative that first-tier genetic testing for IRDs becomes a routine procedure in the Irish health service. This would allow research laboratories to focus on second tier genetic testing and functional analyses, creating more space for innovation and discovery within the research setting. It would also result in timely access to genetic opinions or to genetic testing at the clinic level which will facilitate faster diagnoses, treatments, and interventions for service users. The Target 5000 team recommend that the following are vital in developing a national IRD programme as outlined in Stephenson *et al*., 1. Identification of national centres of ophthalmic expertise that are appropriately staffed and resourced. 2. Engage and partner with support agencies and patient organisations. 3. Concentrate clinical genetics and support services at these sites. 4. Encourage referrals of all diagnosed or suspected IRD patients to these centres with regional follow-up. 5. Implementation of a robust pathway for funded genetic testing and resolution of contentious variants. 6. Routinely use and regularly update variant classification and collaborative databases. 7. Become members of international alliances (e.g. ERN-EYE). 8. Establish therapeutic pathways nationally and internationally. 9. Create bespoke care plans for each patients/pedigree and finally 10. Audit outcomes for patients[116]. It is also imperative that research is embedded into a national IRD strategy. Essential research endeavours include but not limited novel variant evaluation, functional analyses, novel methodology development and the application of artificial intelligence. A

report published by Health Research Charities Ireland on an Irish Health Research Forum event held in May 2022 discusses embedding research in the Irish health service[356]. Making research an integral part of the Irish health service would facilitate opportunities such as the improvement of patient outcomes, better opportunities for public and patient involvement, streamlining ethics approval, and improved research outcome dissemination.

It is imperative that users themselves are central to the creation of a genomics service, that it adequately reflects their needs and views them as key collaborators. Studies have investigated the patient perspective on an ideal genetic service. Patients emphasised the importance of genetic screening throughout the diagnosis process[357–360]. They noted that both pre and post-test counselling were extremely important to prepare for and to deal with a spectrum of results. The continuous need for excellent communication of results was stressed, including information on disease progression, underlying genetic cause, inheritance patterns and current research ongoing in the field[357,361].

Additionally it should also be highlighted that as of yet, there is no public Irish genome sequencing project. A partly government funded Irish population genome sequencing study was announced by Genuity Science (previously Genomics Medicine Ireland). They plan to obtain genomic data from 1 in 10 people in Ireland, including patients with common and rare diseases, making this study one of the world's most comprehensive genome sequencing programmes[362]. WGS data from unaffected control individuals with highly similar genetic backgrounds to patients would greatly aid in the interpretation of genetic variants in Ireland by allowing for the identification of common Irish polymorphisms and determination of carrier frequency. Unfortunately these data remains publicly unavailable. An open source Irish genome database would be an excellent step forward for the Irish health system and genomics community, and would be welcomed as part of the governments newly initiated strategy.

7.4 The health economics of IRDs in Ireland

IRDs are individually classed as rare in the European Union[363], historically limiting funding for diagnostic and therapeutic research due to the relatively low number of individuals affected. This has a significant impact on patients, families and care givers[364,365.] IRD COUNTS is a consortium of patient organisations and industry partners that aim to establish the prevalence, burden and impact of IRDs in the Republic of Ireland (ROI) and the UK[366]. A study published in 2020 funded by IRD COUNTS presented data to define the economic burden of IRDs, highlighting the need for investment in IRD services and to provide a basis for policy and planning from a patient-centric perspective[14].

This landmark study has provided valuable insights into the personal and societal economic impact of IRDs and will be discussed in detail below from the ROI perspective.

This study determined that costs attributable to IRDs in the ROI were ~€48.1 million in 2019. These costs can be divided into 5 categories; 1. Wellbeing costs (years of healthy life lost to IRD); 2. Productivity losses (reduced workforce participation and presenteeism); 3. Formal and informal care costs; 4. Costs to the health system and 5. Society wide efficiency losses. Starting with wellbeing as it is centre to the World Health Organisations definition of health: "a state of complete physical, mental and social wellbeing and not merely the absence of disease or infirmity". A study by Garcia *et al*., found that profound vision loss impairs psychological wellbeing in young and middle-aged individuals[367]. This further highlights the cycle that a lower quality of life as a result of visual impairment results in the inability to work and thus lower socioeconomic status[368]. Wellbeing costs made up the largest subsection of total cost, at ~€15.6 million[14].

Some individuals living with/affected by an IRD find it difficult to actively participate in the workforce relative to the general population. This is due to a number of factors including disadvantages in job seeking and lack of support in the workplace. Productivity costs of IRDs in the ROI were estimated to be ~€9.1 million, with individuals with an IRD 55.7% less likely to be in paid employment than the general public. Further contributing to this figure was the finding that IRDs resulted in a 9.6% reduction in productivity while at work. Strikingly 24% of individuals affected by IRD and in paid employment stated that they did not use visual aids at work because they were not available to them[14].

IRDs can have a significant impact not only the affected individual but also on people in their support system. 7.9% of survey respondents indicated that they received formal care, that is, from private health care staff or assistance with childcare, housekeeping and shopping. Formal care costs in the ROI were estimated to be ~€1.7 million, with an average of 14.1 hours per week. Informal care is generally that which is carried out by family members and friends and can include organising and timing of medication, cooking and laundry as well as cleaning and monitoring the patients wellbeing. An average of 16.3 hours informal care was reported and incurred costs of ~€6.5 million in 2019. This level of reliance on informal care has significant impacts on those with IRDs and their carers.

Health system costs were ~€2.1 million in the ROI. The majority of these costs were associated with secondary health care at ~€0.8 million. Losses due to lack of taxation revenue or expenditure on government programs were calculated by applying the marginal

burden of taxation to the total of lost taxation and government expenditures. These losses are estimated to be ~€8 million, resulting in a ~€4 million loss in taxation revenue[14].

Currently, lack of awareness and investment in IRDs has resulted in significant personal costs to those affected by an IRD and economic losses to the Irish economy. These costs may be alleviated by investing in diagnostic and therapeutic research, further education of professionals involved in the care of those with IRDs, access to genetic screening and counselling and reassessment of policy for the reimbursing those in the IRD community affected by these costs. This study by Galvin *et al*.,[14] highlights the need to view costs of IRDs from a societal rather than a healthcare system perspective to improve the lives of those affected by an IRD through limiting economic burden and associated negative psychological and psychosocial effects.

While there is a significant knowledge gap in terms of the cost of embedding a genomics service in the HSE, an open letter published in Health Research Board Open Research does discuss the future of genomics in Ireland from a health economics perspective[369]. Initially genome sequencing was hugely expensive and inaccessible. However, there is now a myriad of studies to suggest that genomics results in decreased costs when applied in the clinical setting[370–374]. These savings stem from avoiding explorative repeated genetic testing, reducing use of ineffective treatments given patient genomic context and shorter hospital stays, thus lowering time on waiting lists[375,376]. Such savings could be better invested in care and supports for those with genetic diseases. In terms of the cost of genetic testing for IRD, Stephenson *et al*., outline an example of 1522 IRD patients. Genetic testing for a proband and relative is estimated to be €2.18 million if performed akin to Target 5000 testing[116]. The cost of illness outlined by IRD COUNTS and evidence of success in other countries underscores the potential benefits that utilising genomics in the clinical setting can provide for a health system and its patients.

7.5 The personal impact of genetic testing for IRDs

While many studies have illustrated the value of genomic testing in research settings, the impact of testing on patients involved is not well explored. It is imperative that we do not simply view genetic testing as something "about" a patient, but rather a process where the patient and their wellbeing is central. It is important to explore why individuals want to undergo genetic testing, particularly for rare diseases where treatments are still in the pre-clinical stage. Patients have both personal and altruistic reasons for undergoing genetic screening, with patients on the UK National Health Service reporting feeling empowered by their genetic test results[361,377]. Patients are optimistic about future

potential treatments and access to novel therapeutic options for themselves and others[378]. As gene-based therapies for IRD are still widely at a research and clinical trials stage, the other benefits of testing are likely to stem from elsewhere, including more accurate diagnoses and understanding of disease inheritance and progression. Patients diagnosed with RP have also displayed interest in prenatal testing[378–380], with case studies describing the use of both prenatal diagnosis and pre-implantation diagnosis for IRDs[381,382].

Research in the area of the personal impact of genetic testing has centred on those diagnosed with an inherited predisposition to cancer[383,384]. A study by Meiser *et al*., found that the majority of those who underwent genetic screening for BRCA1/BRCA2 variants had a positive experience[384], but also notes alongside a study by Heshka *et al*., that *those with a* positive genetic test result (one where a pathogenic *no* variant was detected) experienced increased levels of anxiety and depression[383]. This same psychological impact from genetic testing has been seen in those affected by an IRD, where some noted that increased awareness of future progression of disease and worsening vision can be difficult to come to terms with[358].

A study by McVeigh *et al.,* coalated the existing data on the psychosocial and service delivery impact of IRDs and conducted interviews with IRD patients who had undergone genetic testing to understand their experiences. They explored the experience of 23 IRD patients in the UK. The results of this study identified three key themes. The first is the journey towards a genetic diagnosis. Respondents detailed feeling loss of control, loss of independence and learning to adapt. Genetic testing was described as a milestone, without which it was difficult to accept their diagnosis or to think about treatment options. The second theme was the impact that genetic testing had. All participants were highly keen to undergo genetic testing. Motives included a need for closure, increased knowledge in terms of disease progression and risk to family members. All individuals reported both positive and negative aspects to the testing process. Disadvantages to testing included wait times, percieved secondary findings such as the potential for a syndromic aetiology and worry surrounding recieveing a negative or result of unknown significance. Experience differed between those who received a positive result, a VUS/carrier result or a negative test result. Those with positive results felt empowered by the knowledge. Patients also felt that being able to name their specific condition facilitated access to networks and support groups. Those who were carriers or had a VUS did feel dissapointed, but had hope that they would receive a positive result in the future. Those with a negative result reported feelings of dissapointment and uncertainty. However, research has demonstrated that continued contact with these

individuals and their families indicated they felt encouraged to continue with genetic testing[360]. This finding underpins the importance to follow up with those who *recieve* negative test results. However, even when presented with inconclusive results some patients feel they obtained benefit from genetic screening and that they wanted to be informed of non-actionable results pertaining to the IRD diagnosis. This has also been reported in other studies on other genetic disease cohorts[357,385]. The final theme was that of the experience of service delivery. Respondents viewed genetic testing as a way to access treatment and further knowedlge in the field. They hoped that with medical advancements additional causes would be detected and new treatments would be developed. Many participants reported positive experiences within the genomic service, including information provided by consultants and support provided by genomic counsellors. Some patients reported that focus was often placed on future generations rather than themselves at that moment in time, which could feel like their immediate needs were not being addressed. Valuably, participants also described suggested improvements to the service, such as providing updates on services and research on a regular basis.

These data supports the view that there should be a genetic testing provision for IRDs in Ireland. Patients currently enrolled in Target 5000 may otherwise not have been able to avail of genetic screening. Benefits of testing include patient empowerment, better understanding of disease prognosis, identifying risks to other family members, reproductive empowerment and potential access to future treatment where appropriate. However, it is also important to note that with genetic testing comes the potential for increased anxiety in the affected individual and the family. Therefore, it is paramount that genetic testing is only carried out in conjunction with a specifically qualified genetic counsellor and other genetic specialists.

7.6 Current gene based therapies/clinical trials for IRD.

This section will focus on a subset of gene-based medicines that are currently in use or are being investigated for their efficacy in treating IRD. Those still undergoing clinical trial and the most recently completed trials for these therapies are summarised in Table 7.1. IDs starting with “NC” are provided for reference on clinicaltrials.gov[386]. This is a non-exhaustive list and focuses on a subset IRDs, disease mechanisms and variant types that have been disucssed in detail in this thesis. Given that therapies for IRD is a wide-reaching and extensive topic, therapies such as cell therapy, neurotrophic factors for neuroprotection, stem-cell-based therapy, optogenetics and retinal implants will not be

discussed in detail here, but it should be noted that these are also extremely important theapeutic endeavours given the genetic heterogeneity associated with IRDs.

Gene augmentation therapies involve delivering a functional version of DNA or RNA to supplement what is missing or disfunctional in the affected individual. Luxturna*TM*, the Food and Drug Administration and European Medicines Agency approved treatment for pathogenic biallelic *RPE65* variants is a prime example of this. A functional copy of *RPE65* is delivered sub-retinally using an AAV vector. Pre-clinical work focused on gene supplementation to the rd12 mouse and a canine model[387,388]. Once benefit was observed, human clinical trials commenced. The phase III trial used a primary efficacy endpoint of 1-year change in a standardised multi-luminance mobility test, measuring functional vision at specified light levels[389]. 13 (65%) of 20 intervention participants, but no control participants, passed the multi-luminance mobility test at the lowest luminance level tested. No related serious adverse events were reported. Approval of Luxturna*TM* was monumentos for the field as it was the first instance of a commercially approved AAV gene therapy for use in humans[390]. A number of follow up studies have been carried out. One study has illustrated that improvements to persist at least three-four years after administration of the therapy[391]. However, perifoveal atrophy was noted in a large majority of study participants (18 eyes of 20) starting in some 5 months post delivery[392]. A larger follow up study of 41 patients found that 77 eyes displayed decreased foveal thickness. No significant change in vision was found in adults however some children displayed a trend of improvement. Following this, 29% of eyes from the paediatric cohort improved by a minimum of two lines[393]. Importantly, another study investigated the long-term effect of the therapy on photoreceptor function and incidences of post-operative complications in 27 eyes from 14 patients. These complications include increased intraocular pressure in 59% of cases, intraocular inflammation in 15% of cases and vitreious opacities in 26% of cases. However, the study found that these complications resolved over a number of months. Additional studies should be performed to ensure the tolerability and safety of the medicine in the real-world setting[394].

Table 7.1 | Summary table of a number of clinical trials for IRD currently underway.

Condition	Gene	Biologic	Mechanism	Responsible party	Current phase	ID	Enrollment	Ages eligible (years)	Sexes elligible	Status
CHM	*CHM*	BIIB111 (rAAV2-REP)	Replacement	Biogen	Phase III	NCT03496012	170	>18	Male	Completed
CHM	*CHM*	AAV2-hCHM	Replacement	Spark Therapeutics	Phase I/II	NCT02341807	15	>18	Male	Completed
STGD1	*ABCA4*	SAR422459	Replacement	Sanofi	Phase I/II	NCT01736592	27	>6	All	Active - not recruiting
Type 10 LCA	*CEP290*	QR-110	Splice modulation	ProQR Therapeutics	Phase II/III	NCT03913143, NCT04855045	36, 15	>8, <8	All, All	Active - not recruiting, Recruiting
XLRP	*RPGR*	AAV2/5-RPGR	Replacement	MeiraGTx UK	Phase I/II	NCT04312672	36	>5	Male	Recruiting
XLRP	*RPGR*	AGTC-501 (rAAV2tYF-GRK1-hRPGRco)	Replacement	Applied Genetic Technologies Corp	Phase II/III	NCT03316560	42	6-50	Male	Recruiting

In phase I/II clinical trials BIIB111 (rAAV2-REP) promisingly improved visual acuity in choroideremia patients however adverse effects were reported. A study reporting 24 month results two patients had retinal hole over non-functional retina[395]. Another report after 24 months described one patient with a localised intra rerinal immune response[396]. In a following phase II trial, results were promising in terms of safety[397]. However, a following phase III trial did not reach it's primary endpoint of 15 ETDRS letter improvement[398].

Another gene therapy for choroideremia, AAV2-hCHM has undergone a phase I/II clinical trial. The therapy was injected unilaterally in patients with a BVCA of greater than 20/200. A report detailed that in 8/9 patients the cone photoreceptor mosaic resettled on the RPE following subretinal injection. Widepread cone loss was not reported which indicated that these photoreceptors are not lost in the area of the retina where detachment occurs[399]. Two year results of the phase I/II trial reported that visual acuity was within 15 letters of baseline following subfoveal injection in 13 of 15 patients. One patient experienced acute foveal thinning with unchanges perfoveal function and another patient experienced a macular hole thaht suggests foveal vulnerability to subretinal injection. Longer-term follow-up studies are needed to determine differences between patients and between intervention and control eyes[400].

SAR422459 is a lentiviral mediated treatment for STGD1. A phase I/II clinical trial examining this treatment is currently active but not recruiting. Early poitive safety data was reported. However, it was terminated due to sponsorship loss. Three-year safety results have been reported describing that sub retinal injections were tolerated well with only one case of ocular hypertension. No significant changes in visual function were reported. However, 27% of eyes did display worsened RPE atrophy on FAF. One patient who recieved the highest dose had a significant reduction in macular flecks. Future studies are needed to explore the efficacy of this medicine[401].

RNA therapeutics are being developed to address gene defects resulting from pathogenic variants causing mis-splicing of IRD genes, such as the examples outlined in the results chapters of this doctoral thesis. AONs are promising tools under exploration for their ability to correct abberant splicing patterns. They are small DNA/RNA oligonucleotides that can be designed to be complementary to pre-mRNA with modified chemistries to increase longevity. With respect to IRDs, AON-based therapies have been most extensively applied to a DIV in *CEP290,* c.2991+1655A>G causative of LCA10. This variant creates a cryptic splice donor site resulting in a pseduoexon inclusion. Pre-

clinical studies were promising, with significant rescue of CEP290 protein levels in patient derived fibroblasts and lymphoblastoid cells[248,249]. Promising rescue was also achieved in organoids and murine models of disease[250,285]. A phase 1b/2 clinical trial had a primary outcome measure was evaluation of safety and tolerability according to the frequency and severity of ocular adverse events[402]. 11 individuals received the treatment, with key findings including increased mild adverse events with higher dosase of the drug. Statistically significant improvements in visual acuity and retinal sensitivity were reported, indicating that this treatment should continue to be investigated. However, a phase II/III trial (NCT03913143) did not meet its primary outcome measure of a change in best-corrected visual acuity (BCVA) relative to baseline after 12 months of treatment versus a sham-procedure. The drug was generally well tolerated. However, additional post-hoc analysis did reveal that when the effect in the treated eye was compared to the untreated contralateral eye in the same patient At 12 months, a benefit in vision was observed as a mean change from baseline in BCVA of -0.12 logMAR in the treated groups. A phase II/III trial (NCT04855045) investigating the therapy in children under 8 years of age is currently recruiting[403]. In August 2022, ProQR announced that it will focus exclusively on ADAR editing and is seeking to partner its ophthalmology programmes including the *CEP290* and *USH2A* trials[404].

In addition to the discussed trials a myriad of pre-clinical studies are investigating novel and innovative approaches to the genetic treatment of IRDs. These include but are not limited to ADAR-based RNA editing, post transciptional gene silencing, CRISPR interferance and activation and translational read through inducing drugs. With current successes and despite setbacks and challenges, the future looks promising for the treatment of these complex and heteroegnous diseases.

7.7 Future perspectives on genetic diagnoses for IRD

Incredible progress has been made in diagnostic DNA sequencing and genetic therapies over the last 10 years. Targeted panel sequencing, followed by more comprehensive techniques such as whole gene or WGS has uncovered a complex genetic architecture which is actively being addressed by numerous studies and clinical trials. With a positive diagnostic rate of approximately 70%, IRDs represent ideal candidates for routine genetic analysis in the clinical setting. Hopefully, in the not so distant future we see implementation of routine DNA sequencing for IRD patients within the HSE, with research endeavours focusing on those who have not yet received a diagnosis via currently employed software and sequencing techniques.

This poses the question – what is causing IRDs that remain genetically undiagnosed? The studies in this PhD thesis provide insight into this, highlighting the distinct role that splice-altering variants have in the causation of IRDs and how they may be overlooked. Continued exploration of regions that harbour these variants and their disease mechanisms will undoubtedly provide additional patients with diagnoses. This section will focus on other potential sources/mechanisms of disease causation and how we can work towards implementing systems that will aid in their discovery.

Pathogenic variants in the promoter regions of IRD genes have been detected. Radziwon *et al*. described two patients harbouring upstream variants in CHM[405]. Luciferase reporter assays were employed to examine the effect of *CHM* c. −98: C>A and C>T on gene expression, with both leading to decreased activity. They also described the promoter region for *CHM* as c. −119 to c. −76. It is likely that other IRD patients harbour similar undiscovered variants and improvements in their detection is imperative. Insuring that defined promoter regions are included in sequencing panels would facilitate this on a larger scale. Examining large numbers of patients with clear clinical diagnoses of conditions with single gene aetiologies for promoter variants by comparison with controls may be a good place to begin. This is actively being explored in Target 5000.

Untranslated region variants may also represent a significant cause of IRD. Coppieters *et al*., set out to ascertain a genetic diagnosis for a consanguneous family with LCA in which no pathogenic coding variants were detected. Sequence exploration revealed a novel homozygous mutation in the 5' UTR of *NMNAT1*, c.-70A-T[406]. Following this, they also discovered an adjacent variant, c.-69C>T in another family. Examination of the effect of these variants using mRNA from patient derived lymphocytes and using a luciferase assay in human RPE-1 cells demonstrated decreased mRNA abundance and decreased activity respectively.

CNVs and SVs are certainly under-represented causes of IRDs, given that the most commonly employed methods of sequencing (panel sequencing and WES) are not optimised for their detection, particularly when these variations are heterozygous in nature or if their breakpoints lie outside captured regions. An extensive meta-analysis of CNV prevalence in IRD was performed in 2018 revealing 317 CNVs had been reported across 81 IRD genes[407]. Deleterious SVs were detected in a multitude of IRD-associated genes, including *BEST1, EYS, MERTK, USH2A* and many more from the aforementioned study alone. Large deletions have also been reported in *RPGR*[408], *CHM*[13], *OPN1LW/OPN1MW*[69], *USH2A*[56], *OAT*[118] and *USH1C*[118]. The disconcordance between the ability of first-tier sequencing procedures to detect all CNVs and SVs and incidence

rates of variants in the literature indicate that they may be missed in a routine clinical sequencing setting. Duplications within IRD genes represent a significant challenge as sequencing procedures involving amplification may mask any change in read depth that could represent the presence of a duplication event. As WGS becomes more readily employed, it is likely that patients who are as of now genetically unresolved may be diagnosed with an IRD associated CNV or SV.

Creation or disruption of topologicaly associated domains (TADs) may also represent a significant cause of IRD. A recent study De Bruijn *et al*., demonstrated that an SV in the RP17 locus created new TADs and ectopic retinal enhancer-gene contact in previously unresolved families with autosomal dominant RP[409]. Eight different complex SVs were characterized in 22 autosomal dominant RP-affected families with >300 affected individuals. A Hi-C map of the RP17 locus in retinal organoids revealed that a neo-TAD with with ectopic contacts between GDPD1 and retinal enhancers was created. Therefore, altered TAD structure resulting in increased retinal expression of GDPD1 is likely the underlying disease mechanism. This landmark study excellently illustrated the potential contribution of epigenetic abberations to IRD causation. It is likely that similar phenomena will be observed in other IRD-associated genes in the future. More elusive epigenetic mechanisms may also affect disease outcome. Gene expression is partly regulated by processes such as DNA methylation and histone modifications[410,411]. It has been illustrated that in DNA isolated from PPCs that the promoters of an array of IRD-associated genes (*NR2E3, PDE6A, RHO* and *CNGB1* among others) are hypermethylated. This level of methylation significantly reduced during differentiation into photoreceptors thus increasing the expression of these genes[412,413]. One could hypothesise that any perturbation to this demethylation mechanism could cause photoreceptor dysfunction and thus disease.

Mobile elements have also been implicated in the causation of IRD. *Alu* sequences fall into this category of variation and over 1 million such repeats *are* present throughout the human genome[36]. Reports of *Alu* elements in IRD-associated genes are present in the literature. For example, a homozygous 353-bp *Alu* insertion in exon 9 of the *MAK* gene has been described, with a high AF in individuals of Ashkenazi Jewish descent[414]. Additionally, a founder 328-bp *Alu* insertion has been identified in exon 4 of the *RP1* gene in 26 Japanese patients[289]. Another interesting study by Delvallée *et al*., described the presence of a SVA F retrotransposon insertion in exon 13 of the BBS1 gene[415]. WGS as well as de novo assembly and SNP array analysis revealed that this variant was 2435-bp in length and displayed hallmarks of LINE1 mediated retrotransposition. These studies

demonstrate the role that mobile elements play in the causation of IRDs, and all authors highlight that improvements are needed to current bioinformatic pipelines to aid in their identification. Indeed, such studies are mirrored in other neurodegenerative disorders with an renewed focus on mobile elements as causative of some nerudegenerations[416].

Another overlooked source of disease may be where variation lies on a spectrum of severity. For example, there are 80 conditions categorised as syndromic that include a retinal defect as a hallmark[417]. However, if we strictly view IRDs as syndromic vs non-syndromic, we may risk making assumptions about the underlying genetic cause, particularly when designing sequencing and virtual panels. Hypomorphic variants are examples of this. Take the *HGSNAT* gene, pathogenic variants in which are associated with both non-syndromic autosomal recessive RP and recessive mucopolysaccharidosis. A common hypomorhic variant, c.1843G>A,p.(Ala615Thr) is a frequent cause of RP alone, without systemic features of disease. It could be argued that we are only aware of this variant given its high AF and prominence in *HGSNAT* related disease. Another example includes variants detected in *DYNC2H1*, a gene typically associated with two severe cilliopathies; Jeune asphyxiating thoracic dystrophy (MIM: #613091) and short-rib polydactyly (MIM: #613091). A study by Vig *et al.,* described four novel variants in this gene that either affected a retina specific transcript or are hypomorphic thus causing non-syndromic RP only[418]. It is possible that patients tested prior to this new genotype-phenotype correlation have not had this gene sequenced and thus variants may be missed. It is imperative to ask if we are overlooking non-syndromic IRD caused by rare variation in genes more typically associated with severe syndromic phenotypes. It is important that this is also taken into account when designing virtual and physical sequencing panels.

IRDs as a disease group were traditionally viewed as highly Mendelian. This is somewhat true, as for many IRD cases primary disease causation (without taking disease modifiers into account) can be attributed to singular pathogenic variants. Also as detailed above, there are certainly Mendelian sources of disease that are being missed in those not provided with a genetic diagnosis by the current techniques employed. However, the field is moving towards a more complex view involving poly/omni-genicity as well as environmental effects on disease outcome. There is a strong likelihood that some of the individuals who remain genetically undiagnosed have a much more complex disease aetiology than was previously thought.

Using macular diseases as an example, they can be broadly defined into two categories. That is, inherited macular diseases (IMD) and AMD. IMDs present at earlier ages and are relatively rare, while AMD is more common and mostly affects those in later

decades of life. Given that both diseases affect the macula, they have a number of overlapping clinical features. Unfortunately, a large proportion of individuals who are clinically diagnosed with macular disease remain genetically unresolved (for example the CRD patients in section 4 of this thesis). This may suggest that the underlying genetic architecture of macular disease as a whole is more complex than we think. A preprint by Hitti-Malin *et al*., in medrxiv sequenced the coding regions of 105 genes and non-coding or regulatory loci associated with both IMD and AMD in 379 probands[419]. Findings include the discovery of a homozygous variant in *CNGB3* and a heterozygous variant in *CNGA3* in the same proband. The authors indicate that this heterozygous variant may also influence the phenotype, as has been reported previously[420]. This example highlights the possibility that recessive disease may not be fully explained by identifying bi-allelic variants in associated genes, but rather that modifier variants or multiple variants in genes converging on the same mechanism may be causing the overall phenotype. Studies such as these will continue to explore macular disease in greater numbers. It may be revealed that some AMD patients have very late-onset forms of what are now known as IMD, or some late-onset suspected IMDs are in fact more similar to a complex AMD phenotype caused by genetic makeup across multuple loci as well as environmental factors. There is likely more cross talk between the molecular pathways implicated in IMD and AMD than we currently understand.

Continuing, disease modifier variants undoubtedly play a role in IRDs and the ultimate outcome of the associated phenotype. Prime examples of this are cis modifiers in *ABCA4.* A study by Lee and Zernant *et al.,* (2021) illustrated that c.789-784C>T is a major *cis*-acting modifier of c.5882G>A p.(Gly1961Glu), the most common pathogenic *ABCA4* variant[421]. This study was undertaken by sequencing the entire *ABCA4* locus in a large cohort of bi-allelic patients with *ABCA4*-related disease. Subsequently, 140 individuals carrying p.(Gly1961Glu) compound heterozygously and 10 homozygous individuals were genetically and clinically analysed. When c.789-784C>T and p.(Gly1961Glu) are present in a complex allele versus p.(Gly1961Glu) alone, there is increased clinical severity of disease. In addition, penetrance of disease in p.(Gly1961Glu) homozygotes is also increased by the presence of c.789-784C>T. c.789-784C>T was not detected homozygously in isolation in affected individuals. In terms of trans modifiers, *ABCA4* disease again serves as a good example. Zernant *et al*., described the potential role of *ROM1* and *PRPH2* in the modification of *ABCA4*-related disease[422]. They identified that the presense of *ROM1* and *PRPH2* variants was increased in *ABCA4*-patients versus controls. They concluded that the findings concerning *ROM1* needed to be studied in

larger cohorts. However, the presence of the common *PRPH2* p.Asp388 variant was significantly increased in patients who presented with late-onset STGD1 associated with the *ABCA4* p.Asn1868Ile variant, therefore increasing the penetrance of this *ABCA4* variant. The authors suggest the mechanism for this is possibly variable defects in photoreceptor outer segment membrane structure resulting from dysfunctional *PRPH2.* Therefore they suggest that these variants in *ROM1/PRPH2* do not modify disease significantly on a clinical level, but rather contribute to the overall penetrance of disease through some sort of disease threshold mechanism. Other potential modifier variants have been reported in *PDZD7/USH2A*[423]*, AHI1/CEP290*[424] *and CFH/PRDM13*[425]*, CNOT3/PRPF31*[426] among others, with variant in the former genes acting as modifiers. True identification of modifier alleles will require cross-collaboration and large cohorts of IRD patients and should be a central focus of the community and should be given consideration in the design and application of therapeutic endeavours.

These examples make it clear that current sequencing techniques are capable of detecting these more elusive variants at the sequence level. However, improved awareness of these mutation types is crucial to their detection, as well as accessible and bespoke *in silico* analysis tools and bioinformatic pipelines. Additionally, potential modifier variants highlight the need for increased awareness and investigation of the true genetic complexity of IRDs. The future of genetic diagnoses for IRDs should focus on continuing efforts to accurately detect and describe these variants and disease mechanisms. This will provide patients who are currently diagnosed with more comprehensive genotyping and those who are undiagnosed with an underlying explanation for their disease. The future of genetic testing for IRDs is bright with exciting posibilities for further discovery of variant types we are aware of and even variant types that we do not have knowledge of – yet.

7.8 Beyond genomics

Widescale availability of DNA sequencing has facilitated the generation of an astronomical volume of genomic data globally. While interrogation of these data to the fullest in order to identify potential disease causing variants is extremely important, exploring the functional consequences of these variants is imperative, as demonstrated by the functional analysis of multiple splice-altering variants in this doctoral thesis. IRDs, given their underlying genetic heterogeneity, are prime candidates for this scientific endeavour. This section will discuss briefly a number of disease mechanisms associated with IRDs and provide examples of viable strategies to explore as of yet undefined pathological pathways.

As mentioned previously, the inheritance pattern of several IRDs can be modified dependent upon the type and location of the variant in a given gene. One such example is pathogenic variation in the *GUCY2D* gene. Most frequently disease causing variants (of all types) in this gene result in autosomal recessive LCA. However, some in-frame aberrations and missense variants result in autosomal dominant CRD. Sharon *et al.,* describe that the recessive form of disease is primarily caused by absence or near-absence of functional guanylate cyclase via a number of mechanisms such as an inability to synthesize sufficient cGMP or protein misfolding and subsequent degradation. In contrast, the dominant form of disease is usually as a result of a shifting Ca2+-sensitivity curve[427].

It is also possible that some variants may have dual disease mechanisms. For example, a study by Abrahams *et al.,* described how some nonsense variants can alter the splicing process, a process known as nonsense-associated alternative splicing[428]. Nonsense variants may disrupt splicing in the same manner that other variants do, i.e., they may disrupt existing essential spice motifs. This study demonstrated that exonic splice enhancer motifs are susceptible to nonsense mutations due to the nature of their purine rich sequences. They employed machine learning tools to predict splice disruption as a result of these variants and by comparison with other pathogenic variants, nonsense variants appear to be enriched in exonic splice enhancers. These splice altering nonsense variants appear to be associated with exon skipping defects. This skipping may or may not be protective if the associated exon is a multiple of 3 bases long (i.e. if the skipping eliminates the nonsense mutation and a small encoded protein is well tolerated). However if the exon is out of frame then the downstream effect is akin to a frameshift variant. Discovery of this additional mechanism of disease is important. Given that splicing is not binary, some aberrant transcripts may contain the nonsense variants while others will contain the splice defect. Given that stop variants (with the exception of those at the extreme 3' end) are automatically assigned a pathogenic status under ACMG-AMP guidelines, additional disease mechanisms such as these may be overlooked.

Dominant acting variants are particularly interesting with regard to underlying disease mechanisms. They may cause disease through haploinsufficiency, a dominant-negative effect or gain-of-function. Schneider *et al.,* illustrated that IRD genes associated with autosomal dominant inheritance such as *RHO* appear to have a high ratio of pathogenic missense variants indicating a propensity toward dominant-negative/gain-of-function disease mechanisms. Other IRD genes associated with autosomal dominant inheritance, such as *PRPF31*, have a lower ratio of pathogenic missense variation and studies have demonstrated a haploinsufficiency-based disease mechanism. These are

extremely important considerations for therapeutic intervention. If the underlying mechanism is haploinsufficiency gene augmentation therapy may be beneficial. However, a dominant-negative/gain of function defect may require suppression and replacement of the mutant allele to achieve benefit.

Methods by which we can investigate these disease mechanisms are numerous. Variants that are likely to alter enzymatic function, such as those in *RPE65*[429], may be applicable to assessment *in vitro* by employing enzyme function assays[429–431]. Alternatively, if predicted to perturb splicing having undergone *in silico* interrogation[432], these variants could be assessed by midigene analysis, stem cell or organoid technology as demonstrated and discussed throughout this thesis.

A viable strategy for the assessment of novel variants is to introduce them into model organisms with similar or conserved genes. One significant advantage of such model organisms is the ability to rapidly generate models of disease[433]. This is particularly relevant currently, as recent advances in genome editing technologies has made this approach more readily accessible and successful than ever before[434]. Although not all model organisms fully represent the visual system that exists in humans, many organisms can offer useful information of the molecular aetiology of disease ranging from protein mislocalisation[435] to optokinetic response[436]. By introducing a VUS into the appropriate system, additional evidence can be gained to better evaluate the functional impact of the variant in human ocular pathologies and test potential therapeutic approaches. Similarly as outlined above and elsewhere in the thesis, patient derived induced pluripotent stem cell and organoid models represent powerful model systems in which to interrogate disease mechanisms and to assess candidate therapies.

Investigation of disease mechanisms via single cell sequencing is a promising methodology. Multiple studies have used single cell RNA sequencing to investigate perturbation of cellular homeostasis in specific cell types. For example, Voigt *et al.,* employed single cell transcriptomics on human RPE and choroid and identified that *RGCC,* a gene that responds to complement activation and induces endothelial apoptosis was specifically expressed in choriocapillaris endothelial cells[437]. They also found that this gene was most upregulated in a donor diagnosed with AMD. Another study by Karademir *et al.,* utilised single cell RNA sequencing in the retina of the *rd10* RP mouse model[438]. They found that the *Egr1* gene is upregulated in two consecutive phases of rod degeneration at P21. Metabolic changes, neuroprotective changes and modifications to synapses and microtubules were identified in degenerating rod cells. *Egr1* was also strongly upregulated in cone cells which demonstrated early respiratory dysfunction and

changes in signalling pathways. These findings illustrate that *Erg1* may be a potential target for broad treatment. The identification of overarching mechanisms of disease such as these is vitally important in order t o develop therapeutic interventions that can be applied to many IRD types given the significant challenge of gene and allelic heterogeneity in this group of conditions. Furthermore, single cell sequencing will enable exploration at a cell level the potential heterogeneity in perturbed cellular homeostasis that may characterise many disorders, with significant variations in cellular profiles (transcriptomics, proteomics etc) between cells of the same cell type in a single patient. Within in any disease cell population or tissue there may be great variance in the level of perturbed cellular pathway and dysfunction present at a single cell level, and therefore, in principle, there may be subpopulations of cells in a patient whose function may be easier to rescue than others. Investigations at a single cell level of resolution will enable us to more fully elucidate disease mechanisms.

Novel methods of gene prioritisation are being developed for IRDs given the distinct phenotypic features associated with these diseases. Experts learn to recognise these features over time using various modern retinal imaging technologies. This deep phenotyping allows for the prediction of which gene or which subset of genes the disease driving variation lies. Eye2Gene is a deep-learning model composed of neural networks that predicts the perturbed IRD gene based on retinal imaging. With a top-5 accuracy of 85.6%, which is comparable to retinal disease experts who took part in a study validating the tool[439]. Unsupervised machine learning of datasets such as the one used in this study may be a viable strategy to identify as of yet uncovered IRD-associated genes. While there have been enormous strides particularly over the past two decades towards elucidating the genetic causes of disease as outlined in this PhD thesis for IRDs, it is clear that identification of the enormous pathogenic variation underlying disease is only still beginning to unfold. True understanding of the mechanisms by which these variants drive disease in the retina is needed to develop effective and efficient therapeutics that broadly address the complexity of IRDs. In the next era of genomic medicine, increased focus and resolution of these processes is paramount.

8. Conclusion

The results of Target 5000 to date have highlighted the huge value of genetic screening for those affected by an IRD. Target 5000 has reached important milestones in determining the underlying genetic architecture of IRDs in Ireland, with many patients receiving actionable results from their participation in the study. These findings highlight the value of a designated first tier genetic testing service for IRD patients with the Irish health care system, and hopefully this is supported by the publication of a national genetic and genomic strategy later in 2022. Target 5000 continues to grow and progress, implementing new sequencing procedures for cases that remain unresolved. Pilot studies for whole genome and whole gene sequencing carried out as part of this doctoral thesis attest to the need for more comprehensive sequencing measures on a case dependent basis. The Target 5000 studies illustrated in this PhD thesis also highlight the continued need for an interface between the clinic and the research team given the identification of many novel coding and non-coding variants which require significant additional studies to elucidate whether they may be pathogenic or benign. Furthermore, the vital utility of this interface for IRDs is further emphasised by the overall candidate variant detection rate of 70% in the current study, which while in line with the higher rates globally, highlights that 30% of IRDs remain elusive and will require additional focus beyond standard clinical testing from research teams if they are to be resolved. Now that we have the power to sequence the whole genome of an individual, resources must be invested in interpreting this genomic information in the most optimal way for patients, including open source data sharing and functional analysis of variants of unknown significance. Additionally, significant research is needed into variants that have traditionally remained elusive to date. The Target 5000 team aims to do everything it can to provide patients with an accurate genetic diagnosis through a standardised clinical care pathway and research study arm. Studies have highlighted that patients benefit greatly from a genetic diagnosis in a number of ways, including a feeling of empowerment, ability to make more informed life decisions, a greater understanding of their disease prognosis as well as facilitating an optimistic outlook on future potential treatment. These diagnoses have never been more important in what is quickly becoming the age of gene-based medicine with multiple clinical trials for IRDs underway. In the hopefully not-so-distant future, an array of comprehensive and bespoke genetic tests and precise treatment options for IRD patients can become a reality

References

1. Purves, D. *et al.* Anatomical Distribution of Rods and Cones. *Neurosci. 2nd Ed.* (2001).
2. Kimbrel, E. A. & Lanza, R. Current status of pluripotent stem cells: moving the first therapies to the clinic. *Nat. Rev. Drug Discov.* **14**, 681–692 (2015).
3. Zele, A. J. & Cao, D. Vision under mesopic and scotopic illumination. *Front. Psychol.* **5**, (2015).
4. Hartveit, E. & Veruki, M. L. Electrical synapses between AII amacrine cells in the retina: Function and modulation. *Brain Res.* **1487**, 160–172 (2012).
5. Euler, T., Haverkamp, S., Schubert, T. & Baden, T. Retinal bipolar cells: elementary building blocks of vision. *Nat. Rev. Neurosci.* **15**, 507–519 (2014).
6. Tanaka, M. & Tachibana, M. Independent control of reciprocal and lateral inhibition at the axon terminal of retinal bipolar cells. *J. Physiol.* **591**, 3833–3851 (2013).
7. Masland, R. H. The neuronal organization of the retina. *Neuron* **76**, 266–280 (2012).
8. Tsin, A., Betts-Obregon, B. & Grigsby, J. Visual cycle proteins: Structure, function, and roles in human retinal disease. *J. Biol. Chem.* **293**, 13016–13021 (2018).
9. Schneider, N. *et al.* Inherited retinal diseases: Linking genes, disease-causing variants, and relevant therapeutic modalities. *Prog. Retin. Eye Res.* **89**, 101029 (2022).
10. Retinitis pigmentosa: for professionals. *Gene Vision* https://gene.vision/knowledge-base/retinitis-pigmentosa-for-doctors/.
11. Cone/Cone-rod dystrophy: for professionals. *Gene Vision* https://gene.vision/knowledge-base/cone-cone-rod-dystrophy-for-professionals/.
12. Werdich, X. Q., Place, E. M. & Pierce, E. A. Systemic Diseases Associated with Retinal Dystrophies. *Semin. Ophthalmol.* **29**, 319–328 (2014).
13. Whelan, L. *et al.* Findings from a Genotyping Study of over 1000 People with Inherited Retinal Disorders in Ireland. *Genes* **11**, 105 (2020).
14. Galvin, O. *et al.* The Impact of Inherited Retinal Diseases in the Republic of Ireland (ROI) and the United Kingdom (UK) from a Cost-of-Illness Perspective. *Clin. Ophthalmol. Auckl. NZ* **14**, 707–719 (2020).
15. RetNet: Summaries. https://sph.uth.edu/retnet/sum-dis.htm.
16. Retinal disorders (Version 4.10). https://panelapp.genomicsengland.co.uk/panels/307/.
17. Mizobuchi, K. *et al.* Genotype-Phenotype Correlations in RP1-Associated Retinal Dystrophies: A Multi-Center Cohort Study in JAPAN. *J. Clin. Med.* **10**, 2265 (2021).
18. Milenkovic, A., Milenkovic, V. M., Wetzel, C. H. & Weber, B. H. F. BEST1 protein stability and degradation pathways differ between autosomal dominant Best disease and autosomal recessive bestrophinopathy accounting for the distinct retinal phenotypes. *Hum. Mol. Genet.* **27**, 1630–1641 (2018).
19. Farrar, G. J. *et al.* Toward an elucidation of the molecular genetics of inherited retinal degenerations. *Hum. Mol. Genet.* **26**, R2–R11 (2017).

20. Cremers, F. P. *et al.* Autosomal recessive retinitis pigmentosa and cone-rod dystrophy caused by splice site mutations in the Stargardt's disease gene ABCR. *Hum. Mol. Genet.* **7**, 355–362 (1998).
21. Maugeri, A. *et al.* Mutations in the ABCA4 (ABCR) Gene Are the Major Cause of Autosomal Recessive Cone-Rod Dystrophy. *Am. J. Hum. Genet.* **67**, 960–966 (2000).
22. Maugeri, A. *et al.* The 2588G-->C mutation in the ABCR gene is a mild frequent founder mutation in the Western European population and allows the classification of ABCR mutations in patients with Stargardt disease. *Am. J. Hum. Genet.* **64**, 1024–1035 (1999).
23. Khan, M. *et al.* Resolving the dark matter of ABCA4 for 1,054 Stargardt disease probands through integrated genomics and transcriptomics. *bioRxiv* 817767 (2019) doi:10.1101/817767.
24. Becker, T. *et al.* Genetic information from discordant sibling pairs points to ESRP2 as a candidate trans-acting regulator of the CF modifier gene SCNN1B. *Sci. Rep.* **10**, 22447 (2020).
25. Bandres-Ciga, S. *et al.* Large-scale pathway specific polygenic risk and transcriptomic community network analysis identifies novel functional pathways in Parkinson disease. *Acta Neuropathol. (Berl.)* **140**, 341–358 (2020).
26. Sanger, F. Sequences, sequences, and sequences. *Annu. Rev. Biochem.* **57**, 1–28 (1988).
27. Wu, R. & Kaiser, A. D. Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. *J. Mol. Biol.* **35**, 523–537 (1968).
28. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467 (1977).
29. Maxam, A. M. & Gilbert, W. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 560–564 (1977).
30. Staden, R. A strategy of DNA sequencing employing computer programs. *Nucleic Acids Res.* **6**, 2601–2610 (1979).
31. Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F. & Petersen, G. B. Nucleotide sequence of bacteriophage lambda DNA. *J. Mol. Biol.* **162**, 729–773 (1982).
32. Smith, L. M. *et al.* Fluorescence detection in automated DNA sequence analysis. *Nature* **321**, 674–679 (1986).
33. Prober, J. M. *et al.* A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* **238**, 336–341 (1987).
34. DeAngelis, M. M., Wang, D. G. & Hawkins, T. L. Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Res.* **23**, 4742–4743 (1995).
35. Zhang, J. *et al.* Use of non-cross-linked polyacrylamide for four-color DNA sequencing by capillary electrophoresis separation of fragments up to 640 bases in length in two hours. *Anal. Chem.* **67**, 4589–4593 (1995).
36. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
37. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
38. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).

39. Ruparel, H. *et al.* Design and synthesis of a 3′-O-allyl photocleavable fluorescent nucleotide as a reversible terminator for DNA sequencing by synthesis. *Proc. Natl. Acad. Sci.* **102**, 5932–5937 (2005).
40. Seo, T. S. *et al.* Four-color DNA sequencing by synthesis on a chip using photocleavable fluorescent nucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 5926–5931 (2005).
41. Barnes, C., Balasubramanian, S., Liu, X., Swerdlow, H. & Milton, J. Labelled nucleotides. (2006).
42. Shendure, J. *et al.* Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–1732 (2005).
43. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
44. DNA Sequencing Costs: Data. *Genome.gov* https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data.
45. Levene, M. J. *et al.* Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299**, 682–686 (2003).
46. Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138 (2009).
47. Akeson, M., Branton, D., Church, G. & Deamer, D. W. Characterization of individual polymer molecules based on monomer-interface interactions. (2012).
48. Lionel, A. C. *et al.* Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genet. Med.* **20**, 435–443 (2018).
49. Perea-Romero, I. *et al.* Genetic landscape of 6089 inherited retinal dystrophies affected cases in Spain and their therapeutic and extended epidemiological implications. *Sci. Rep.* **11**, 1526 (2021).
50. Sharon, D. *et al.* A nationwide genetic analysis of inherited retinal diseases in Israel as assessed by the Israeli inherited retinal disease consortium (IIRDC). *Hum. Mutat.* **41**, 140–149 (2020).
51. Fadaie, Z. *et al.* Whole genome sequencing and in vitro splice assays reveal genetic causes for inherited retinal diseases. *Npj Genomic Med.* **6**, 1–11 (2021).
52. Carss, K. J. *et al.* Comprehensive Rare Variant Analysis via Whole-Genome Sequencing to Determine the Molecular Pathology of Inherited Retinal Disease. *Am. J. Hum. Genet.* **100**, 75–90 (2017).
53. BioRender. https://biorender.com/.
54. Khan, M. *et al.* Resolving the dark matter of ABCA4 for 1054 Stargardt disease probands through integrated genomics and transcriptomics. *Genet. Med. Off. J. Am. Coll. Med. Genet.* (2020) doi:10.1038/s41436-020-0787-4.
55. Hiatt, J. B., Pritchard, C. C., Salipante, S. J., O'Roak, B. J. & Shendure, J. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. *Genome Res.* **23**, 843–854 (2013).

56. Reurink, J. *et al.* Molecular Inversion Probe-Based Sequencing of USH2A Exons and Splice Sites as a Cost-Effective Screening Tool in USH2 and arRP Cases. *Int. J. Mol. Sci.* **22**, 6419 (2021).
57. Consugar, M. B. *et al.* Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing. *Genet. Med.* **17**, 253–261 (2015).
58. Patel, A. *et al.* The Oculome Panel Test: Next-Generation Sequencing to Diagnose a Diverse Range of Genetic Developmental Eye Disorders. *Ophthalmology* **126**, 888–907 (2019).
59. Stone, E. M. *et al.* Clinically Focused Molecular Investigation of 1000 Consecutive Families with Inherited Retinal Disease. *Ophthalmology* **124**, 1314–1331 (2017).
60. Riera, M. *et al.* Whole exome sequencing using Ion Proton system enables reliable genetic diagnosis of inherited retinal dystrophies. *Sci. Rep.* **7**, (2017).
61. Sharon, D. *et al.* A nationwide genetic analysis of inherited retinal diseases in Israel as assessed by the Israeli inherited retinal disease consortium (IIRDC). *Hum. Mutat.* **n/a**,.
62. Haer-Wigman, L. *et al.* Diagnostic exome sequencing in 266 Dutch patients with visual impairment. *Eur. J. Hum. Genet. EJHG* **25**, 591–599 (2017).
63. Ellingford, J. M. *et al.* Whole Genome Sequencing Increases Molecular Diagnostic Yield Compared with Current Diagnostic Testing for Inherited Retinal Disease. *Ophthalmology* **123**, 1143–1150 (2016).
64. Radziwon, A. *et al.* Single-base substitutions in the CHM promoter as a cause of choroideremia. *Hum. Mutat.* **38**, 704–715 (2017).
65. Bauwens, M. *et al.* ABCA4-associated disease as a model for missing heritability in autosomal recessive disorders: novel noncoding splice, cis-regulatory, structural, and recurrent hypomorphic variants. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **21**, 1761–1771 (2019).
66. Research, C. for B. E. and. LUXTURNA. *FDA* (2019).
67. Thompson, J. A. *et al.* The genetic profile of Leber congenital amaurosis in an Australian cohort. *Mol. Genet. Genomic Med.* **5**, 652–667 (2017).
68. Motta, F. L., Martin, R. P., Filippelli-Silva, R., Salles, M. V. & Sallum, J. M. F. Relative frequency of inherited retinal dystrophies in Brazil. *Sci. Rep.* **8**, (2018).
69. Liu, X., Tao, T., Zhao, L., Li, G. & Yang, L. Molecular diagnosis based on comprehensive genetic testing in 800 Chinese families with non-syndromic inherited retinal dystrophies. *Clin. Experiment. Ophthalmol.* (2020) doi:10.1111/ceo.13875.
70. Gao, F.-J. *et al.* Genetic and Clinical Findings in a Large Cohort of Chinese Patients with Suspected Retinitis Pigmentosa. *Ophthalmology* **126**, 1549–1556 (2019).
71. Liu, X. Z., Li, Y. Y. & Yang, L. P. [Comparison study of whole exome sequencing and targeted panel sequencing in molecular diagnosis of inherited retinal dystrophies]. *Beijing Da Xue Xue Bao* **52**, 836–844 (2020).
72. Huang, L. *et al.* Mutation screening in genes known to be responsible for Retinitis Pigmentosa in 98 Small Han Chinese Families. *Sci. Rep.* **7**, 1948 (2017).

73. Dan, H., Huang, X., Xing, Y. & Shen, Y. Application of targeted panel sequencing and whole exome sequencing for 76 Chinese families with retinitis pigmentosa. *Mol. Genet. Genomic Med.* **8**, e1131 (2020).
74. Wang, L. *et al.* Application of Whole Exome and Targeted Panel Sequencing in the Clinical Molecular Diagnosis of 319 Chinese Families with Inherited Retinal Dystrophy and Comparison Study. *Genes* **9**, (2018).
75. Avela, K. *et al.* The genetic aetiology of retinal degeneration in children in Finland - new founder mutations identified. *Acta Ophthalmol. (Copenh.)* **97**, 805–814 (2019).
76. Weisschuh, N. *et al.* Genetic architecture of inherited retinal degeneration in Germany: A large cohort study from a single diagnostic center over a 9-year period. *Hum. Mutat.* **41**, 1514–1527 (2020).
77. Birtel, J. *et al.* Clinical and genetic characteristics of 251 consecutive patients with macular and cone/cone-rod dystrophy. *Sci. Rep.* **8**, 4824 (2018).
78. Tayebi, N. *et al.* Targeted next generation sequencing reveals genetic defects underlying inherited retinal disease in Iranian families. *Mol. Vis.* **25**, 106–117 (2019).
79. Koyanagi, Y. *et al.* Genetic characteristics of retinitis pigmentosa in 1204 Japanese patients. *J. Med. Genet.* **56**, 662–670 (2019).
80. Numa, S. *et al.* EYS is a major gene involved in retinitis pigmentosa in Japan: genetic landscapes revealed by stepwise genetic screening. *Sci. Rep.* **10**, (2020).
81. Surl, D. *et al.* Copy number variations and multiallelic variants in Korean patients with Leber congenital amaurosis. *Mol. Vis.* **26**, 26–35 (2020).
82. Kim, M. S. *et al.* Genetic Mutation Profiles in Korean Patients with Inherited Retinal Diseases. *J. Korean Med. Sci.* **34**, e161 (2019).
83. Zenteno, J. C. *et al.* Extensive genic and allelic heterogeneity underlying inherited retinal dystrophies in Mexican patients molecularly analyzed by next-generation sequencing. *Mol. Genet. Genomic Med.* **8**, (2020).
84. Holtan, J. P., Selmer, K. K., Heimdal, K. R. & Bragadóttir, R. Inherited retinal disease in Norway - a characterization of current clinical and genetic knowledge. *Acta Ophthalmol. (Copenh.)* **98**, 286–295 (2020).
85. Wawrocka, A. *et al.* Novel variants identified with next-generation sequencing in Polish patients with cone-rod dystrophy. *Mol. Vis.* **24**, 326–339 (2018).
86. Vincent, A. L. *et al.* Next-generation sequencing targeted disease panel in rod-cone retinal dystrophies in Māori and Polynesian reveals novel changes and a common founder mutation. *Clin. Experiment. Ophthalmol.* **45**, 901–910 (2017).
87. Martin-Merida, I. *et al.* Genomic Landscape of Sporadic Retinitis Pigmentosa: Findings from 877 Spanish Cases. *Ophthalmology* **126**, 1181–1188 (2019).
88. Gonzàlez-Duarte, R. *et al.* Scaling New Heights in the Genetic Diagnosis of Inherited Retinal Dystrophies. *Adv. Exp. Med. Biol.* **1185**, 215–219 (2019).

89. Diñeiro, M. *et al.* Comprehensive genomic diagnosis of inherited retinal and optical nerve disorders reveals hidden syndromes and personalized therapeutic options. *Acta Ophthalmol. (Copenh.)* **98**, e1034–e1048 (2020).
90. Chen, Z.-J. *et al.* Mutation spectrum and genotype-phenotype correlation of inherited retinal dystrophy in Taiwan. *Clin. Experiment. Ophthalmol.* **48**, 486–499 (2020).
91. Habibi, I. *et al.* Genetic spectrum of retinal dystrophies in Tunisia. *Sci. Rep.* **10**, (2020).
92. Khan, A. O. PHENOTYPE-GUIDED GENETIC TESTING OF PEDIATRIC INHERITED RETINAL DISEASE IN THE UNITED ARAB EMIRATES. *Retina Phila. Pa* **40**, 1829–1837 (2020).
93. Patel, N. *et al.* Mutations in known disease genes account for the majority of autosomal recessive retinal dystrophies. *Clin. Genet.* **94**, 554–563 (2018).
94. Jiman, O. A. *et al.* Diagnostic yield of panel-based genetic testing in syndromic inherited retinal disease. *Eur. J. Hum. Genet. EJHG* **28**, 576–586 (2020).
95. Shah, M. *et al.* Next generation sequencing using phenotype-based panels for genetic testing in inherited retinal diseases. *Ophthalmic Genet.* **41**, 331–337 (2020).
96. Lenassi, E. *et al.* Clinical utility of genetic testing in 201 preschool children with inherited eye disorders. *Genet. Med.* **22**, 745–751 (2020).
97. Taylor, R. L. *et al.* Panel-Based Clinical Genetic Testing in 85 Children with Inherited Retinal Disease. *Ophthalmology* **124**, 985–991 (2017).
98. Goetz, K. E. *et al.* Genetic testing for inherited eye conditions in over 6,000 individuals through the eyeGENE network. *Am. J. Med. Genet. C Semin. Med. Genet.* **184**, 828–837 (2020).
99. Bryant, L., Lozynska, O., Maguire, A. M., Aleman, T. S. & Bennett, J. Prescreening whole exome sequencing results from patients with retinal degeneration for variants in genes associated with retinal degeneration. *Clin. Ophthalmol. Auckl. NZ* **12**, 49–63 (2018).
100. Pontikos, N. *et al.* Genetic Basis of Inherited Retinal Disease in a Molecularly Characterized Cohort of More Than 3000 Families from the United Kingdom. *Ophthalmology* **127**, 1384–1394 (2020).
101. Méjécase, C., Kozak, I. & Moosajee, M. The genetic landscape of inherited eye disorders in 74 consecutive families from the United Arab Emirates. *Am. J. Med. Genet. C Semin. Med. Genet.* **184**, 762–772 (2020).
102. Peter, V. G. *et al.* The first genetic landscape of inherited retinal dystrophies in Portuguese patients identifies recurrent homozygous mutations as a frequent cause of pathogenesis. *PNAS Nexus* **2**, pgad043 (2023).
103. Jurica, M. S. & Roybal, G. A. RNA Splicing. in *Encyclopedia of Biological Chemistry (Second Edition)* (eds. Lennarz, W. J. & Lane, M. D.) 185–190 (Academic Press, 2013). doi:10.1016/B978-0-12-378630-2.00674-5.
104. Aebi, M., Hornig, H., Padgett, R. A., Reiser, J. & Weissmann, C. Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. *Cell* **47**, 555–565 (1986).
105. Jurica, M. S. & Moore, M. J. Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* **12**, 5–14 (2003).

106. Sheth, N. *et al.* Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res.* **34**, 3955–3967 (2006).
107. Kango-Singh, M. Vogel and Motulsky's human genetics--problems and approaches. *Hum. Genomics* **5**, 73–74 (2010).
108. Aukrust, I. *et al.* The intronic ABCA4 c.5461-10T>C variant, frequently seen in patients with Stargardt disease, causes splice defects and reduced ABCA4 protein level. *Acta Ophthalmol. (Copenh.)* **95**, 240–246 (2017).
109. Fadaie, Z. *et al.* BBS1 branchpoint variant is associated with non-syndromic retinitis pigmentosa. *J. Med. Genet.* (2021) doi:10.1136/jmedgenet-2020-107626.
110. Riazuddin, S. A. *et al.* A Splice-Site Mutation in a Retina-Specific Exon of BBS8 Causes Nonsyndromic Retinitis Pigmentosa. *Am. J. Hum. Genet.* **86**, 805–812 (2010).
111. Weisschuh, N., Buena-Atienza, E. & Wissinger, B. Splicing mutations in inherited retinal diseases. *Prog. Retin. Eye Res.* **80**, 100874 (2021).
112. Albert, S. *et al.* Identification and Rescue of Splice Defects Caused by Two Neighboring Deep-Intronic ABCA4 Mutations Underlying Stargardt Disease. *Am. J. Hum. Genet.* **102**, 517–527 (2018).
113. Khan, A. O. *et al.* A deep intronic CLRN1 (USH3A) founder mutation generates an aberrant exon and underlies severe Usher syndrome on the Arabian Peninsula. *Sci. Rep.* **7**, 1411 (2017).
114. Li, Y. *et al.* Aberrant RNA splicing is the major pathogenic effect in a knock-in mouse model of the dominantly inherited c.1430A>G human RPE65 mutation. *Hum. Mutat.* **40**, 426–443 (2019).
115. Liquori, A. *et al.* Whole USH2A Gene Sequencing Identifies Several New Deep Intronic Mutations. *Hum. Mutat.* **37**, 184–193 (2016).
116. Stephenson, K. A. J. *et al.* Target 5000: a standardized all-Ireland pathway for the diagnosis and management of inherited retinal degenerations. *Orphanet J. Rare Dis.* **16**, 200 (2021).
117. Carrigan, M. *et al.* Panel-Based Population Next-Generation Sequencing for Inherited Retinal Degenerations. *Sci. Rep.* **6**, 33248 (2016).
118. Dockery, A. *et al.* Target 5000: Target Capture Sequencing for Inherited Retinal Degenerations. *Genes* **8**, 304 (2017).
119. Hanany, M., Rivolta, C. & Sharon, D. Worldwide carrier frequency and genetic prevalence of autosomal recessive inherited retinal diseases. *Proc. Natl. Acad. Sci.* **117**, 2710–2716 (2020).
120. Dockery, A. *et al.* Target 5000: Target Capture Sequencing for Inherited Retinal Degenerations. *Genes* **8**, 304 (2017).
121. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
122. Picard Tools - By Broad Institute. http://broadinstitute.github.io/picard/ (2018).
123. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. *ArXiv12073907 Q-Bio* (2012).

124. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
125. Liu, X., Wu, C., Li, C. & Boerwinkle, E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum. Mutat.* **37**, 235–241 (2018).
126. Dong, C. *et al.* Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* **24**, 2125–2137 (2015).
127. Jagadeesh, K. A. *et al.* M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat. Genet.* **48**, 1581–1586 (2016).
128. Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* **15**, R84 (2014).
129. Chen, X. *et al.* Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).
130. Krumm, N. *et al.* Copy number variation detection and genotyping from exome sequence data. *Genome Res.* **22**, 1525–1532 (2012).
131. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **17**, 405–424 (2015).
132. Jarvik, G. P. & Browning, B. L. Consideration of Cosegregation in the Pathogenicity Classification of Genomic Variants. *Am. J. Hum. Genet.* **98**, 1077–1081 (2016).
133. Romanet, P. *et al.* Proposition of adjustments to the ACMG-AMP framework for the interpretation of MEN1 missense variants. *Hum. Mutat.* **40**, 661–674 (2019).
134. Ioannidis, N. M. *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* **99**, 877–885 (2016).
135. Jian, X. & Liu, X. In Silico Prediction of Deleteriousness for Nonsynonymous and Splice-Altering Single Nucleotide Variants in the Human Genome. in *In Vitro Mutagenesis* (ed. Reeves, A.) vol. 1498 191–197 (Springer New York, 2017).
136. Li, J. *et al.* Improved Diagnosis of Inherited Retinal Dystrophies by High-Fidelity PCR of ORF15 followed by Next-Generation Sequencing. *J. Mol. Diagn. JMD* **18**, 817–824 (2016).
137. Boyle, E. A., O'Roak, B. J., Martin, B. K., Kumar, A. & Shendure, J. MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. *Bioinformatics* **30**, 2670–2672 (2014).
138. Sullivan, L. S. *et al.* Prevalence of disease-causing mutations in families with autosomal dominant retinitis pigmentosa: a screen of known genes in 200 families. *Invest. Ophthalmol. Vis. Sci.* **47**, 3052–3064 (2006).

139. Lenassi, E. *et al.* A detailed clinical and molecular survey of subjects with nonsyndromic USH2A retinopathy reveals an allelic hierarchy of disease-causing variants. *Eur. J. Hum. Genet. EJHG* **23**, 1318–1327 (2015).
140. Klevering, B. J. *et al.* Microarray-based mutation analysis of the ABCA4 (ABCR) gene in autosomal recessive cone-rod dystrophy and retinitis pigmentosa. *Eur. J. Hum. Genet. EJHG* **12**, 1024–1032 (2004).
141. Allikmets, R. *et al.* A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy. *Nat. Genet.* **15**, 236–246 (1997).
142. Sangermano, R. *et al.* Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. *Genet. Med.* **21**, 1751–1760 (2019).
143. Khan, M. *et al.* Cost-effective molecular inversion probe-based ABCA4 sequencing reveals deep-intronic variants in Stargardt disease. *Hum. Mutat.* **40**, 1749–1759 (2019).
144. Gene Vision. *Gene Vision* https://gene.vision/.
145. gnomAD. https://gnomad.broadinstitute.org/.
146. RPE65 gene. *Gene Vision* https://gene.vision/knowledge-base/rpe65-for-doctors/.
147. Bowne, S. J. *et al.* A dominant mutation in *RPE65* identified by whole-exome sequencing causes retinitis pigmentosa with choroidal involvement. *Eur. J. Hum. Genet.* **19**, 1074–1081 (2011).
148. Hull, S., Mukherjee, R., Holder, G. E., Moore, A. T. & Webster, A. R. The clinical features of retinal disease due to a dominant mutation in RPE65. *Mol. Vis.* **22**, 626–635 (2016).
149. Li, S. *et al.* RPE65 mutation frequency and phenotypic variation according to exome sequencing in a tertiary centre for genetic eye diseases in China. *Acta Ophthalmol. (Copenh.)* (2019) doi:10.1111/aos.14181.
150. Jauregui, R., Park, K. S. & Tsang, S. H. Two-year progression analysis of RPE65 autosomal dominant retinitis pigmentosa. *Ophthalmic Genet.* **39**, 544–549 (2018).
151. Chiabrando, D. *et al.* Mutations in the Heme Exporter FLVCR1 Cause Sensory Neurodegeneration with Loss of Pain Perception. *PLoS Genet.* **12**, (2016).
152. Castori, M. *et al.* Posterior column ataxia with retinitis pigmentosa coexisting with sensory-autonomic neuropathy and leukemia due to the homozygous p.Pro221Ser FLVCR1 mutation. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **174**, 732–739.
153. Ishiura, H. *et al.* Posterior column ataxia with retinitis pigmentosa in a Japanese family with a novel mutation in FLVCR1. *neurogenetics* **12**, 117–121 (2011).
154. Glöckle, N. *et al.* Panel-based next generation sequencing as a reliable and efficient technique to detect mutations in unselected patients with retinal dystrophies. *Eur. J. Hum. Genet.* **22**, 99–104 (2014).
155. Yusuf, I. H., Shanks, M. E., Clouston, P. & MacLaren, R. E. A splice-site variant in FLVCR1 produces retinitis pigmentosa without posterior column ataxia. *Ophthalmic Genet.* **39**, 263–267 (2018).
156. Tiwari, A. *et al.* Next generation sequencing based identification of disease-associated mutations in Swiss patients with retinal dystrophies. *Sci. Rep.* **6**, 28755 (2016).

157. Dockery, A. *et al.* A Novel FLVCR1 Variant Implicated in Retinitis Pigmentosa. *Adv. Exp. Med. Biol.* **1185**, 203–207 (2019).
158. Allikmets, R. *et al.* Mutation of the Stargardt disease gene (ABCR) in age-related macular degeneration. *Science* **277**, 1805–1807 (1997).
159. Strauss, R. W. *et al.* The Natural History of the Progression of Atrophy Secondary to Stargardt Disease (ProgStar) Studies: Design and Baseline Characteristics: ProgStar Report No. 1. *Ophthalmology* **123**, 817–828 (2016).
160. Fishman, G. A., Farbman, J. S. & Alexander, K. R. Delayed rod dark adaptation in patients with Stargardt's disease. *Ophthalmology* **98**, 957–962 (1991).
161. Cornelis, S. S. *et al.* In Silico Functional Meta-Analysis of 5,962 ABCA4 Variants in 3,928 Retinal Dystrophy Cases. *Hum. Mutat.* **38**, 400–408 (2017).
162. Li, J., Edwards, P. C., Burghammer, M., Villa, C. & Schertler, G. F. X. Structure of bovine rhodopsin in a trigonal crystal form. *J. Mol. Biol.* **343**, 1409–1438 (2004).
163. Keel, S. B. *et al.* A heme export protein is required for red blood cell differentiation and iron homeostasis. *Science* **319**, 825–828 (2008).
164. Tailor, C. S., Willett, B. J. & Kabat, D. A putative cell surface receptor for anemia-inducing feline leukemia virus subgroup C is a member of a transporter superfamily. *J. Virol.* **73**, 6500–6505 (1999).
165. Quigley, J. G. *et al.* Identification of a human heme exporter that is essential for erythropoiesis. *Cell* **118**, 757–766 (2004).
166. Higgins, J. J., Morton, D. H., Patronas, N. & Nee, L. E. An autosomal recessive disorder with posterior column ataxia and retinitis pigmentosa. *Neurology* **49**, 1717–1720 (1997).
167. Marlhens, F. *et al.* Mutations in RPE65 cause leber's congenital amaurosis. *Nat. Genet.* **17**, 139–141 (1997).
168. Marlhens, F. *et al.* Autosomal recessive retinal dystrophy associated with two novel mutations in the RPE65 gene. *Eur. J. Hum. Genet.* **6**, 527–531 (1998).
169. Hamel, C. P., Griffoin, J.-M., Lasquellec, L., Bazalgette, C. & Arnaud, B. Retinal dystrophies caused by mutations in RPE65: Assessment of visual functions. *Br. J. Ophthalmol.* **85**, 424–427 (2001).
170. Maguire, A. M. *et al.* Efficacy, Safety, and Durability of Voretigene Neparvovec-rzyl in RPE65 Mutation-Associated Inherited Retinal Dystrophy: Results of Phase 1 and 3 Trials. *Ophthalmology* **126**, 1273–1285 (2019).
171. Choi, E. H. *et al.* Insights into the pathogenesis of dominant retinitis pigmentosa associated with a D477G mutation in RPE65. *Hum. Mol. Genet.* **27**, 2225–2243 (2018).
172. Li, S. *et al.* Rescue of enzymatic function for disease-associated RPE65 proteins containing various missense mutations in non-active sites. *J. Biol. Chem.* **289**, 18943–18956 (2014).
173. Kenna, P. F. *et al.* Advanced late-onset retinitis pigmentosa with dominant-acting D477G RPE65 mutation is responsive to oral synthetic retinoid therapy. *BMJ Open Ophthalmol.* **5**, e000462 (2020).

174. Collins, R. L. *et al.* A structural variation reference for medical and population genetics. *Nature* **581**, 444–451 (2020).
175. Migeon, B. R. X-linked diseases: susceptible females. *Genet. Med.* **22**, 1156–1174 (2020).
176. Simunovic, M. P. *et al.* The Spectrum of CHM Gene Mutations in Choroideremia and Their Relationship to Clinical Phenotype. *Invest. Ophthalmol. Vis. Sci.* **57**, 6033–6039 (2016).
177. Buermans, H. P. J. & den Dunnen, J. T. Next generation sequencing technology: Advances and applications. *Biochim. Biophys. Acta* **1842**, 1932–1941 (2014).
178. Lin, X. *et al.* Applications of targeted gene capture and next-generation sequencing technologies in studies of human deafness and other genetic disabilities. *Hear. Res.* **288**, 67–76 (2012).
179. Choi, M. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19096–19101 (2009).
180. Tucker, T., Marra, M. & Friedman, J. M. Massively parallel sequencing: the next big thing in genetic medicine. *Am. J. Hum. Genet.* **85**, 142–154 (2009).
181. Lewis, C. A. *et al.* Tubby-like protein 1 homozygous splice-site mutation causes early-onset severe retinal degeneration. *Invest. Ophthalmol. Vis. Sci.* **40**, 2106–2114 (1999).
182. Stojic, J., Stöhr, H. & Weber, B. H. F. Three novel ABCC5 splice variants in human retina and their role as regulators of ABCC5 gene expression. *BMC Mol. Biol.* **8**, 42 (2007).
183. Haer-Wigman, L. *et al.* Diagnostic exome sequencing in 266 Dutch patients with visual impairment. *Eur. J. Hum. Genet. EJHG* **25**, 591–599 (2017).
184. Belkadi, A. *et al.* Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc. Natl. Acad. Sci.* **112**, 5473–5478 (2015).
185. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **17**, 405–424 (2015).
186. Weisschuh, N. *et al.* Molecular and clinical analysis of 27 German patients with Leber congenital amaurosis. *PloS One* **13**, e0205380 (2018).
187. Krumm, N. *et al.* Copy number variation detection and genotyping from exome sequence data. *Genome Res.* **22**, 1525–1532 (2012).
188. Verbakel, S. K. *et al.* The identification of a RNA splice variant in TULP1 in two siblings with early-onset photoreceptor dystrophy. *Mol. Genet. Genomic Med.* **7**, (2019).
189. Okonechnikov, K., Conesa, A. & García-Alcalde, F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinforma. Oxf. Engl.* **32**, 292–294 (2016).
190. Farek, J. *et al.* xAtlas: Scalable small variant calling across heterogeneous next-generation sequencing experiments. *bioRxiv* 295071 (2018) doi:10.1101/295071.
191. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
192. Francioli, L. C. *et al.* Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat. Genet.* **46**, 818–825 (2014).
193. Erikson, G. A. *et al.* Whole-Genome Sequencing of a Healthy Aging Cohort. *Cell* **165**, 1002–1011 (2016).

194. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
195. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
196. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
197. Grantham, R. Amino acid difference formula to help explain protein evolution. *Science* **185**, 862–864 (1974).
198. Jaganathan, K. *et al.* Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **176**, 535-548.e24 (2019).
199. Boeva, V. *et al.* Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **28**, 423–425 (2012).
200. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
201. Firth, H. V. *et al.* DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am. J. Hum. Genet.* **84**, 524–533 (2009).
202. Dolzhenko, E. *et al.* Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome Res.* **27**, 1895–1903 (2017).
203. Allaire: RStudio: integrated development environment for R - Google Scholar. https://scholar.google.com/scholar_lookup?title=RStudio%3A%20integrated%20development%20environment%20for%20R&journal=RStudio&volume=770&publication_year=2012&author=Allaire%2CJ.
204. Sangermano, R. *et al.* ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome Res.* **28**, 100–110 (2018).
205. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
206. Wallis, Y., Payne, S., McAnulty, C., Bodmer, D. & Sister, E. Practice Guidelines for the Evaluation of Pathogenicity and the Reporting of Sequence Variants in Clinical Molecular Genetics. 16 (2013).
207. Gardiner, S. L. *et al.* Large normal-range TBP and ATXN7 CAG repeat lengths are associated with increased lifetime risk of depression. *Transl. Psychiatry* **7**, e1143 (2017).
208. Stingl, K. *et al.* CDHR1 mutations in retinal dystrophies. *Sci. Rep.* **7**, 6992 (2017).
209. Khan, M. *et al.* Detailed Phenotyping and Therapeutic Strategies for Intronic ABCA4 Variants in Stargardt Disease. *Mol. Ther. - Nucleic Acids* **21**, 412–427 (2020).
210. Braun, T. A. *et al.* Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. *Hum. Mol. Genet.* **22**, 5136–5145 (2013).
211. Fadaie, Z. *et al.* Identification of splice defects due to noncanonical splice site or deep-intronic variants in ABCA4. *Hum. Mutat.* (2019) doi:10.1002/humu.23890.
212. Weisschuh, N. *et al.* Deep-intronic variants in CNGB3 cause achromatopsia by pseudoexon activation. *Hum. Mutat.* **41**, 255–264 (2020).

213. Weisschuh, N., Buena-Atienza, E. & Wissinger, B. Splicing mutations in inherited retinal diseases. *Prog. Retin. Eye Res.* **80**, 100874 (2021).
214. Di Scipio, M. *et al.* Phenotype Driven Analysis of Whole Genome Sequencing Identifies Deep Intronic Variants that Cause Retinal Dystrophies by Aberrant Exonization. *Invest. Ophthalmol. Vis. Sci.* **61**, 36 (2020).
215. Di, Y. *et al.* Whole-exome Sequencing Analysis Identifies Mutations in the EYS Gene in Retinitis Pigmentosa in the Indian Population. *Sci. Rep.* **6**, 19432 (2016).
216. Fedele, A. O. & Hopwood, J. J. Functional analysis of the HGSNAT gene in patients with mucopolysaccharidosis IIIC (Sanfilippo C Syndrome). *Hum. Mutat.* **31**, E1574–E1586 (2010).
217. Schiff, E. R. *et al.* A genetic and clinical study of individuals with nonsyndromic retinopathy consequent upon sequence variants in HGSNAT, the gene associated with Sanfilippo C mucopolysaccharidosis. *Am. J. Med. Genet. C Semin. Med. Genet.* **184**, 631–643 (2020).
218. González-del Pozo, M. *et al.* Unmasking Retinitis Pigmentosa complex cases by a whole genome sequencing algorithm based on open-access tools: hidden recessive inheritance and potential oligogenic variants. *J. Transl. Med.* **18**, 73 (2020).
219. Bhatia, S., Goyal, S., Singh, I. R., Singh, D. & Vanita, V. A novel mutation in the PRPF31 in a North Indian adRP family with incomplete penetrance. *Doc. Ophthalmol. Adv. Ophthalmol.* **137**, 103–119 (2018).
220. Rose, A. M. & Bhattacharya, S. S. Variant haploinsufficiency and phenotypic non-penetrance in PRPF31-associated retinitis pigmentosa. *Clin. Genet.* **90**, 118–126 (2016).
221. Rose, A. M. *et al.* Expression of PRPF31 and TFPT: regulation in health and retinal disease. *Hum. Mol. Genet.* **21**, 4126–4137 (2012).
222. Brambillasca, F. *et al.* Promoter Analysis of TFPT (FB1), a Molecular Partner of TCF3 (E2A) in Childhood Acute Lymphoblastic Leukemia. *Biochem. Biophys. Res. Commun.* **288**, 1250–1257 (2001).
223. Rak, M. & Rustin, P. Supernumerary subunits NDUFA3, NDUFA5 and NDUFA12 are required for the formation of the extramembrane arm of human mitochondrial complex I. *FEBS Lett.* **588**, 1832–1838 (2014).
224. Rose, A. M., Mukhopadhyay, R., Webster, A. R., Bhattacharya, S. S. & Waseem, N. H. A 112 kb deletion in chromosome 19q13.42 leads to retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.* **52**, 6597–6603 (2011).
225. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care — Preliminary Report | NEJM. https://www.nejm.org/doi/full/10.1056/NEJMoa2035790.
226. Runhart, E. H. *et al.* Association of Sex With Frequent and Mild ABCA4 Alleles in Stargardt Disease. *JAMA Ophthalmol.* **138**, 1035–1042 (2020).
227. Maggi, J. *et al.* De Novo Assembly-Based Analysis of RPGR Exon ORF15 in an Indigenous African Cohort Overcomes Limitations of a Standard Next-Generation Sequencing (NGS) Data Analysis Pipeline. *Genes* **11**, 800 (2020).
228. Blacharski, P. A. & Newsome, D. A. Bilateral Macular Holes after Nd:YAG Laser Posterior Capsulotomy. *Am. J. Ophthalmol.* **105**, 417–418 (1988).

229. Bauwens, M. *et al.* An augmented ABCA4 screen targeting noncoding regions reveals a deep intronic founder variant in Belgian Stargardt patients. *Hum. Mutat.* **36**, 39–42 (2015).
230. Schulz, H. L. *et al.* Mutation Spectrum of the ABCA4 Gene in 335 Stargardt Disease Patients From a Multicenter German Cohort—Impact of Selected Deep Intronic Variants and Common SNPs. *Invest. Ophthalmol. Vis. Sci.* **58**, 394–403 (2017).
231. Zhang, K. *et al.* A novel mutation in the ABCR gene in four patients with autosomal recessive Stargardt disease. *Am. J. Ophthalmol.* **128**, 720–724 (1999).
232. Zernant, J. *et al.* Analysis of the ABCA4 genomic locus in Stargardt disease. *Hum. Mol. Genet.* **23**, 6797–6806 (2014).
233. Bax, N. M. *et al.* Heterozygous deep-intronic variants and deletions in ABCA4 in persons with retinal dystrophies and one exonic ABCA4 variant. *Hum. Mutat.* **36**, 43–47 (2015).
234. Maugeri, A. *et al.* The 2588G-->C mutation in the ABCR gene is a mild frequent founder mutation in the Western European population and allows the classification of ABCR mutations in patients with Stargardt disease. *Am. J. Hum. Genet.* **64**, 1024–1035 (1999).
235. Genome of the Netherlands. http://www.nlgenome.nl/.
236. Cremers, F. P. *et al.* Autosomal recessive retinitis pigmentosa and cone-rod dystrophy caused by splice site mutations in the Stargardt's disease gene ABCR. *Hum. Mol. Genet.* **7**, 355–362 (1998).
237. Zernant, J. *et al.* Frequent hypomorphic alleles account for a significant fraction of ABCA4 disease and distinguish it from age-related macular degeneration. *J. Med. Genet.* **54**, 404–412 (2017).
238. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
239. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
240. Schwarz, J. M., Rödelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7**, 575–576 (2010).
241. Home - LOVD - An Open Source DNA variation database system. https://www.lovd.nl/3.0/home.
242. Stenson, P. D. *et al.* The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum. Genet.* **136**, 665–677 (2017).
243. Sangermano, R. *et al.* Photoreceptor Progenitor mRNA Analysis Reveals Exon Skipping Resulting from the ABCA4 c.5461-10T→C Mutation in Stargardt Disease. *Ophthalmology* **123**, 1375–1385 (2016).
244. Sangermano, R. *et al.* ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome Res.* **28**, 100–110 (2018).
245. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).

246. Poplin, R. *et al.* Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* 201178 (2018) doi:10.1101/201178.

247. Rivera, A. *et al.* A comprehensive survey of sequence variation in the ABCA4 (ABCR) gene in Stargardt disease and age-related macular degeneration. *Am. J. Hum. Genet.* **67**, 800–813 (2000).

248. Collin, R. W. *et al.* Antisense Oligonucleotide (AON)-based Therapy for Leber Congenital Amaurosis Caused by a Frequent Mutation in CEP290. *Mol. Ther. - Nucleic Acids* **1**, e14 (2012).

249. Gerard, X. *et al.* AON-mediated Exon Skipping Restores Ciliation in Fibroblasts Harboring the Common Leber Congenital Amaurosis CEP290 Mutation. *Mol. Ther. - Nucleic Acids* **1**, e29 (2012).

250. Garanto, A. *et al.* In vitro and in vivo rescue of aberrant splicing in CEP290-associated LCA by antisense oligonucleotide delivery. *Hum. Mol. Genet.* **25**, 2552–2563 (2016).

251. Duijkers, L. *et al.* Antisense Oligonucleotide-Based Splicing Correction in Individuals with Leber Congenital Amaurosis due to Compound Heterozygosity for the c.2991+1655A>G Mutation in CEP290. *Int. J. Mol. Sci.* **19**, (2018).

252. Dulla, K. *et al.* Splice-Modulating Oligonucleotide QR-110 Restores CEP290 mRNA and Function in Human c.2991+1655A>G LCA10 Models. *Mol. Ther. Nucleic Acids* **12**, 730–740 (2018).

253. Bonifert, T. *et al.* Antisense Oligonucleotide Mediated Splice Correction of a Deep Intronic Mutation in OPA1. *Mol. Ther. - Nucleic Acids* **5**, e390 (2016).

254. Garanto, A., van der Velde-Visser, S. D., Cremers, F. P. M. & Collin, R. W. J. Antisense Oligonucleotide-Based Splice Correction of a Deep-Intronic Mutation in CHM Underlying Choroideremia. *Adv. Exp. Med. Biol.* **1074**, 83–89 (2018).

255. Slijkerman, R. W. *et al.* Antisense Oligonucleotide-based Splice Correction for USH2A-associated Retinal Degeneration Caused by a Frequent Deep-intronic Mutation. *Mol. Ther. Nucleic Acids* **5**, e381 (2016).

256. Garanto, A., Duijkers, L., Tomkiewicz, T. Z. & Collin, R. W. J. Antisense Oligonucleotide Screening to Optimize the Rescue of the Splicing Defect Caused by the Recurrent Deep-Intronic ABCA4 Variant c.4539+2001G>A in Stargardt Disease. *Genes* **10**, 452 (2019).

257. Tomkiewicz, T. Z., Suárez-Herrera, N., Cremers, F. P. M., Collin, R. W. J. & Garanto, A. Antisense Oligonucleotide-Based Rescue of Aberrant Splicing Defects Caused by 15 Pathogenic Variants in ABCA4. *Int. J. Mol. Sci.* **22**, 4621 (2021).

258. Commissioner, O. of the. FDA Approves Targeted Treatment for Rare Duchenne Muscular Dystrophy Mutation. *FDA* https://www.fda.gov/news-events/press-announcements/fda-approves-targeted-treatment-rare-duchenne-muscular-dystrophy-mutation-0 (2021).

259. Ambrosini, G., Dreos, R., Kumar, S. & Bucher, P. The ChIP-Seq tools and web server: a resource for analyzing ChIP-seq and other types of genomic data. *BMC Genomics* **17**, 938 (2016).

260. Göndör, A., Rougier, C. & Ohlsson, R. High-resolution circular chromosome conformation capture assay. *Nat. Protoc.* **3**, 303–313 (2008).
261. ENCODE: Encyclopedia of DNA Elements – ENCODE. https://www.encodeproject.org/.
262. Bronstein, R. *et al.* A combined RNA-seq and whole genome sequencing approach for identification of non-coding pathogenic variants in single families. *Hum. Mol. Genet.* doi:10.1093/hmg/ddaa016.
263. Dyka, F. M., Molday, L. L., Chiodo, V. A., Molday, R. S. & Hauswirth, W. W. Dual ABCA4-AAV Vector Treatment Reduces Pathogenic Retinal A2E Accumulation in a Mouse Model of Autosomal Recessive Stargardt Disease. *Hum. Gene Ther.* **30**, 1361–1370 (2019).
264. Lenis, T. L. *et al.* Expression of ABCA4 in the retinal pigment epithelium and its implications for Stargardt macular degeneration. *Proc. Natl. Acad. Sci.* **115**, E11120–E11127 (2018).
265. Quazi, F., Lenevich, S. & Molday, R. S. ABCA4 is an N-retinylidene-phosphatidylethanolamine and phosphatidylethanolamine importer. *Nat. Commun.* **3**, 925 (2012).
266. Quazi, F. & Molday, R. S. Differential phospholipid substrates and directional transport by ATP-binding cassette proteins ABCA1, ABCA7, and ABCA4 and disease-causing mutants. *J. Biol. Chem.* **288**, 34414–34426 (2013).
267. Quazi, F. & Molday, R. S. ATP-binding cassette transporter ABCA4 and chemical isomerization protect photoreceptor cells from the toxic accumulation of excess 11-cis-retinal. *Proc. Natl. Acad. Sci.* **111**, 5024–5029 (2014).
268. Burke, T. R. *et al.* Quantitative fundus autofluorescence in recessive Stargardt disease. *Invest. Ophthalmol. Vis. Sci.* **55**, 2841–2852 (2014).
269. Sparrow, J. R. & Boulton, M. RPE lipofuscin and its role in retinal pathobiology. *Exp. Eye Res.* **80**, 595–606 (2005).
270. Sparrow, J. R. *et al.* The bisretinoids of retinal pigment epithelium. *Prog. Retin. Eye Res.* **31**, 121–135 (2012).
271. Sparrow, J. R., Hicks, D. & Hamel, C. P. The retinal pigment epithelium in health and disease. *Curr. Mol. Med.* **10**, 802–823 (2010).
272. Jaakson, K. *et al.* Genotyping microarray (gene chip) for the ABCR (ABCA4) gene. *Hum. Mutat.* **22**, 395–403 (2003).
273. Maia-Lopes, S. *et al.* ABCA4 mutations in Portuguese Stargardt patients: identification of new mutations and their phenotypic analysis. *Mol. Vis.* **15**, 584–591 (2009).
274. Schulz, H. L. *et al.* Mutation Spectrum of the ABCA4 Gene in 335 Stargardt Disease Patients From a Multicenter German Cohort-Impact of Selected Deep Intronic Variants and Common SNPs. *Invest. Ophthalmol. Vis. Sci.* **58**, 394–403 (2017).
275. Zernant, J. *et al.* Extremely hypomorphic and severe deep intronic variants in the ABCA4 locus result in varying Stargardt disease phenotypes. *Cold Spring Harb. Mol. Case Stud.* **4**, (2018).
276. Zhang, Z. *et al.* Noisy splicing, more than expression regulation, explains why some exons are subject to nonsense-mediated mRNA decay. *BMC Biol.* **7**, 23 (2009).

277. Caminsky, N., Mucaki, E. J. & Rogan, P. K. Interpretation of mRNA splicing mutations in genetic disease: review of the literature and guidelines for information-theoretical analysis. *F1000Research* **3**, (2014).
278. Sangermano, R. *et al.* Photoreceptor Progenitor mRNA Analysis Reveals Exon Skipping Resulting from the ABCA4 c.5461-10T→C Mutation in Stargardt Disease. *Ophthalmology* **123**, 1375–1385 (2016).
279. Fakin, A. *et al.* The Effect on Retinal Structure and Function of 15 Specific ABCA4 Mutations: A Detailed Examination of 82 Hemizygous Patients. *Invest. Ophthalmol. Vis. Sci.* **57**, 5963–5973 (2016).
280. Carrigan, M. *et al.* Panel-Based Population Next-Generation Sequencing for Inherited Retinal Degenerations. *Sci. Rep.* **6**, 33248 (2016).
281. Cornelis, S. S. *et al. Genetic risk estimates for offspring of patients with Stargardt disease*. 2021.08.11.21261888 https://www.medrxiv.org/content/10.1101/2021.08.11.21261888v1 (2021) doi:10.1101/2021.08.11.21261888.
282. Lee, W. *et al.* Complex inheritance of ABCA4 disease: four mutations in a family with multiple macular phenotypes. *Hum. Genet.* **135**, 9–19 (2016).
283. Tanna, P., Strauss, R. W., Fujinami, K. & Michaelides, M. Stargardt disease: clinical features, molecular genetics, animal models and therapeutic options. *Br. J. Ophthalmol.* **101**, 25–30 (2017).
284. Lois, N., Holder, G. E., Bunce, C., Fitzke, F. W. & Bird, A. C. Phenotypic Subtypes of Stargardt Macular Dystrophy–Fundus Flavimaculatus. *Arch. Ophthalmol.* **119**, 359–369 (2001).
285. Parfitt, D. A. *et al.* Identification and Correction of Mechanisms Underlying Inherited Blindness in Human iPSC-Derived Optic Cups. *Cell Stem Cell* **18**, 769–781 (2016).
286. Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q. & Krainer, A. R. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* **31**, 3568–3571 (2003).
287. Rehman, A. U. *et al.* Exploring the Genetic Landscape of Retinal Diseases in North-Western Pakistan Reveals a High Degree of Autozygosity and a Prevalent Founder Mutation in ABCA4. *Genes* **11**, (2019).
288. Motta, F. L. *et al.* Pathogenicity Reclasssification of RPE65 Missense Variants Related to Leber Congenital Amaurosis and Early-Onset Retinal Dystrophy. *Genes* **11**, (2019).
289. Nishiguchi, K. M. *et al.* A founder Alu insertion in RP1 gene in Japanese patients with retinitis pigmentosa. *Jpn. J. Ophthalmol.* **64**, 346–350 (2020).
290. Auslender, N. *et al.* Four USH2A founder mutations underlie the majority of Usher syndrome type 2 cases among non-Ashkenazi Jews. *Genet. Test.* **12**, 289–294 (2008).
291. Khan, A. O., Bergmann, C., Eisenberger, T. & Bolz, H. J. A TULP1 founder mutation, p.Gln301*, underlies a recognisable congenital rod-cone dystrophy phenotype on the Arabian Peninsula. *Br. J. Ophthalmol.* **99**, 488–492 (2015).
292. Avela, K. *et al.* A founder mutation in CERKL is a major cause of retinal dystrophy in Finland. *Acta Ophthalmol. (Copenh.)* **96**, 183–191 (2018).
293. Consultations - Department of Foreign Affairs. https://www.dfa.ie/global-irish/consultations/.

294. Government of Canada, S. C. Census in Brief: Ethnic and cultural origins of Canadians: Portrait of a rich heritage. https://www12.statcan.gc.ca/census-recensement/2016/as-sa/98-200-x/2016016/98-200-x2016016-eng.cfm (2017).
295. Statistics, c=AU; o=Commonwealth of A. ou=Australian B. of. Main Features - Australians overall claim more than 250 ancestries, speak 400 languages at home: Census. https://www.abs.gov.au/ausstats/abs@.nsf/7d12b0f6763c78caca257061001cc588/5a47791aa683b719ca257306000d536c!OpenDocument (2007).
296. Runhart, E. H. *et al.* The Common ABCA4 Variant p.Asn1868Ile Shows Nonpenetrance and Variable Expression of Stargardt Disease When Present in trans With Severe Variants. *Invest. Ophthalmol. Vis. Sci.* **59**, 3220–3231 (2018).
297. Kong, J. *et al.* Correction of the disease phenotype in the mouse model of Stargardt disease by lentiviral gene therapy. *Gene Ther.* **15**, 1311–1320 (2008).
298. Phase I/II Study of SAR422459 in Patients With Stargardt's Macular Degeneration - Full Text View - ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/NCT01367444.
299. McClements, M. E. *et al.* An AAV Dual Vector Strategy Ameliorates the Stargardt Phenotype in Adult Abca4−/− Mice. *Hum. Gene Ther.* **30**, 590–600 (2018).
300. Sub-retinal Transplantation of hESC Derived RPE(MA09-hRPE)Cells in Patients With Stargardt's Macular Dystrophy - Full Text View - ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/NCT01345006.
301. Schwartz, S. D. *et al.* Human embryonic stem cell-derived retinal pigment epithelium in patients with age-related macular degeneration and Stargardt's macular dystrophy: follow-up of two open-label phase 1/2 studies. *Lancet Lond. Engl.* **385**, 509–516 (2015).
302. Commissioner, O. of the. FDA approves first drug for spinal muscular atrophy. *FDA* https://www.fda.gov/news-events/press-announcements/fda-approves-first-drug-spinal-muscular-atrophy (2020).
303. RetNet - Retinal Information Network. https://sph.uth.edu/RetNet/ (2018).
304. OMIM - Online Mendelian Inheritance in Man. https://www.omim.org/.
305. Miraglia, S. *et al.* A novel five-transmembrane hematopoietic stem cell antigen: isolation, characterization, and molecular cloning. *Blood* **90**, 5013–5021 (1997).
306. Yang, Z. *et al.* Mutant prominin 1 found in patients with macular degeneration disrupts photoreceptor disk morphogenesis in mice. *J. Clin. Invest.* **118**, 2908–2916 (2008).
307. Bhattacharya, S. *et al.* Prominin-1 Is a Novel Regulator of Autophagy in the Human Retinal Pigment Epithelium. *Invest. Ophthalmol. Vis. Sci.* **58**, 2366–2387 (2017).
308. ClinVar. https://www.ncbi.nlm.nih.gov/clinvar/.
309. gnomAD. https://gnomad.broadinstitute.org/.
310. Zhang, M. Q. Statistical Features of Human Exons and Their Flanking Regions. *Hum. Mol. Genet.* **7**, 919–932 (1998).
311. Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals | Journal of Computational Biology.

https://www.liebertpub.com/doi/10.1089/1066527041410418?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%200pubmed.

312. Reese, M. G., Eeckman, F. H., Kulp, D. & Haussler, D. Improved Splice Site Detection in Genie. *J. Comput. Biol.* **4**, 311–323 (1997).

313. Pertea, M., Lin, X. & Salzberg, S. L. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res.* **29**, 1185–1190 (2001).

314. da Palma, M. M. *et al.* Synonymous Variant in the CHM Gene Causes Aberrant Splicing in Choroideremia. *Invest. Ophthalmol. Vis. Sci.* **61**, 38 (2020).

315. Davidson, A. E. *et al.* A synonymous codon variant in two patients with autosomal recessive bestrophinopathy alters in vitro splicing of BEST1. *Mol. Vis.* **16**, 2916–2922 (2010).

316. Littink, K. W. *et al.* Homozygosity Mapping in Patients with Cone–Rod Dystrophy: Novel Mutations and Clinical Characterizations. *Invest. Ophthalmol. Vis. Sci.* **51**, 5943–5951 (2010).

317. Michaelides, M. *et al.* An autosomal dominant bull's-eye macular dystrophy (MCDR2) that maps to the short arm of chromosome 4. *Invest. Ophthalmol. Vis. Sci.* **44**, 1657–1662 (2003).

318. Kniazeva, M. *et al.* Clinical and genetic studies of an autosomal dominant cone-rod dystrophy with features of Stargardt disease. *Ophthalmic Genet.* **20**, 71–81 (1999).

319. Permanyer, J. *et al.* Autosomal recessive retinitis pigmentosa with early macular affectation caused by premature truncation in PROM1. *Invest. Ophthalmol. Vis. Sci.* **51**, 2656–2663 (2010).

320. Pras, E. *et al.* Cone-rod dystrophy and a frameshift mutation in the PROM1 gene. *Mol. Vis.* **15**, 1709–1716 (2009).

321. Maw, M. A. *et al.* A frameshift mutation in prominin (mouse)-like 1 causes human retinal degeneration. *Hum. Mol. Genet.* **9**, 27–34 (2000).

322. Michaelides, M. *et al.* The PROM1 mutation p.R373C causes an autosomal dominant bull's eye maculopathy associated with rod, rod-cone, and macular dystrophy. *Invest. Ophthalmol. Vis. Sci.* **51**, 4771–4780 (2010).

323. Del Pozo-Valero, M. *et al.* Expanded Phenotypic Spectrum of Retinopathies Associated with Autosomal Recessive and Dominant Mutations in PROM1. *Am. J. Ophthalmol.* **207**, 204–214 (2019).

324. Cehajic-Kapetanovic, J. *et al.* Clinical and Molecular Characterization of PROM1-Related Retinal Degeneration. *JAMA Netw. Open* **2**, e195752 (2019).

325. Millington-Ward, S. *et al.* Suppression and replacement gene therapy for autosomal dominant disease in a murine model of dominant retinitis pigmentosa. *Mol. Ther. J. Am. Soc. Gene Ther.* **19**, 642–649 (2011).

326. Lentz, J. J. *et al.* Rescue of hearing and vestibular function by antisense oligonucleotides in a mouse model of human deafness. *Nat. Med.* **19**, 345–350 (2013).

327. Mykytyn, K. *et al.* Identification of the gene (BBS1) most commonly involved in Bardet-Biedl syndrome, a complex human obesity syndrome. *Nat. Genet.* **31**, 435–438 (2002).

328. Mykytyn, K. *et al.* Evaluation of complex inheritance involving the most common Bardet-Biedl syndrome locus (BBS1). *Am. J. Hum. Genet.* **72**, 429–437 (2003).

329. Weihbrecht, K. *et al.* Keeping an Eye on Bardet-Biedl Syndrome: A Comprehensive Review of the Role of Bardet-Biedl Syndrome Genes in the Eye. *Med. Res. Arch.* **5**, (2017).
330. Estrada-Cuzcano, A. *et al.* BBS1 mutations in a wide spectrum of phenotypes ranging from nonsyndromic retinitis pigmentosa to Bardet-Biedl syndrome. *Arch. Ophthalmol. Chic. Ill 1960* **130**, 1425–1432 (2012).
331. Abu Safieh, L. *et al.* Clinical and molecular characterisation of Bardet-Biedl syndrome in consanguineous populations: the power of homozygosity mapping. *J. Med. Genet.* **47**, 236–241 (2010).
332. Nachury, M. V. *et al.* A core complex of BBS proteins cooperates with the GTPase Rab8 to promote ciliary membrane biogenesis. *Cell* **129**, 1201–1213 (2007).
333. Anna, A. & Monika, G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J. Appl. Genet.* **59**, 253–268 (2018).
334. Gao, K., Masuda, A., Matsuura, T. & Ohno, K. Human branch point consensus sequence is yUnAy. *Nucleic Acids Res.* **36**, 2257–2267 (2008).
335. McCulloch, D. L. *et al.* ISCEV Standard for full-field clinical electroretinography (2015 update). *Doc. Ophthalmol. Adv. Ophthalmol.* **130**, 1–12 (2015).
336. Rueden, C. T. *et al.* ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* **18**, 529 (2017).
337. Kuivenhoven, J. A. *et al.* An intronic mutation in a lariat branchpoint sequence is a direct cause of an inherited human disorder (fish-eye disease). *J. Clin. Invest.* **98**, 358–364 (1996).
338. Burrows, N. P. *et al.* A point mutation in an intronic branch site results in aberrant splicing of COL5A1 and in Ehlers-Danlos syndrome type II in two British families. *Am. J. Hum. Genet.* **63**, 390–398 (1998).
339. De Klein, A. *et al.* A G-->A transition creates a branch point sequence and activation of a cryptic exon, resulting in the hereditary disorder neurofibromatosis 2. *Hum. Mol. Genet.* **7**, 393–398 (1998).
340. Khan, S. G. *et al.* Two essential splice lariat branchpoint sequences in one intron in a xeroderma pigmentosum DNA repair gene: mutations result in reduced XPC mRNA levels that correlate with cancer risk. *Hum. Mol. Genet.* **13**, 343–352 (2004).
341. Khan, S. G. *et al.* XPC branch-point sequence mutations disrupt U2 snRNP binding, resulting in abnormal pre-mRNA splicing in xeroderma pigmentosum patients. *Hum. Mutat.* **31**, 167–175 (2010).
342. Ece Solmaz, A. *et al.* Targeted multi-gene panel testing for the diagnosis of Bardet Biedl syndrome: Identification of nine novel mutations across BBS1, BBS2, BBS4, BBS7, BBS9, BBS10 genes. *Eur. J. Med. Genet.* **58**, 689–694 (2015).
343. Castro-Sánchez, S. *et al.* Exploring genotype-phenotype relationships in Bardet-Biedl syndrome families. *J. Med. Genet.* **52**, 503–513 (2015).
344. Azari, A. A. *et al.* Retinal Disease Expression in Bardet-Biedl Syndrome-1 (BBS1) Is a Spectrum from Maculopathy to Retina-Wide Degeneration. *Invest. Ophthalmol. Vis. Sci.* **47**, 5004–5010 (2006).

345. ProQR Therapeutics. *A Double-Masked, Randomized, Controlled, Multiple-Dose Study to Evaluate the Efficacy, Safety and Tolerability of QR-421a in Subjects With Retinitis Pigmentosa (RP) Due to Mutations in Exon 13 of the USH2A Gene With Advanced Vision Loss*. https://clinicaltrials.gov/ct2/show/NCT05158296 (2022).
346. Abou Tayoun, A. N. *et al.* Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum. Mutat.* **39**, 1517–1524 (2018).
347. ACGS Best Practice Guidelines for Variant Classification 2019. *The Association for Clinical Genomic Science* https://www.acgs.uk.com/news/acgs-best-practice-guidelines-for-variant-classification-2019/.
348. Bournazos, A. M. *et al.* Standardized practices for RNA diagnostics using clinically accessible specimens reclassifies 75% of putative splicing variants. *Genet. Med.* **24**, 130–145 (2022).
349. Brnich, S. E. *et al.* Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med.* **12**, 3 (2019).
350. Ellingford, J. M. *et al.* Recommendations for clinical interpretation of variants found in non-coding regions of the genome. *Genome Med.* **14**, 73 (2022).
351. D'haene, E. & Vergult, S. Interpreting the impact of noncoding structural variation in neurodevelopmental disorders. *Genet. Med.* **23**, 34–46 (2021).
352. Ward, L. D. & Kellis, M. Interpreting non-coding variation in complex disease genetics. *Nat. Biotechnol.* **30**, 1095–1106 (2012).
353. Abu-Safieh, L. *et al.* A large deletion in the adRP gene PRPF31: evidence that haploinsufficiency is the cause of disease. *Mol. Vis.* **12**, 384–388 (2006).
354. Variant Co-Occurrence (Phasing) Information in gnomAD | gnomAD browser. https://gnomad.broadinstitute.org/news/2021-07-variant-co-occurrence-phasing-information-in-gnomad/.
355. HSE initiates a National Genetics and Genomics Strategy for Ireland. *HSE.ie* https://www.hse.ie/eng/services/news/media/pressrel/hse-initiates-a-national-genetics-and-genomics-strategy-for-ireland.html.
356. 6477 IHRF May22 Meeting Report_v2 single pages.pdf. *Google Docs* https://drive.google.com/file/d/1SIcA7QoQg5XqRHzuDLbcJw3S9beZRs9u/view?usp=embed_facebook.
357. Clift, K. E. *et al.* Patients' views on incidental findings from clinical exome sequencing. *Appl. Transl. Genomics* **4**, 38–43 (2015).
358. Combs, R. *et al.* Understanding the impact of genetic testing for inherited retinal dystrophy. *Eur. J. Hum. Genet.* **21**, 1209–1213 (2013).
359. Facio, F. M., Lee, K. & O'Daniel, J. M. A genetic counselor's guide to using next-generation sequencing in clinical practice. *J. Genet. Couns.* **23**, 455–462 (2014).
360. McVeigh, E., Jones, H., Black, G. & Hall, G. The psychosocial and service delivery impact of genomic testing for inherited retinal dystrophies. *J. Community Genet.* **10**, 425–434 (2019).

361. Halverson, C. M., Clift, K. E. & McCormick, J. B. Was it worth it? Patients' perspectives on the perceived value of genomic-based individualized medicine. *J. Community Genet.* **7**, 145–152 (2016).
362. Coyle, D. Genomics research: Chinese group to lead $400m investment. *The Irish Times* https://www.irishtimes.com/business/health-pharma/genomics-research-chinese-group-to-lead-400m-investment-1.3711337.
363. Sergouniotis, P. I. Inherited Retinal Disorders: Using Evidence as a Driver for Implementation. *Ophthalmol. J. Int. Ophtalmol. Int. J. Ophthalmol. Z. Augenheilkd.* **242**, 187–194 (2019).
364. Endocrinology, T. L. D. &. Spotlight on rare diseases. *Lancet Diabetes Endocrinol.* **7**, 75 (2019).
365. The socioeconomic impact of inherited retinal dystrophies | Deloitte Australia | Deloitte Access Economics. *Deloitte Australia* https://www2.deloitte.com/au/en/pages/economics/articles/socioeconomic-impact-inherited-retinal-dystrophies.html.
366. IRD COUNTS – the impact of inherited retinal degenerations. *Fighting Blindness* https://www.fightingblindness.ie/ird-counts/.
367. Garcia, G. A. *et al.* Profound vision loss impairs psychological well-being in young and middle-aged individuals. *Clin. Ophthalmol. Auckl. NZ* **11**, 417–427 (2017).
368. Cumberland, P. M. & Rahi, J. S. Visual Function, Social Position, and Health and Life Chances: The UK Biobank Study. *JAMA Ophthalmol.* **134**, 959–966 (2016).
369. Seoighe, C. *et al.* The future of genomics in Ireland – focus on genomics for health. Preprint at https://doi.org/10.12688/hrbopenres.13187.1 (2020).
370. Pennell, N. A. *et al.* Economic Impact of Next-Generation Sequencing Versus Single-Gene Testing to Detect Genomic Alterations in Metastatic Non–Small-Cell Lung Cancer Using a Decision Analytic Model. *JCO Precis. Oncol.* 1–9 (2019) doi:10.1200/PO.18.00356.
371. Sabatini, L. M. *et al.* Genomic Sequencing Procedure Microcosting Analysis and Health Economic Cost-Impact Analysis: A Report of the Association for Molecular Pathology. *J. Mol. Diagn.* **18**, 319–328 (2016).
372. Shen, T., Lee, A., Shen, C. & Lin, C. J. The long tail and rare disease research: the impact of next-generation sequencing for rare Mendelian disorders. *Genet. Res.* **97**, e15 (2015).
373. Stark, Z. *et al.* Prospective comparison of the cost-effectiveness of clinical whole-exome sequencing with that of usual care overwhelmingly supports early use and reimbursement. *Genet. Med.* **19**, 867–874 (2017).
374. Williams, M. S. *et al.* Patient-Centered Precision Health In A Learning Health Care System: Geisinger's Genomic Medicine Experience. *Health Aff. (Millwood)* **37**, 757–764 (2018).
375. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Health Sciences Policy, & Roundtable on Genomics and Precision Health. *Implementing and Evaluating Genomic Screening Programs in Health Care Systems: Proceedings of a Workshop*. (National Academies Press (US), 2018).

376. Wurcel, V. *et al.* The Value of Diagnostic Information in Personalised Healthcare: A Comprehensive Concept to Facilitate Bringing This Technology into Healthcare Systems. *Public Health Genomics* **22**, 8–15 (2019).
377. McAllister, M., Dunn, G. & Todd, C. Empowerment: qualitative underpinning of a new clinical genetics-specific patient-reported outcome. *Eur. J. Hum. Genet.* **19**, 125–130 (2011).
378. Mezer, E. *et al.* Attitudes regarding predictive testing for retinitis pigmentosa. *Ophthalmic Genet.* **28**, 9–15 (2007).
379. Furu, T., Kääriäinen, H., Sankila, E. M. & Norio, R. Attitudes towards prenatal diagnosis and selective abortion among patients with retinitis pigmentosa or choroideremia as well as among their relatives. *Clin. Genet.* **43**, 160–165 (1993).
380. Pawlowitzki, I. H., Rüther, K., Brunsmann, F. & von Gizycki, R. Acceptability of prenatal diagnosis for retinitis pigmentosa. *Lancet Lond. Engl.* **2**, 1394–1395 (1986).
381. Schwartz, M. & Rosenberg, T. Prenatal diagnosis of choroideremia. *Acta Ophthalmol. Scand. Suppl.* 33–36 (1996) doi:10.1111/j.1600-0420.1996.tb00381.x.
382. Sohrab, M. A., Allikmets, R., Guarnaccia, M. M. & Smith, R. T. Preimplantation genetic diagnosis for stargardt disease. *Am. J. Ophthalmol.* **149**, 651-655.e2 (2010).
383. Heshka, J. T., Palleschi, C., Howley, H., Wilson, B. & Wells, P. S. A systematic review of perceived risks, psychological and behavioral impacts of genetic testing. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **10**, 19–32 (2008).
384. Meiser, B. *et al.* Psychological impact of genetic testing in women from high-risk breast cancer families. *Eur. J. Cancer Oxf. Engl. 1990* **38**, 2025–2031 (2002).
385. Cornelis, C. *et al.* Whole-exome sequencing in pediatrics: parents' considerations toward return of unsolicited findings for their child. *Eur. J. Hum. Genet.* **24**, 1681–1687 (2016).
386. Home - ClinicalTrials.gov. https://clinicaltrials.gov/.
387. Acland, G. M. *et al.* Gene therapy restores vision in a canine model of childhood blindness. *Nat. Genet.* **28**, 92–95 (2001).
388. Pang, J. *et al.* Self-complementary AAV-mediated gene therapy restores cone function and prevents cone degeneration in two models of Rpe65 deficiency. *Gene Ther.* **17**, 815–826 (2010).
389. Russell, S. *et al.* Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. *Lancet Lond. Engl.* **390**, 849–860 (2017).
390. Dias, M. F. *et al.* Molecular genetics and emerging therapies for retinitis pigmentosa: Basic research and clinical perspectives. *Prog. Retin. Eye Res.* **63**, 107–131 (2018).
391. Maguire, A. M. *et al.* Durability of Voretigene Neparvovec for Biallelic RPE65-Mediated Inherited Retinal Disease: Phase 3 Results at 3 and 4 Years. *Ophthalmology* **128**, 1460–1468 (2021).
392. Gange, W. S. *et al.* Perifoveal Chorioretinal Atrophy after Subretinal Voretigene Neparvovec-rzyl for RPE65-Mediated Leber Congenital Amaurosis. *Ophthalmol. Retina* **6**, 58–64 (2022).

393. Sengillo, J. D. *et al.* Visual Acuity, Retinal Morphology, and Patients' Perceptions after Voretigene Neparovec-rzyl Therapy for RPE65-Associated Retinal Disease. *Ophthalmol. Retina* **6**, 273–283 (2022).
394. Deng, C. *et al.* Real-world outcomes of voretigene neparvovec treatment in pediatric patients with RPE65-associated Leber congenital amaurosis. *Graefes Arch. Clin. Exp. Ophthalmol. Albrecht Von Graefes Arch. Klin. Exp. Ophthalmol.* **260**, 1543–1550 (2022).
395. Lam, B. L. *et al.* Choroideremia Gene Therapy Phase 2 Clinical Trial: 24-Month Results. *Am. J. Ophthalmol.* **197**, 65–73 (2019).
396. Dimopoulos, I. S. *et al.* Two-Year Results After AAV2-Mediated Gene Therapy for Choroideremia: The Alberta Experience. *Am. J. Ophthalmol.* **193**, 130–142 (2018).
397. NightstaRx Ltd, a Biogen Company. *An Open-Label Safety Study of Retinal Gene Therapy for Choroideremia With Bilateral, Sequential Administration of Adeno-Associated Viral Vector (AAV2) Encoding Rab Escort Protein 1 (REP1)*. https://clinicaltrials.gov/ct2/show/NCT03507686 (2023).
398. NightstaRx Ltd, a Biogen Company. *A Randomised, Open Label, Outcomes-Assessor Masked, Prospective, Parallel Controlled Group, Phase 3 Clinical Trial of Retinal Gene Therapy for Choroideremia Using an Adeno-Associated Viral Vector (AAV2) Encoding Rab Escort Protein 1 (REP1)*. https://clinicaltrials.gov/ct2/show/NCT03496012 (2023).
399. Short-term Assessment of Subfoveal Injection of Adeno-Associated Virus-Mediated hCHM Gene Augmentation in Choroideremia Using Adaptive Optics Ophthalmoscopy | Genetics and Genomics | JAMA Ophthalmology | JAMA Network. https://jamanetwork.com/journals/jamaophthalmology/fullarticle/2789795.
400. Aleman, T. S. *et al.* Adeno-Associated Virus Serotype 2-hCHM Subretinal Delivery to the Macula in Choroideremia: Two-Year Interim Results of an Ongoing Phase I/II Gene Therapy Trial. *Ophthalmology* **129**, 1177–1191 (2022).
401. Parker, M. A. *et al.* Three-year safety results of SAR422459 (EIAV-ABCA4) gene therapy in patients with ABCA4-associated Stargardt disease: an open-label dose-escalation phase I/IIa clinical trial, cohorts 1-5. *Am. J. Ophthalmol.* **240**, 285–301 (2022).
402. Russell, S. R. *et al.* Intravitreal antisense oligonucleotide sepofarsen in Leber congenital amaurosis type 10: a phase 1b/2 trial. *Nat. Med.* **28**, 1014–1021 (2022).
403. ProQR Therapeutics. *An Open-Label, Dose Escalation and Double-Masked, Randomized, Controlled Study to Evaluate the Safety and Tolerability of Sepofarsen in Pediatric Subjects <8 Years of Age With Leber Congenital Amaurosis Type 10 (LCA10) Due to the c.2991 +1655A>G (p.Cys998X) Mutation.* https://clinicaltrials.gov/ct2/show/NCT04855045 (2022).
404. ProQR to Focus Exclusively on Axiomer RNA-editing Technology and Partner Ophthalmology Programs | ProQR Therapeutics. https://www.proqr.com/press-releases/proqr-to-focus-exclusively-on-axiomer-rna-editing-technology-and-partner-ophthalmology-programs.
405. Radziwon, A. *et al.* Single-base substitutions in the CHM promoter as a cause of choroideremia. *Hum. Mutat.* **38**, 704–715 (2017).

406. Coppieters, F. *et al.* Hidden Genetic Variation in LCA9-Associated Congenital Blindness Explained by 5′UTR Mutations and Copy-Number Variations of NMNAT1. *Hum. Mutat.* **36**, 1188–1196 (2015).
407. Van Schil, K. *et al.* Mapping the genomic landscape of inherited retinal disease genes prioritizes genes prone to coding and noncoding copy-number variations. *Genet. Med.* **20**, 202–213 (2018).
408. Li, H.-P., Yuan, S.-Q., Wang, X.-G., Sheng, X.-L. & Li, X.-R. Myopia with X-linked retinitis pigmentosa results from a novel gross deletion of RPGR gene. *Int. J. Ophthalmol.* **13**, 1306–1311 (2020).
409. de Bruijn, S. E. *et al.* Structural Variants Create New Topological-Associated Domains and Ectopic Retinal Enhancer-Gene Contact in Dominant Retinitis Pigmentosa. *Am. J. Hum. Genet.* **107**, 802–814 (2020).
410. Zhou, V. W., Goren, A. & Bernstein, B. E. Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* **12**, 7–18 (2011).
411. Corso-Díaz, X., Jaeger, C., Chaitankar, V. & Swaroop, A. Epigenetic control of gene regulation during development and disease: A view from the retina. *Prog. Retin. Eye Res.* **65**, 1–27 (2018).
412. Dvoriantchikova, G., Seemungal, R. J. & Ivanov, D. DNA Methylation Dynamics During the Differentiation of Retinal Progenitor Cells Into Retinal Neurons Reveal a Role for the DNA Demethylation Pathway. *Front. Mol. Neurosci.* **12**, 182 (2019).
413. Dvoriantchikova, G., Seemungal, R. J. & Ivanov, D. The epigenetic basis for the impaired ability of adult murine retinal pigment epithelium cells to regenerate retinal tissue. *Sci. Rep.* **9**, 3860 (2019).
414. Tucker, B. A. *et al.* Exome sequencing and analysis of induced pluripotent stem cells identify the cilia-related gene male germ cell-associated kinase (MAK) as a cause of retinitis pigmentosa. *Proc. Natl. Acad. Sci. U. S. A.* **108**, E569-576 (2011).
415. Delvallée, C. *et al.* A BBS1 SVA F retrotransposon insertion is a frequent cause of Bardet-Biedl syndrome. *Clin. Genet.* **99**, 318–324 (2021).
416. Savage, A. L. *et al.* Retrotransposons in the development and progression of amyotrophic lateral sclerosis. *J. Neurol. Neurosurg. Psychiatry* **90**, 284–293 (2019).
417. Tatour, Y. & Ben-Yosef, T. Syndromic Inherited Retinal Diseases: Genetic, Clinical and Diagnostic Aspects. *Diagn. Basel Switz.* **10**, E779 (2020).
418. Vig, A. *et al.* DYNC2H1 hypomorphic or retina-predominant variants cause nonsyndromic retinal degeneration. *Genet. Med.* **22**, 2041–2051 (2020).
419. Hitti-Malin, R. J. *et al.* Using single molecule Molecular Inversion Probes as a cost-effective, high-throughput sequencing approach to target all genes and loci associated with macular diseases. 2022.03.17.22272534 Preprint at https://doi.org/10.1101/2022.03.17.22272534 (2022).
420. Burkard, M. *et al.* Accessory heterozygous mutations in cone photoreceptor CNGA3 exacerbate CNG channel-associated retinopathy. *J. Clin. Invest.* **128**, 5663–5675 (2018).

421. Lee, W. *et al.* Cis-acting modifiers in the ABCA4 locus contribute to the penetrance of the major disease-causing variant in Stargardt disease. *Hum. Mol. Genet.* **30**, 1293–1304 (2021).
422. Zernant, J. *et al.* Rare and common variants in ROM1 and PRPH2 genes trans-modify Stargardt/ABCA4 disease. *PLOS Genet.* **18**, e1010129 (2022).
423. Ebermann, I. *et al. PDZD7* is a modifier of retinal disease and a contributor to digenic Usher syndrome. *J. Clin. Invest.* **120**, 1812–1823 (2010).
424. Coppieters, F. *et al.* Genetic screening of LCA in Belgium: predominance of CEP290 and identification of potential modifier alleles in AHI1 of CEP290-related phenotypes. *Hum. Mutat.* **31**, E1709–E1766 (2010).
425. Namburi, P. *et al.* A unique PRDM13-associated variant in a Georgian Jewish family with probable North Carolina macular dystrophy and the possible contribution of a unique CFH variant. *Mol. Vis.* **26**, 299–310 (2020).
426. Venturini, G., Rose, A. M., Shah, A. Z., Bhattacharya, S. S. & Rivolta, C. CNOT3 Is a Modifier of PRPF31 Mutations in Retinitis Pigmentosa with Incomplete Penetrance. *PLoS Genet.* **8**, (2012).
427. Sharon, D., Wimberg, H., Kinarty, Y. & Koch, K.-W. Genotype-functional-phenotype correlations in photoreceptor guanylate cyclase (GC-E) encoded by GUCY2D. *Prog. Retin. Eye Res.* **63**, 69–91 (2018).
428. Abrahams, L. *et al.* Evidence in disease and non-disease contexts that nonsense mutations cause altered splicing via motif disruption. *Nucleic Acids Res.* **49**, 9665–9685 (2021).
429. Uppal, S., Liu, T., Poliakov, E., Gentleman, S. & Redmond, T. M. The dual roles of RPE65 S-palmitoylation in membrane association and visual cycle function. *Sci. Rep.* **9**, 5218 (2019).
430. Redmond, T. M. *et al.* Rpe65 is necessary for production of 11- cis -vitamin A in the retinal visual cycle. *Nat. Genet.* **20**, 344–351 (1998).
431. Redmond, T. M. *et al.* Mutation of key residues of RPE65 abolishes its enzymatic role as isomerohydrolase in the visual cycle. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 13658–13663 (2005).
432. Ellingford, J. M. *et al.* Functional and in-silico interrogation of rare genomic variants impacting RNA splicing for the diagnosis of genomic disorders. *bioRxiv* 781088 (2019) doi:10.1101/781088.
433. Dickinson, D. J. & Goldstein, B. CRISPR-Based Methods for Caenorhabditis elegans Genome Engineering. *Genetics* **202**, 885–901 (2016).
434. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 1–1 (2019) doi:10.1038/s41586-019-1711-4.
435. Dam, T. J. P. van *et al.* CiliaCarta: An integrated and validated compendium of ciliary genes. *PLOS ONE* **14**, e0216705 (2019).
436. Messchaert, M. *et al.* Eyes shut homolog is important for the maintenance of photoreceptor morphology and visual function in zebrafish. *PLoS ONE* **13**, (2018).
437. Voigt, A. P. *et al.* Single-cell RNA sequencing in vision research: Insights into human retinal health and disease. *Prog. Retin. Eye Res.* 100934 (2020) doi:10.1016/j.preteyeres.2020.100934.

438. Karademir, D., Todorova, V., Ebner, L. J. A., Samardzija, M. & Grimm, C. Single-cell RNA sequencing of the retina in a model of retinitis pigmentosa reveals early responses to degeneration in rods and cones. *BMC Biol.* **20**, 86 (2022).
439. Pontikos, N. *et al. Eye2Gene: prediction of causal inherited retinal disease gene from multimodal imaging using deep-learning*. https://www.researchsquare.com/article/rs-2110140/v1 (2022) doi:10.21203/rs.3.rs-2110140/v1.

Appendix 1

genes

MDPI

Article

Findings from a Genotyping Study of over 1000 People with Inherited Retinal Disorders in Ireland

Laura Whelan [1,*], Adrian Dockery [1], Niamh Wynne [2], Julia Zhu [3], Kirk Stephenson [3], Giuliana Silvestri [4,5], Jacqueline Turner [3], James J. O'Byrne [3], Matthew Carrigan [1], Peter Humphries [1], David Keegan [3], Paul F. Kenna [1,2] and G. Jane Farrar [1]

1 The School of Genetics & Microbiology, Trinity College Dublin, D02 VF25 Dublin, Ireland; dockerya@tcd.ie (A.D.); carrigma@tcd.ie (M.C.); pete.humphries@tcd.ie (P.H.); Paul.Kenna@tcd.ie (P.F.K.); gjfarrar@tcd.ie (G.J.F.)
2 The Research Foundation, Royal Victoria Eye and Ear Hospital, D02 XK51 Dublin, Ireland; wynnenc@tcd.ie
3 Clinical Genetics Centre for Ophthalmology, The Mater Misericordiae University Hospital, D07 R2WY Dublin, Ireland; julia.q.zhu@gmail.com (J.Z.); kirkstephenson@hotmail.com (K.S.); jturner@mater.ie (J.T.); jamesobyrne@mater.ie (J.J.O.); dkeegan@mater.ie (D.K.)
4 Department of Ophthalmology, The Royal Victoria Hospital, Belfast BT12 6BA, Northern Ireland, UK; Julie.Silvestri@belfasttrust.hscni.net
5 Centre for Experimental Medicine, Queen's University Belfast, Belfast BT7 1NN, Northern Ireland, UK
* Correspondence: whelanl1@tcd.ie

Received: 16 November 2019; Accepted: 13 January 2020; Published: 16 January 2020

check for updates

Abstract: The Irish national registry for inherited retinal degenerations (Target 5000) is a clinical and scientific program to identify individuals in Ireland with inherited retinal disorders and to attempt to ascertain the genetic cause underlying the disease pathology. Potential participants first undergo a clinical assessment, which includes clinical history and analysis with multimodal retinal imaging, electrophysiology, and visual field testing. If suitable for recruitment, a sample is taken and used for genetic analysis. Genetic analysis is conducted by use of a retinal gene panel target capture sequencing approach. With over 1000 participants from 710 pedigrees now screened, there is a positive candidate variant detection rate of approximately 70% (495/710). Where an autosomal recessive inheritance pattern is observed, an additional 9% (64/710) of probands have tested positive for a single candidate variant. Many novel variants have also been detected as part of this endeavor. The target capture approach is an economic and effective means of screening patients with inherited retinal disorders. Despite the advances in sequencing technology and the ever-decreasing associated processing costs, target capture remains an attractive option as the data produced is easily processed, analyzed, and stored compared to more comprehensive methods. However, with decreasing costs of whole genome and whole exome sequencing, the focus will likely move towards these methods for more comprehensive data generation.

Keywords: inherited eye disease; ophthalmic genetics; genomics; next generation sequencing; retinopathy; rare variants; novel variants

1. Introduction

Inherited retinal degenerations (IRDs) are a broad set of clinically and genetically diverse conditions that represent the leading cause of visual dysfunction in those of working age. IRDs are typically caused by improper development or death of photoreceptor cells and have a substantial effect on both the quality of life of those affected and health economics. Inheritance patterns include autosomal recessive, autosomal dominant and X-linked, as well as rarer mitochondrial and digenic forms [1,2].

Since the discovery of the first IRD genes via linkage analysis [3–5], a comprehensive repertoire of disease-causing genes has been described.

To date, over 270 genes have been associated with a wide variety of IRDs across a phenotypic spectrum [6], which are routinely divided into large subcategories in order to readily distinguish the specific retinal regions and cell types affected; for example, retinitis pigmentosa (RP) versus macular dystrophy, which initially cause predominantly peripheral and central vision loss, respectively. In the majority of IRDs, ocular manifestations occur in isolation. However, many forms of syndromic IRD that result in systemic disease have also been described [7]; more common forms include Usher Syndrome and Bardet-Biedl syndrome. RP is the most frequent IRD presentation with an incidence of approximately 1 in 3000 [8]. Over 70 genes are implicated as causative of RP, making it one of the most diverse and genetically heterogeneous Mendelian conditions described to date [9]. Other IRDs are less heterogeneous—autosomal recessive Stargardt disease, for example, which is caused by pathogenic variations in *ABCA4* in approximately 95% of cases [10–13]. Furthermore, IRDs such as X-linked retinoschisis are characterised by a single gene aetiology [14].

Using the incidence rates previously mentioned, it can be estimated that over 2.5 million individuals are affected by RP alone worldwide, and diagnoses are often based solely on clinical presentation [15]. The importance of identifying the genetic causes of IRDs is becoming increasingly clear, allowing for more accurate prognosis of disease progression and risk assessment for other family members. Importantly, a genetic diagnosis can also facilitate a better understanding of the disease for those affected. Furthermore, multiple clinical trials to treat these diseases are ongoing, with the first gene therapy for IRD, Luxturna, having been approved for use by both the US Food and Drug Administration and the European Medicines Agency [16]. Typically, such trials and treatments require that a causative variant has been identified in order to participate. Here, we present an update on Target 5000, a next-generation sequencing (NGS) study that aims to provide a large cohort of Irish patients with an accurate genetic diagnosis, whilst further elucidating the underlying genetic architecture of IRDs in Ireland.

Target 5000 is the Irish national registry for those affected by an IRD. It is both a research and clinical study that aims to elucidate the underlying genetic architecture of IRDs in Ireland and to provide an accurate genetic diagnosis for those affected. Target 5000 utilises target capture next generation sequencing, a time and cost-effective screening method that boasts more even coverage across captured regions, positively affecting diagnostic yield, in comparison with whole exome sequencing (WES) [17]. Exonic regions of over 250 IRD-associated genes as well as a number of intronic regions harbouring previously identified pathogenic variants were captured. Although whole genome sequencing (WGS) has become more accessible in recent years and facilitates the discovery of both deep intronic variants and novel genotype-phenotype correlations [18], significant challenges remain in terms of data handling, variant interpretation, and cost. Target 5000 significantly highlights the benefits of target capture NGS as a preliminary diagnostic measure, before employing more comprehensive sequencing methods such as WES or WGS.

Here we present novel findings from both newly recruited participants and retrospective analysis of 750 individuals previously screened, resulting over 1000 individuals across 710 pedigrees genetically screened to date. 458 individuals have been sequenced for variants in 210 genes, 689 individuals have been sequenced for variants in 254 genes, and confirmation sequencing is ongoing for other affected family members. The average age of recruited participants to Target 5000 is 55 (Figure S1). Target 5000 has identified positive candidate variants in almost 70% (495/710) of cases, including both previously reported pathogenic and novel potentially pathogenic variants. This includes novel structural variants, single nucleotide variants, and splice variants. Furthermore, 9% (64/710) of pedigrees were found to carry a pathogenic/likely pathogenic variant on one allele where the genotype and phenotype were associated with a recessive inheritance pattern.

2. Materials and Methods

2.1. Patient Identification and Recruitment

Probands and other family members are recruited nationwide and primarily assessed at the Research Foundation of the Royal Victoria Eye and Ear Hospital (Dublin, Ireland), the Mater Misericordiae University Hospital (Dublin, Ireland), and the Royal Victoria Hospital (Belfast, Northern Ireland, UK). This allows for the inclusion of participants from both The Republic of Ireland and Northern Ireland. Following informed consent, the applicable combination of tests were carried out to accurately assess the patients' phenotype (Table 1).

Table 1. Table of typical assessments, instruments, and manufacturers routinely used as part of Target 5000 recruitment.

Analyses	Method	Manufacturer
Best-corrected visual acuity	Revised 2000 early treatment diabetic retinopathy study (ETDRS) charts	Precision Vision, La Salle, IL, USA
Colour vision	Lanthony desaturated D-15 panel under standardised lighting conditions	Gulden Ophthalmics, Elkins Park, PA, USA
Peripheral visual fields	Goldmann perimeter 940 (iv4e, i4e and 04e targets)	Haag-Streit AG, Köniz, Switzerland
Full field electroretinograms	Roland Consult RETI-port retiscan.	Brandenburg an der Havel, Germany
Fundus colour and autofluorescence photography	Topcon CRC50DX/Optos Daytona	Topcon Great Britain Ltd., Berkshire, England/Optos plc, Dunfermline, Scotland
Spectral domain optical coherence tomography	Cirrus HD-OCT	Carl Zeiss Meditec, Berlin, Germany

2.2. DNA Acquisition and Next Generation Sequencing

DNA was isolated from either blood (DNA Blood Maxi Kit, Qiagen, Hilden, Germany) or saliva (Oragene-DNA, DNA Genotek, ON, Canada) samples from participants. Sample preparation was carried out using a hybridisation-based target capture sequencing method previously described [15]. The average read coverage achieved was 125× per captured region. All genomic locations refer to the Hg38 reference genome (Homo sapiens GRCh38).

2.3. Variant Confirmation Sequencing

To validate variants identified by NGS, relevant genomic loci containing the mutations were amplified by polymerase chain reaction (PCR) and investigated by direct sequencing. Primers were procured from Sigma-Aldrich (Gillingham, England, UK). DNA products were standardly amplified using Q5 High-Fidelity 2× Master Mix (New England Biolabs Inc., Ipswich, MA, USA). Segregation analyses for additional family members were amplified similarly or alternatively, directly from blood (Phusion Blood Direct PCR Kit, Thermo Scientific, MA, USA) or saliva (Phusion Human Specimen Direct PCR Kit, Thermo Scientific) where applicable. The annealing temperatures for reactions were optimised for each variant; all other details were executed as per the supplier's recommendations. Sanger sequencing was performed by Eurofins Genomics (Ebersberg, Germany).

2.4. Choroideremia Deletion Confirmation

Primers were designed to probe for the exonic regions of *CHM* (MIM: 300390; ChrX: 85861180-86047558) and the neighbouring genes, *POF1B* (MIM: 300603; ChrX: 85277396-85379709) and

DACH2 (MIM: 300608; ChrX: 86148451-86832602), as well as the intergenic space between. A full list of the primers utilised is available in Supplementary Materials (Table S1).

2.5. Sequencing of RPGR ORF15

A previously reported sequencing strategy of RPGR was employed in order to cover the highly repetitive sequences of ORF15, a major cause of X-linked RP [19]. Primers used to probe this region can be found in Figure S2.

2.6. Single-Molecule Molecular Inversion Probe (smMIP)-Based Sequencing of ABCA4

smMIPs-based whole gene sequencing of *ABCA4* was carried out using a NextSeq 500 as part of a large-scale study by Khan et al. [13]. 3866 single-molecule molecular inversion probes (smMIPs) were designed as described previously [20] and employed to capture 110nt increments of both the sense and antisense strands of the *ABCA4* gene, as well as 40 kb of flanking sequences. Sample processing, sequencing, data analysis, and variant interpretation was carried out, as described by Khan et al. [13].

2.7. Data Analysis and Variant Interpretation of Target Capture NGS Data

Raw sequencing data were demultiplexed and mapped to the IRD-relevant regions of the human genome (Hg38) as previously described [15]. The American College of Medical Genetics and Genomics (ACMG) criteria for classifying pathogenic variants was utilised to interpret variants [21]. Other proposed modifications to the ACMG guidelines to further quantitate certain lines of evidence were also employed during the study. These modifications effect three lines of evidence that can be applied. These adjustments consider using segregation data in a quantitative manner (code: PP1) [22], quantitating evidence that considers the rare incidence of variants (code: PS4) and the value of accurately phenotyping conditions with single gene aetiologies (code: PP4) [23]. To further implement these recommendations, REVEL and dbscSNV RF scores were added to the existing suite of ensemble predictors to more stringently detect agreement between predictors of pathogenicity [24,25]. Similarly, Manta was added to our structural variant pipeline to assist in the detection of genomic rearrangements [26].

2.8. Ethical Approval

Ethical approval for this study was awarded by the Research and Medical Ethics committee of the Royal Victoria Eye and Ear Hospital (13-06-2011: HRA-POR201097) and by the Institutional Review Board of the Mater Misericordiae University Hospital and Mater Private Hospital (MMUH IRB 1/378/1358), Dublin, Ireland, prior to commencement. All work was carried out in accordance with the approved guidelines. All patients have given informed consent before recruitment to the study.

3. Results

3.1. Clinical Presentation and Positive Candidate Detection Rates

Thus far, 710 pedigrees have been analysed as part of this study by target capture sequencing. 458 individuals have been sequenced for variants in 210 genes, 689 individuals have been sequenced for variants in 254 genes, and confirmation sequencing is ongoing for other affected family members. A spectrum of IRDs were observed in this study, as can be seen in Figure 1. The positive candidate dectection rate has improved from previous reports [15,27], now reaching approximately 70% (495/710). Furthermore, a single candidate variant only has been identified in an additional 9% (64/710) of pedigrees diagnosed with an IRD that typically exhibits a recessive inheritance pattern (Figure 2). In addition, a number of novel missense variants have been detected in the course of this study (Table 2).

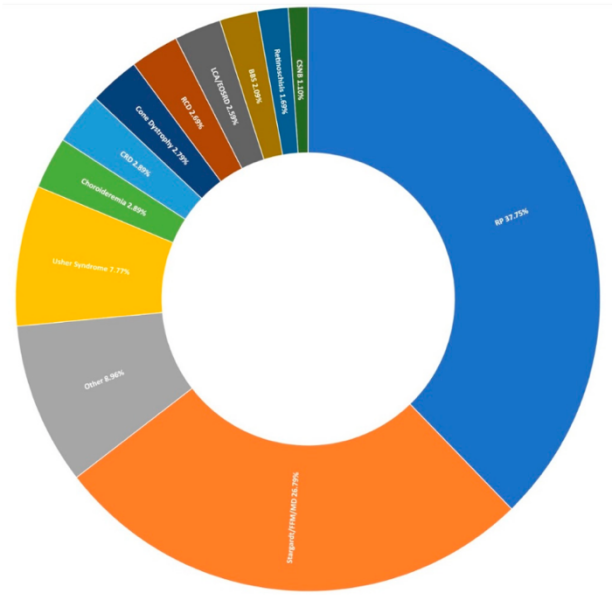


Figure 1. Clinical presentation of 1004 recruited Target 5000 pedigrees. RP: Retinitis Pigmentosa, FFM: Fundus Flavimaculatus, MD: Macular Dystrophy, CRD: Cone-Rod Dystrophy, RCD: Rod-Cone Dystrophy, LCA: Leber Congenital Amaurosis, EOSRD: Early-Onset Retinal Degeneration, BBS: Bardet-Biedl Syndrome, CSNB: Congenital Stationary Night Blindness (Table S2).

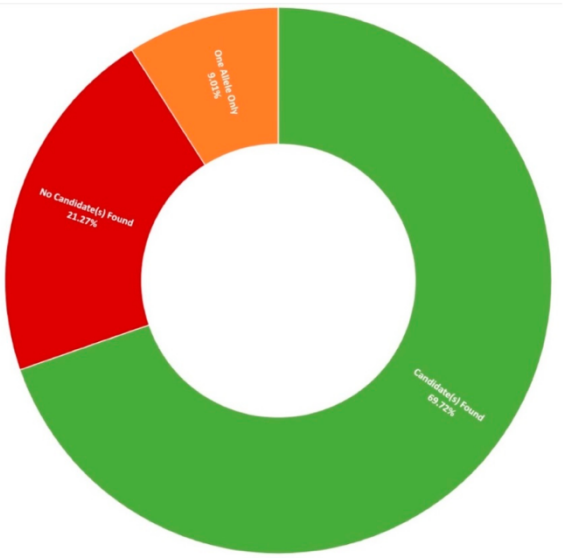


Figure 2. Diagnostic yield rates for 710 Target 5000 pedigrees utilising target capture next-generation sequencing of the exonic regions of over 250 genes and previously identified pathogenic intronic variants (Table S3).

Table 2. List of 19 novel missense variants discovered in Target 5000. This table does not include novel structural variants reported below. To the best of our knowledge, mutations listed here have not been previously associated with an IRD. American College of Medical Genetics and Genomics (ACMG) guidelines have been applied to all variants [21], with modifications made in accordance with recommendations by Romanet et al. [23]. (S): reclassified as strong evidence. VUS: Variant of unknown significance.

Gene	Condition	Transcript ID	Chr	Genomic Location	Nucleotide Change	Protein Change	ACMG Criteria	ACMG Classification	MetaLR Score	M-CAP Score	REVEL Score	Observed With
ABCA4	Stargardt Disease	NM_000350.2	chr1	g.94062649	c.1865G>A	p.Ser622Asn	PM2, PP3, PP4(S)	Likely Pathogenic	0.668	0.065	0.279	c.161G>A
ABCA4	Macular Dystophy (Recessive)	NM_000350.2	chr1	g.94111517	c.223T>C	p.Cys75Arg	PM2, PM5, PP3	VUS	0.995	0.711	0.967	c.4253+43G>A
ABCA4	Stargardt Disease	NM_000350.2	chr1	g.94029516	c.4468T>C	p.Cys1490Arg	PM2, PP3, PP4(S)	Likely Pathogenic	0.952	0.723	0.932	c.3305A>T
ABCA4	Macular Dystophy (Recessive)	NM_000350.2	chr1	g.94014674	c.5329A>G	p.Met1777Val	PM2, PP3	VUS	0.663	0.066	0.758	c.1037A>C
ABCA4	Stargardt Disease	NM_000350.2	chr1	g.93996182	c.6743T>C	p.Phe2248Ser	PM2, PP3, PP4(S)	Likely Pathogenic	0.861	0.528	0.935	c.302+68C>T
AHI1	Retinitis Pigmentosa (Recessive)	NM_001134830.1	chr6	g.135428670	c.2582G>A	p.Gly861Glu	PM2, PM3, PP1, PP3	Likely Pathogenic	0.583	0.159	0.816	c.1267C>T
AIPL1	Leber Congenital Amaurosis	NM_014336.4	chr17	g.6425759	c.856G>C	p.Asp286His	PM2, PM3, PP1, PP3	Likely Pathogenic	0.877	0.290	0.791	c.834G>A
BBS10	Retinitis Pigmentosa (Recessive)	NM_024685.3	chr12	g.76348204	c.155G>A	p.Gly52Asp	PM1, PM2, PP3	VUS	0.527	0.050	0.591	c.2119_2120delGT
BEST1	Best Vitelliform Macular Dystrophy (Dominant)	NM_001139443.1	chr11	g.61955783	c.133C>G	p.Arg45Gly	PM2, PP3, PP4(S)	Likely Pathogenic	0.966	0.699	0.945	/
BEST1	Best Vitelliform Macular Dystrophy (Dominant)	NM_001139443.1	chr11	g.61955157	c.23A>G	p.Tyr8Cys	PM2, PP3, PP4(S)	Likely Pathogenic	0.976	0.710	0.917	/
CLRN1	USH Type III	NM_001195794.1	chr3	g.150928039	c.635A>G	p.His212Arg	PM2, PM3, PP1, PP3	Likely Pathogenic	0.573	0.055	0.799	c.118T>G
CNGA3	Achromatopsia	NM_001298.2	chr2	g.98391904	c.607T>C	p.Trp203Arg	PM2, PP3	VUS	0.960	0.316	0.950	c.1694C>T
MYO7A	USH Type I	NM_000260.3	chr11	g.77192227	c.4101C>G	p.Ile1367Met	PM2, PP3	VUS	0.542	0.220	0.672	c.2904G>T
PEX7	Uncategorised Syndrome	NM_000288.3	chr6	g.136822742	c.77C>T	p.Pro26Leu	PM2, PP3, PP4	VUS	0.962	0.947	0.864	c.875T>A
PRPF31	Retinitis Pigmentosa (Simplex)	NM_015629.3	chr19	g.54124558	c.757G>A	p.Gly253Arg	PM2, PP3	VUS	0.946	0.911	0.957	/
PRPH2	Retinitis Pigmentosa (Dominant)	NM_000322.4	chr6	g.42721878	c.457A>G	p.Lys153Glu	PM2, PP3	VUS	0.628	0.180	0.864	/
PRPH2	Retinitis Pigmentosa (Dominant)	NM_000322.4	chr6	g.42721871	c.464C>T	p.Thr155Ile	PM2, PP3	VUS	0.518	0.100	0.649	/
RDH5	Fundus Albipunctatus	NM_001199771.1	chr12	g.55721276	c.92T>C	p.Val31Ala	PM2, PP3	VUS	0.670	0.143	0.737	c.712G>T
USH2A	USH Type II	NM_206933.2	chr1	g.216247097	c.2297G>A	p.Cys766Tyr	PM2, PM3, PM5, PP3	Likely Pathogenic	0.945	0.669	0.952	c.2299delG

3.2. *Retinitis Pigmentosa*

The most prevalent clinical presentation was retinitis pigmentosa (RP; MIM: 268000), accounting for 37.75% (379/1004) of all recruited pedigrees (Figure 1). These figures incorporate various phenotypic presentations, including atypical, inverse, and paravenous RP. The total number of pedigrees analysed and clinically reported as having a family history of RP was 184. Of these, 131 have been genetically resolved. The total number of pedigrees analysed and clinically reported as having simplex RP was 79. 48 of these pedigrees have been genetically resolved. This gives a genetic diagnosis rate of 71% and 61%, respectively. Of these resolved cases, 81 exhibited an autosomal dominant inheritance pattern, 71 exhibited an autosomal recessive inheritance pattern, and 27 exhibited an X-linked inheritance pattern. Autosomal dominant was the most common inheritance pattern, accounting for over 45% (81/178) of all sequenced RP pedigrees in this cohort (Figure 3). This figure is larger than reported in other studies [28]. Variants in *RHO* (MIM: 180380) were the most common candidate variants for RP in this cohort, accounting for over 14% (26/178) of all sequenced RP pedigrees, and over 32% (26/81) of dominant pedigrees alone. It is important to note that there is likely an even higher prevalence of *RHO*-linked RP in the Irish IRD cohort, as many pedigrees involved in single-gene studies were excluded from this study due to the discovery of a causative variant previously [15,27]. Of these variants, namely c.533A>G (p.Tyr178Cys), c.541G>A (p.Glu181Lys), and c.620T>G (p.Met207Arg) collectively account for candidate *RHO* mutations across 13 different pedigrees to date.



Figure 3. A summary of the genetic architecture of retinitis pigmentosa (RP) in Target 5000 participants (Table S4).

The most commonly observed candidate gene for autosomal recessive RP was *USH2A* (MIM: 608400). Two variants in this gene were found in over 21% (15/71) of pedigrees with this phenotype (Figure 3). As variants in *USH2A* are also the primary cause of Type II Usher Syndrome, it

is important to note that participants included in this section presented with recessive RP without syndromic disease manifestation. Non-syndromic *USH2A*-linked RP has been consistently reported in other IRD cohorts [29]. Interestingly, a proportion of recessive RP pedigrees (>9%) (7/71) carry variants in *ABCA4*, typically associated with Stargardt Disease, a form of macular dystrophy. In the Target 5000 patient cohort, it has become apparent that an atypical RP phenotype appears to be caused by variants in *ABCA4*, as previously described in other cohorts (Figures 4 and 5) [10,30].

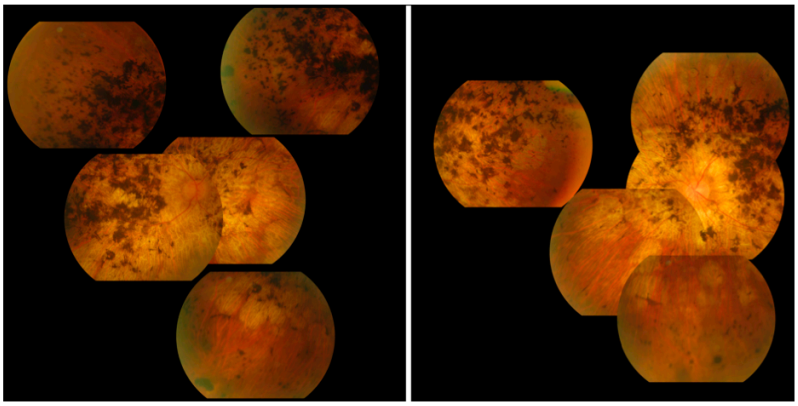


Figure 4. Collage of fundus images of a patient initially presenting with a Stargardt phenotype that progressed to phenotype suggestive of retinitis pigmentosa. Extensive pigmentation can be seen at both maculae and also in the periphery of both eyes (**left**—right eye, **right**—left eye). There are also distinct punched-out areas in the pigment epithelium bilaterally.

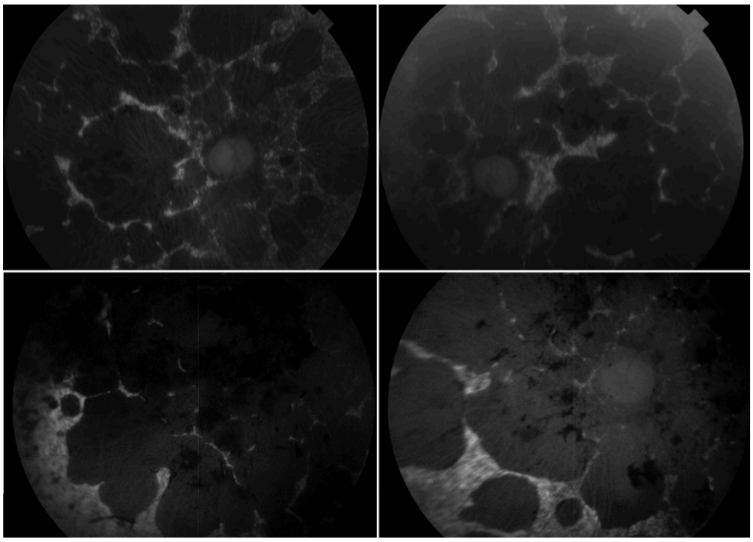


Figure 5. Montage of autofluorescence fundus images of the patients initially presenting with a Stargardt phenotype that progressed to phenotype suggestive of retinitis pigmentosa. The right eyes are shown in the **left** column; the left eyes are shown in the **right** column. Patient example 1 is the first row and patient example 2 is the second row. All eyes demonstrate a dark choroid and large nummular regions of hypoautofluorescence throughout the retinae.

Pathogenic variants in *RPGR* (MIM: 312610) were the most frequent cause of X-Linked RP, explaining over 88% (23/26) of sequenced pedigrees (Figure 3). Employment of a bespoke sequencing strategy of ORF15 has enhanced the success of sequencing this region, which typically presents a diagnostic challenge due to low coverage and poor variant detection [19].

3.3. Stargardt Disease and Other Macular Dystrophies

Stargardt disease (STGD1) is the second largest phenotypic presentation in this study (Figure 1). As mentioned previously, variants in *ABCA4* [31] were the most common cause, accounting for over 97% (109/112) of resolved STGD1 pedigrees in this cohort (Figure 6). Six potentially novel coding variants in *ABCA4* were detected as part of this study, NM_000350.2: c.1865G>A (p.Ser622Asn), c.223T>C (p.Cys75Arg), c.4468T>C (p.Cys1490Arg), c.5329A>G (p.Met1777Val), c.5351T>C (p.Leu1784Pro), and c.6743T>C (Phe2248Ser). All of these variants, with the exception of c.5351T>C (p.Leu1784Pro), were observed in individual pedigrees where another known pathogenic *ABCA4* variant was also observed. Variant c.5351T>C (p.Leu1784Pro) was observed in three individuals of two not knowingly related families also carrying a known pathogenic variant. The probands of these pedigrees do not appear to share any other rare variants in common in the regions captured.

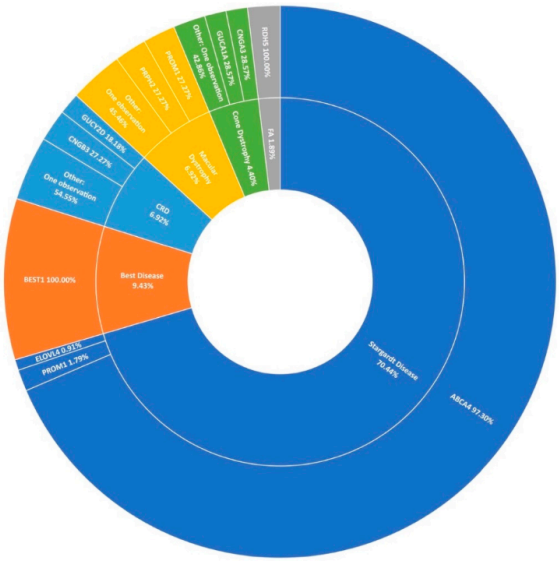


Figure 6. A summary of the genetic architecture of macular dystrophies in Target 5000 participants. CRD: Cone-Rod Dystrophy FA: Fundus Albipunctatus (Table S5).

It is becoming increasingly clear that intronic variants play a major role in the causation of STGD1. Nine pathogenic intronic variants were included in the capture panel of this study and found in 26 cases in 20 pedigrees with another *ABCA4* variant. In recent years, single-molecule molecular inversion probe (smMIP) based sequencing of the whole *ABCA4* gene including intronic sequences has dramatically increased positive candidate detection rates for STGD1 [32,33]. Thirty-six unresolved probands from Target 5000 presenting as STGD1 or cone-rod dystophy included in this cohort were analysed as part of a landmark study on the genetic landscape of ABCA4 in over 1000 probands, with a positive candidate detection rate of 44%. (16/36) [13]. This has increased the rate at which *ABCA4* is identified as the causal gene of STGD1 in this cohort (Figure 6). Of note from Irish participants in

3.6. Novel Variants

Here we present 19 novel missense variants that, to the best of our knowledge, have not yet been associated with an IRD (Table 2). In silico tools including MetaLR, M-CAP and REVEL were utilised to predict the pathogenicity of these variants. Segregation analysis was carried out when possible and was used as evidence according to the American College of Medical Genetics and Genomics (ACMG) guidelines where appropriate [21]. Two of the variants listed here, BBS10 c.155G>A (p.Gly52Asp) and PRPH2 c.464C>T (p.Thr155Ile), had been allocated dbSNP IDs; however, this is likely due to their detection in population sequencing studies, such as gnomAD [34], where they have not previously been associated with an IRD. Ten of these variants have been classified as variants of unknown significance (VUS) although having strong in silico predicted pathogenicities. This highlights the importance of functional analysis in vitro and in vivo in order to confidently call such variants likely pathogenic.

3.7. RPE65

In Ireland, the prevailing RPE65 phenotype has been that associated with the specific variant, (NM_000329.2) c.1430A>G (p.Asp477Gly). The disease phenotype is comparatively much milder than *RPE65*-LCA and closely resembles the clinical manifestations of choroideremia. This amino acid position has been shown to be highly conserved across multiple species [35]. This variant is one that continues to be detected in additional cases in the ongoing Target 5000 study. Initial reports and phenotype characterisation of this mutation were based on several pedigrees originating from Ireland [35,36]. The dominant c.1430A>G (p.Asp477Gly) variant remains to be probed for, yet is undetected, in large RPE65 screening studies (n > 2000) of non-Irish ethnicity [37]. Currently in the Target 5000 cohort, there are 23 genetically confirmed affected individuals, with many patients from these pedigrees that are pending recruitment. These 23 affected patients span 7 pedigrees although likely originate from a single source as observed in previous studies [38]. An example of segregation of the c.1430A>G (p.Asp477Gly) variant in one such pedigree is shown in Figure 9.

Genotype	Current Phenotype
+/-	Affected
+/-	Affected
+/-	Affected
+/-	Affected
+/-	Affected
+/-	Affected
+/+	Unaffected
+/+	Unaffected
+/+	Unaffected
+/+	Unaffected
+/+	Unaffected

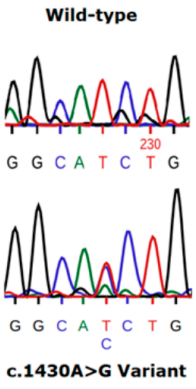


Figure 9. An illustration of patients harbouring the RPE65 c.1430A>G (p.D477G) variant. In this pedigree, each individual's phenotype corresponds to the expected genotype (left image). Also shown is a representative Sanger sequence trace from an unaffected (wild-type) family member (right image, top). The variant sequence shown below is that from an affected family member (right image, bottom). Sequences were read from the sense strand so the observed change above is T>G (GGCA[T>G]CTG).



Figure 8. A summary of the genetic architecture of less common phenotypes encountered in the Target 5000 study. LCA: Leber Congenital Amaurosis, LHON: Leber Hereditary Optic Neuropathy (Table S7).

Many other rare forms of IRD are also included in the Target 5000 study. Sixteen pedigrees with a diagnosis with Leber congenital amaurosis (LCA) have been genotyped with a spectrum of genes implicated as positive candidates (Figure 8). Other rare IRDs examined in this study include retinoschisis (n = 10), achromatopsia (n = 5), Stickler syndrome (n = 4), Leber hereditary optic neuropathy (n = 3), and optic atrophy (n = 2), among others. In total, 90 pedigrees with positive candidate variants are encompassed by Figure 8, illustrating the wide spectrum of disease presentation included in Target 5000.

3.8. FLVCR1

Typically, variants in FLVCR1 have been associated with a neurological syndrome, posterior column ataxia with retinitis pigmentosa (PCARP; MIM: 609033) [39–41], and more recently a specific splice variant (c.1092 + 5G>A) has been reported multiple times to be associated with non-syndromic RP [42–44]. Through Target 5000, substantial evidence has been obtained that suggests the first incidence of a protein coding FLVCR1 variant c.1022A>G (p.Tyr341Cys) implicated in non-syndromic RP [45]. RP is the most common clinical diagnosis for participants in the Target 5000 study, where the clinical diagnosis of RP accounts for nearly 40% (379/1004) of total pedigrees enrolled to date (Figure 1). Patients with this variant present with typical RP, without any extraocular features (Figure 10). Since its initial detection, this variant has been observed and deemed to segregate with the condition in three additional not knowingly related families in the Target 5000 cohort. Each of the genotyped affected individuals in these three pedigrees are homozygous for this variant.

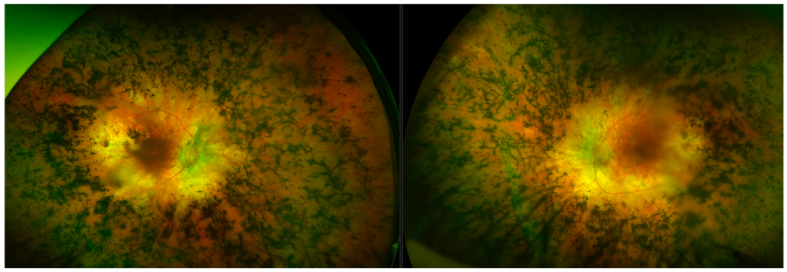


Figure 10. Montage of wide-field fundus images from a retinitis pigmentosa patient with a *FLVCR1* genotype. The *FLVCR1* genotype results in a phenotype (left- right eye, right- left eye) that appears to bear a resemblance to classical retinitis pigmentosa, with features such as masses of bony spicules in the periphery, attenuated blood vessels, and waxy disc pallor. There is also a relative preservation of the maculae.

3.9. Choroideremia

Choroideremia is an X-Linked recessive chorioretinal degenerative condition with progressive atrophy of various retinal cell types and the surrounding blood retinal barrier. Here we describe a novel deletion in the *CHM* gene found in two Irish pedigrees. This nearly 500 kb deletion represents the largest as yet detected IRD-associated gene deletion in Ireland (Figure 11). Two members of a large X-linked Retinitis Pigmentosa pedigree (Figure 12) clinically presented with choroideremia and tested negative for the segregating *RPGR* variant found in other affected members of this pedigree. Both males were analysed with target capture sequencing and found to possess large deletions spanning the *CHM* gene, approximating 500 kb. The observation of two IRDs in this pedigree highlights the significant value of NGS-based diagnostics for IRDs (Figure 12).

The same *CHM* deletion has also been detected in a second Irish pedigree since its initial discovery. Two additional males and two carrier females from this second pedigree were all found to be affected with progressive choroideremia. In this pedigree, the incidence of choroideremia could be traced back 5 generations, 4 of which have been assessed by the clinical team. A third Irish pedigree with a large *CHM* deletion was also detected previously as part of Target 5000 [27]. This mutation spanned approximately 6.5 kb and encompassed exons 3 and 4 of *CHM* (Figure 11). In each instance, target capture NGS detected the presence of the deletion and breakpoints were approximated by tiled PCR analyses. Interestingly, some probing PCRs in the intergenic regions surrounding *CHM* suggested that genomic DNA was present between deleted regions. Upon analysis, these PCR products were likely the result of large regions of homology that exist within this genomic area and other regions of the

genome. Several structural variants in this *CHM*-proximal intergenic region have also been observed in control samples (Figure 11).

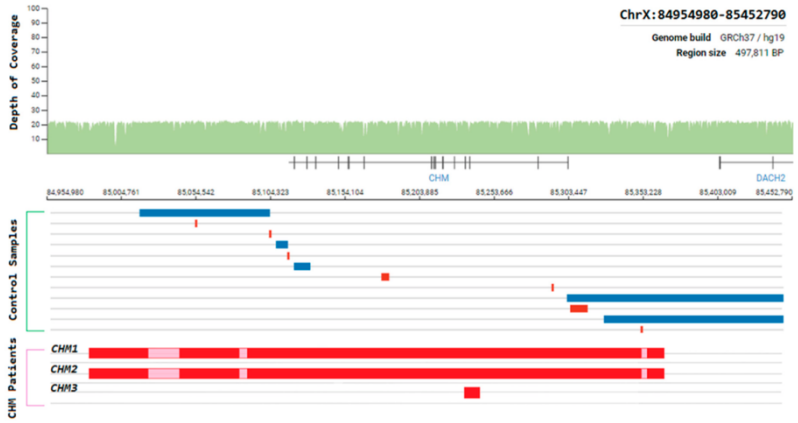


Figure 11. Schematic representation of the genomic region surrounding the CHM gene. Whole-genome coverage data from a control population database (https://gnomad.broadinstitute.org/). Structural variants detected from controls samples and Target 5000 patients are aligned to this region to illustrate the instability of this genomic region. Blue = duplication, Red = deletion, Pink = template positive results likely due to highly similar sequence elsewhere in the genome.

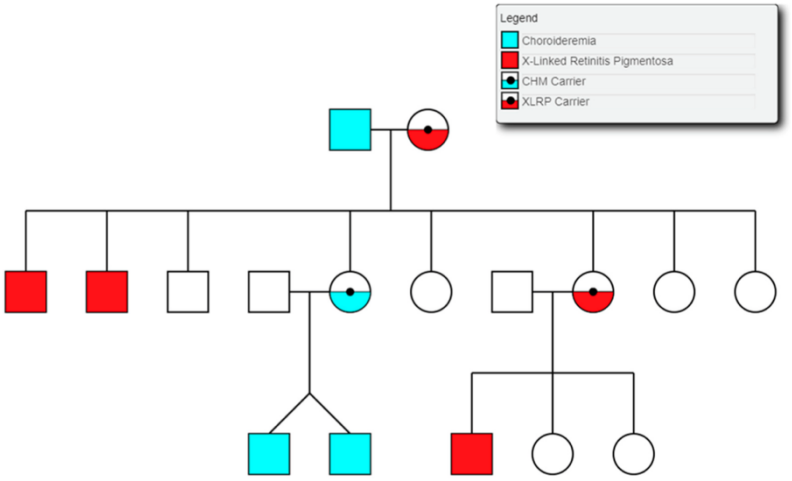


Figure 12. Partial pedigree tree illustrating the segregation of two X-linked retinal conditions within one family.

4. Discussion

Up to now, over 1000 individuals have been genotyped as part of Target 5000, accounting for over 20% of the estimated Irish IRD cohort. The results of this study thus far highlight the unique genetic architecture of IRDs in Ireland, cumulatively resulting in over 89 novel variants identified to

date [15,27], 19 of which were missense mutations in addition to 1 novel structural variant identified in the most recent analysis (Table 2, Figure 11). With a positive candidate detection rate of almost 70% (495/710 pedigrees), Target 5000 highlights the value of sequencing the exons of 254 IRD-associated genes as well as some known pathogenic intronic regions. This is consistent with other sequencing studies of various IRD cohorts, which report a diagnostic yield of between 38–75% [17,46–49]. This range can be expected due a number of factors. Firstly, deep-phenotyping to accurately categorise IRDs greatly assists with genotype-phenotype assessment. Assessor variability will exist between clinics and studies, contributing greatly to the variability of diagnostic yield. Secondly, there are differences between the numbers of genes and regions sequenced in these cohorts, where smaller gene panels will not capture some of the known genes associated with IRDs [46]. Lastly, the employment of additional analyses such as WES or WGS in other studies has increased the detection of candidate variants in previously unresolved cases [18,49–51]. The candidate variant detection rate of Target 5000 demonstrates that target capture sequencing is a cost and time-effective first-tier approach for genetic screening of those affected by IRDs in Ireland.

This study also highlights the importance of an accurate genetic diagnosis in addition to a clinical diagnosis. Genetic diagnoses continue to become more relevant as gene-based medicines move towards the forefront of IRD treatment given a first approved gene therapy for an IRD and an array of gene therapies for other forms of IRD in clinical and preclinical development. Furthermore, genetic diagnoses facilitate a better understanding of disease progression and manifestation for both patients and clinicians.

Stargardt disease (STGD1) is an autosomal recessive disorder caused almost exclusively by variants in the ATP-binding cassette subfamily A member 4 (*ABCA4*) gene (MIM: 601691) [52]. It is the most frequent form of macular dystrophy with an estimated prevalence of approximately 1 in 10,000 individuals [53]. STGD1 is characterised by progressive bilateral central vision loss, colour vision defects, delayed dark adaptation, and flecks in the retinal pigmentary epithelium [54]. A spectrum of disease severity underlies STGD1, with age of onset and progression governed by the specific combination of variants an individual carries, highlighting again the value of a specific genetic diagnosis for disease progression and risk assesment. An extensive list of 5962 likely pathogenic *ABCA4* variants in 3928 cases was published in 2017 [55], providing a valuable resource to aid in the interpretation of variants discovered as part of this study. In addition to coding variants captured, it has become increasingly clear that many of the unresolved one allele only (Figure 2) cases may likely carry a pathogenic intronic variant where the disease is typically associated with a recessive inheritance pattern. This has been elegantly demonstrated by deep intronic variants in *ABCA4* causing STGD1, such as c.4539+2028C>T (p.[=,Arg1514Leufs*36]). This variant has been observed as a candidate in 10 Target 5000 individuals, of which 5 were disovered as part of a landmark study of *ABCA4*-linked STGD1 employing smMIPs based whole gene sequencing [13] and 5 of which were discovered through target capture sequencing runs. This highlights a significant enrichment of this variant in the Irish IRD patient cohort. c.4539+2028C>T (p.[=,Arg1514Leufs*36]) was previously functionally analysed in photoreceptor precursor cells, and was shown to result in a 345 nucleotide pseudoexon inclusion due to strengthening of exonic splice enhancers. It is notable that such variants are an attractive target for antisense oligonucleotide-based splice correction therapy [56].

Identified variants proximal to the *RHO* p.Met207 region appear predominant in our sub-cohort of autosomal dominant RP patients, as reported in Section 3.2. Amino acids in this region of the protein form part of the fifth transmembrane domain of rhodopsin. This domain has been shown to be vital to retinal binding in bovine rhodopsin [57]. It is possible that undesirable changes in the protein folding of this region result in steric hindrance that prevents the apoprotein from successfully binding retinal. This impaired binding capacity may possibly underpin the reason for the prevalence of pathogenic variants found in the coding region of this protein domain.

Feline leukaemia virus subgroup C cellular receptor 1 (*FLVCR1*; MIM: 609144) is a transmembrane protein involved in erythropoiesis and heme transport [58]. *FLVCR1* was first discovered for its role in

aplastic anaemia in domestic cats and subsequently erythroblast destruction in vitro [59]. This was also the first study that theorised that *FLVCR1* was a receptor for an organic anion and identified its component domains as strikingly similar to other ancient Major Facilitator Superfamily (MFS) members. In a later study it was shown this receptor protein provided the first description of a mammalian heme transporter [60]. The same group proceeded to investigate its effect in knockout mouse models, which resulted in embryonic lethality [58]. Candidate variant detection in *FLVCR1* has significantly increased in recently recruited pedigrees. It is notable the c.1022A>G (p.Tyr341Cys) variant is only the second disease-associated variant to be found in *FLVCR1* associated with RP without posterior column degeneration [43]. Moreover, it is the first protein coding variant found in this gene to be affiliated with non-syndromic RP [45]. There were no extraocular features associated with this variant (Figure 10). It is possible that the distinct families in this cohort with this variant share a common ancestry. However, they do not share any rare variants in the regions captured and originate from distinct geographical locations in Ireland. The age of onset for the symptoms of ataxia in PCARP has been reported as typically in the third decade of life. Of note in the Target 5000 cohort, all patients carrying this variant have reached this age, with the oldest patient currently in their seventh decade of life. It is unclear yet as to whether the disease pathology associated with this variant will remain completely non-syndromic throughout the entirety of a patient's lifespan or will result in a milder, later onset of additional symptoms compared to other pathogenic variants found in this gene. Fortunately, the small size of this gene (2.6 kb) in principle makes it suitable for inclusion in AAV vectors for use as a gene therapy. Such a therapy may help to alleviate the toxic effects of intracellular free-heme [39]. The serotype of AAV could be chosen based on the presence or absence of systemic phenotypes. For example, if only the retina was to be targeted, an AAV 2/5 or 2/8 might be effective serotypes; on the other hand, if the whole central nervous system was to be targeted, then AAV9 or AAVB1 would be more valuable [61]. Many additional bespoke serotypes are now emerging as a result of site-directed evolution of AAV serotypes [62–65]. Given the recent results here on the role of FLVCR1 in this form of non-syndromic RP, FLVCR1 along with other IRD genes, becomes an increasingly interesting candidate for exploration of AAV-mediated gene therapies.

The RPE65 enzyme is active in the retinoid cycle that recycles retinoids in the RPE, which are utilised by photoreceptor cells. Mutations in the RPE65 gene were initially found to be causative of some cases of autosomal recessive LCA, a severe and clinically distinct IRD [66–68]. This was subsequently expanded to include a milder, dominant form of retinopathy and more recently, to include less severe autosomal recessive retinopathies [35,37]. Luxturna, the first ocular gene therapy approved by the FDA, is for biallelic RPE65 retinal disease [16]. RPE65 c.1430A>G (p.Asp477Gly) remains the predominant RPE65 genotype observed in the Irish population. This variant has since been detected in other population studies alongside the clinical presentation of similar phenotypes, largely described as a choroideremia phenocopy [36,38]. However, until recently the exact disease mechanism caused by this variant has evaded researchers. It was initially believed that this variant produced an abnormal protein structure that may affect enzymatic function [35]. More recently, it has been shown in a mouse model that although this variant is not located in a critical functional domain, the variant induces sufficient change to alter the physiochemical properties of the p.D477 loop to produce an aggregation-prone surface. This surface gains the aberrant function of enabling abnormal protein-protein interactions [69]. One hypothesised type of interaction is with ubiquitin ligases, which would likely result in proteasomal degradation as seen in several other RPE65 mutations [70]. This theory may support the variability observed phenotypically as severity might then become dependent on several other factors such as ubiquitination rates and proteasome response. It is of note that aberrant RNA splicing has been shown to have a major pathogenic effect in a knock-in mouse model with this variant [71]. Given the success of the gene therapy treatment of biallelic RPE65 retinopathies with Luxturna [16], an AAV-based therapy, it is interesting to speculate whether the therapy might have any utility for monoallelic dominant RPE65 cases.

The *CHM* gene hosts the largest collection of pathogenic structural variants detected in any IRD-related gene as part of the Target 5000 study to date. The presence of several structural variants in the general population may also indicate a natural instability in this genomic region. Additionally, female carriers of *CHM* mutations typically show mild stationary signs with no symptoms, while males are severely affected. In this instance, some females were more severely affected than expected with advanced signs of degeneration and progressive visual decline. One female carrier showed radial pigmentation at the retinal pigment epithelium level and scattered choroidal atrophy, whereas another showed more marked reactive pigmentation and extensive symmetrical atrophy of the choroid and outer retina. These manifestations were apparent both anatomically and functionally. A more severe phenotype was also associated with whole *CHM* gene deletion in a previous genotype-phenotype choroideremia study [72]. However, in this study, the phenotype was described as choroideremia plus, due to additional syndromic features. These additional symptoms, such as deafness, were attributed to the extent of the deletion encompassing several surrounding genes, which is not the case for the whole gene deletion patients described here.

Going forward with the Target 5000 study, more efforts will be focused on the detection of additional variants in unresolved IRD cases and also in the functional assessment of variants of unknown significance identified thus far. In the age of budding emerging gene therapies, it has never been more important to stringently identify the pathogenic variants that underpin a condition. It is therefore vital that the variants located outside of exonic regions causing disease pathologies are detected, and that these participants are also considered for treatment, should an applicable gene therapy become available in the future. A whole exome or whole genome approach will be more extensively employed in the future to increase the identification of novel variants. Recent studies have highlighted that molecular diagnostic yield is increased when WGS is employed by comparison with other methods [51]. It is expected that for many of the currently unresolved cases, pathogenic variants may be present in deep-intronic, promoter, enhancer regions or in genes that are not yet associated with an IRD [73,74]. Recent studies such as that in STGD1, highlighting the relative prevalence of deep intronic mutations in the *ABCA4* gene, attest to the need for more extensive sequence analysis [13]. These regions have been very poorly covered in patients screened by target capture protocols, unless specifically designed to do so. Additionally, WGS data will offer higher resolution in terms of data associated with copy number abnormalities and large structural rearrangements. It has been reported that there are over 1300 CNVs associated with IRD genes reported in the literature [75]. Exon-based target capture methods have a significantly reduced capacity to detect CNVs and SVs. This is due in part to the range of read-depth at captured regions and the inability to detect breakpoints that occur outside of captured loci. Notably, similarly targeted methods, such as non-WES/WGS studies, which have specifically probed intronic and exonic regions of genes have been shown to successfully detect large structural variants [13]. This illustrates that the enormous data volumes generated by WGS studies may be surplus to requirement if the SVs/CNVs to be searched for are relevant to a small genomic region.

Table 2 illustrates that many variants with seemingly strong evidence of pathogenicity may still be classified as VUS. Variants that are currently of unknown significance likely to alter enzyme function, such as those in *RPE65* [76], may be applicable to assessment in vitro by employing enzyme function assays. Alternatively, if predicted to perturb splicing having undergone in silico interrogation [77], these variants could be assessed by midigene analysis. Midigene analysis refers to the incorporation of variants of interest into vectors containing several exons of the relevant IRD gene, enabling the in vitro interrogation of the effect(s) of that variant directly at an RNA level. This approach is particularly effective for testing exonic and intronic variants that are likely to disrupt correct splicing, as these variants will have a distinct impact on the RNA product when compared to the wild-type equivalent. This method is also particularly useful for the interpretation of variants that are located within introns, yet outside of canonical splice sites, and additionally to examine rare synonymous exonic variants proximal to exon junctions [32,78].

A viable in vivo strategy for the assessment of novel missense variants is to introduce these VUS into model organisms with similar or conserved genes. The significant advantages of such model organisms is the ability to rapidly generate models of disease [79]. This is particularly relevant currently, as recent advances in genome editing technologies has made this approach more readily accessible and successful than ever before [80]. Although not all model organisms fully represent the visual system that exists in humans, many organisms can offer useful information ranging from protein mislocalisation [81] to optokinetic response [82]. By introducing a VUS into the appropriate system, additional evidence can be gained to better evaluate the functional impact of the VUS in human ocular pathologies.

Thus far, genetic analysis of IRD patients has helped to resolve ambiguous phenotypes and to identify causative mutations in nearly 70% (495/710) of sequenced pedigrees. The continuous expansion of our cohort has enabled us to better interrogate the sequencing data and interpret the potential pathogenicity of novel variants when detected. In addition to this, the growing body of data from NGS studies of IRDs globally should facilitate better correlations between genotype and phenotype and further refine methods for diagnoses and prognoses. Furthermore, whole gene and WGS analyses are highlighting the significant role of non-coding variants as causative of some IRDs. Target 5000 aims to provide actionable outcomes empowering patients with genetic diagnoses and potentially future access to clinical trials or approved treatments where appropriate. Given the rapid development of the field of ocular therapeutics, it is clear that genetically characterising those affected by an IRD has become a diagnostic imperative of the utmost importance.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/11/1/105/s1, Table S1: Primers used to probe exonic regions of CHM (MIM: 300390; ChrX: 85861180-86047558) and the neighbouring genes, POF1B (MIM: 300603; ChrX: 85277396-85379709) and DACH2 (MIM: 300608; ChrX: 86148451-86832602) as well as the intergenic space between. Table S2: Tabular summary of the clinical presentation of 1004 recruited Target 5000 pedigrees. RP: Retinitis Pigmentosa, FFM: Fundus Flavimaculatus, MD: Macular Dystrophy, CRD: Cone-Rod Dystrophy, RCD: Rod-Cone Dystrophy, LCA: Leber Congenital Amaurosis, EOSRD: Early-Onset Retinal Degeneration, BBS: Bardet-Biedl Syndrome, CSNB: Congenital Stationary Night Blindness, Table S3: Tabular summary of the diagnostic yield rates for 710 Target 5000 pedigrees utilising target capture next-generation sequencing of the exonic regions of over 250 genes and previously identified pathogenic intronic variants, Table S4: Tabular summar of the genetic architecture of retinitis pigmentosa (RP) in Target 5000 participants, Table S5: Tabular summary of the genetic architecture of macular dystrophies in Target 5000 participants. CRD: Cone-Rod Dystrophy FA: Fundus Albipunctatus, Table S6: Tabular summary of the genetic architecture of Usher Syndrome (USH) in Target 5000 participants, Table S7: Tabular summary of the genetic architecture of less common phenotypes encountered in the Target 5000 study. LCA: Leber Congenital Amaurosis, LHON: Leber Hereditary Optic Neuropathy, Figure S1: The available age distribution of recruited Target 5000 participants, Figure S2: Sequencing coverage from three affected X-linked retinitis pigmentosa individuals (H41, P1 and K1) from three distinct pedigrees as seen on IGV software. Alligned to this region is sequencing primers (R4,R7,R8,R9,F10 and R5) and PCR primers (F3,R6 and P9) as shown by forward (blue) and reverse (red) orientations.

Author Contributions: Conceptualization, P.H., D.K., P.K. and G.J.F.; Data curation, L.W. and A.D.; Formal analysis, L.W. and A.D.; Funding acquisition, G.J.F.; Investigation, L.W., A.D., N.W., J.Z., K.S., G.S., J.T., J.J.O., D.K., P.F.K. and G.J.F.; Methodology, A.D., M.C. and G.J.F.; Project administration, G.J.F.; Resources, G.J.F.; Software, A.D. and M.C.; Supervision, A.D., M.C. and G.J.F.; Validation, L.W. and A.D.; Visualization, L.W. and A.D.; Writing—original draft, L.W. and A.D.; Writing—review & editing, L.W. and A.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by grant awards from Fighting Blindness Ireland (FB Irl; FB16FAR, FB18CRE), The Health Research Board of Ireland (HRB; POR/2010/97) in conjunction with Health Research Charities Ireland (HRCI; MRCG-2013-8, MRCG-2016-14), the Irish Research Council (IRC; GOIPG/2017/1631), and Science Foundation Ireland (SFI; 16/1A/4452).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mei, S.; Huang, X.; Cheng, L.; Peng, S.; Zhu, T.; Chen, L.; Wang, Y.; Zhao, J. A missense mutation in *OPA1* causes dominant optic atrophy in a Chinese family. *J. Ophthalmol.* **2019**, *2019*, 1424928. [CrossRef]
2. Gao, F.-J.; Zhang, S.-H.; Chen, J.-Y.; Xu, G.-Z.; Wu, J.-H. Digenic heterozygous mutations in *EYS/LRP5* in a Chinese family with retinitis pigmentosa. *Int. J. Ophthalmol.* **2017**, *10*, 325–328.

3. Bhattacharya, S.S.; Wright, A.F.; Clayton, J.F.; Price, W.H.; Phillips, C.I.; McKeown, C.M.; Jay, M.; Bird, A.C.; Pearson, P.L.; Southern, E.M. Close genetic linkage between X-linked retinitis pigmentosa and a restriction fragment length polymorphism identified by recombinant DNA probe L1.28. *Nature* **1984**, *309*, 253–255. [CrossRef]
4. Farrar, G.J.; Kenna, P.; Jordan, S.A.; Kumar-Singh, R.; Humphries, M.M.; Sharp, E.M.; Sheils, D.M.; Humphries, P. A three-base-pair deletion in the peripherin–*RDS* gene in one form of retinitis pigmentosa. *Nature* **1991**, *354*, 478–480. [CrossRef]
5. Kumar-Singh, R.; Bradley, D.G.; Farrar, G.J.; Lawler, M.; Jordan, S.A.; Humphries, P. Autosomal dominant retinitis pigmentosa: A new multi-allelic marker (D3S621) genetically linked to the disease locus (RP4). *Hum. Genet.* **1991**, *86*, 502–504. [CrossRef] [PubMed]
6. RetNet: Summaries. Available online: https://sph.uth.edu/retnet/sum-dis.htm (accessed on 29 October 2019).
7. Werdich, X.Q.; Place, E.M.; Pierce, E.A. Systemic diseases associated with retinal dystrophies. *Semin. Ophthalmol.* **2014**, *29*, 319–328. [CrossRef] [PubMed]
8. Ferrari, S.; Di Iorio, E.; Barbaro, V.; Ponzin, D.; Sorrentino, F.S.; Parmeggiani, F. Retinitis pigmentosa: Genes and disease mechanisms. *Curr. Genomics* **2011**, *12*, 238–249. [PubMed]
9. Farrar, G.J.; Carrigan, M.; Dockery, A.; Millington-Ward, S.; Palfi, A.; Chadderton, N.; Humphries, M.; Kiang, A.S.; Kenna, P.F.; Humphries, P. Toward an elucidation of the molecular genetics of inherited retinal degenerations. *Hum. Mol. Genet.* **2017**, *26*, R2–R11. [CrossRef]
10. Cremers, F.P.; van de Pol, D.J.; van Driel, M.; den Hollander, A.I.; van Haren, F.J.; Knoers, N.V.; Tijmes, N.; Bergen, A.A.; Rohrschneider, K.; Blankenagel, A.; et al. Autosomal recessive retinitis pigmentosa and cone-rod dystrophy caused by splice site mutations in the Stargardt's disease gene *ABCR*. *Hum. Mol. Genet.* **1998**, *7*, 355–362. [CrossRef]
11. Maugeri, A.; Klevering, B.J.; Rohrschneider, K.; Blankenagel, A.; Brunner, H.G.; Deutman, A.F.; Hoyng, C.B.; Cremers, F.P.M. Mutations in the *ABCA4* (*ABCR*) gene are the major cause of autosomal recessive cone-rod dystrophy. *Am. J. Hum. Genet.* **2000**, *67*, 960–966. [CrossRef]
12. Maugeri, A.; van Driel, M.A.; van de Pol, D.J.; Klevering, B.J.; van Haren, F.J.; Tijmes, N.; Bergen, A.A.; Rohrschneider, K.; Blankenagel, A.; Pinckers, A.J.; et al. The 2588G→C mutation in the *ABCR* gene is a mild frequent founder mutation in the Western European population and allows the classification of *ABCR* mutations in patients with Stargardt disease. *Am. J. Hum. Genet.* **1999**, *64*, 1024–1035. [CrossRef] [PubMed]
13. Khan, M.; Cornelis, S.S.; Pozo-Valero, M.; del Whelan, L.; Runhart, E.H.; Mishra, K.; Bults, F.; AlSwaiti, Y.; AlTabishi, A.; Baere, E.D.; et al. Resolving the dark matter of *ABCA4* for 1054 Stargardt disease probands through integrated genomics and transcriptomics. *bioRxiv* **2019**, 817767. [CrossRef]
14. Kim, D.Y.; Mukai, S. X-linked Juvenile Retinoschisis (XLRS): A review of genotype-phenotype relationships. *Semin. Ophthalmol.* **2013**, *28*, 392–396. [CrossRef] [PubMed]
15. Dockery, A.; Stephenson, K.; Keegan, D.; Wynne, N.; Silvestri, G.; Humphries, P.; Kenna, P.F.; Carrigan, M.; Farrar, G.J. Target 5000: Target capture sequencing for inherited retinal degenerations. *Genes* **2017**, *8*, 304. [CrossRef] [PubMed]
16. Maguire, A.M.; Russell, S.; Wellman, J.A.; Chung, D.C.; Yu, Z.-F.; Tillman, A.; Wittes, J.; Pappas, J.; Elci, O.; Marshall, K.A.; et al. Efficacy, Safety, and Durability of Voretigene Neparvovec-rzyl in RPE65 Mutation-Associated Inherited Retinal Dystrophy: Results of Phase 1 and 3 Trials. *Ophthalmology* **2019**, *126*, 1273–1285. [CrossRef]
17. Consugar, M.B.; Navarro-Gomez, D.; Place, E.M.; Bujakowska, K.M.; Sousa, M.E.; Fonseca-Kelly, Z.D.; Taub, D.G.; Janessian, M.; Wang, D.Y.; Au, E.D.; et al. Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing. *Genet. Med.* **2015**, *17*, 253–261. [CrossRef]
18. Carss, K.J.; Arno, G.; Erwood, M.; Stephens, J.; Sanchis-Juan, A.; Hull, S.; Megy, K.; Grozeva, D.; Dewhurst, E.; Malka, S.; et al. Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. *Am. J. Hum. Genet.* **2017**, *100*, 75–90. [CrossRef]
19. Li, J.; Tang, J.; Feng, Y.; Xu, M.; Chen, R.; Zou, X.; Sui, R.; Chang, E.Y.; Lewis, R.A.; Zhang, V.W.; et al. Improved diagnosis of inherited retinal dystrophies by high-fidelity PCR of ORF15 followed by next-generation sequencing. *J. Mol. Diagn.* **2016**, *18*, 817–824. [CrossRef]
20. Boyle, E.A.; O'Roak, B.J.; Martin, B.K.; Kumar, A.; Shendure, J. MIPgen: Optimized modeling and design of molecular inversion probes for targeted resequencing. *Bioinformatics* **2014**, *30*, 2670–2672. [CrossRef]

21. Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody, W.W.; Hegde, M.; Lyon, E.; Spector, E.; et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **2015**, *17*, 405–424. [CrossRef]
22. Jarvik, G.P.; Browning, B.L. Consideration of cosegregation in the pathogenicity classification of genomic variants. *Am. J. Hum. Genet.* **2016**, *98*, 1077–1081. [CrossRef] [PubMed]
23. Romanet, P.; Odou, M.-F.; North, M.-O.; Saveanu, A.; Coppin, L.; Pasmant, E.; Mohamed, A.; Goudet, P.; Borson-Chazot, F.; Calender, A.; et al. Proposition of adjustments to the ACMG-AMP framework for the interpretation of *MEN1* missense variants. *Hum. Mutat.* **2019**, *40*, 661–674. [CrossRef] [PubMed]
24. Ioannidis, N.M.; Rothstein, J.H.; Pejaver, V.; Middha, S.; McDonnell, S.K.; Baheti, S.; Musolf, A.; Li, Q.; Holzinger, E.; Karyadi, D.; et al. REVEL: An ensemble method for predicting the pathogenicity of rare missense variants. *Am. J. Hum. Genet.* **2016**, *99*, 877–885. [CrossRef] [PubMed]
25. Jian, X.; Liu, X. In silico prediction of deleteriousness for nonsynonymous and splice-altering single nucleotide variants in the human genome. In *Vitro Mutagenesis*; Reeves, A., Ed.; Springer: New York, NY, USA, 2017; Volume 1498, pp. 191–197. ISBN 978-1-4939-6470-3.
26. Chen, X.; Schulz-Trieglaff, O.; Shaw, R.; Barnes, B.; Schlesinger, F.; Källberg, M.; Cox, A.J.; Kruglyak, S.; Saunders, C.T. Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **2016**, *32*, 1220–1222. [CrossRef]
27. Carrigan, M.; Duignan, E.; Malone, C.P.G.; Stephenson, K.; Saad, T.; McDermott, C.; Green, A.; Keegan, D.; Humphries, P.; Kenna, P.F.; et al. Panel-based population next-generation sequencing for inherited retinal degenerations. *Sci. Rep.* **2016**, *6*, 33248. [CrossRef]
28. Sullivan, L.S.; Bowne, S.J.; Birch, D.G.; Hughbanks-Wheaton, D.; Heckenlively, J.R.; Lewis, R.A.; Garcia, C.A.; Ruiz, R.S.; Blanton, S.H.; Northrup, H.; et al. Prevalence of disease-causing mutations in families with autosomal dominant retinitis pigmentosa: A screen of known genes in 200 families. *Investig. Ophthalmol. Vis. Sci.* **2006**, *47*, 3052–3064. [CrossRef]
29. Lenassi, E.; Vincent, A.; Li, Z.; Saihan, Z.; Coffey, A.J.; Steele-Stallard, H.B.; Moore, A.T.; Steel, K.P.; Luxon, L.M.; Héon, E.; et al. A detailed clinical and molecular survey of subjects with nonsyndromic *USH2A* retinopathy reveals an allelic hierarchy of disease-causing variants. *Eur. J. Hum. Genet.* **2015**, *23*, 1318–1327. [CrossRef]
30. Klevering, B.J.; Yzer, S.; Rohrschneider, K.; Zonneveld, M.; Allikmets, R.; van den Born, L.I.; Maugeri, A.; Hoyng, C.B.; Cremers, F.P.M. Microarray-based mutation analysis of the *ABCA4 (ABCR)* gene in autosomal recessive cone-rod dystrophy and retinitis pigmentosa. *Eur. J. Hum. Genet.* **2004**, *12*, 1024–1032. [CrossRef]
31. Allikmets, R.; Singh, N.; Sun, H.; Shroyer, N.F.; Hutchinson, A.; Chidambaram, A.; Gerrard, B.; Baird, L.; Stauffer, D.; Peiffer, A.; et al. A photoreceptor cell-specific ATP-binding transporter gene (*ABCR*) is mutated in recessive Stargardt macular dystrophy. *Nat. Genet.* **1997**, *15*, 236–246. [CrossRef]
32. Sangermano, R.; Garanto, A.; Khan, M.; Runhart, E.H.; Bauwens, M.; Bax, N.M.; van den Born, L.I.; Khan, M.I.; Cornelis, S.S.; Verheij, J.B.G.M.; et al. Deep-intronic *ABCA4* variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. *Genet. Med.* **2019**, *21*, 1751–1760. [CrossRef]
33. Khan, M.; Cornelis, S.S.; Khan, M.I.; Elmelik, D.; Manders, E.; Bakker, S.; Derks, R.; Neveling, K.; Vorst, M.; van de Gilissen, C.; et al. Cost-effective molecular inversion probe-based *ABCA4* sequencing reveals deep-intronic variants in Stargardt disease. *Hum. Mutat.* **2019**, *40*, 1749–1759. [CrossRef] [PubMed]
34. GnomAD. Available online: https://gnomad.broadinstitute.org/ (accessed on 14 November 2019).
35. Bowne, S.J.; Humphries, M.M.; Sullivan, L.S.; Kenna, P.F.; Tam, L.C.S.; Kiang, A.S.; Campbell, M.; Weinstock, G.M.; Koboldt, D.C.; Ding, L.; et al. A dominant mutation in *RPE65* identified by whole-exome sequencing causes retinitis pigmentosa with choroidal involvement. *Eur. J. Hum. Genet.* **2011**, *19*, 1074–1081. [CrossRef] [PubMed]
36. Hull, S.; Mukherjee, R.; Holder, G.E.; Moore, A.T.; Webster, A.R. The clinical features of retinal disease due to a dominant mutation in *RPE65*. *Mol. Vis.* **2016**, *22*, 626–635.
37. Li, S.; Xiao, X.; Yi, Z.; Sun, W.; Wang, P.; Zhang, Q. RPE65 mutation frequency and phenotypic variation according to exome sequencing in a tertiary centre for genetic eye diseases in China. *Acta Ophthalmol.* **2019**. [CrossRef] [PubMed]

38. Jauregui, R.; Park, K.S.; Tsang, S.H. Two-year progression analysis of *RPE65* autosomal dominant retinitis pigmentosa. *Ophthalmic Genet.* **2018**, *39*, 544–549. [CrossRef] [PubMed]
39. Chiabrando, D.; Castori, M.; di Rocco, M.; Ungelenk, M.; Gießelmann, S.; Di Capua, M.; Madeo, A.; Grammatico, P.; Bartsch, S.; Hübner, C.A.; et al. Mutations in the heme exporter *FLVCR1* cause sensory neurodegeneration with loss of pain perception. *PLoS Genet.* **2016**, *12*, e1006461. [CrossRef]
40. Castori, M.; Morlino, S.; Ungelenk, M.; Pareyson, D.; Salsano, E.; Grammatico, P.; Tolosano, E.; Kurth, I.; Chiabrando, D. Posterior column ataxia with retinitis pigmentosa coexisting with sensory-autonomic neuropathy and leukemia due to the homozygous p.Pro221Ser *FLVCR1* mutation. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **2017**, *174*, 732–739. [CrossRef]
41. Ishiura, H.; Fukuda, Y.; Mitsui, J.; Nakahara, Y.; Ahsan, B.; Takahashi, Y.; Ichikawa, Y.; Goto, J.; Sakai, T.; Tsuji, S. Posterior column ataxia with retinitis pigmentosa in a Japanese family with a novel mutation in *FLVCR1*. *Neurogenetics* **2011**, *12*, 117–121. [CrossRef]
42. Glöckle, N.; Kohl, S.; Mohr, J.; Scheurenbrand, T.; Sprecher, A.; Weisschuh, N.; Bernd, A.; Rudolph, G.; Schubach, M.; Poloschek, C.; et al. Panel-based next generation sequencing as a reliable and efficient technique to detect mutations in unselected patients with retinal dystrophies. *Eur. J. Hum. Genet.* **2014**, *22*, 99–104. [CrossRef]
43. Yusuf, I.H.; Shanks, M.E.; Clouston, P.; MacLaren, R.E. A splice-site variant in *FLVCR1* produces retinitis pigmentosa without posterior column ataxia. *Ophthalmic Genet.* **2018**, *39*, 263–267. [CrossRef]
44. Tiwari, A.; Bahr, A.; Bähr, L.; Fleischhauer, J.; Zinkernagel, M.S.; Winkler, N.; Barthelmes, D.; Berger, L.; Gerth-Kahlert, C.; Neidhardt, J.; et al. Next generation sequencing based identification of disease-associated mutations in Swiss patients with retinal dystrophies. *Sci. Rep.* **2016**, *6*, 28755. [CrossRef] [PubMed]
45. Dockery, A.; Carrigan, M.; Wynne, N.; Stephenson, K.; Keegan, D.; Kenna, P.F.; Farrar, G.J. A novel *FLVCR1* variant implicated in retinitis pigmentosa. In *Retinal Degenerative Diseases*; Advances in Experimental Medicine and Biology; Springer: New York, NY, USA, 2019; Volume 1185.
46. Patel, A.; Hayward, J.D.; Tailor, V.; Nyanhete, R.; Ahlfors, H.; Gabriel, C.; Jannini, T.B.; Abbou-Rayyah, Y.; Henderson, R.; Nischal, K.K.; et al. The Oculome panel test: Next-generation sequencing to diagnose a diverse range of genetic developmental eye disorders. *Ophthalmology* **2019**, *126*, 888–907. [CrossRef] [PubMed]
47. Stone, E.M.; Andorf, J.L.; Whitmore, S.S.; DeLuca, A.P.; Giacalone, J.C.; Streb, L.M.; Braun, T.A.; Mullins, R.F.; Scheetz, T.E.; Sheffield, V.C.; et al. Clinically focused molecular investigation of 1000 consecutive families with inherited retinal disease. *Ophthalmology* **2017**, *124*, 1314–1331. [CrossRef] [PubMed]
48. Riera, M.; Navarro, R.; Ruiz-Nogales, S.; Méndez, P.; Burés-Jelstrup, A.; Corcóstegui, B.; Pomares, E. Whole exome sequencing using Ion Proton system enables reliable genetic diagnosis of inherited retinal dystrophies. *Sci. Rep.* **2017**, *7*, 42078. [CrossRef]
49. Sharon, D.; Ben-Yosef, T.; Goldenberg-Cohen, N.; Pras, E.; Gradstein, L.; Soudry, S.; Mezer, E.; Zur, D.; Abbasi, A.H.; Zeitz, C.; et al. A nationwide genetic analysis of inherited retinal diseases in Israel as assessed by the Israeli inherited retinal disease consortium (IIRDC). *Hum. Mutat.* **2020**, *41*, 140–149. [CrossRef]
50. Haer-Wigman, L.; van Zelst-Stams, W.A.; Pfundt, R.; van den Born, L.I.; Klaver, C.C.; Verheij, J.B.; Hoyng, C.B.; Breuning, M.H.; Boon, C.J.; Kievit, A.J.; et al. Diagnostic exome sequencing in 266 Dutch patients with visual impairment. *Eur. J. Hum. Genet.* **2017**, *25*, 591–599. [CrossRef]
51. Ellingford, J.M.; Barton, S.; Bhaskar, S.; Williams, S.G.; Sergouniotis, P.I.; O'Sullivan, J.; Lamb, J.A.; Perveen, R.; Hall, G.; Newman, W.G.; et al. Whole genome sequencing increases molecular diagnostic yield compared with current diagnostic testing for inherited retinal disease. *Ophthalmology* **2016**, *123*, 1143–1150. [CrossRef]
52. Allikmets, R.; Shroyer, N.F.; Singh, N.; Seddon, J.M.; Lewis, R.A.; Bernstein, P.S.; Peiffer, A.; Zabriskie, N.A.; Li, Y.; Hutchinson, A.; et al. Mutation of the Stargardt disease gene (*ABCR*) in age-related macular degeneration. *Science* **1997**, *277*, 1805–1807. [CrossRef]
53. Strauss, R.W.; Ho, A.; Muñoz, B.; Cideciyan, A.V.; Sahel, J.-A.; Sunness, J.S.; Birch, D.G.; Bernstein, P.S.; Michaelides, M.; Traboulsi, E.I.; et al. The natural history of the progression of atrophy secondary to Stargardt disease (ProgStar) studies: Design and baseline characteristics: ProgStar Report No. 1. *Ophthalmology* **2016**, *123*, 817–828. [CrossRef]
54. Fishman, G.A.; Farbman, J.S.; Alexander, K.R. Delayed rod dark adaptation in patients with Stargardt's disease. *Ophthalmology* **1991**, *98*, 957–962. [CrossRef]

55. Cornelis, S.S.; Bax, N.M.; Zernant, J.; Allikmets, R.; Fritsche, L.G.; den Dunnen, J.T.; Ajmal, M.; Hoyng, C.B.; Cremers, F.P.M. In silico functional meta-analysis of 5,962 *ABCA4* variants in 3,928 retinal dystrophy cases. *Hum. Mutat.* **2017**, *38*, 400–408. [CrossRef]
56. Albert, S.; Garanto, A.; Sangermano, R.; Khan, M.; Bax, N.M.; Hoyng, C.B.; Zernant, J.; Lee, W.; Allikmets, R.; Collin, R.W.J.; et al. Identification and rescue of splice defects caused by two neighboring deep-intronic *ABCA4* mutations underlying Stargardt disease. *Am. J. Hum. Genet.* **2018**, *102*, 517–527. [CrossRef] [PubMed]
57. Li, J.; Edwards, P.C.; Burghammer, M.; Villa, C.; Schertler, G.F.X. Structure of bovine rhodopsin in a trigonal crystal form. *J. Mol. Biol.* **2004**, *343*, 1409–1438. [CrossRef] [PubMed]
58. Keel, S.B.; Doty, R.T.; Yang, Z.; Quigley, J.G.; Chen, J.; Knoblaugh, S.; Kingsley, P.D.; De Domenico, I.; Vaughn, M.B.; Kaplan, J.; et al. A heme export protein is required for red blood cell differentiation and iron homeostasis. *Science* **2008**, *319*, 825–828. [CrossRef] [PubMed]
59. Tailor, C.S.; Willett, B.J.; Kabat, D. A putative cell surface receptor for anemia-inducing feline leukemia virus subgroup C is a member of a transporter superfamily. *J. Virol.* **1999**, *73*, 6500–6505. [CrossRef] [PubMed]
60. Quigley, J.G.; Yang, Z.; Worthington, M.T.; Phillips, J.D.; Sabo, K.M.; Sabath, D.E.; Berg, C.L.; Sassa, S.; Wood, B.L.; Abkowitz, J.L. Identification of a human heme exporter that is essential for erythropoiesis. *Cell* **2004**, *118*, 757–766. [CrossRef] [PubMed]
61. Choudhury, S.R.; Fitzpatrick, Z.; Harris, A.F.; Maitland, S.A.; Ferreira, J.S.; Zhang, Y.; Ma, S.; Sharma, R.B.; Gray-Edwards, H.L.; Johnson, J.A.; et al. In vivo selection yields AAV-B1 capsid for central nervous system and muscle gene therapy. *Mol. Ther.* **2016**, *24*, 1247–1257. [CrossRef] [PubMed]
62. Herrmann, A.-K.; Grosse, S.; Börner, K.; Krämer, C.; Wiedtke, E.; Gunkel, M.; Grimm, D. Impact of the assembly-activating protein (AAP) on molecular evolution of synthetic Adeno-associated virus (AAV) capsids. *Hum. Gene Ther.* **2018**, *30*, 21–35. [CrossRef]
63. Tordo, J.; O'Leary, C.; Antunes, A.S.L.M.; Palomar, N.; Aldrin-Kirk, P.; Basche, M.; Bennett, A.; D'Souza, Z.; Gleitz, H.; Godwin, A.; et al. A novel adeno-associated virus capsid with enhanced neurotropism corrects a lysosomal transmembrane enzyme deficiency. *Brain J. Neurol.* **2018**, *141*, 2014–2031. [CrossRef]
64. Smith, J.K.; Agbandje-McKenna, M. Creating an arsenal of Adeno-associated virus (AAV) gene delivery stealth vehicles. *PLoS Pathog.* **2018**, *14*, e1006929. [CrossRef]
65. Paulk, N.K.; Pekrun, K.; Zhu, E.; Nygaard, S.; Li, B.; Xu, J.; Chu, K.; Leborgne, C.; Dane, A.P.; Haft, A.; et al. Bioengineered AAV capsids with combined high human liver transduction in vivo and unique humoral seroreactivity. *Mol. Ther.* **2018**, *26*, 289–303. [CrossRef] [PubMed]
66. Marlhens, F.; Bareil, C.; Griffoin, J.-M.; Zrenner, E.; Amalric, P.; Eliaou, C.; Liu, S.-Y.; Harris, E.; Redmond, T.M.; Arnaud, B.; et al. Mutations in *RPE65* cause Leber's congenital amaurosis. *Nat. Genet.* **1997**, *17*, 139–141. [CrossRef] [PubMed]
67. Marlhens, F.; Griffoin, J.-M.; Bareil, C.; Arnaud, B.; Claustres, M.; Hamel, C.P. Autosomal recessive retinal dystrophy associated with two novel mutations in the *RPE65* gene. *Eur. J. Hum. Genet.* **1998**, *6*, 527–531. [CrossRef] [PubMed]
68. Hamel, C.P.; Griffoin, J.-M.; Lasquellec, L.; Bazalgette, C.; Arnaud, B. Retinal dystrophies caused by mutations in *RPE65*: Assessment of visual functions. *Br. J. Ophthalmol.* **2001**, *85*, 424–427. [CrossRef] [PubMed]
69. Choi, E.H.; Suh, S.; Sander, C.L.; Hernandez, C.J.O.; Bulman, E.R.; Khadka, N.; Dong, Z.; Shi, W.; Palczewski, K.; Kiser, P.D. Insights into the pathogenesis of dominant retinitis pigmentosa associated with a D477G mutation in *RPE65*. *Hum. Mol. Genet.* **2018**, *27*, 2225–2243. [CrossRef] [PubMed]
70. Li, S.; Izumi, T.; Hu, J.; Jin, H.H.; Siddiqui, A.-A.A.; Jacobson, S.G.; Bok, D.; Jin, M. Rescue of enzymatic function for disease-associated RPE65 proteins containing various missense mutations in non-active sites. *J. Biol. Chem.* **2014**, *289*, 18943–18956. [CrossRef]
71. Li, Y.; Furhang, R.; Ray, A.; Duncan, T.; Soucy, J.; Mahdi, R.; Chaitankar, V.; Gieser, L.; Poliakov, E.; Qian, H.; et al. Aberrant RNA splicing is the major pathogenic effect in a knock-in mouse model of the dominantly inherited c.1430A>G human *RPE65* mutation. *Hum. Mutat.* **2019**, *40*, 426–443. [CrossRef]
72. Simunovic, M.P.; Jolly, J.K.; Xue, K.; Edwards, T.L.; Groppe, M.; Downes, S.M.; MacLaren, R.E. The spectrum of *CHM* gene mutations in choroideremia and their relationship to clinical phenotype. *Investig. Ophthalmol. Vis. Sci.* **2016**, *57*, 6033–6039. [CrossRef]

73. Bauwens, M.; Garanto, A.; Sangermano, R.; Naessens, S.; Weisschuh, N.; De Zaeytijd, J.; Khan, M.; Sadler, F.; Balikova, I.; Van Cauwenbergh, C.; et al. *ABCA4*-associated disease as a model for missing heritability in autosomal recessive disorders: Novel noncoding splice, cis-regulatory, structural, and recurrent hypomorphic variants. *Genet. Med.* **2019**, *21*, 1761–1771. [CrossRef]
74. Radziwon, A.; Arno, G.K.; Wheaton, D.; McDonagh, E.M.; Baple, E.L.; Webb-Jones, K.G.; Birch, D.; Webster, A.R.; MacDonald, I.M. Single-base substitutions in the *CHM* promoter as a cause of choroideremia. *Hum. Mutat.* **2017**, *38*, 704–715. [CrossRef]
75. Van Schil, K.; Naessens, S.; Van de Sompele, S.; Carron, M.; Aslanidis, A.; Van Cauwenbergh, C.; Kathrin Mayer, A.; Van Heetvelde, M.; Bauwens, M.; Verdin, H.; et al. Mapping the genomic landscape of inherited retinal disease genes prioritizes genes prone to coding and noncoding copy-number variations. *Genet. Med.* **2018**, *20*, 202–213. [CrossRef] [PubMed]
76. Uppal, S.; Liu, T.; Poliakov, E.; Gentleman, S.; Redmond, T.M. The dual roles of RPE65 S-palmitoylation in membrane association and visual cycle function. *Sci. Rep.* **2019**, *9*, 5218. [CrossRef] [PubMed]
77. Ellingford, J.M.; Thomas, H.B.; Rowlands, C.; Arno, G.; Beaman, G.; Gomes-Silva, B.; Campbell, C.; Gossan, N.; Hardcastle, C.; Webb, K.; et al. Functional and in-silico interrogation of rare genomic variants impacting RNA splicing for the diagnosis of genomic disorders. *bioRxiv* **2019**, 781088. [CrossRef]
78. Verbakel, S.K.; Fadaie, Z.; Klevering, B.J.; van Genderen, M.M.; Feenstra, I.; Cremers, F.P.M.; Hoyng, C.B.; Roosing, S. The identification of a RNA splice variant in TULP1 in two siblings with early-onset photoreceptor dystrophy. *Mol. Genet. Genomic Med.* **2019**, *7*, e660. [CrossRef]
79. Dickinson, D.J.; Goldstein, B. CRISPR-based methods for *Caenorhabditis elegans* genome engineering. *Genetics* **2016**, *202*, 885–901. [CrossRef]
80. Anzalone, A.V.; Randolph, P.B.; Davis, J.R.; Sousa, A.A.; Koblan, L.W.; Levy, J.M.; Chen, P.J.; Wilson, C.; Newby, G.A.; Raguram, A.; et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **2019**, *576*, 149–157. [CrossRef]
81. Dam, T.J.P.; van Kennedy, J.; Lee, R.; van der Vrieze, E.; de Wunderlich, K.A.; Rix, S.; Dougherty, G.W.; Lambacher, N.J.; Li, C.; Jensen, V.L.; et al. CiliaCarta: An integrated and validated compendium of ciliary genes. *PLoS ONE* **2019**, *14*, e0216705.
82. Messchaert, M.; Dona, M.; Broekman, S.; Peters, T.A.; Corral-Serrano, J.C.; Slijkerman, R.W.N.; van Wijk, E.; Collin, R.W.J. Eyes shut homolog is important for the maintenance of photoreceptor morphology and visual function in zebrafish. *PLoS ONE* **2018**, *13*, e0200789. [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

© American College of Medical Genetics and Genomics

ARTICLE Genetics inMedicine

Resolving the dark matter of *ABCA4* for 1054 Stargardt disease probands through integrated genomics and transcriptomics

A full list of authors and affiliations appears at the end of the paper.

Purpose: Missing heritability in human diseases represents a major challenge, and this is particularly true for *ABCA4*-associated Stargardt disease (STGD1). We aimed to elucidate the genomic and transcriptomic variation in 1054 unsolved STGD and STGD-like probands.

Methods: Sequencing of the complete 128-kb *ABCA4* gene was performed using single-molecule molecular inversion probes (smMIPs), based on a semiautomated and cost-effective method. Structural variants (SVs) were identified using relative read coverage analyses and putative splice defects were studied using in vitro assays.

Results: In 448 biallelic probands 14 known and 13 novel deep-intronic variants were found, resulting in pseudoexon (PE) insertions or exon elongations in 105 alleles. Intriguingly, intron 13 variants c.1938-621G>A and c.1938-514G>A resulted in dual PE insertions consisting of the same upstream, but different downstream PEs. The intron 44 variant c.6148-84A>T resulted in two PE insertions and flanking exon deletions. Eleven distinct large deletions were found, two of which contained small inverted segments. Uniparental isodisomy of chromosome 1 was identified in one proband.

Conclusion: Deep sequencing of *ABCA4* and midigene-based splice assays allowed the identification of SVs and causal deep-intronic variants in 25% of biallelic STGD1 cases, which represents a model study that can be applied to other inherited diseases.

Genetics in Medicine (2020) 22:1235–1246; https://doi.org/10.1038/s41436-020-0787-4

Keywords: *ABCA4*; Stargardt disease; smMIPs; deep-intronic variants; structural variants

INTRODUCTION

High-throughput genome sequencing has made a huge impact in biology and is considered the most powerful genetic test to elucidate inherited human diseases.[1] It allows the unbiased detection of a wide spectrum of genetic variants including coding and noncoding single-nucleotide variants (SNVs), as well as structural variants (SVs). However, sequencing and data storage costs as well as the possibility of secondary genetic findings hamper the use of genome sequencing.

Based on the advantages and limitations mentioned above, genome sequencing is not the best method to perform sequence analysis of one or a few genes that are associated with a clinically distinct condition. This is illustrated by autosomal recessive Stargardt disease (STGD1), which is caused by variants in the *ABCA4* gene. STGD1 is the most frequently inherited macular dystrophy with an estimated prevalence of 1/10,000.[2] Thus far, 1180 unique *ABCA4* variants have been reported in 8777 alleles of 6684 cases (www.lovd.nl/ABCA4).[3] A large proportion of the variants affect noncanonical splice site (NCSS) sequences, with variable effects on messenger RNA (mRNA) processing,[4–6] and several deep-intronic (DI) variants have been identified.[5,7–13] Most of these DI variants strengthen cryptic splice sites resulting in the insertion of pseudoexons (PEs) in the mature *ABCA4* mRNA. SVs seem to be rare in *ABCA4*,[7,10,12,14] although systematic copy-number variant (CNV) analyses have not been performed in most STGD1 cases.

Due to the relatively large size of the *ABCA4* gene (50 exons; 128,313 bp), variant screening initially was restricted to the scanning of the exons and flanking splice sites with poor sensitivity, leaving 50–70% of STGD1 probands genetically unsolved.[14–17] Recently, sequence analysis of the entire 128-kb gene was performed using next-generation sequencing platforms using Raindance microdroplet polymerase chain reaction (PCR) target enrichment or Illumina TruSeq Custom Amplicon target enrichment,[10] HaloPlex-based sequence enrichment,[7,9] or genome sequencing.[1,9]

Identification of two pathogenic alleles is important to confirm the clinical diagnosis because several promising clinical trials are underway based on RNA modulation with antisense oligonucleotides,[7,9,18] drug based therapies,[19] as well as gene augmentation[20] and stem cell therapies.[21] STGD1

Correspondence: Frans P. M. Cremers (frans.cremers@radboudumc.nl)
These senior authors contributed equally: Claire-Marie Dhaenens, Frans P. M. Cremers

Submitted 26 November 2019; accepted: 18 March 2020
Published online: 20 April 2020

Appendix 2

cases will only be eligible for these therapies if both causal alleles are known. In addition, recent studies identified alleles carrying a coding variant in *cis* with a DI variant, and only these combinations represented fully penetrant alleles,[7,9] pointing toward the importance of analyzing noncoding regions in the STGD1 cases.

Recently, we reported on the use of 483 single-molecule molecular inversion probes (smMIPs) to sequence the 50 exons and 12 intronic regions carrying 11 pathogenic DI variants of 412 genetically unsolved STGD1 cases.[5] In this study, we aimed to design a semiautomated, high-throughput, cost-effective, and comprehensive sequence analysis of the entire *ABCA4* gene, which could serve as a model study to investigate human inherited diseases due to variants in one or a few genes. Using 3866 smMIPs we sequenced 1054 genetically unsolved STGD or STGD-like probands and 138 biallelic controls carrying known *ABCA4* variants. Novel NCSS and DI variants were tested in vitro for splice defects. Additionally, a very high and reproducible read coverage allowed us to perform CNV analysis.

MATERIALS AND METHODS

Samples

Twenty-one international and four national centers ascertained 1054 genetically unsolved probands in whom STGD was part of the differential diagnosis as determined by the local ophthalmologists specializing in inherited retinal diseases. Since *ABCA4* disease is known for its clinical heterogeneity, a spectrum of (overlapping) *ABCA4*-associated phenotypes were part of this study, as well as a STGD1 phenocopy: central areolar choroidal dystrophy (CACD). The clinical findings specific to a certain clinical diagnosis and the main phenotypic characteristics used in the differential diagnosis are described in Table S1. Also, 19 cases with a clinical diagnosis of macular dystrophy without further specification were included.

Among 1054 cases 833 probands were previously screened by employing different screening methods, i.e., exome sequencing, targeted gene panel sequencing including all *ABCA4* coding regions, and Sanger sequencing of all coding *ABCA4* exons. Details are provided in Table S2.

We discerned two patient groups. The first patient group consisted of 993 genetically unsolved probands who carried one ($n = 345$) or no ($n = 648$) *ABCA4* allele. For two subjects, DNA was not available and both parents of the probands were studied, assuming autosomal recessive inheritance. The second patient group consisted of 61 "partially solved" probands, carrying the c.5603A>T (p.Asn1868Ile) variant in *trans* with other alleles. This last group was also investigated as it was suspected that there could be unidentified DI variants in *cis* with c.5603A>T, as the penetrance of c.5603A>T, when in *trans* with a severe *ABCA4* variant, was ~5% in the population.[22,23]

This study was approved by the Medical Ethical Committee 2010-359 (Protocol nr. 2009-32; NL nr. 34152.078.10) and the Commissie Mensgebonden Onderzoek Arnhem-Nijmegen (Dossier no. 2015-1543; dossier code sRP4h). All samples were collected according to the tenets of the Declaration of Helsinki and written informed consent was obtained for all patients participating in the study.

smMIPs design and *ABCA4* sequence analysis

Detailed information on the smMIPs-based *ABCA4* sequencing, selection of candidate splice variants, and inclusion criteria is provided in the Supplementary Materials and Methods.

Midigene-based splice assay

The effect of nine NCSS variants and 58 DI variants was assessed by midigene-based splicing assays employing 23 wild-type (WT) BA clones previously described[4] and the newly designed BA32, BA33, BA34, and BA35. WT and mutant constructs were transfected in HEK293T cells and the extracted total RNA was subjected to reverse transcription (RT)-PCR as described previously.[4] Details are provided in Supplementary Materials and Methods.

Identification of CNVs and assessment of the underlying mechanism

An Excel-based script was employed to detect CNVs using smMIP read number. Microhomology at the breakpoints was assessed using ClustalW, breakpoint regions were analyzed for non-B motifs by tool (nBMST and QGRS Mapper) (for details see Supplementary Materials and Methods).

Semiquantification of RT-PCR products

To quantify the ratios between correct and aberrant RT-PCR products, densitometric analysis was performed using ImageJ software.

Uniparental disomy detection

To test the presence of uniparental disomy (UPD), haplotype analysis was performed in one STGD1 case (DNA14-33085) using exome sequencing data.

RESULTS

smMIPs performance and *ABCA4* sequencing

A pilot sequencing study was conducted using 15 STGD1 samples and five DNA samples of control individuals, revealing all 34 previously identified variants (Table S3). The average number of reads for the 20 DNA samples ranged from 10 to 152,500 per smMIP, with an overall average coverage of 933× for each smMIP.

In total 1192 DNA samples were analyzed for variants in *ABCA4* using six NextSeq500 runs. The average number of reads of the 3866 smMIPs was 377×. As most nucleotide positions are targeted with two smMIPs, the effective average coverage was ~700×. To determine the coverage of *ABCA4* in more detail, we calculated the average coverage of each nucleotide position for runs 1 to 5 combined (Table S4). To visualize the results, nucleotide positions that were not covered or poorly covered (≤10 reads), moderately covered

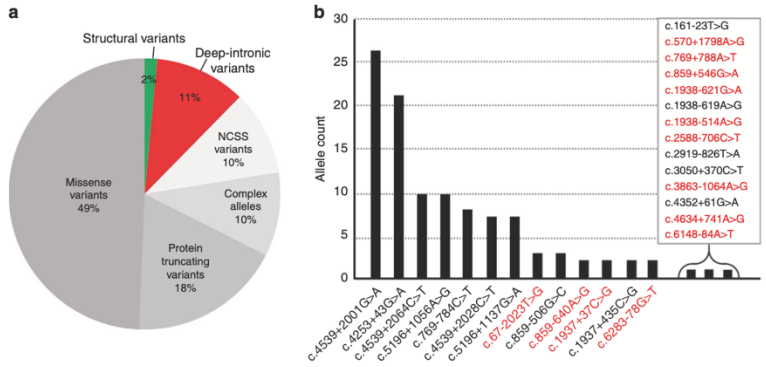


Fig. 1 Distribution of different types of alleles and deep-intronic variants in *ABCA4*. **a** The contribution of each type of variant or allele in biallelic and monoallelic cases except those carrying c.5603A>T is represented. Protein truncating variants comprise nonsense, frameshift, and canonical splice site variants. The 10% complex alleles only consist of combinations of missense variants, the most frequent of which were c.[1622T>C;3113C>T] ($n = 30$; 27% of all complex alleles) and c.[4469G>A;5603A>T] ($n = 27$; 25% of all complex alleles). They do not include the complex alleles that contain noncanonical splice site (NCSS) variants, deep-intronic variants, or protein truncating variants, when present in *cis* with other variants. If these would have been included, 16% of the alleles would consist of complex alleles. **b** Deep-intronic variant allele count in this study. Novel deep-intronic variants are highlighted in red. One hundred seventeen causal deep-intronic variants were identified. The deep-intronic variants c.4539+2001G>A ($n = 26$) and c.4253+43G>A ($n = 21$) were found most frequently. Most of the novel deep-intronic variants were found in single STGD1 probands.

(11–49 reads), or well covered (≥50 reads) are depicted in Fig. S1. From the 128,366 nt of *ABCA4*, 1980 nt (1.5%) were not or poorly covered, 1410 nt (1.1%) were moderately covered, and 124,976 nt (97.4%) were well covered. Although *ABCA4* introns carry several repetitive elements (Fig. S1), they only had a small effect on smMIPs design. Several larger repeats are present in up- and downstream regions of *ABCA4*, which resulted in the absence—or poor performance—of smMIPs. Sequencing of 1192 samples yielded a total of 7756 unique *ABCA4* variants that are listed in Table S5.

Sensitivity and specificity of the smMIPs-based sequencing

To assess the sensitivity of the new smMIPs sequencing platform, we tested 123 previously genotyped samples[5,9] in three series (runs 2, 3, and 6) (Table S6) as well as 15 control DNA samples carrying 13 different SVs spread throughout the *ABCA4* gene (run 6) (Table S7). All previously known SNVs ($n = 300$) and 13 SVs could be identified, yielding a sensitivity of 100%. Six additional variants were found due to low coverage in the previous studies, and three variants had not been annotated correctly previously.

ABCA4 gene sequencing and identification of variants

ABCA4 sequencing was performed for 1054 genetically unsolved STGD and STGD-like patients. This revealed 323 unique (likely) pathogenic SNVs and 11 SVs in 1144 alleles. Sixty-four of 323 SNVs (26%) and all 11 SVs were novel (Table S8). Detailed in silico analysis of novel SNVs is provided in Table S9. Thirteen percent of these alleles were represented by DI variants and SVs and another 10% accounted for NCSS variants (Fig. 1a). All variants and the respective cases were uploaded into the *ABCA4* variant and STGD1 cases database LOVD at www.lovd.nl/ABCA4.

Two (likely) pathogenic variants were found in 323 probands, three probands carried p.Asn1868Ile in a homozygous manner, and one (likely) causal variant in *trans* with p.Asn1868Ile was found in 125 probands. Only one (likely) causal variant was identified in 174 probands. Additionally, in 65 probands, the p.Asn1868Ile variant was the only identified variant (Table S10). No (likely) causal variants were found in 364 cases.

Among the SNVs, the most common causal alleles were c.5603A>T ($n = 134$), c.5882G>A ($n = 84$), c.[5461-10T>C;5603A>T] ($n = 44$), c.[1622T>C;3113C>T] ($n = 30$), c.[4469G>A;5603A>T] ($n = 27$), c.4539+2001G>A ($n = 26$), c.6079C>T ($n = 23$), and c.4253+43G>A ($n = 21$) (Table S8). To visualize the relative frequency of causal STGD1-causing alleles, we excluded 65 heterozygous c.5603A>T alleles that were found as the only *ABCA4* allele in these cases, as they were most likely present because of its high allele frequency (0.06) in the general population (Fig. S2).[23,24]

Splice defects due to noncanonical splice site variants

The effect on splicing of nine NCSS variants was tested in nine wild-type splice constructs previously described[4] (Fig. S3). All of the nine tested novel NCSS variants showed a splice defect when tested in HEK293T cells. Severity was assigned according to the percentage of remaining WT mRNA, as described previously.[4] Five NCSS variants were deemed severe as they showed ≤30 of WT mRNA, three were

considered to have a moderate effect with WT RNA present between >30 and ≤70% correct RNA and only one was mild as it showed >70% of WT RNA (Table S11, Fig. S4).

Deep-intronic variants identification and functional characterization

Based on the defined selection criteria, 58 DI variants were selected for splice assays. To test their effects, 27 WT midigene splice constructs were employed, 23 of which were described previously,[4] and four of which were new (Fig. S3). Thirteen of 58 tested DI variants showed a splice defect upon RT-PCR and Sanger validation (Figs. 2 and 3). For the variants that did not show any splice defect, RT-PCR results are shown in Fig. S5.

Six of the novel DI variants, i.e., variants c.570+1798A>G, c.769-788A>T, c.859-640A>G, c.1938-514A>G, c.2588-706C>T, and c.4634+741A>G, resulted in out-of-frame PE inclusions in the RNA and were deemed severe (Figs. 2 and 3). Variants c.67-2023T>G and c.859-546G>A were classified to have a moderate effect as 33% and 36% of the WT RNA products were present, respectively. As predicted due to the presence of a downstream cryptic splice donor site (SDS), variant c.1937+37C>G led to an elongation of exon 13 by 36 nucleotides, which resulted in the introduction of a premature stop codon (p.Phe647*). Moreover, two intron 13 variants, c.1938-621G>A and c.1938-514A>G, showed a complex splice pattern that led to the generation of two mutant transcripts each (Fig. 3**a**–**c**). Each of these products contained a shared PE of 134 nt (PE1) as well as variant-specific PEs, denoted PE2 (174 nt) or PE3 (109 nt) for c.1938-621G>A and c.1938-514A>G, respectively (Fig. S6). For variant c.1938-621G>A only 7% of the total complementary DNA (cDNA) product showed PE inclusion whereas for c.1938-514A>G, 87% of the cDNA products were mutant. To investigate the nature of the PE1 insertions, we studied the exon 12–17 segment of the mRNA obtained from photoreceptor progenitor cells (PPCs) derived from a control person. As depicted in Fig. S7, transcripts containing PE1 or PE1 and PE2 were identified when PPCs were grown under nonsense-mediated decay–suppressing conditions. The sum quantity of these two products was 2.9% of total mRNA suggesting that there are small amounts of PE insertions involving PE1 in the healthy retina.

Intriguingly, DI variant c.6148-84A>T showed four RNA splice products, namely a normal spliced RNA, the skipping of exon 45, the insertion of a 221-nt PE (pe1a) coupled with the deletion of exon 44, and finally, the insertion of a 173-nt PE (pe1b) that consist of the same SDS as pe1a but a different splice acceptor site (SAS) (Fig. 3**d**–**f**). Finally, variant c.3863-1064A>G showed a complex splice pattern compared with the WT and variant c.6283-78G>T led to the insertion of a 203-nt PE in intron 45 (Fig. S6). However, the exact boundaries of the presumed PE for variant c.3863-1064A>G could not yet be determined due to technical difficulties.

Overall, 13 novel DI variants were found in 18 alleles. Next to the novel variants, 14 previously reported pathogenic DI variants[7–9,13,18] were found in a total of 99 alleles, details of which are shown in Fig. 1**b** and Table S8.

Identification of novel structural variants in STGD1 cases

Among 1054 STGD and STGD-like patients analyzed, we identified 11 unique novel heterozygous SVs, all exon-spanning deletions, in 16 patients. The corresponding deletions encompass between 1 and 33 exons, ranging from 411 bp to 55.7 kb (Fig. 4, Tables S12–S17). All deletions were found in a heterozygous state in single cases, except the smallest (c.699_768+341del), which encompassed 70 bp of exon 6 and 341 bp of intron 6, and was found in six unrelated patients of Spanish origin. Deletion breakpoints were determined employing genomic PCR and Sanger sequencing for 9 of the 11 deletions. Two deletion junctions (deletions 7 and 11) could not be amplified as the 3' breakpoints were located downstream of the gene beyond the regions targeted by smMIPs. Surprisingly, Sanger sequencing revealed two complex rearrangements as deletions 5 and 6 carried inverted fragments of 279 and 224 bp respectively, residing between large deletions. These small inversions could not be identified with the CNV detection tool.

Microhomologies, repetitive elements, and non-B DNA conformation at deletion breakpoints

The breakpoints of the deletions were subjected to bioinformatic analysis to find elements underlying their formation. The presence of microhomology, repetitive elements, and non-B DNA conformations was investigated except for deletions 7 and 11 as exact boundaries could not be determined by Sanger sequencing. All other studied SVs presented microhomology at the breakpoint junctions, ranging in size from 1 to 6 bp (Fig. S8), four of which presented short insertions (Table S18). In 8 of 11 (72.7%) of the deletion breakpoints, a known repetitive element was observed, including seven non–long terminal repeats (non-LTR) retrotransposons, among which there were one short interspersed nuclear element (SINE) and four long interspersed nuclear elements (LINEs), three DNA transposons from the *hAT* superfamily, and two retrotransposons from the LTR superfamily. However, no breakpoint was part of a known element belonging to the same class and no *Alu* sequence was observed at the breakpoint junctions. Finally, the most prevalent non-B conformations observed among our breakpoints are Oligo(G)n tracts as 21 of these repeats were found in seven SVs (Tables S18, S19). Inverted repeats were observed in five breakpoint regions. No direct repeats or mirror repeats have been detected, therefore excluding triplex and slipped hairpin structure formation, respectively.

Uniparental isodisomy of chromosome 1

In STGD1 proband DNA14-33085, a causal homozygous DI variant, c.859-506G>C (p.[Phe287Thrfs*32,=]), was identified. Segregation analysis revealed this variant to be present in his unaffected father, but not in his unaffected mother. To test the possibility that the mother carried a deletion

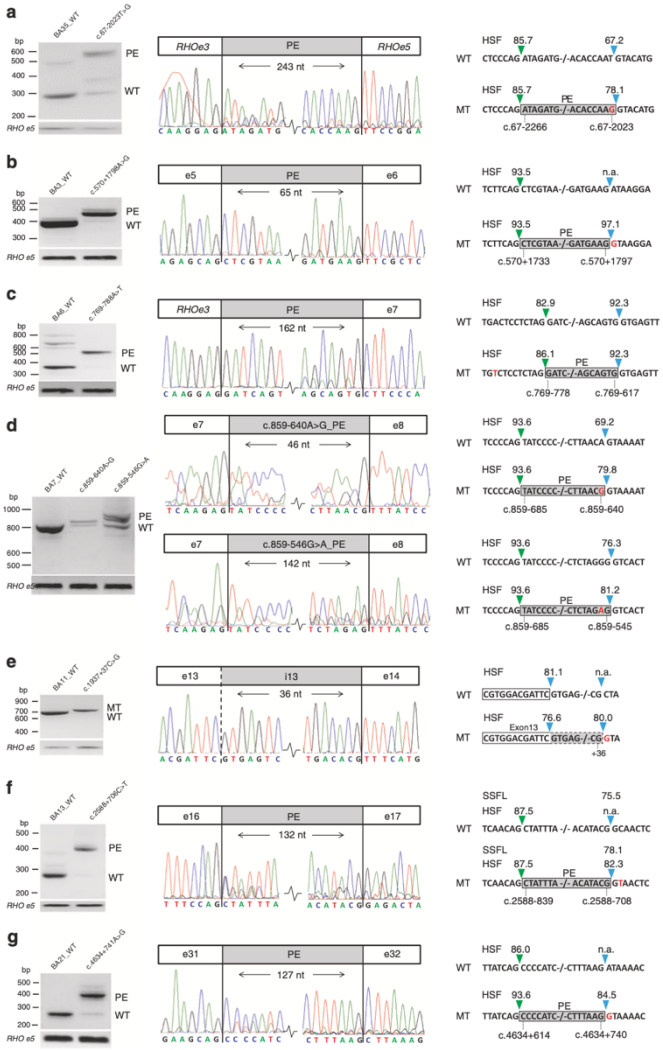


Fig. 2 Novel splice defects due to deep-intronic *ABCA4* variants. Wild-type (WT) and mutant (MT) midigenes were transfected in HEK293T cells and the extracted RNA was subjected to reverse-transcription polymerase chain reaction (RT-PCR). Left panels show the *ABCA4*-specific RT-PCR products with Rhodopsin exon 5 (*RHO* e5) RT-PCR as a transfection efficiency control. In the middle panels, Sanger sequencing results of the RT-PCR products are given. At the right side, pseudoexons (PEs) and an exon elongation are depicted with splice site strength predictions for WT and MT sequences, with green rectangles representing the splice acceptor sites and blue rectangles representing the splice donor sites. Red highlighted nucleotides represent the variants. Except for c.1937+37C>G (**e**), which resulted in a 36-nt exon 13 elongation, all deep-intronic variants lead to PEs (**a**–**d**, **f**, **g**). The intron 7 variants in (**d**) result in partially overlapping PEs that share the same splice acceptor site at position c.859-685. *HSF* Human Splicing Finder, *na* not applicable, *PE* pseudoexon, *SSFL* splice site finder–like.

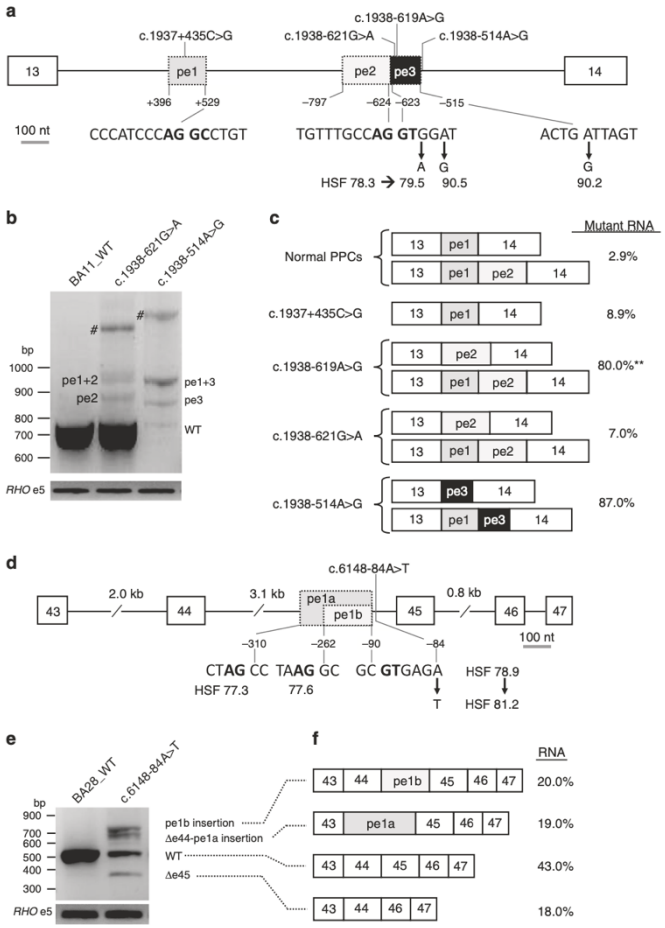


Fig. 3 Splice defects due to variants in *ABCA4* introns 13 and 44. **a** Genomic structure of intron 13 containing three pseudoexons (PEs) due to four deep-intronic variants. PE2 and PE3 share a splice donor site (SDS) (for PE2) and splice acceptor site (for PE3). Variants c.1938-621G>A and c.1938-619A>G strengthen the same cryptic SDS of PE2 slightly or strongly, respectively, as based on the Human Splicing Finder (HSF). Variant c.1938-514A>G creates a new strong SDS of PE3. The canonical and putative canonical splice sequences are given in bold lettering. The first and last positions of the PEs are provided. **b** Agarose gel analysis of reverse-transcription polymerase chain reaction (RT-PCR) products for intron 13 variants upon HEK293T cell splice assays. PE2 and PE3 were observed as single insertions, but also in combination with PE1. #Heteroduplex fragments of the lower bands. **c** Schematic representation of all mutant transcripts identified upon RT-PCR in HEK293T cell splice assays and of PE1 and PE1/PE2 observed as naturally occurring PEs when analyzing photoreceptor progenitor cells (PPCs) derived from a healthy individual. Interestingly, PE1 was previously shown to be induced by variant c.1937+435C>G (*Sangermano et al.)[9] and also can be part of mutant transcripts, together with PE2 or PE3. This is surprising as it is located far upstream of the other causal variants. **Reported by Fadaie et al.[13] **d** Variant c.6148-84A>T strengthens a SDS and results in PE1a or PE1b by employing upstream or downstream splice acceptor sites, respectively. These splice acceptor sites are comparable in predicted strength based on HSF. The canonical splice sequences are given in bold. **e** Agarose gel analysis of RT-PCR products due to c.6148-84A>T. **f** The largest fragment shows a 173-nt PE insertion between exons 44 and 45. The second largest band contains a 221-nt PE insertion (PE1a) and skipping of exon 44. The third-largest fragment represents the WT messenger RNA (mRNA) and the smallest fragment misses exon 45. The relative amounts of the products are listed at the right side.

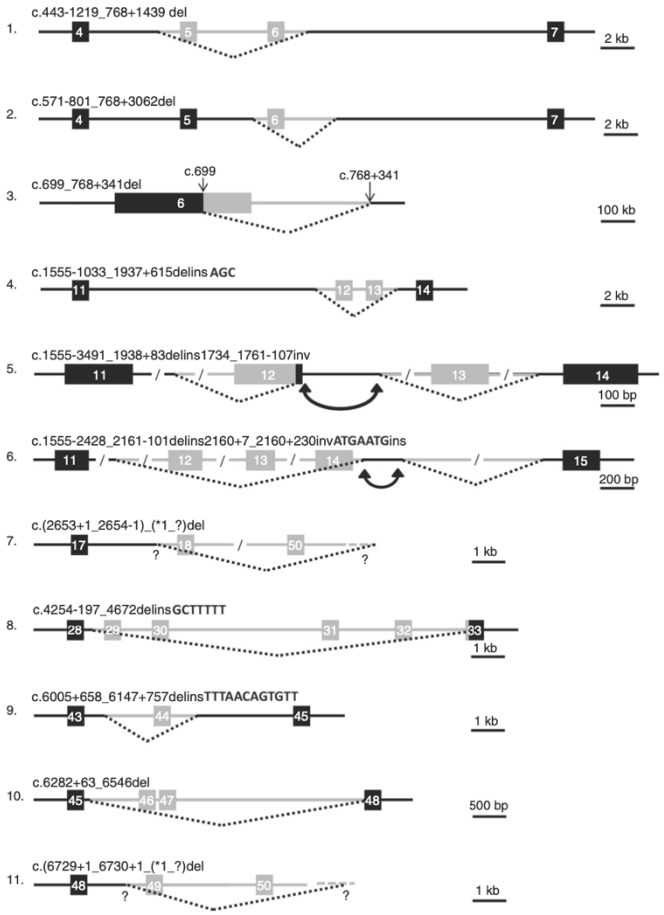


Fig. 4 Novel heterozygous structural variants in *ABCA4*. Schematic representation of the 11 structural variants identified. Exons are represented as boxes, black when they are not deleted and gray when they are deleted. Introns are represented as continuous lines, whereas stippled lines depict the deleted regions. Question marks denote that the exact location of breakpoints were unknown. Inverted double arrows represent inverted sequences.

spanning this variant, we performed CNV analysis in the proband's *ABCA4* gene. No deletion was identified (Table S17, column AU) and no heterozygous SNPs were observed in or near *ABCA4* in the proband's DNA. To test whether the chromosome 1 of the father carrying the c.859-506G>C *ABCA4* variant was passed on to the proband as two copies (UPD), exome sequencing was conducted for the proband's DNA. As shown in Fig. S9, chromosome 1 of the proband carries only homozygous SNPs, strongly suggesting the occurrence of UPD.

DISCUSSION

Employing 3866 smMIPs, 97.4% of the 128-kb *ABCA4* gene could be sequenced robustly in 1054 genetically unsolved probands with a STGD or a STGD-like phenotype. In this way, 448 (42.5%) of the probands could be genetically solved. We not only identified nine novel NCSS variants and 13 novel DI variants, but also 11 novel heterozygous SVs. The large setup of this study allowed us to provide a "landscape" overview of the different variant types underlying STGD1. As depicted in Fig. 1a, we can appreciate that DI variants

constitute a significant cause of STGD1, i.e., 11.7% of the alleles in biallelic cases, identified in 22.5% of biallelic probands. Deletions constitute 1.8% of alleles and were found in 3.5% of biallelic cases. Seven probands carried two DI variants or one DI variant and one SV. Taken together, "dark matter" alleles were found in 113/448 (25.2%) biallelic STGD1 probands. Together, these results strongly argue for a complete sequence analysis of the *ABCA4* gene to fully appreciate its mutational landscape.

Complex splice defects due to intron 13 and 44 variants

Interestingly, the two intron 13 DI variants, i.e., c.1938-621G>A and c.1938-514G>A, were in close vicinity of two previously described variants, c.1937+435C>G[9] and c.1938-619A>G.[10,13] As shown in Fig. **3a–c** the PE resulting from c.1937-514A>G (PE3) is located adjacent to PE2 as they share a dual SAS/SDS (Fig. **3a–c**). The involvement of PE1, located 491, 493, and 775 nt upstream of variants c.1937-621G>A, c.1937-619A>G, and c.1937-514A>G, respectively, is very surprising. Control PPCs also show a small percentage (2.9%) of mRNAs containing PE1 or PE1–PE2. Interestingly, the SDS of PE1 also can be employed as a SAS, which, in theory could render this intronic SAS/SDS a target for recursive splicing.[25] Together, these findings suggest that there is a "natural sensitivity" for PE1 to be recognized as a PE even if the splice defect is located far downstream. Intron 44 variant c.6148-84A>T interestingly resulted in three abnormal splice products involving different PE insertions with or without flanking exon 44 or 45 deletions. Follow-up studies employing patient-derived retinal-like cells are required to validate these complex splicing patterns.

In Table S20, we listed all published 353 DI variant alleles.[5–13,18,26,27] The three most frequent are c.4253+43G>A ($n = 100$), c.4539+2001G>A ($n = 64$), and c.5196+1137G>A ($n = 47$). For some DI variants, the splice defects in HEK293T cells or patient-derived PPCs are very small (c.769-784C>T, c.1937+435G>C, c.1937-621G>A)[9,28] (this study) or smaller than expected (c.4539+2001G>A, c.4539 +2028C>T). We hypothesize that retina-specific splice factors play roles that are largely missing (HEK293T cells) or underrepresented (PPCs) compared with the normal retina.

Current state of knowledge on structural variants in *ABCA4*

In this study, 11 unique SVs with sizes ranging from 411 bp to 55.7 kb were readily identified employing an easy-to-use visual detection tool taking advantage of the high number of reads obtained from smMIPs-based sequencing. Although this tool needs further automation to increase its performance for the detection of smaller deletions or duplications, it demonstrated its efficiency for deletions as small as 411 bp. To our knowledge, 47 different SVs have been identified in STGD1 patients (Table S21), 25 of which have been published elsewhere. Forty SVs are deletions, ranging in size from 23 bp to complete deletion of the *ABCA4* gene. There are six duplications, ranging from 24 bp to 26 kb, two indels, and one small insertion of 24 bp. As shown in Fig. S10, these SVs are spread over the entire gene. All SVs are rare, except for a 23-bp deletion affecting the splicing of exons 28 and 29 in 15 Israeli probands, as well as deletions spanning exons 20–22 and exon 6, both found in 6 probands, in Belgium/Germany/Netherlands and from Iberic origin, respectively, suggesting founder effects.

This genomic instability could be explained by the local genomic architecture (the presence of microhomology, repetitive elements, sequences forming non-B DNA conformations, and sequence motifs), leading to genomic rearrangements by impairing the replication process. For example, a microhomology of 1–4 bp may facilitate nonhomologous end joining (NHEJ)[29] and longer microhomologies of between 5 and 25 bp may favor microhomology-mediated end joining (MMEJ).[30] The assessment of the local architecture of deletions identified in this study lead us to rule out the non-allelic homologous recombination (NAHR) hypothesis (as no *Alu* sequence or L1 at any breakpoint was observed) and to propose the NHEJ or replication slippage models as the main implicated mechanisms (Table S18, Fig. S8). Indeed, the presence of microhomologies <5 bp in most of the junctions, and of scars characterized by insertion of several random nucleotides, could be a signature for NHEJ. Alternatively, several examples of an impaired replication fork have been noted that supports the replicative-based repair model. Indeed, despite the absence of repetitive elements of the same class at both sides of the breakpoints, their presence may initiate the formation of secondary structures, as repetitive elements could be more difficult to replicate, leading to an increased chance of replication fork stalling or collapsing.[31] Finally, Oligo(G)n tracts displayed a significant overrepresentation in the breakpoint regions. Such structures can induce tetraplex formation[32] and could also trigger rearrangement.

Uniparental isodisomy chromosome 1

UPD was found in one STGD1 case in this study, which represents the third STGD1 case showing UPD thus far reported.[33,34] UPD is a rare event, with an estimated occurrence of 1 in 5000 or even fewer individuals.[35] UPD was also described in six other inherited retinal dystrophy patients in which chromosomes 1, 2, and 6 were implicated.[36–41] We cannot exclude that there are additional UPD cases in our cohort as segregation analysis was not performed for all homozygous cases. Our finding stresses the importance of segregation analysis in the parents' DNAs as the recurrence risk for future offspring is very low in UPD families.

Missing heritability

In 174/1054 (16.5%) of probands, we identified only one (likely) causal allele. In view of the high carrier frequency of *ABCA4* variants in the general population, estimated to be ~5%,[3,14] about one-third of these monoallelic cases may be explained in this way. This may even be higher as we intentionally recruited monoallelic STGD and STGD-like probands for this study. Some causal variants may have

escaped our attention. First, we have not focused on variants affecting transcription regulation. Thus far, there is limited evidence for *ABCA4* variants affecting transcription,[7] but the reported putative regulatory variants were not found in this study. As in silico tools (Alamut algorithms, SpliceAI)[42] may not predict retina-specific splice defects, we may have missed some causal variants. Also, smMIPs-based sequencing may miss heterozygous deletions smaller than ~400 bp and will not detect insertions or inversions larger than ~40 bp. In addition, more refined functional tests of coding and noncoding *ABCA4* variants are needed to understand the full genetic landscape of STGD1.

The major advantages of smMIPs-based *ABCA4* sequencing compared with genome sequencing are that it (1) is at least an order of magnitude cheaper than genome sequencing, (2) results in much smaller data storage, and (3) requires no separate informed consent regarding secondary findings. Disadvantages of smMIPs are that (1) it is restricted to one or a few genes if including introns, (2) it is more cost-effective when large series are analyzed, (3) the analysis is suitable for the detection of CNVs but not for inversions and insertions, and (4) the sequencing procedure and variant calling require a specialized setup.

In our study a significant fraction of probands carried one (likely) causal variant or c.5603A>T as a single allele (239; 22.7%) or no causal variant (364; 34.5%). A more comprehensive smMIPs-based screening platform for these STGD-like cases would likely require the sequence analysis of an additional ~80 genes associated with inherited central vision defects.

As shown in this study, smMIPs-based analysis of the complete sequence(s) of one or a few genes implicated in clinically well-defined human diseases may allow the (re) analysis of hundreds to thousands of samples, in particular by targeting cohorts in developing countries in which low-cost analysis is crucial. A similar approach can be applied to all other frequent monogenic disorders to find missing variants in noncoding regions to provide a genetic diagnosis.

In conclusion, comprehensive sequence analysis of *ABCA4* in 1054 unsolved STGD and STGD-like probands, splice assays in HEK293T cells, and SV analysis resulted in the identification of "dark matter" variants in 25% of biallelic STGD1 probands. Novel complex types of splice defects were identified for intron 13 and 44 variants. Together with published causal DI variants and SVs, a detailed genomic and transcriptomic landscape of *ABCA4*-associated STGD1 was thereby established.

SUPPLEMENTARY INFORMATION

The online version of this article (https://doi.org/10.1038/s41436-020-0787-4) contains supplementary material, which is available to authorized users.

ACKNOWLEDGEMENTS

We thank Ellen Blokland, Duaa Elmelik, Emeline Gorecki, Marlie Jacobs-Camps, Charlene Piriou, Mariateresa Pizzo, and Saskia van der Velde-Visser for technical assistance. We thank Béatrice Bocquet, Dominique Bonneau, Krystyna H. Chrzanowska, Hélene Dollfus, Isabelle Drumare, Monika Heusipp, Takeshi Iwata, Beata Kocyła-Karczmarewicz, Atsushi Mizota, Nobuhisa Nao-i, Adrien Pagin, Valérie Pelletier, Rafal Ploski, Agnieszka Rafalska, Rosa Riveiro, Malgorzata Rydzanicz, Blanca Garcia Sandoval, Kei Shinoda, Francesco Testa, Kazushige Tsunoda, Shinji Ueno, and Catherine Vincent-Delorme for their cooperation and ascertaining STGD1 cases. We thank Rolph Pfundt for his assistance in exome sequencing data analysis. We are grateful to the Eichler and Shendure labs (Department of Genome Sciences, University of Washington), for assistance with the initial MIP protocol. We thank the European Reference Network (ERN)-EYE and European Retinal Disease Consortium (ERDC) networks, the Japan Eye Genetics Consortium, and the East Asian Inherited Retinal Disease Society.

This work was supported by the RetinaUK, grant number GR591 (to F.P.M.C.); a Fighting Blindness Ireland grant, grant number FB18CRE (to F.P.M.C., G.J.F.); a Horizon 2020, Marie Sklodowska-Curie Innovative Training Network entitled European Training Network to Diagnose, Understand and Treat Stargardt Disease; Frequent Inherited Blinding Disorder-StarT (813490) (to E.D.B., F.P.M.C., S.B., G.J.F.); Foundation Fighting Blindness USA, grant number PPA-0517-0717-RAD (to F.P.M.C.); the Rotterdamse Stichting Blindenbelangen, the Stichting Blindenhulp, and the Stichting tot Verbetering van het Lot der Blinden (to F.P.M.C.); and by the Landelijke Stichting voor Blinden en Slechtzienden, Macula Degeneratie fonds and the Stichting Blinden-Penning that contributed through Uitzicht 2016-12 (to F.P.M.C.). This work was also supported by the Algemene Nederlandse Vereniging ter Voorkoming van Blindheid and Landelijke Stichting voor Blinden en Slechtzienden that contributed through UitZicht 2014-13, together with the Rotterdamse Stichting Blindenbelangen, Stichting Blindenhulp, and the Stichting tot Verbetering van het Lot der Blinden (to F.P.M.C.). This work was also supported by Groupement de Coopération Sanitaire Interrégional G4 qui réunit les Centres Hospitaliers Universitaires Amiens, Caen, Lille et Rouen (GCS G4) and by the Fondation Stargardt France (to C.-M.D.), Federal Ministry of Education and Research (BMBF), grant numbers 01GM0851 and 01GM1108B (to B.H.F.W.), programs SVV 260516, UNCE 204064, and PROGRES-Q26/LF1 of the Charles University (to B.K., L.D., P.L.). This work was supported by grant AZV NU20-07-00182 (to P.L., B.K. and L.D.). The work of A.D. was supported by Fighting Blindness Ireland, Health Research Board of Ireland and the Medical Research Charities Group (MRCG-2016-14) (to G.J.F.). This work was supported by grant AZV NU20-07-00182 (to P.L., B.K., L.D.). This work was also supported by the Ghent University Research Fund (BOF15/GOA/011), by the Research Foundation Flanders (FVO) G0C6715N, by the Hercules foundation AUGE/13/023 and JED Foundation (to E.D.B.). M.B. was PhD fellow of the FWO and recipient of a grant of the funds for Research in Ophthalmology (FRO). E.D.B. is Senior Clinical Investigator of the FWO (1802215N; 1802220N). The work of M. D.P-.V. is supported by the Conchita Rábago Foundation and the Boehringer Ingelheim Fonds. The work of C.A. is supported by grants PI16/0425 from ISCIII partially supported by the European

Regional Development Fund (ERDF), RAREGenomics-CM (CAM, B2017/BMD-3721), ONCE, and Ramon Areces Foundation. This work was supported by the Peace for Sight grant (to D.S., A.A.). The work of L.R. and R.R. was supported by Retina South Africa and the South African Medical Research Council (MRC). This work was also supported by the Foundation Fighting Blindness, grant/award number BR-GE-0214–0639-TECH and BRGE-0518–0734-TECH (to T.B.-Y., D.S., H.N.); the Israeli Ministry of Health, grant/award number 3-12583Q4 (to T.B.-Y., D.S., H.N.); Olive Young Fund, University Hospital Foundation, Edmonton (to I.M.M.); the National Science Center (Poland) grant number N N402 591640 (5916/B/P01/2011/40) (to M.O.); and UMO-2015/19/D/NZ2/03193 (to A.M.T.). This work was supported by the Italian Fondazione Roma (to S.B., F.S.), the Italian Telethon Foundation (to S.B.), and the Ministero dell'Istruzione del l'Università e della Ricerca (MIUR) under PRIN 2015 (to S.B., F.S.). M.B.G. and A.M. were supported by the Daljit S. and Elaine Sarkaria Charitable Foundation. The funding organizations had no role in the design or conduct of this research, and provided unrestricted grants.

DISCLOSURE

The authors declare no conflicts of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Carss KJ, Arno G, Erwood M, et al. Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. Am J Hum Genet. 2017;100:75–90.
2. Blacharski PA, Newsome DA. Bilateral macular holes after Nd:YAG laser posterior capsulotomy. Am J Ophthalmol. 1988;105:417–418.
3. Cornelis SS, Bax NM, Zernant J, et al. In silico functional meta-analysis of 5,962 ABCA4 variants in 3,928 retinal dystrophy cases. Hum Mutat. 2017;38:400–408.
4. Sangermano R, Khan M, Cornelis SS, et al. ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. Genome Res. 2018;28:100–110.
5. Khan M, Cornelis SS, Khan MI, et al. Cost-effective molecular inversion probe-based ABCA4 sequencing reveals deep-intronic variants in Stargardt disease. Hum Mutat. 2019;40:1749–1759.
6. Schulz HL, Grassmann F, Kellner U, et al. Mutation spectrum of the ABCA4 gene in 335 Stargardt disease patients from a multicenter German cohort—impact of selected deep intronic variants and common SNPs. Invest Ophthalmol Vis Sci. 2017;58:394–403.
7. Bauwens M, Garanto A, Sangermano R, et al. ABCA4-associated disease as a model for missing heritability in autosomal recessive disorders: novel noncoding splice, cis-regulatory, structural, and recurrent hypomorphic variants. Genet Med. 2019;21:1761–1771.
8. Braun TA, Mullins RF, Wagner AH, et al. Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. Hum Mol Genet. 2013;22:5136–5145.
9. Sangermano R, Garanto A, Khan M, et al. Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. Genet Med. 2019;21:1751–1760.
10. Zernant J, Xie YA, Ayuso C, et al. Analysis of the ABCA4 genomic locus in Stargardt disease. Hum Mol Genet. 2014;23:6797–6806.
11. Bauwens M, De Zaeytijd J, Weisschuh N, et al. An augmented ABCA4 screen targeting noncoding regions reveals a deep intronic founder variant in Belgian Stargardt patients. Hum Mutat. 2015;36:39–42.
12. Bax NM, Sangermano R, Roosing S, et al. Heterozygous deep-intronic variants and deletions in ABCA4 in persons with retinal dystrophies and one exonic ABCA4 variant. Hum Mutat. 2015;36:43–47.
13. Fadaie Z, Khan M, Del Pozo-Valero M, et al. Identification of splice defects due to noncanonical splice site or deep-intronic variants in ABCA4. Hum Mutat. 2019;40:2365–2376.
14. Maugeri A, van Driel MA, van de Pol DJR, et al. The 2588G -> C mutation in the ABCR gene is a mild frequent founder mutation in the western European population and allows the classification of ABCR mutations in patients with Stargardt disease. Am J Hum Genet. 1999;64:1024–1035.
15. Rivera A, White K, Stohr H, et al. A comprehensive survey of sequence variation in the ABCA4 (ABCR) gene in Stargardt disease and age-related macular degeneration. Am J Hum Genet. 2000;67:800–813.
16. Jaakson K, Zernant J, Kulm M, et al. Genotyping microarray (gene chip) for the ABCR (ABCA4) gene. Hum Mutat. 2003;22:395–403.
17. Maia-Lopes S, Aguirre-Lamban J, Castelo-Branco M, Riveiro-Alvarez R, Ayuso C, Silva ED. ABCA4 mutations in Portuguese Stargardt patients: identification of new mutations and their phenotypic analysis. Mol Vis. 2009;15:584–591.
18. Albert S, Garanto A, Sangermano R, et al. Identification and rescue of splice defects caused by two neighboring deep-intronic ABCA4 mutations underlying Stargardt disease. Am J Hum Genet. 2018;102:517–527.
19. Charbel Issa P, Barnard AR, Herrmann P, Washington I, MacLaren RE. Rescue of the Stargardt phenotype in Abca4 knockout mice through inhibition of vitamin A dimerization. Proc Natl Acad Sci U S A. 2015;112:8415–8420.
20. Allocca M, Doria M, Petrillo M, et al. Serotype-dependent packaging of large genes in adeno-associated viral vectors results in effective gene delivery in mice. J Clin Invest. 2008;118:1955–1964.
21. Lu B, Malcuit C, Wang S, et al. Long-term safety and function of RPE from human embryonic stem cells in preclinical models of macular degeneration. Stem Cells. 2009;27:2126–2135.
22. Cremers FPM, Cornelis SS, Runhart EH, Astuti GDN. Author response: penetrance of the ABCA4 p.Asn1868Ile allele in Stargardt disease. Invest Ophthalmol Vis Sci. 2018;59:5566–5568.
23. Runhart EH, Sangermano R, Cornelis SS, et al. The common ABCA4 variant p.Asn1868Ile shows nonpenetrance and variable expression of Stargardt disease when present in trans with severe variants. Invest Ophthalmol Vis Sci. 2018;59:3220–3231.
24. Zernant J, Lee W, Collison FT, et al. Frequent hypomorphic alleles account for a significant fraction of ABCA4 disease and distinguish it from age-related macular degeneration. J Med Genet. 2017;54:404–412.
25. Hafez M, Hausner G. Convergent evolution of twintron-like configurations: one is never enough. RNA Biol. 2015;12:1275–1288.
26. Zernant J, Lee W, Nagasaki T, et al. Extremely hypomorphic and severe deep intronic variants in the ABCA4 locus result in varying Stargardt disease phenotypes. Cold Spring Harb Mol Case Stud. 2018;4:a002733.
27. Nassisi M, Mohand-Said S, Andrieu C, et al. Prevalence of ABCA4 deep-intronic variants and related phenotype in an unsolved "one-hit" cohort with Stargardt disease. Int J Mol Sci. 2019;20:5053.
28. Runhart EH, Valkenburg D, Cornelis SS, et al. Late-onset Stargardt disease due to mild, deep-intronic ABCA4 alleles. Invest Ophthalmol Vis Sci. 2019;60:4249–4256.
29. Lieber MR. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu Rev Biochem. 2010;79:181–211.
30. McVey M, Lee SE. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet. 2008;24:529–538.
31. Vissers LE, Bhatt SS, Janssen IM, et al. Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genomic architecture. Hum Mol Genet. 2009;18:3579–3593.
32. Bacolla A, Wells RD. Non-B DNA conformations, genomic rearrangements, and human disease. J Biol Chem. 2004;279:47411–47414.
33. Fingert JH, Eliason DA, Phillips NC, Lotery AJ, Sheffield VC, Stone EM. Case of Stargardt disease caused by uniparental isodisomy. Arch Ophthalmol. 2006;124:744–745.
34. Riveiro-Alvarez R, Valverde D, Lorda-Sanchez I, et al. Partial paternal uniparental disomy (UPD) of chromosome 1 in a patient with Stargardt disease. Mol Vis. 2007;13:96–101.
35. Liehr T. Cytogenetic contribution to uniparental disomy (UPD). Mol Cytogenet. 2010;3:8.
36. Rivolta C, Berson EL, Dryja TP. Paternal uniparental heterodisomy with partial isodisomy of chromosome 1 in a patient with retinitis pigmentosa without hearing loss and a missense mutation in the Usher syndrome type II gene USH2A. Arch Ophthalmol. 2002;120:1566–1571.

37. Roosing S, van den Born LI, Hoyng CB, et al. Maternal uniparental isodisomy of chromosome 6 reveals a TULP1 mutation as a novel cause of cone dysfunction. Ophthalmology. 2013;120: 1239–1246.
38. Thompson DA, Gyurus P, Fleischer LL, et al. Genetics and phenotypes of RPE65 mutations in inherited retinal degeneration. Invest Ophthalmol Vis Sci. 2000;41:4293–4299.
39. Thompson DA, McHenry CL, Li Y, et al. Retinal dystrophy due to paternal isodisomy for chromosome 1 or chromosome 2, with homoallelism for mutations in RPE65 or MERTK, respectively. Am J Hum Genet. 2002; 70:224–229.
40. Wiszniewski W, Lewis RA, Lupski JR. Achromatopsia: the CNGB3 p. T383fsX mutation results from a founder effect and is responsible for the visual phenotype in the original report of uniparental disomy 14. Hum Genet. 2007;121:433–439.
41. Souzeau E, Thompson JA, McLaren TL, et al. Maternal uniparental isodisomy of chromosome 6 unmasks a novel variant in TULP1 in a patient with early onset retinal dystrophy. Mol Vis. 2018;24:478–484.
42. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, et al. Predicting splicing from primary sequence with deep learning. Cell. 2019;176:535–48 e524.

Mubeen Khan, MSc[1,2], Stéphanie S. Cornelis, MSc[1,2], Marta Del Pozo-Valero, MSc[1,3], Laura Whelan, BSc[4], Esmee H. Runhart, MD[2,5], Ketan Mishra, MSc[1,2], Femke Bults, BSc[1], Yahya AlSwaiti, MD[6], Alaa AlTalbishi, MD, PhD[6], Elfride De Baere, MD, PhD[7], Sandro Banfi, MD[8], Eyal Banin, MD, PhD[9], Miriam Bauwens, PhD[7], Tamar Ben-Yosef, PhD[10], Camiel J. F. Boon, MD, PhD[11,12], L. Ingeborgh van den Born, MD, PhD[13,14], Sabine Defoort, MD[15], Aurore Devos, MSc[16], Adrian Dockery, PhD[4], Lubica Dudakova, PhD[17], Ana Fakin, MD, PhD[18], G. Jane Farrar, PhD[4], Juliana Maria Ferraz Sallum, MD, PhD[19,20], Kaoru Fujinami, MD, PhD[21,22,23,24], Christian Gilissen, PhD[1,25], Damjan Glavač, PhD[26], Michael B. Gorin, MD, PhD[27,28], Jacquie Greenberg, PhD[29], Takaaki Hayashi, MD, PhD[30], Ymkje M. Hettinga, MD[31], Alexander Hoischen, PhD[1], Carel B. Hoyng, MD, PhD[2,5], Karsten Hufendiek, MD[32], Herbert Jägle, MD[33], Smaragda Kamakari, PhD[34], Marianthi Karali, PhD[8], Ulrich Kellner, MD[35,36], Caroline C. W. Klaver, MD, PhD[5,37,38], Bohdan Kousal, MD[17,39], Tina M. Lamey, PhD[40,41], Ian M. MacDonald, MD, CM[42], Anna Matynia, PhD[27,28], Terri L. McLaren, BSc[40,41], Marcela D. Mena, PhD[43], Isabelle Meunier, MD, PhD[44], Rianne Miller, BSc[1], Hadas Newman, MD[45,46], Buhle Ntozini, BSc[29], Monika Oldak, MD, PhD[47], Marc Pieterse, MSc[1], Osvaldo L. Podhajcer, PhD[43], Bernard Puech, MD[15], Raj Ramesar, MBA, PhD[29], Klaus Rüther, MD, PhD[48], Manar Salameh, BSc[6], Mariana Vallim Salles, MD[19,20], Dror Sharon, PhD[9], Francesca Simonelli, MD, PhD[49], Georg Spital, MD[50], Marloes Steehouwer, BSc[1], Jacek P. Szaflik, MD, PhD[51], Jennifer A. Thompson, PhD[41], Caroline Thuillier, MSc[52], Anna M. Tracewska, PhD[53], Martine van Zweeden, BSc[1], Andrea L. Vincent, MBChB, MD[54,55], Xavier Zanlonghi, MD, PhD[56], Petra Liskova, MD, PhD[17,39], Heidi Stöhr, PhD[57], John N. De Roach, PhD[40,41], Carmen Ayuso, MD, PhD[3], Lisa Roberts, PhD[29], Bernhard H. F. Weber, PhD[57], Claire-Marie Dhaenens, MD, PhD[1,16] and Frans P. M. Cremers, PhD[1,2]

[1]Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands; [2]Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands; [3]Department of Genetics, IIS-Fundación Jiménez Díaz, CIBERER, Madrid, Spain; [4]The School of Genetics & Microbiology, Trinity College Dublin, Dublin, Ireland; [5]Department of Ophthalmology, Radboud University Medical Center, Nijmegen, The Netherlands; [6]St John of Jerusalem Eye Hospital Group, East Jerusalem, Palestine; [7]Center for Medical Genetics Ghent, Ghent University and Ghent University Hospital, Ghent, Belgium; [8]Department of Precision Medicine, University of Campania Luigi Vanvitelli, Naples and Telethon Institute of Genetics and Medicine (TIGEM), Pozzuoli, Italy; [9]Department of Ophthalmology, Hadassah Medical Center, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel; [10]Ruth and Bruce Rappaport Faculty of Medicine, Technion-Israel Institute of Technology, Haifa, Israel; [11]Department of Ophthalmology, Leiden University Medical Center, Leiden, The Netherlands; [12]Department of Ophthalmology, Amsterdam University Medical Centers, Amsterdam, The Netherlands; [13]The Rotterdam Eye Hospital, Rotterdam, The Netherlands; [14]The Rotterdam Ophthalmic Institute, Rotterdam, The Netherlands; [15]Service d'exploration de la vision et neuro-ophtalmologie, Centre Hospitalier Universitaire de Lille, Lille, France; [16]Univ. Lille, Inserm, CHU Lille, U1172 - LilNCog - Lille Neuroscience & Cognition, F-59000 Lille, France; [17]Research Unit for Rare Diseases, Department of Paediatrics and Adolescent Medicine, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic; [18]Eye Hospital, University Medical Centre Ljubljana, Ljubljana, Slovenia; [19]Department of Ophthalmology and Visual Sciences, Universidade Federal de São Paulo, São Paulo, SP, Brazil; [20]Instituto de Genética Ocular, São Paulo, SP, Brazil; [21]UCL Institute of Ophthalmology, London, UK; [22]Laboratory of Visual Physiology, Division of Vision Research, National Institute of Sensory Organs, National Hospital Organization Tokyo Medical Center, Tokyo, Japan; [23]Graduate School of Health Management, Keio University, Tokyo, Japan; [24]Moorfields Eye Hospital, London, UK; [25]Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands; [26]Department of Molecular Genetics, Institute of Pathology, University of

Ljubljana, Ljubljana, Slovenia; [27]Department of Ophthalmology, David Geffen School of Medicine, Stein Eye Institute, University of California–Los Angeles, Los Angeles, CA, USA; [28]Department of Human Genetics, David Geffen School of Medicine, University of California–Los Angeles, Los Angeles, CA, USA; [29]University of Cape Town/MRC Genomic and Precision Medicine Research Unit, Division of Human Genetics, Department of Pathology, Institute of Infectious Disease and Molecular Medicine (IDM), Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa; [30]Department of Ophthalmology, The Jikei University School of Medicine, Tokyo, Japan; [31]Bartiméus Diagnostic Center for Complex Visual Disorders, Zeist, The Netherlands; [32]University Eye Hospital Hannover Medical School, Hannover, Germany; [33]Department of Ophthalmology, University Hospital, University Regensburg, Regensburg, Germany; [34]Ophthalmic Genetics Unit, OMMA Ophthalmological Institute of Athens, Athens, Greece; [35]Rare Retinal Disease Center, AugenZentrum Siegburg, MVZ ADTC Siegburg GmbH, Siegburg, Germany; [36]RetinaScience, Bonn, Germany; [37]Department of Ophthalmology, Erasmus Medical Centre, Rotterdam, The Netherlands; [38]Department of Epidemiology, Erasmus Medical Centre, Rotterdam, The Netherlands; [39]Department of Ophthalmology, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic; [40]Centre for Ophthalmology and Visual Science, The University of Western Australia, Nedlands, WA, Australia; [41]Australian Inherited Retinal Disease Registry and DNA Bank, Department of Medical Technology and Physics, Sir Charles Gairdner Hospital, Nedlands, WA, Australia; [42]Departments of Ophthalmology and Medical Genetics, University of Alberta, Edmonton, AB, Canada; [43]Laboratory of Molecular and Cellular Therapy, Fundacion Instituto Leloir-CONICET, Buenos Aires, Argentina; [44]Institut des Neurosciences de Montpellier, INSERM, Université de Montpellier, Montpellier, France; [45]Department of Ophthalmology, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel; [46]Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; [47]Department of Histology and Embryology, Medical University of Warsaw, Warsaw, Poland; [48]Augenarztpraxis, Dorotheenstraße, Berlin, Germany; [49]Eye Clinic, Multidisciplinary Department of Medical, Surgical and Dental Sciences, University of Campania Luigi Vanvitelli, Naples, Italy; [50]Department of Ophthalmology, St. Franziskus-Hospital, Münster, Germany; [51]Department of Ophthalmology, Medical University of Warsaw SPKSO Ophthalmic University Hospital, Warsaw, Poland; [52]CHU Lille, Institut de Génétique Médicale, Lille, France; [53]DNA Analysis Unit, ŁUKASIEWICZ Research Network—PORT Polish Center for Technology Development, Wroclaw, Poland; [54]Department of Ophthalmology, New Zealand National Eye Centre, Faculty of Medical and Health Sciences, The University of Auckland, Grafton, Auckland, New Zealand; [55]Eye Department, Greenlane Clinical Centre, Auckland District Health Board, Auckland, New Zealand; [56]Centre de Compétence Maladie Rare, Clinique Jules Verne, Nantes, France; [57]Institute of Human Genetics, University of Regensburg, Regensburg, Germany.

Appendix 3

Vision science

Original research

BBS1 branchpoint variant is associated with non-syndromic retinitis pigmentosa

Zeinab Fadaie ,[1,2] Laura Whelan ,[3] Adrian Dockery ,[3] Catherina H Z Li ,[2,4] L Ingeborgh van den Born ,[5] Carel B Hoyng ,[2,4] Christian Gilissen ,[1,6] Jordi Corominas ,[1,6] Charlie Rowlands ,[7,8] Roly Megaw ,[9,10] Anne K Lampe ,[11] Frans P M Cremers ,[1,2] Gwyneth Jane Farrar ,[3] Jamie M Ellingford ,[7,8] Paul F. Kenna ,[3,12] Susanne Roosing [1,2]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/jmedgenet-2020-107626).

For numbered affiliations see end of article.

Correspondence to
Susanne Roosing, Department of Human Genetics, Radboudumc, Nijmegen 6525 GA, The Netherlands; Susanne.Roosing@radboudumc.nl

Received 1 December 2020
Revised 11 February 2021
Accepted 26 February 2021
Published Online First 28 April 2021

Check for updates

© Author(s) (or their employer(s)) 2022. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Fadaie Z, Whelan L, Dockery A, *et al*. *J Med Genet* 2022;**59**:438–444.

ABSTRACT

Background Inherited retinal diseases (IRDs) can be caused by variants in >270 genes. The Bardet-Biedl syndrome 1 (*BBS1*) gene is one of these genes and may be associated with syndromic and non-syndromic autosomal recessive retinitis pigmentosa (RP). Here, we identified a branchpoint variant in *BBS1* and assessed its pathogenicity by in vitro functional analysis.

Methods Whole genome sequencing was performed for three unrelated monoallelic *BBS1* cases with non-syndromic RP. A fourth case received MGCM 105 gene panel analysis. Functional analysis using a midigene splice assay was performed for the putative pathogenic branchpoint variant in *BBS1*. After confirmation of its pathogenicity, patients were clinically re-evaluated, including assessment of non-ocular features of Bardet-Biedl syndrome.

Results Clinical assessments of probands showed that all individuals displayed non-syndromic RP with macular involvement. Through detailed variant analysis and prioritisation, two pathogenic variants in *BBS1*, the most common missense variant, c.1169T>G (p.(Met390Arg)), and a branchpoint variant, c.592-21A>T, were identified. Segregation analysis confirmed that in all families, probands were compound heterozygous for c.1169T>G and c.592-21A>T. Functional analysis of the branchpoint variant revealed a complex splicing defect including exon 8 and exon 7/8 skipping, and partial in-frame deletion of exon 8.

Conclusion A putative severe branchpoint variant in *BBS1*, together with a mild missense variant, underlies non-syndromic RP in four unrelated individuals. To our knowledge, this is the first report of a pathogenic branchpoint variant in IRDs that results in a complex splice defect. In addition, this research highlights the importance of the analysis of non-coding regions in order to provide a conclusive molecular diagnosis.

INTRODUCTION

The Bardet-Biedl syndrome 1 (*BBS1*) gene is associated with both syndromic Bardet-Biedl syndrome (BBS; OMIM 209900)[1 2] and non-syndromic retinitis pigmentosa (RP) (OMIM: #268000).[3 4] Pathogenic variants in *BBS1* are an important cause of BBS, which is the second most frequent syndromic form of RP after Usher syndrome.[4] In the majority of cases, two severe pathogenic variants result in syndromic forms of RP, while a combination of severe and mild pathogenic variants result in non-syndromic RP with no, or very few, non-ocular characteristics, depending on the remaining protein function. This is not unique to *BBS1*. The phenomenon is also observed in other genes, such as *ARL6* (*BBS3*) (OMIM: 600151), *TTC8* (*BBS8*) (OMIM: 615985), *BBS12* (OMIM: 610683) or *CEP290* (*BBS14*) (OMIM: 615991), which can be associated with non-syndromic retinal degeneration or retinal degeneration with minimal non-ocular characteristics.[3 5]

BBS1 is located on chromosome 11q13.2 and contains 17 exons and encodes 593 amino acids. BBS1, together with other BBS proteins, forms a complex called the BBSome which is located at the basal body and connecting cilium and is thought to be required for membrane trafficking to the outer segments of photoreceptors.[4 6] Based on previous mouse and zebrafish model studies, pathogenic variants likely cause either accumulation of toxic BBS1 protein or loss of protein function in the cell body. This accumulation or dysfunction leads to BBS or non-syndromic RP based on the severity of the pathogenic variant(s) and characterisation of the phenotype.[4] To date, 129 pathogenic or likely pathogenic variants have been reported in this gene (https://www.ncbi.nlm.nih.gov/clinvar). Moreover, 80 *BBS1* variants have been reported in the Leiden Open Variation Database (LOVD), of which 13 variants are pathogenic or likely pathogenic (https://www.lovd.nl/BBS1).

A myriad of pathogenic splice site variants have been identified in inherited retinal disease (IRD)–associated genes that lead to RP or other retinal dystrophies.[7 8] The splicing process is a complex phenomenon which involves a large number of proteins with various interactions between *cis* and *trans* elements.[9] It requires highly conserved sequences in the canonical acceptor and donor splice site sequences (98.7% contain GT in +1and +2and 100% contain AG in −1 and −2 positions, respectively). These conserved sequences, together with a consensus branchpoint sequence (BPS), are a prerequisite to ensure that splicing factors bind to proper positions in pre-mRNAs.[10–12] In addition, partially conserved non-canonical splice site sequences are located at the first and last three

nucleotides of an exon as well as the −3 to −14 nucleotides from the splice acceptor site (SAS), and +3 to +6 from the splice donor site (SDS). All are essential for an accurate splicing process.[13]

Pathogenic variants in these conserved sequences will disturb the process of splicing in the mRNA, leading to dysfunction or deficiency of a protein. Likewise, pathogenic variants in the non-canonical splice site can cause splicing defects. Deep-intronic variants can create new or activate existing splice sites, known as cryptic splice sites, which are analogous to the consensus sequences and can be used as alternative splice sites which in most cases result in the insertion of a pseudoexon into the mRNA and protein truncation.[9 14]

In this study, we describe the pathogenicity of a near-exon aberrant RNA (NEAR) splice variant, c.592-21A>T in *BBS1*, which alters the BPS upstream of exon 8 at its most critical position. We explored the effect of c.592-21A>T using a midigene splice assay and show the relevance of assessing branchpoint motif regions in IRDs.

METHODS AND MATERIALS

Patient cohort and their clinical evaluations

Patients with autosomal recessive retinal degenerations, two Dutch, one Irish and one British, were recruited from the Radboud University Medical Center and The Rotterdam Eye Hospital (Nijmegen, The Netherlands), the Research Foundation of the Royal Victoria Eye and Ear Hospital (Dublin, Ireland), and the Manchester Centre for Genomic Medicine (Manchester, UK), respectively. The clinical diagnoses were based on ophthalmological examinations, including assessment of best-corrected visual acuity, Goldmann visual field testing, slit-lamp biomicroscopy, detailed fundoscopic examination and fundus photography. Furthermore, electrophysiological evaluations, including full-field flash electroretinography (ERG) and multifocal ERG, were performed according to the guidelines of the International Society of Clinical Electrophysiology of Vision.[15]

Patients and patient records were re-evaluated to identify clinical characteristics associated with BBS, including the primary features, postaxial polydactyly, truncal obesity, cognitive impairment, hypogonadism and renal abnormalities as well as secondary features including speech disorders, developmental delay, behavioural abnormalities, orodental abnormalities, brachydactyly/syndactyly, ataxia, diabetes mellitus, cardiovascular anomalies, hepatic involvement, mild hypertonia and additional eye abnormalities such as strabismus, cataracts and astigmatism were also investigated. However, features such as renal and hepatic abnormalities and cardiovascular anomalies were not specifically examined when the general health of patients was not indicative of these conditions.

Genetic analysis

Genomic DNA was isolated from peripheral lymphocytes based on the standard procedures. Whole genome sequencing (WGS) was performed for three cases from three families (A–C). The patients included in this study remained genetically undiagnosed after genetic assessment through routine diagnostic whole exome sequencing (families A and B) or Target5000 panel sequencing (family C). Family D was screened using the MCGM 105 gene inherited retinal disease panel (family D) and did not receive WGS.[16] For families A–C, WGS was performed by BGI (Hong Kong, China) on a BGISeq500 using a 2×100 bp paired-end reads, with a minimal median coverage per genome of 30-fold. Burrows-Wheeler Aligner was used to map WGS data to the human genome (hg19). The quality of the sequencing data was based on insert size, percentage mapped reads, percentage duplicated mapped reads, coverage, bases with >20× coverage and error rate, using evaluated using Qualimap V.2.2.1.[17]

Variant calling was carried out using xAtlas V.0.1 and annotated using an in-house developed pipeline. Variant annotation was performed using the Variant Effect Predictor (VEP V.91) and Gencode V.34lift37 basic gene annotations. Among others, annotations consist of chromosomal location and position, gene component (eg, intron, exon, splice site, 5′UTR, 3′UTR, intragenic), various population frequency databases (eg, gnomAD, GoNL, Wellderly, 1000genomes, in-house variant frequency WES/WGS data), all transcripts including protein effect, various in silico prediction scores (eg, Grantham, PhyloP, CADD_PHRED, SpliceAI), gene ontology description, gene and disease OMIM description, and gene regulation and expression data. Runs of homozygosity were detected using plink V.1.07[18] with the following parameters: homozyg-window-het=3, homozyg-snp=50 and homozyg-kb=300. Structural variants (SVs) were detected based on paired-end and split-read evidence using Manta Structural Variant Caller V.1.1.0 (Illumina) with default parameters.[19] CNVs were called using Control-FREEC, which detects copy number changes and allelic imbalances based on read depth.[20] SVs and CNVs were annotated using an in-house developed pipeline based on ANNOVAR and Gencode V.34lift37 basic gene annotations. This contains annotations for chromosomal location and position, gene(s) and their component (eg, intronic, exonic), type of SV (gain/loss), percentage overlap and frequency of various population frequency databases (eg, gnomAD, GoNL, Decipher, Wellderly, 1000genomes, in-house SV frequency data), all transcripts including protein effect, gene ontology description and gene and disease OMIM description. Short tandem repeats (STRs) for 33 known pathogenic sites were detected using Expansion Hunter V.3.1.2. using default settings.[21] Variant specific data may be available on request.

Variant prioritisation and selection

Initially, SVs and CNVs were filtered based on a minor allele frequency of <1% in the gnomAD database. Prioritisation of candidate SVs and CNVs was based on the assessment of all exonic CNVs and SVs followed by those that were intronic in a customised panel of 446 IRD and ocular defect–associated genes and 163 ciliopathy-associated genes based on their potential effects on open reading frame or regulatory regions. In addition, SNVs were filtered based on a minor allele frequency of <1% in the gnomAD database.[22] Subsequently, in the presence of a heterozygous pathogenic allele in one of the IRD-associated genes, all the NEAR and intronic variants in that gene were assessed to identify a second pathogenic allele.

The pathogenicity of missense variants was evaluated based on score thresholds of in silico prediction tools: PhyloP (range −14.1_6.4; predicted pathogenic ≥2.7),[23] CADD-PHRED (range 1_99; predicted pathogenic ≥15)[24] and Grantham (range 0_215; predicted pathogenic ≥80).[25] Variants that met all three criteria were investigated initially, followed by those that met the threshold scores of two in silico tools.

Non-coding splice site, NEAR splice and deep-intronic variants were evaluated as described previously[14] using algorithms including SpliceSiteFinder-like, MaxEntScan, GeneSplicer and Human Splicing Finder embedded in the Alamut Visual software V.2.10 (Interactive Biosoftware, Rouen, France; http://www.interactive-biosoftware.com).[26–30] In addition, the SpliceAI algorithm[31] was used to prioritise non-coding variants with at least

one delta-score above the threshold of 0.01 for gain or loss in a SDS or SAS for a putative in vitro functional analysis.[32] The input sequences for SpliceAI analysis included 500 bp upstream and downstream of a variant. Moreover, the pathogenicity score of the candidate non-coding variants was evaluated using functional analysis through Hidden Markov Models with extended features (FATHMM-XF V.2.3, range 0–1) to predict if a variant is functional or non-functional in inherited diseases.[33] A score >0.5 is predicted to be deleterious and score <0.5 are considered benign.[33]

In vitro splice assay

The functional analysis of a putative causal splice site variant, c.592-21A>T in *BBS1* (GenBank: NM_024649.4), was carried out using midigene splice assays. The wild-type and mutant constructs, with a 7.3 kb insert, were designed according to the previously described protocol.[34] In short, *BBS1* exons 4 to 9 of a genomic control DNA sample were amplified using primers, which both contain attB1 and attB2 tags at their 5′ ends to facilitate Gateway cloning. The generated wild-type construct served as a template to generate the mutant construct with c.592-21A>T by site-directed mutagenesis PCR. Subsequently, wild-type and mutant constructs were independently incorporated into the *pCI-NEO-RHO* Gateway-adapted vector to generate wild-type and mutant midigenes. Ultimately, both midigenes were separately transfected into HEK293T (Human embryonic kidney 293) cells and, after 48 hours of incubation, RNA was collected, and the transcripts were analysed by reverse transcription–polymerase chain reaction (RT-PCR) with primers in exons 6 and 9. All the primers designed for midigene splice assays are listed in online supplemental table S1. In the presence of multiple fragments after gel electrophoresis, they were quantified using Fiji software. Furthermore, the protein structure of the BBSome complex (PDB entry 6XT9), the BBS1 reference protein (Uniprot accession code Q8NFJ9), and the 10 amino acid and 44 amino acid deletions in the BBS1 were examined in silico using the YASARA tool (V.20.12.24) (http://www.yasara.org/).

RESULTS

Clinical findings

Pedigrees of the four studied families are shown in figure 1. Two families were of Dutch origin, one was from Irish and one from British ancestries, and all four probands presented with non-syndromic RP. An overview of clinical characteristics is provided in figure 2 and online supplemental table S2. Three out of four patients (patients A-II-2, B-II-1 and D-II-1) presented RP with early macular involvement indicated by loss of visual acuity followed by rod degeneration characteristics. Patient C-II-1 was diagnosed with RP. No additional BBS1-associated features[3] have been observed in patients A-II-2, C-II-1 and D-II-1, whereas patient B-II-1 showed a mild cognitive impairment (special-needs education and living with his parents). Furthermore, all four patients had no complaints related to cardiac, hepatic or renal features of BBS. Further phenotypic characteristics of four affected individuals can be found in supplemental data (online supplemental figures S1–S4).

Genetic findings

To investigate the presence of potential candidate variants in IRD-associated genes, we performed WGS for the probands of families A–C. After an initial comprehensive analysis of the coding and splice site regions in WGS data, we detected a common heterozygous c.1169T>G (p.(Met390Arg)) missense variant in *BBS1* in all probands, confirming the pre-screening findings in these cases. The c.1169T>G variant had a phyloP score of 6.411, a CADD_PHRED score of 26.1 and a Grantham Score of 91. In addition, the in silico missense prediction tools SIFT and MutationTaster considered this pathogenic variant as "Deleterious" and "Disease causing", respectively. The variant is the most common pathogenic *BBS1* variant with a total allele frequency of 0.001570 in GnomAD (allele frequency of 0.002773 in European non-Finnish population) and, when occurring in a homozygous state, is associated with a wide spectrum of phenotypes ranging from BBS to non-syndromic RP.[3] Since no other pathogenic coding or splice site candidate variants in other IRD-associated genes were observed, we further investigated *BBS1*. To do this, we performed an extensive analysis of the remaining 35 NEAR and intronic variants in *BBS1* in patients A-II-2, B-II-1 and C-II-1, leading to the identification of an overlapping heterozygous intronic variant, c.592-21A>T, which uniquely met the defined criteria. Based on the splice score prediction algorithms embedded in Alamut 2.10, the effect of c.592-21A>T on the nearby SAS is not damaging (a 0.1 lower score in NNSPLICE only [range: 0-1]) and the FATHMM-XF score was calculated 0.07 for this variant, which predicted to be benign. Interestingly, c.592-21A>T alters the recognition score for the Branch Point algorithm (range 0–100) from 73.5 in wild-type to zero in the mutant as a result of the removal of the branchpoint "A" nucleotide (figure 3A). Moreover, the SpliceAI algorithm predicts a loss of the SAS and SDS of exon 8 with delta scores of 0.44 and 0.04, respectively (online supplemental table S3). Moreover, we identified c.592-21A>T in a fourth case, family D (figure 1), who also carried c.1169T>G as a pathogenic variant identified through the MGCM 105 gene panel analysis. Segregation analysis confirmed that in the participating families, probands were compound heterozygous for c.1169T>G and

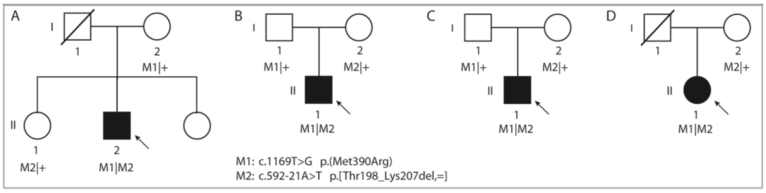


Figure 1 Pedigrees of four unrelated individuals in this study. The *BBS1* variants c.1169T>G and c.592-21A>T segregate with the disease in the four studied families. The arrow indicates the proband in each family.

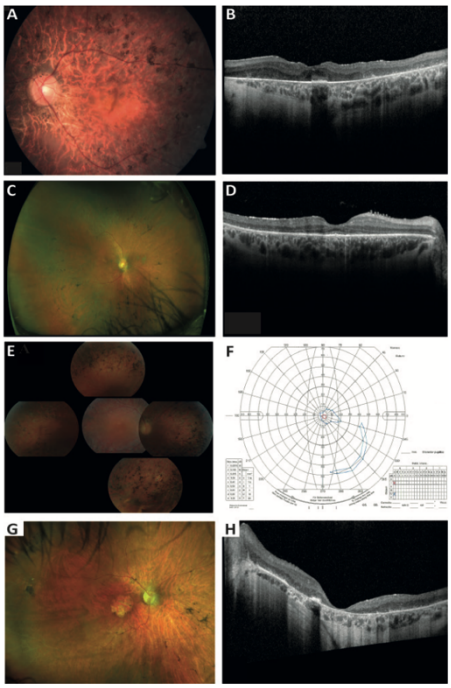


Figure 2 Clinical characterisations of four unrelated affected individuals included in this study. (A) In patient A-II-2, fundus image of the left eye showed attenuated retinal vessels and extensive atrophy in the posterior pole and periphery with intra-retinal pigment deposits. (B) In the same patient (A-II-2), spectral-domain optical coherence tomography (SD-OCT) of left eye reveals extensive retinal atrophy. (C) In patient B-II-1, funduscopy image showed optic disc pallor, attenuated blood vessels and bone spicule pigmentation in the right eye. (D) The SD-OCT of the right eye reveals extensive retinal atrophy in this patient (B-II-1). (E) In the third proband (patient C-II-1), ophthalmoscopy image showed disc pallor, extensive bone-spicule intra-retinal pigmentary deposits and vascular attenuation in the right eye. (F) Goldmann perimetry demonstrates very extensive visual field loss, with gross concentric constriction to the IV4e target and preservation of a narrow island of field temporally in the right eye of the patient C-II-1. (G) Optos colour fundus image showed symmetrical central macular atrophy, spicule formation, blood vessel attenuation, optic disc pallor and a left posterior subcapsular lens capacity in left eye of patient D-II-1. (H) In the same patient (D-II-1), SD-OCT showed central photoreceptor and retinal pigment epithelium loss bilaterally of the left eye.

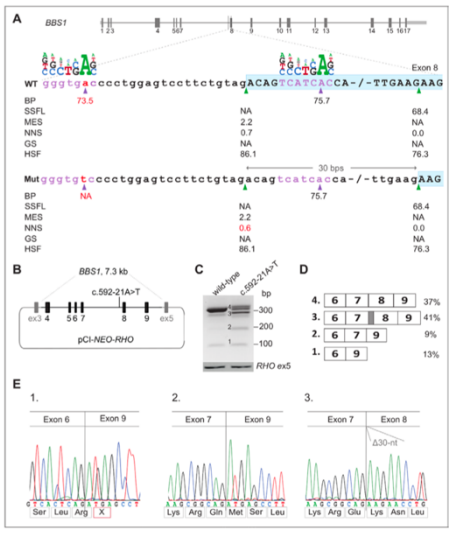


Figure 3 Molecular genetic characterisation of variant c.592-21A>T in *BBS1*. (A) Schematic representation of the *BBS1* gene and enlargement of the exon–intron boundary of exon 8 in the wild-type and mutant sequences. The SpliceSiteFinder-like (SSFL, range 0–100), MaxEntSCan (MES, range 0–12), GeneSplicer (GS, range 0–24) and Human Splicing Finder (HSF, range 0–100) scores for the splice acceptor site, and Branch Points (BP, 0–100) scores for the branchpoint sequence are indicated below each sequence. The green and purple triangles indicate the position of splice acceptor site and the branchpoint in both wild-type and mutant sequences. The purple sequences highlight the 7-mer branchpoint motif and the sequence above that shows the general branchpoint motif. The red "t" highlights the c.592-21A>T variant in the mutant sequence. The red numbers represent the altered scores compared to wild-type. (B) Schematic representation of the mutant *pCI-NEO-RHO* vector, containing exons 4–9 of *BBS1* flacked by *RHO* exons 3 and 5, which was used to transfect HEK293T cells with wild-type and mutant constructs, respectively. (C, D) The gel image of RT-PCR products of the wild-type midigene reveals the expected 300 bp wild-type fragment along with 95 bp fragment of natural exon 7/8 skipping. In the mutant midigene, two additional fragments were observed in which the larger one corresponds to a 270 bp fragment representing a 30 bp deletion of exon 8 and a smaller fragment of 168 bp corresponding to the complete deletion of exon 8. RT-PCR analysis of *RHO* exon 5 was performed as a control for efficient transfection. Fiji software was used for quantification of each fragment in mutant construct. (E) Sanger sequence analysis of the RT-PCR fragments confirmed the wild-type and exon 7/8 deletion in the wild-type midigene as well as exon 8 deletion, and the 30 bp deletion of exon 8 in the mutant. Mut, mutant; WT, wild-type.

c.592-21A>T, while available unaffected parents carried one of these two variants in a heterozygous state.

To assess pathogenicity of c.592-21A>T, midigene splice assays were performed using wild-type and mutant midigenes. After RNA isolation and RT-PCR analysis of wild-type midigenes, we detected a predicted 300 bp fragment containing *BBS1* wild-type mRNA (78% of PCR product quantified using Fiji software). In addition, we observed a 95 bp fragment (22% of PCR product) corresponding with exon 7/8 skipping of *BBS1* mRNA. This apparent natural exon skipping results in a stop-gain after the last amino acid of exon 6 (p.(Phe174*)).

Compared with the wild-type construct, two additional mRNA fragments, measuring 270 bp and 168 bp, were detected on transfection of the mutant midigene. The 300 bp wild-type fragment comprised 37% and the 95 bp fragment with exon 7/8 deletion

13% of the total RNA after quantification analysis using Fiji software. The mutant-specific fragments comprised 41% (270 bp) and 9% (168 bp) of total RNA, respectively (figure 3B–D and online supplemental table S4). Sanger sequencing verified that the 270 bp fragment corresponded to a 30-nucleotide deletion of exon 8 (p.(Thr198_Lys207del)), likely due to the use of a cryptic BPS at the c.601 position, while the 168 bp fragment corresponded to complete exon 8 skipping (p.(Thr198_Lys241del)).

We examined the effect of the in-frame deletion on the protein structure in silico using the YASARA tool (V.20.12.24) (http://www.yasara.org/). The 10 amino acids (p.(Thr198_Lys207del)) and 44 amino acids (p.(Thr198_Lys241del)) lacking due to the branchpoint variant are localised to the β-propeller domain of the BBS1 protein (online supplemental figure S5A). The residues are located in the core of the protein, and lack of these residues will likely have a strong effect and is very plausible to disrupt local protein folding. Specifically, the β-propeller domain may not be properly formed due to p.(Thr198_Lys207del) or p.(Thr198_Lys241del), suggesting loss of function or loss of interaction (online supplemental figure S5B–D). BBS1 is part of the BBSome complex (subunits BBS1, BBS4, BBS8, BBS9 and BBS18) and is known to directly interact with BBS4. Moreover, BBS3/ARL6 is shown to interact with the β-propeller of BBS1 and thereby essential for activation of the BBSome complex and its membrane recruitment into the cilium.[35]

DISCUSSION AND CONCLUSION

In this study, we identified a branchpoint variant, c.592-21A>T, in *BBS1*, which causes a complex splice defect. Branchpoint variants have been described previously for genes such as *COL5A1*, *LCAT*, *NF2* and *XPC*, involved in Ehlers-Danlos syndrome, Norum disease, neurofibromatosis type II and Xeroderma pigmentosum, respectively.[36–40] To our knowledge, this study details the first pathogenic variant disrupting a BPS in an IRD-associated gene. The c.592-21A>T variant was investigated through a midigene splice assay and shown to result in partial and complete deletion of exon 8 in the RNA (p.[Thr198_Lys207del,=]). Furthermore, natural exon 7/8 skipping was observed in both wild-type and mutant midigenes, further demonstrating that the impact of variants on splicing is not always binary. This information may be an important factor for consideration of disease presentation and prognosis. For example, the level of remaining wild-type transcript for the splicing gene *PRPF31* correlates with the level of disease severity in other retinal disorders[41] and, in some genes, the pathogenicity of genomic variants may be influenced by the presence of other genomic variants impacting gene expression levels.[42]

Thus far, only a few causative variants disrupting BPSs have been identified as such variants appear to be rare and, to date, difficult to recognise as the branchpoint is localised between 9 and 400 nucleotides upstream of the SAS. The branchpoint is embedded in a yUnAy-consensus sequence in humans, in which the branchpoint conserved A is underlined and the lowercase pyrimidines ('y') are less conserved compared with the U and A.[43] Pathogenic variants affecting numerous nucleotides of the consensus sequences of splice sites have been shown to lead to exon skipping.[40] Pathogenic branchpoint variants have also been shown to lead to the recognition of an alternative BPS with subsequent intron retention in the upstream region.[36 38] Conversely, newly recognised branchpoint consensus variants in the presence of a cryptic acceptor splice site and donor splice site downstream resulted in the inclusion of a pseudoexon.[38]

In our study, we hypothesise that c.592-21A>T would lead to exon 8 skipping based on the loss of the branchpoint 'A', which is supported by the SpliceAI prediction. Analysis of skipping of exon 8 in 9% of total RNA as well as the more abundant fragment (41%) showed a 30-nucleotide deletion of exon 8 due to the usage of a cryptic BPS within exon 8 and a naturally present nearby SAS within exon 8. We observed a moderately severe effect in the midigene splice assay for c.592-21A>T as 37% of wild-type fragment remained alongside the mutant fragments after quantification analysis. However, to determine the actual severity, photoreceptor-like cells derived from individuals harbouring this variant would have to be assessed. In a previous study, the analysis of photoreceptor precursor cells differentiated from stem cells derived from a patient with Stargardt disease showed aberrant *ABCA4* RNA (pseudo-exon insertion in mRNA) which was not visible in patient's fibroblasts nor in mutated midigenes transfected in HEK293T cells.[44] Likewise, a pseudo-exon insertion into the *CEP290* mRNA in a patient with Leber congenital amaurosis was more pronounced in patient-derived retinal organoids compared with fibroblasts.[45] The complex splice defect, including partial and complete exon deletion maintaining the reading frame, has not been previously observed as an effect attributed to a branchpoint variant. In addition, to the best of our knowledge, this is the first report of natural skipping of exons 7 and 8 in *BBS1* mRNA.

BBS1 variants have been associated with a spectrum of phenotypes ranging from non-syndromic RP to BBS.[3] Genotype–phenotype correlation between the type of pathogenic variant and the severity of the phenotype indicates that combinations of frameshift and/or nonsense variants lead to a more severe phenotype.[46] Generally, cases with *BBS1* pathogenic variants have a relatively mild phenotype compared with cases with pathogenic variants in other *BBS* genes.[47] For *BBS1*, the only established genotype–phenotype correlation is described for p.(Met390Arg), which is associated with mild eye abnormalities, non-syndromic RP or mild BBS,[3] whereas other BBS1 pathogenic variants may have a more severe phenotype.[1 2 47] The affected probands in our study can be classified as non-syndromic RP with macular involvement, with no additional non-ocular features in three individuals and only mild intellectual impairment in patient B-II-1. Similar to previous studies, we observe clinical features corresponding to non-syndromic RP in our patients all carrying p.(Met390Arg) and observe cognitive impairment indicating mild BBS in patient B-II:1. Also, cataract and nystagmus noted in patient A-II-2 and cataract in patient B-II-1 have been previously reported in cases with the homozygous p.(Met390Arg) variant.[3 47] In addition, we have observed macular involvement in our p.(Met390Arg) compound heterozygous cases, as is previously described by others.[48]

A previous study suggested a putative pathogenic variant in family A, being a hemizygous missense variant in *SYLT4*.[49] This was based on an in silico analysis. However, DNA from family members was not available at the time to confirm or exclude segregation. In the current study, we confirmed that *BBS1* is the cause of the disease in this family and *SYLT4* is no longer considered as the gene underlying RP.

The genetic finding in family D was enabled by detection of the c.592-21A>T variant using the MCGM 105 panel. Retrospectively, the c.592-21A>T could be detected in the pre-screening data of all cases (both WES and Target5000). However, it was unrecognised or remained unnoticed due to the strict criteria of variant prioritisation used in routine diagnostic genetic analysis. The c.592-21A>T pathogenic variant was identified in four unrelated families from three different origins and does not appear as a shared haplotype (data not shown). However, as it is a recurrent variant it may be considered a suitable target for genetic therapy to redirect the splicing machinery and rescue the clinical phenotype. One of these approaches is modified small nuclear RNAs (snRNAs), which have been shown to be an effective technique to rescue exon skipping in mature mRNA. As

an example, modified U1 snRNAs bind to the complementary sequence in intronic region downstream of the exon and rescue splicing defects which are associated with exon skipping.[50] However, application of modified snRNAs molecules to rescue exon skipping defects caused by branchpoint pathogenic variants still needs to be further investigated.

In conclusion, we have identified a complex splice defect caused by a pathogenic branchpoint variant c.592-21A>T in *BBS1* in *trans* with the most frequent pathogenic *BBS1* missense variant p.(Met390Arg) in four unrelated individuals. To our knowledge, it is the first report of a pathogenic branchpoint variant in an IRD-associated gene leading to a non-syndromic RP phenotype, as well as the first report of natural exon 7/8 skipping event in *BBS1* mRNA. Our study stresses the relevance of the identification of causative non-coding variants, in particular branchpoint variants, and confirmation of their pathogenicity in order to provide a genetic diagnosis in patients with an inherited retinal dystrophy.

Author affiliations
[1]Department of Human Genetics, Radboudumc, Nijmegen, The Netherlands
[2]Donders Institute for Brain Cognition and Behaviour, RadboudUMC, Nijmegen, The Netherlands
[3]School of Genetics & Microbiology, Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland
[4]Department of Ophthalmology, Radboudumc, Nijmegen, The Netherlands
[5]Eye Hospital Rotterdam, Rotterdam Ophthalmic Institute, Rotterdam, The Netherlands
[6]Radboud Institute of Molecular Life Sciences, RadboudUMC, Nijmegen, The Netherlands
[7]North West Genomic Laboratory Hub, Manchester Centre for Genomic Medicine, Manchester University Hospitals NHS Foundation Trust, St Mary's Hospital, Manchester, UK
[8]Division of Evolution and Genomic Sciences, Neuroscience and Mental Health Domain, The University of Manchester Faculty of Biology Medicine and Health, Manchester, UK
[9]MRC Human Genetics Unit, University of Edinburgh Western General Hospital, Edinburgh, UK
[10]Princess Alexandra Eye Pavilion, Department of Ophthalmology, University of Edinburgh Western General Hospital, Edinburgh, UK
[11]South East of Scotland Clinical Genetics Service, Western General Hospital, Edinburgh, UK
[12]Research Foundation, Royal Victoria Eye and Ear Hospital, Dublin, Ireland

Correction notice This article has been corrected since it was published Online First. The name of author Dr L Ingeborgh van den Born has been corrected.

Twitter Susanne Roosing @susanneroosing

Acknowledgements We thank the affected individuals and their families for participating in this study and our funding bodies.

Contributors ZF performed data analyses, sequencing analyses, RNA expression studies. CHZL, LIvdB, CBH, CR, AKL and PK collected clinical cases and performed clinical examinations of patients. CG and JC provided infrastructure, and bioinformatic expert input. JF, FPMC, AD and JME provided strategic support, expert input and supervision for the project. ZF, LW, FPMC and SR contributed significantly to design of the study. ZF, LW, PK and SR wrote the manuscript. All authors reviewed and approved the manuscript.

Funding The work of ZF and CHZL is funded by the Foundation Fighting Blindness USA Project Program Award, grant no. PPA-0517-0717-RAD (to FPMC, SR and CBH). The research was supported by the Algemene Nederlandse Vereniging ter Voorkoming van Blindheid, Oogfonds, Landelijke Stichting voor Blinden en Slechtzienden; Rotterdamse Stichting Blindenbelangen, Stichting Blindenhulp, Stichting tot Verbetering van het Lot der Blinden and Stichting Blinden-Penning (to SR and FPMC). The work of LW is supported by Fighting Blindness Ireland, grant FB18CRE (to FPMC, SR and JF). RM is supported by the Wellcome Trust (Fellowship number 219607/Z/19/Z) and the Royal College of Surgeons (Edinburgh). JME is funded by a postdoctoral research fellowship from the Health Education England Genomics Education Programme (HEE GEP).

Disclaimer The views expressed in this publication are those of the authors and not necessarily those of the HEE GEP.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The study adhered to the tenets of the Declaration of Helsinki and was approved by the local ethics committees of Radboud University Medical Center, Nijmegen, The Netherlands; Rotterdam Eye Hospital, Rotterdam, The Netherlands; The Research Foundation, The Royal Victoria Eye and Ear Hospital, Dublin, Ireland; and the Manchester Centre for Genomic Medicine, Manchester, UK. Informed consent was obtained from patients prior to DNA analysis in this study.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. All data relevant to the study are included in the article or uploaded as online supplemental information. The pathogenic variant data is submitted to Leiden Open Variation Database (LOVD).

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

ORCID iDs
Zeinab Fadaie http://orcid.org/0000-0003-0440-3846
Laura Whelan http://orcid.org/0000-0002-0537-4349
Adrian Dockery http://orcid.org/0000-0002-3423-2542
Catherina H Z Li http://orcid.org/0000-0003-4449-9360
L Ingeborgh van den Born http://orcid.org/0000-0002-6375-3339
Carel B Hoyng http://orcid.org/0000-0002-1358-5963
Christian Gilissen http://orcid.org/0000-0003-1693-9699
Jordi Corominas http://orcid.org/0000-0003-1736-1278
Charlie Rowlands http://orcid.org/0000-0001-5333-4846
Roly Megaw http://orcid.org/0000-0001-5605-4540
Anne K Lampe http://orcid.org/0000-0002-8517-8240
Frans P M Cremers http://orcid.org/0000-0002-4954-5592
Gwyneth Jane Farrar http://orcid.org/0000-0002-9959-1992
Jamie M Ellingford http://orcid.org/0000-0003-1137-9768
Paul F. Kenna http://orcid.org/0000-0002-3964-4167
Susanne Roosing http://orcid.org/0000-0001-9038-0067

REFERENCES

1 Mykytyn K, Nishimura DY, Searby CC, Shastri M, Yen H-jan, Beck JS, Braun T, Streb LM, Cornier AS, Cox GF, Fulton AB, Carmi R, Lüleci G, Chandrasekharappa SC, Collins FS, Jacobson SG, Heckenlively JR, Weleber RG, Stone EM, Sheffield VC. Identification of the gene (BBS1) most commonly involved in Bardet-Biedl syndrome, a complex human obesity syndrome. *Nat Genet* 2002;31:435–8.
2 Mykytyn K, Nishimura DY, Searby CC, Beck G, Bugge K, Haines HL, Cornier AS, Cox GF, Fulton AB, Carmi R, Iannaccone A, Jacobson SG, Weleber RG, Wright AF, Riise R, Hennekam RCM, Lüleci G, Berker-Karauzum S, Biesecker LG, Stone EM, Sheffield VC. Evaluation of complex inheritance involving the most common Bardet-Biedl syndrome locus (BBS1). *Am J Hum Genet* 2003;72:429–37.
3 Estrada-Cuzcano A, Koenekoop RK, Senechal A, De Baere EBW, de Ravel T, Banfi S, Kohl S, Ayuso C, Sharon D, Hoyng CB, Hamel CP, Leroy BP, Ziviello C, Lopez I, Bazinet A, Wissinger B, Sliesoraityte I, Avila-Fernandez A, Littink KW, Vingolo EM, Signorini S, Banin E, Mizrahi-Meissonnier L, Zrenner E, Kellner U, Collin RWJ, den Hollander AI, Cremers FPM, Klevering BJ. BBS1 mutations in a wide spectrum of phenotypes ranging from nonsyndromic retinitis pigmentosa to Bardet-Biedl syndrome. *Arch Ophthalmol* 2012;130:1425–32.
4 Weihbrecht K, Goar WA, Pak T, Garrison JE, DeLuca AP, Stone EM, Scheetz TE, Sheffield VC. Keeping an eye on Bardet-Biedl syndrome: a comprehensive review of the role of Bardet-Biedl syndrome genes in the eye. *Med Res Arch* 2017;5. doi:10.18103/mra.v5i9.1526. [Epub ahead of print: 18 Sep 2017].
5 Abu Safieh L, Aldahmesh MA, Shamseldin H, Hashem M, Shaheen R, Alkuraya H, Al Hazzaa SAF, Al-Rajhi A, Alkuraya FS. Clinical and molecular characterisation of Bardet-Biedl syndrome in consanguineous populations: the power of homozygosity mapping. *J Med Genet* 2010;47:236–41.
6 Nachury MV, Loktev AV, Zhang Q, Westlake CJ, Peränen J, Merdes A, Slusarski DC, Scheller RH, Bazan JF, Sheffield VC, Jackson PK. A core complex of BBS proteins cooperates with the GTPase Rab8 to promote ciliary membrane biogenesis. *Cell* 2007;129:1201–13.
7 Cremers FP, van de Pol DJ, van Driel M, den Hollander AI, van Haren FJ, Knoers NV, Tijmes N, Bergen AA, Rohrschneider K, Blankenagel A, Pinckers AJ, Deutman AF, Hoyng CB. Autosomal recessive retinitis pigmentosa and cone-rod dystrophy caused by splice site mutations in the Stargardt's disease gene ABCR. *Hum Mol Genet* 1998;7:355–62.

8 Riazuddin SA, Iqbal M, Wang Y, Masuda T, Chen Y, Bowne S, Sullivan LS, Waseem NH, Bhattacharya S, Daiger SP, Zhang K, Khan SN, Riazuddin S, Hejtmancik JF, Sieving PA, Zack DJ, Katsanis N. A splice-site mutation in a retina-specific exon of BBS8 causes nonsyndromic retinitis pigmentosa. *Am J Hum Genet* 2010;86:805–12.
9 Anna A, Monika G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet* 2018;59:253–68.
10 Aebi M, Hornig H, Padgett RA, Reiser J, Weissmann C. Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. *Cell* 1986;47:555–65.
11 Jurica MS, Moore MJ. Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* 2003;12:5–14.
12 Sheth N, Roca X, Hastings ML, Roeder T, Krainer AR, Sachidanandam R. Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res* 2006;34:3955–67.
13 Speicher M, Antonarakis SE, Motulsky AG. *Vogel and Motulsky's human genetics: problems and approaches*. Springer Science & Business Media, 2009: 139–63.
14 Fadaie Z, Khan M, Del Pozo-Valero M, Cornelis SS, Ayuso C, Cremers FPM, Roosing S, The Abca Study Group. Identification of splice defects due to noncanonical splice site or deep-intronic variants in ABCA4. *Hum Mutat* 2019;40:2365–76.
15 McCulloch DL, Marmor MF, Brigell MG, Hamilton R, Holder GE, Tzekov R, Bach M. ISCEV standard for full-field clinical electroretinography (2015 update). *Doc Ophthalmol* 2015;130:1–12.
16 Ellingford JM, Barton S, Bhaskar S, O'Sullivan J, Williams SG, Lamb JA, Panda B, Sergouniotis PI, Gillespie RL, Daiger SP, Hall G, Gale T, Lloyd IC, Bishop PN, Ramsden SC, Black GCM. Molecular findings from 537 individuals with inherited retinal disease. *J Med Genet* 2016;53:761–7.
17 Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 2016;32:292–4.
18 Farek J, Hughes D, Mansfield A, Krasheninina O, Nasser W, Sedlazeck FJ, Khan Z, Venner E, Metcalf G, Boerwinkle E, Muzny DM, Gibbs RA, Salerno W. xAtlas: scalable small variant calling across heterogeneous next-generation sequencing experiments. *bioRxiv* 2018;295071.
19 Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, Cox AJ, Kruglyak S, Saunders CT. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* 2016;32:1220–2.
20 Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, Janoueix-Lerosey I, Delattre O, Barillot E. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* 2012;28:423–5.
21 Dolzhenko E, van Vugt JJFA, Shaw RJ, Bekritsky MA, van Blitterswijk M, Narzisi G, Ajay SS, Rajan V, Lajoie BR, Johnson NH, Kingsbury Z, Humphray SJ, Schellevis RD, Brands WJ, Baker M, Rademakers R, Kooyman M, Tazelaar GHP, van Es MA, McLaughlin R, Sproviero W, Shatunov A, Jones A, Al Khleifat A, Pittman A, Morgan S, Hardiman O, Al-Chalabi A, Shaw C, Smith B, Neo EJ, Morrison K, Shaw PJ, Reeves C, Winterkorn L, Wexler NS, Housman DE, Ng CW, Li AL, Taft RJ, van den Berg LH, Bentley DR, Veldink JH, Eberle MA, US–Venezuela Collaborative Research Group. Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome Res* 2017;27:1895–903.
22 Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Neale BM, Daly MJ, MacArthur DG, Genome Aggregation Database Consortium. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;581:434–43.
23 Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* 2010;20:110–21.
24 Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46:310–5.
25 Grantham R. Amino acid difference formula to help explain protein evolution. *Science* 1974;185:862–4.
26 Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* 1987;15:7155–74.
27 Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol* 1997;4:311–23.
28 Pertea M, Lin X, Salzberg SL. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res* 2001;29:1185–90.
29 Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004;11:377–94.
30 Desmet F-O, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009;37:e67.
31 Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, Kosmicki JA, Arbelaez J, Cui W, Schwartz GB, Chow ED, Kanterakis E, Gao H, Kia A, Batzoglou S, Sanders SJ, Farh KK-H. Predicting splicing from primary sequence with deep learning. *Cell* 2019;176:e24:535–48.
32 Khan M, Cornelis SS, Pozo-Valero MD, Whelan L, Runhart EH, Mishra K, Bults F, AlSwaiti Y, AlTalbishi A, De Baere E, Banfi S, Banin E, Bauwens M, Ben-Yosef T, Boon CJF, van den Born LI, Defoort S, Devos A, Dockery A, Dudakova L, Fakin A, Farrar GJ, Sallum JMF, Fujinami K, Gilissen C, Glavač D, Gorin MB, Greenberg J, Hayashi T, Hettinga YM, Hoischen A, Hoyng CB, Hufendiek K, Jägle H, Kamakari S, Karali M, Kellner U, Klaver CCW, Kousal B, Lamey TM, MacDonald IM, Matynia A, McLaren TL, Mena MD, Meunier I, Miller R, Newman H, Ntozini B, Oldak M, Pieterse M, Podhajcer OL, Puech B, Ramesar R, Rüther K, Salameh M, Salles MV, Sharon D, Simonelli F, Spital G, Steehouwer M, Szaflik JP, Thompson JA, Thuillier C, Tracewska AM, van Zweeden M, Vincent AL, Zanlonghi X, Liskova P, Stöhr H, Roach JND, Ayuso C, Roberts L, Weber BHF, Dhaenens C-M, Cremers FPM. Resolving the dark matter of ABCA4 for 1054 Stargardt disease probands through integrated genomics and transcriptomics. *Genet Med* 2020;22:1235–46.
33 Rogers MF, Shihab HA, Mort M, Cooper DN, Gaunt TR, Campbell C. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics* 2018;34:511–3.
34 Sangermano R, Khan M, Cornelis SS, Richelle V, Albert S, Garanto A, Elmelik D, Qamar R, Lugtenberg D, van den Born LI, Collin RWJ, Cremers FPM. *ABCA4* midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome Res* 2018;28:100–10.
35 Mourão A, Nager AR, Nachury MV, Lorentzen E. Structural basis for membrane targeting of the BBSome by ARL6. *Nat Struct Mol Biol* 2014;21:1035–41.
36 Kuivenhoven JA, Weibusch H, Pritchard PH, Funke H, Benne R, Assmann G, Kastelein JJ. An intronic mutation in a lariat branchpoint sequence is a direct cause of an inherited human disorder (fish-eye disease). *J Clin Invest* 1996;98:358–64.
37 Burrows NP, Nicholls AC, Richards AJ, Luccarini C, Harrison JB, Yates JR, Pope FM. A point mutation in an intronic branch site results in aberrant splicing of COL5A1 and in Ehlers-Danlos syndrome type II in two British families. *Am J Hum Genet* 1998;63:390–8.
38 De Klein A, Riegman PH, Bijlsma EK, Heldoorn A, Muijtjens M, den Bakker MA, Avezaat CJ, Zwarthoff EC. A G-->A transition creates a branch point sequence and activation of a cryptic exon, resulting in the hereditary disorder neurofibromatosis 2. *Hum Mol Genet* 1998;7:393–8.
39 Khan SG, Metin A, Gozukara E, Inui H, Shahlavi T, Muniz-Medina V, Baker CC, Ueda T, Aiken JR, Schneider TD, Kraemer KH. Two essential splice lariat branchpoint sequences in one intron in a xeroderma pigmentosum DNA repair gene: mutations result in reduced XPC mRNA levels that correlate with cancer risk. *Hum Mol Genet* 2004;13:343–52.
40 Khan SG, Yamanegi K, Zheng Z-M, Boyle J, Imoto K, Oh K-S, Baker CC, Gozukara E, Metin A, Kraemer KH. XPC branch-point sequence mutations disrupt U2 snRNP binding, resulting in abnormal pre-mRNA splicing in xeroderma pigmentosum patients. *Hum Mutat* 2010;31:167–75.
41 Wheway G, Douglas A, Baralle D, Guillot E. Mutation spectrum of PRPF31, genotype-phenotype correlation in retinitis pigmentosa, and opportunities for therapy. *Exp Eye Res* 2020;192:107950.
42 Castel SE, Cervera A, Mohammadi P, Aguet F, Reverter F, Wolman A, Guigo R, Iossifov I, Vasileva A, Lappalainen T. Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk. *Nat Genet* 2018;50:1327–34.
43 Gao K, Masuda A, Matsuura T, Ohno K. Human branch point consensus sequence is yUnAy. *Nucleic Acids Res* 2008;36:2257–67.
44 Albert S, Garanto A, Sangermano R, Khan M, Bax NM, Hoyng CB, Zernant J, Lee W, Allikmets R, Collin RWJ, Cremers FPM. Identification and rescue of splice defects caused by two neighboring deep-intronic ABCA4 mutations underlying Stargardt disease. *Am J Hum Genet* 2018;102:517–27.
45 Parfitt DA, Lane A, Ramsden CM, Carr A-JF, Munro PM, Jovanovic K, Schwarz N, Kanuga N, Muthiah MN, Hull S, Gallo J-M, da Cruz L, Moore AT, Hardcastle AJ, Coffey PJ, Cheetham ME. Identification and correction of mechanisms underlying inherited blindness in human iPSC-derived optic cups. *Cell Stem Cell* 2016;18:769–81.
46 Ece Solmaz A, Onay H, Atik T, Aykut A, Cerrah Gunes M, Ozalp Yuregir O, Bas VN, Hazan F, Kirbiyik O, Ozkinay F. Targeted multi-gene panel testing for the diagnosis of Bardet Biedl syndrome: identification of nine novel mutations across BBS1, BBS2, BBS4, BBS7, BBS9, BBS10 genes. *Eur J Med Genet* 2015;58:689–94.
47 Castro-Sánchez S, Álvarez-Satta M, Cortón M, Guillén E, Ayuso C, Valverde D. Exploring genotype–phenotype relationships in Bardet-Biedl syndrome families. *J Med Genet* 2015;52:503–13.
48 Azari AA, Aleman TS, Cideciyan AV, Schwartz SB, Windsor EAM, Sumaroka A, Cheung AY, Steinberg JD, Roman AJ, Stone EM, Sheffield VC, Jacobson SG. Retinal disease expression in Bardet-Biedl syndrome-1 (BBS1) is a spectrum from maculopathy to retina-wide degeneration. *Invest Ophthalmol Vis Sci* 2006;47:5004–10.
49 Astuti GDN, van den Born LI, Khan MI, Hamel CP, Bocquet B, Manes G, Quinodoz M, Ali M, Toomes C, McKibbin M, El-Asrag ME, Haer-Wigman L, Inglehearn CF, Black GCM, Hoyng CB, Cremers FPM, Roosing S. Identification of inherited retinal disease-associated genetic variants in 11 candidate genes. *Genes* 2018;9. doi:10.3390/genes9010021. [Epub ahead of print: 10 Jan 2018].
50 Buratti E, Baralle D. Novel roles of U1 snRNP in alternative splicing regulation. *RNA Biol* 2010;7:412–9.

Appendix 4

Stephenson *et al. Orphanet J Rare Dis* (2021) 16:200
https://doi.org/10.1186/s13023-021-01841-1

Orphanet Journal of Rare Diseases

LETTER TO THE EDITOR Open Access

Target 5000: a standardized all-Ireland pathway for the diagnosis and management of inherited retinal degenerations

Kirk A. J. Stephenson[1*], Julia Zhu[1], Niamh Wynne[2], Adrian Dockery[3], Rebecca M. Cairns[4], Emma Duignan[2], Laura Whelan[3], Conor P. Malone[2], Hilary Dempsey[2], Karen Collins[2], Shana Routledge[1], Rajiv Pandey[1], Elaine Crossan[1,5], Jacqueline Turner[1], James J. O'Byrne[1], Laura Brady[6], Giuliana Silvestri[4], Paul F. Kenna[2,3], G. Jane Farrar[3] and David J. Keegan[1]

Abstract

Introduction: Inherited retinal degenerations (IRD) are rare genetic disorders with > 300 known genetic loci, manifesting variably progressive visual dysfunction. IRDs were historically underserved due to lack of effective interventions. Many novel therapies will require accurate diagnosis (phenotype and genotype), thus an efficient and effective pathway for assessment and management is required.

Methods: Using surveys of existing practice patterns and advice from international experts, an all-Ireland IRD service (Target 5000) was designed. Detailed phenotyping was followed by next generation genetic sequencing in both a research and accredited laboratory. Unresolved pedigrees underwent further studies (whole gene/whole exome/whole genome sequencing). Novel variants were interrogated for pathogenicity (cascade screening, in silico analysis, functional studies). A multidisciplinary team (MDT; ophthalmologists, physicians, geneticists, genetic counsellors) reconciled phenotype with genotype. A bespoke care plan was created for each patient comprising supports, existing interventions, and novel therapies/clinical trials.

Results and discussion: Prior to Target 5000, a significant cohort of patients were not engaged with healthcare/support services due to lack of effective interventions. Pathogenic or likely pathogenic variants in IRD-associated genes were detected in 62.3%, with 11.6% having variants of unknown significance. The genotyping arm of Target 5000 allowed a 42.73% cost saving over independent testing, plus the value of MDT expertise/processing. Partial funding has transferred from charitable sources to government resources.

Conclusion: Target 5000 demonstrates efficacious and efficient clinical/genetic diagnosis, while discovering novel IRD-implicated genes/variants and investigating mechanisms of disease and avenues of intervention. This model could be used to develop similar IRD programmes in small/medium-sized nations.

Keywords: Inherited retinal degenerations, Retinal dystrophy, Ocular genetics, Genetic diagnosis, Clinical diagnostic algorithm, Public and patient involvement

Introduction

Inherited retinal degenerations (IRDs) are a group of rare eye diseases that result in variable and progressive vision loss. IRDs have a heterogeneous phenotype and genotype with over 300 causal genes identified to

*Correspondence: kirkstephenson@hotmail.com
[1] Mater Clinical Ophthalmic Genetics Unit, Mater Misericordiae University Hospital, Dublin, Ireland
Full list of author information is available at the end of the article

© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

date [1], the first IRD-implicated gene (*RHO*) having been identified by co-authors in Dublin [2, 3]. Developed with support from international colleagues [4–7], a systematic all-Ireland plan was proposed for the evaluation and management of IRDs. The goal of this programme was to coordinate medical, academic and financial resources to achieve collaborative validated clinical and genetic diagnosis, to introduce relevant supports and to achieve preparedness for clinical trials and novel therapies both in Ireland [8] and internationally [9–12].

Due to the systematic management of more common eye conditions (i.e., diabetic retinopathy (DR) screening programmes), IRDs have surpassed DR as the leading cause of blind registration in adults < 64 years old in England and Wales [13]. Treatment algorithms have been established for more prevalent eye diseases (e.g. DR, 1.7% of global population [14], age-related macular degeneration, 7.2% of Irish population > 50 years [15]), while IRDs do not have a standardized approach. Despite IRDs being rare diseases, they accrue an €80 million per annum socioeconomic burden [16], with patients often seeking out and paying for expertise abroad at their own initiative. Ophthalmologists (and all eye care professionals) are obliged to advocate for the IRD population (3:10,000) by striving for a similar world-class standard of care with efficient use of available resources.

The ideal pathway is efficient (short turnaround time, i.e., avoiding prolonged delay and the consequent progression of retinal atrophy), accurate (correct clinical and genetic diagnosis) and effective (allowing access to relevant interventions, both supportive and disease-modifying). Combs et al. defined three main output requirements from an IRD focus group. The first was diagnosis and prognosis, allowing relevant family planning and targeting access to clinical research, trials, and treatments. The second requirement was psychological support (i.e., in adjusting to and coping with diagnosis) and the third was practical physical supports (e.g., financial, education, low vision). Respondents to this survey and a recent Irish study expressed dissatisfaction at the lack of integration between health and social care, which may explain lack of representative engagement between people with IRDs and health services [4, 16]. A set of core goals was distilled to best utilise the limited resources available for IRD management in Ireland:

1. Genetically resolve IRD pedigrees,
2. Act on modifiable disease,
3. Offer relevant supports, and
4. Provide access to clinical trials of novel therapeutics when appropriate.

IRDs are often heterogeneous with a similar clinical phenotype caused by mutations in multiple possible genes (Fig. 1). Likewise, a single gene (e.g., *ABCA4* OMIM*601691) may manifest as variable phenotypes (e.g., Cone-rod dystrophy-3, #604116; Fundus flavimaculatus #248200; Retinal dystrophy, early-onset severe, #248200; Retinitis pigmentosa-19, #601718; Stargardt disease-1, #248200; Macular degeneration, age-related-2, #153800). Genetic pleiotropy in IRDs (e.g. *USH2A* gene, *608400) may manifest with syndromic (Usher Syndrome type 2A, #276901) or non-syndromic (retinitis pigmentosa-39, #613809) features dependent on the position and deleteriousness of the variants within the gene [17]. Given the complexity of achieving an accurate diagnosis, a coordinated and qualified multidisciplinary team (MDT) approach was adopted to reach an evidence-based consensus for diagnosis and management of each case. An accurate genetic diagnosis is critical in coordinating career/family-planning choices, determining prognosis and accessing gene-specific therapies [18]. This requires detailed phenotyping (ocular and systemic), affordable and accurate clinically accredited genotyping, clinical genetics oversight, genetic counselling (GC), and connections to international networks for second opinions and access to clinical trials/treatments where available.

Adding to the complexity of accurate IRD diagnosis is the lack of agreed common terminology. Sergounotis et al. have contributed to clarification of this with the updated ocular human phenotype ontology and Orphanet ontology [19, 20]. Similarly, the American College of Medical Genetics (ACMG) grading is used to determine the likelihood of pathogenicity of detected genetic variants, these criteria being the current global standard [21]. International collaborations are required for second opinions from world-experts as well as access to clinical trials. Relationships in Europe were formalised with the founding of the European Reference Network for Rare Eye Disease (ERN-EYE) in which electronic IRD registers can be shared both for clinical and research purposes [22]. Using the above standardized clinical and genetic lexicon, meaningful outputs can be derived from this communal resource. The input of a physician/geneticist is critical, as the focus in syndromic disease is not only on visual prognosis but on detection and intervention for significant systemic disease (e.g., cardiac, metabolic, and renal disease).

For the first time, treatment is available for a small subset of IRDs, the first being approved for biallelic *RPE65* retinal degeneration in America (2017) and Europe (2018) [23] while therapies for some of the more frequent IRD phenotypes (e.g. choroideraemia, achromatopsia, X-linked retinitis pigmentosa, X-linked retinoschisis) are

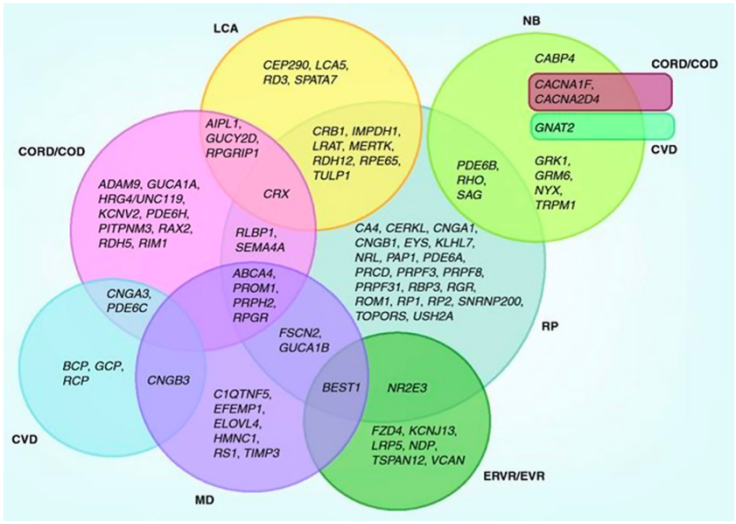


Fig. 1 Reproduced with permission from Sutherland et al. [18]. This figure highlights the pleiotropic nature of genes implicated in IRDs with 1 gene potentially manifesting as multiple/overlapping phenotypes in different individuals. Likewise, a single phenotype may have multiple genetic aetiologies. (CORD/COD Cone-Rod Dystrophy/Cone Dystrophy, CVD colour vision defects, ERVR/EVR Erosive and exudative vitreoretinopathies, LCA Leber Congenital Amaurosis, MD macular dystrophy, NB night blindness, RP retinitis pigmentosa)

in advanced human clinical trials [9–12]. Thus, a major goal of programme participation was "clinical trial/therapy preparedness" (i.e., full clinical and genetic characterization to meet standard inclusion criteria for clinical trials where available). Although it began as a genetic research study, as determination of genotype has become mandatory for both precise diagnosis and eligibility for gene-specific trials of novel treatments, by necessity, Target 5000 has developed into a clinical programme.

This publication outlines the assessment of the pre-existing infrastructure for IRD management in the Republic of Ireland and the use of these data to develop and implement a purpose-built framework to assess and provide world-class care for IRDs on the island of Ireland (Northern Ireland and Republic of Ireland).

Methods

In 2009, an expert working group of Irish IRD specialists (clinicians and scientists), with input from IRD patient groups, met to develop the blueprint of the Target 5000 programme which was to be run at 3 clinical sites (2 in Dublin, Ireland and 1 in Belfast, Northern Ireland) and 1 academic research laboratory (Dublin, Ireland). A consensus decision was made to start a common standardized programme for IRD assessment with a national register and individualized care plans. This genetic characterization study was approved by the institutional review boards of the participating hospitals (Mater Misericordiae University Hospital, Dublin, Ireland; Royal Victoria Eye and Ear Hospital, Dublin, Ireland; and later the Belfast Health and Social Care Trust Hospitals, Belfast, Northern Ireland) and university laboratory (Ocular Genetics Unit, Trinity College Dublin, Ireland). Funding was initially sourced from patient-initiated and other charitable sources (e.g., Health Research Board Ireland, Science Foundation Ireland, Health Research Charities Ireland (HRCI), Fighting Blindness Ireland. Programme funding has now been partly transferred to the national public health service (Health Service Executive, Ireland; HSE).

In 2014, a survey was conducted (Google Forms, Google LLC, Additional file 1) in 12 clinical ophthalmic sites in Ireland to assess the state of the pre-existing infrastructure for IRD assessment, to identify any areas of concern or omission and to inform the creation of a new national IRD pathway.

Patient recruitment was by referral from ophthalmologists, optometrists, and patient organizations to facilitate

open access to participation. Clinical assessment was performed and, where a clinical diagnosis of IRD was suspected, patients were included in the genotyping and follow up arm of the study.

All patients were seen in IRD-specific subspecialty clinics. Meticulous phenotyping was performed including thorough history (ophthalmic, systemic, and family), comprehensive ophthalmic examination (visual acuity, dilated slit lamp biomicroscopy, intraocular pressure measurement), and ancillary clinical tests (multimodal retinal imaging (MMI), visual field assessment and ISCEV-standard electrophysiology). When available, historical data was included for analysis of progression. Initial genotyping was performed in the affiliated research-grade laboratory (Ocular Genetics Unit, Trinity College Dublin) by a panel-based target capture next generation sequencing (NGS) approach assessing all known IRD-associated genes as described in previous publications [24–27]. Genetically unresolved patients were further investigated with whole gene (i.e., phenotypes typically associated with 1 gene, e.g., Stargardt disease/ABCA4, Choroideraemia/CHM, etc.) [28], whole exome (i.e., reanalysis of unresolved pedigrees for new genes) or whole genome sequencing (i.e., phenotypes associated with many genotypes but not resolved with NGS/WES, e.g., retinitis pigmentosa) where appropriate [29]. Prime candidate variants from this initial assessment were confirmed by clinically accredited laboratories (Manchester Centre for Genomic Medicine or Blueprint Genetics, Helsinki Finland).

Confirmed variants were discussed using an MDT approach (ophthalmic clinicians, systemic physicians, geneticists, genetic counsellors). Published evidence, online databases (e.g. RetNet, ClinVar) and the ACMG criteria were used to confirm phenotype-genotype matches in line with the AAO guidelines [30] prior to GC sessions. Gene/condition-specific care plans were created for each genetic diagnosis (see below and Fig. 2).

Results

The survey of the prior IRD standard of care confirmed the national absence of a standardized, coordinated approach to assessment. No formal interdepartmental register, nor electronic medical record, of patients and their diagnoses existed. Patients were often referred, at their own expense, to international experts for workup and management. Extrapolations from the survey of 12/16 (75%) responding ophthalmic centres revealed that 52% of the estimated IRD population were not being reviewed under the pre-existing conditions. Patients were assessed at IRD subspecialty clinics in only 33% of visits (vs 63% in the UK [5]), these centres being the founding collaborator Target 5000 clinical units. These subspecialist units had additional resources in terms of clinical IRD experience, greater time per patient (179% more time per patient visit, mean 47 versus 26.25 min,

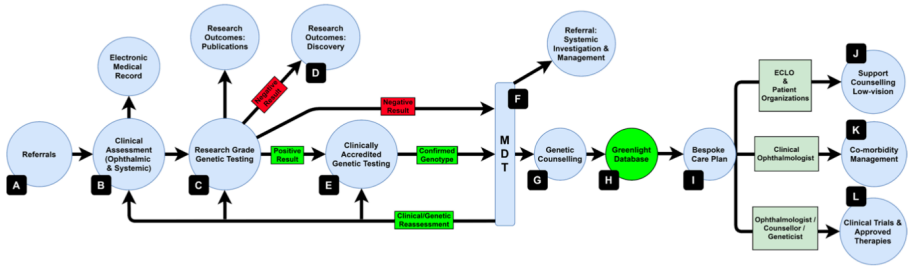


Fig. 2 Schematic diagram of the Target 5000 IRD algorithm. The above figure illustrates the pathway employed by Target 5000 to phenotype and genotype patients with IRDs in Ireland. Following (A) referral and (B) clinical assessment, patients are (C) genotyped via an NGS panel-based approach of all known IRD-implicated genes at a research level. If the NGS panel test is negative they go onto (D) the discovery arm of Target 5000 including whole gene/exome/genome sequencing as appropriate. Novel variants are assessed with tools including cascade screening, in silico analysis and functional studies. If a candidate variant is identified, a biobanked sample is (E) sent to a clinically accredited laboratory for confirmatory testing. The phenotype is reconciled with the confirmed genotype via (F) clinical/genetic MDT discussion. If required, clinical and/or genetic reassessment or referral for systemic investigation/management is initiated at this stage. The clinical genetics team arrange a (G) genetic counselling session with the patient where the most appropriate interventions for each individual are discussed. Individuals meeting the prerequisites for this stage are eligible for entry on (H) the "Greenlight Database" signifying adequate phenotypic/genotypic characterization for (L) existing and novel (i.e., clinical trials) therapies. Underpinning the entire effort is the input of the ECLO who co-ordinates (J) supportive care (e.g., low vision aids, psychological counselling, and technology training). In parallel, the clinical team assess the (K) co-morbidities and introduce therapies such as cataract surgery or treatment for macular oedema. This multifaceted pathway results in (I) a bespoke care plan, which assesses all aspects of their eye condition, ensuring timely access to appropriate novel and established interventions

excluding time for investigations) and access to specific clinical investigations (e.g., MMI, electrophysiology), which were performed during the same visit rather than repeated journeys as in general clinics. Thirty-six per cent of attendees were from different counties than the clinical sites (1–4-h journey) and non-attendance at repeat appointments could be avoided by a comprehensive visit (i.e., full clinical assessment and investigations). These data show that fewer patients must be booked per clinic facilitating accurate clinical phenotyping and discussion with each patient as this has a knock-on effect on subsequent investigations (e.g., determining relevance of detected genetic variants), treatment options (i.e., clinical trial inclusion) and patient decisions.

Informed by these data, a new algorithm was created (Fig. 2) and delivered via a subspecialty service. Thus far, 1482 IRD patients (and unaffected relatives) of an estimated 2600 IRD patients in Ireland (Republic of Ireland and Northern Ireland) [16] have been clinically assessed/phenotyped (see B in Fig. 2). Research-grade NGS (C, Fig. 2) has been completed on 1004 individuals from 710 pedigrees; a causative genetic variant was detected in 495 pedigrees (69.7%), with a single allele variant detected in an additional 64 (9%) of clinically autosomal recessive (AR) pedigrees where an obligatory second pathogenic sequence variant could not be identified [26]. The remaining 478 phenotyped cases are pending shipment to the accredited genetic laboratory, delayed due to a combination of availability of HSE funding and COVID-19 restrictions. In general, the Target 5000 approach is ongoing clinical characterization of new IRD patients/pedigrees to front-load the service with phenotyped cases, which progress to genotyping as further funding becomes available. Gene discovery studies (D, Fig. 2, whole gene, whole exome, and whole genome sequencing) are being carried out on this cohort as well as in pedigrees where no variants in IRD-implicated genes were detected. Novel genetic variants were interrogated for pathogenicity via further studies (e.g., cascade screening, in silico analysis and functional studies). Confirmation of detected variants has been carried out via a clinically accredited genetics laboratory (E, Fig. 2) in 62.9% (440/700) of genetically resolved cases as of 30/06/2020 with the remainder pending. See Table 1.

The MDT consensus for individual patients with a genetic outcome (from a subset of centres) was analysed (440 cases with clinically accredited genotype). ACMG pathogenicity scores for detected causative variants were grade 4 or 5 for 62.3% (n = 274), while 11.6% (n = 51) had at least 1 allele with a variant of unknown significance (VUS, ACMG grade 3). Further genetic confirmation (phase testing, carrier status) was requested in all AR cases. See Table 2 for more detail. MDT consensus has led to GC (either completed or pending) in 99% (n = 434/440) of cases. Of 440 IRD patients recruited at one of the Target 5000 clinical sites (MMUH), 50 patients (11.4%), thus far, are eligible for current therapies or active clinical trials (L, Fig. 2) (e.g., *RS1*, *USH2A*, *RPGR*, *CNGB3*, *CHM*, *RPE65*, *CEP290*).

Table 1 Completion rates of Target 5000 Algorithm Steps

	n =	% total	% category
Phenotyped (B)	1482	100	
Research NGS complete (C)	1004	67.7	
Positive research NGS	700	47.2	69.7
Negative research NGS	304	20.5	30.3
Accredited testing result (E)	440	29.7	62.9
Genetically eligible for trials	50	3.4	11.4

Stage indicated by letter in parentheses refers to Fig. 2

Table 2 Outcomes of accredited genetic testing (n = 440); ACMG grading: Grade 5- pathogenic; Grade 4- likely pathogenic; Grade 3 -VUS

	n =	%
Grade 4 or 5	274	62.3
Grade 3 only	34	7.7
1 allele grade 3 and 1 allele grade 4/5 (AR disease)	17	3.9
No variants detected	19	4.3
Unclassified (regrading)	94	21.4
Benign	2	0.5
Total	440	

Discussion

The rate of blind registration in the UK is 3% compared with 1.1% in Ireland, the age of whom was > 65 years in 66% and 72% of patients in the UK and Ireland respectively [31]. The mean age of all 1482 recruited patients in Target 5000 was 45.54 ± 18.67 years and 20.8% of these were registerable blind by visual acuity criteria (mean VA LogMAR 0.75 ± 0.87) not including those registerable by visual field constriction. Considering that IRDs are the leading cause of blindness in England and Wales in those < 64 years [13], IRDs are not proportionately represented and thus are not being adequately supported. Concurring with this, the internal survey mentioned above estimated 52% of the Irish IRD population was not being seen in eye clinics. This may reflect prior poor patient engagement due to frustration with the lack of accurate diagnoses and effective treatments [4]. As the ability of clinicians to positively impact visual prognosis improves, re-engagement with healthcare professionals

becomes vital to keep abreast of relevant developments and access novel treatments.

Genetic testing in overseas laboratoriess was historically a major component of the cost involved in IRD assessment. Prior to Target 5000, this was not centralized and charges were on an individual basis from either the public departmental resources or private patient funds. One of the main goals of the Target 5000 pathway was to create an equitable standardized level of care for all service users with quality and cost benefits to the health service. Initial funding for Target 5000 was provided from multiple charitable and grant sources (see methods). Genotyping through a common pathway at a new high throughput single accredited laboratory has allowed for significant savings (50.8%, €1000 vs €1968 per test). Using an example of 1522 IRD-affected individuals [16], the cost of genetic testing for all patients and 1 relative would be €3.80 million and €2.18 million outside and inside the Target 5000 programme respectively, saving 42.73%. This estimate does not account for the cost of access to the IRD clinical and genetics testing setup and expertise, nor the ophthalmic genetics MDT, which is unavailable outside the programme. The panel-based NGS approach in a research-based academic laboratory with validation in an accredited laboratory adopted by Target 5000 accrued substantial cost savings [26]; however, due to limited financial, personnel and bioinformatics support inherent in research-based academic laboratories the interval to receive a genetic result was circa 18–24 months. This is due to additional workup including cascade analysis (clinically and genetically testing additional family members and confirming segregation of variants within the family), in silico analysis and functional studies (e.g., investigation of splicing of non-coding variants in the *ABCA4* gene) to explore the pathogenicity of novel variants in IRD-implicated genes. Much of this work has been undertaken at a research-grade level to date highlighting the diverse investigations required to genetically resolve IRDs, the cost of which is difficult to estimate. Delay has significantly reduced with an NGS panel-based genetic result available from a high throughput accredited commercial laboratory within 4 weeks. This enables early recognition of cases suitable for disease modifying therapies before further progressive retinal atrophy has developed, though the contribution of the in-house MDT in correlating genotype to the clinical phenotype remains a vital component of this process. The ongoing development of a domestic accredited genetic laboratory with further capacity and bioinformatics support could further reduce this cost per patient/pedigree. The outcomes of the programme so far have been used as evidence to secure further limited funding from the public health service (HSE Ireland). Given that the cost of whole genome sequencing (WGS), data analysis and storage will likely continue to reduce, it may be that the field will move towards this technique for the investigation of all IRD patients from the outset.

A comprehensive database of available support services was compiled to facilitate access for patients as appropriate (Additional file 2). An on-site eye clinic liaison officer (ECLO) was incorporated into the IRD clinic to assess areas of need and signpost available supports (e.g., social supports, housing, low vision aids, etc.), from first contact and modify them as necessary at subsequent visits.

An MDT approach was adopted, comprising ophthalmic clinicians, systemic physicians, geneticists, and genetic counsellors. This group met weekly to assess each case from a clinical and genetic perspective, with a national multicentre MDT monthly. The goals of the MDT were to:

1. Confirm a genotype/phenotype match,
2. Determine need for phase testing (all AR cases) or further family studies (clinical and/or genetic assessment), and
3. Assess the best available evidence-based interventions (supports and therapeutics).

Specific IRD clinics aim to provide goal-oriented patient interactions with specific progress-dependent goals at each visit. The first visit allows initial phenotyping and DNA sampling for genetic analysis, while early ECLO involvement initiates relevant supports. Affected and unaffected relative involvement can be sought at or following the initial visit. Subsequent visits assess subjective/objective progression, monitor treatment response (e.g., macular oedema, cataract) and update ECLO inputs. When a validated genotype becomes available, following MDT discussion of each case, GC can be performed with the patient (± family). At this stage, a bespoke care plan (curated by the MDT and modified for relevance by an IRD patient focus group) is introduced to highlight relevant areas of intervention (Zhu et al., in prep). As 71.2% of genetically diagnosed IRDs are due to pathogenic variants in 20 genes [32], a small number of care plans apply to the majority of pedigrees, while care plans for less common phenotypes/genotypes can be created and updated as required.

Conclusion

The development of the Target 5000 pathway has allowed timely and accurate MDT-validated clinical and genetic diagnoses for patients with known pathogenic variants while enabling detection and exploration of pathogenicity for novel variants, thus increasing global understanding in this field (i.e., RetNet, ClinVar, etc.). A coordinated

Table 3 Recommendations for developing a national IRD programme

Identify national centres of ophthalmic expertise that are staffed and resourced to see patients with IRDs
Engage and partner with support agencies and patient organisations
Concentrate clinical genetics and support services at these sites
Encourage referrals of all diagnosed or suspected IRD cases to these centres. Follow up can be at regional centres with appropriate clinical testing capability
Have a robust pathway for funded genetic testing and resolution of contentious variants (via MDT and partnerships with other international laboratories)
Routinely use and regularly update variant classification (e.g., in silico and functional analysis for novel variants) and collaborative databases (e.g., ClinVar, RetNet)
Become members of international alliances (e.g., ERN-EYE) to share and gain knowledge and expertise of rare eye disease
Establish therapeutic pathways nationally and internationally
Create bespoke care plans for each patient/pedigree
Audit outcomes for patients

register of the clinically and genetically characterized IRD population becomes increasingly relevant as the scientific and medical community better understand the mechanisms of disease and potential avenues of intervention for these rare diseases. Findings from this pathway can be distilled into accessible and patient-friendly bespoke care plans which guide actionable outcomes for the individual. In addition to streamlined access to supports and standard clinical care (e.g., cataract, cystoid macular lesions, refractive correction), clinical trials and novel therapies are within reach via pharmaceutical industry directly or via the link to ERN-EYE, a first for IRD patients in Ireland [8]. The Target 5000 template could be used to guide development of IRD programmes in small to medium-sized countries in concordance with international best practice (Table 3).

Abbreviations

ACMG: American College of Medical Genetics; AR: Autosomal recessive; DR: Diabetic retinopathy; ECLO: Eye clinic liaison officer; ERN-EYE: European Reference Network for Rare Eye Disease; GC: Genetic counselling; HRCI: Health Research Charities Ireland; HSE: Health Service Executive (Ireland); IRD: Inherited retinal degeneration; ISCEV: International Society for Clinical Electrophysiology of Vision; MDT: Multidisciplinary team; MMI: Multimodal imaging; NGS: Next generation sequencing; VUS: Variant of unknown significance; WGS: Whole genome sequencing.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13023-021-01841-1.

Additional file 1. Survey and analysis of pre-existing infrastructure for assessment and management of IRDs in the Republic of Ireland.

Additional file 2. Services provided by vision impairment support groups/organizations in the Republic of Ireland and other western countries.

Acknowledgements

The authors thank the patients involved in the programme. We thank all the involved staff in MMUH, RVEEH, BHSCT and TCD for their effort and enthusiasm. We thank Fighting Blindness Ireland and the other funding bodies for their ongoing support.

Authors' contributions

KAJS: pathway development, data collection, manuscript creation and revision. JZ: data collection, manuscript revision. NW: data collection, manuscript revision. AD: data processing (genetics), manuscript revision. RMC: data collection, manuscript revision. ED: data collection, manuscript revision. LW: data processing (genetics), manuscript revision, CPM: data collection, manuscript revision. HD: data collection, manuscript revision. KC: data collection, manuscript revision. SR: data collection, manuscript revision. RP: data collection, manuscript revision. EC: pathway development, data collection, manuscript revision. JT: pathway development, data collection, manuscript revision. JJO'B: pathway development, data collection, manuscript revision. LB: pathway development, data collection, manuscript revision. GS: pathway development, data collection, manuscript revision. PFK: pathway development, data collection and processing (clinical and genetic), manuscript revision. GJF: pathway development, data processing (genetic), manuscript revision. DJK: pathway development, data collection, manuscript revision. All authors read and approved the final manuscript.

Funding

Funding was granted from Health Research Board Ireland, Science Foundation Ireland, Irish Research Council, and Health Research Charities Ireland/Fighting Blindness Ireland.

Availability of data and materials

All data generated or analysed during this study are included in this published article and its figures and tables.

Declarations

Ethics approval and consent to participate

This study was approved by the institutional review boards of the Mater Misericordiae University Hospital, Dublin, Ireland, the Royal Victoria Eye and Ear Hospital, Dublin, Ireland and the Belfast Health and Social Care Trust Hospitals, Belfast, Northern Ireland. All subjects whose data is included in this manuscript signed informed consent before participation in the Target 5000 study.

Consent for publication

All subjects whose data is included in this manuscript signed informed consent allowing anonymised use of their data in scientific publications.All subjects whose data is included in this manuscript signed informed consent allowing anonymised use of their data in scientific publications.

Competing interests
The authors have no competing interests to declare relevant to this manuscript.

Author details
[1]Mater Clinical Ophthalmic Genetics Unit, Mater Misericordiae University Hospital, Dublin, Ireland. [2]The Research Foundation, Royal Victoria Eye and Ear Hospital, Dublin, Ireland. [3]Ocular Genetics Unit, Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland. [4]Belfast Health and Social Care Trust Hospitals, Belfast, Northern Ireland. [5]National Council for the Blind of Ireland, Whitworth Road, Dublin 9, Ireland. [6]Fighting Blindness Ireland, Ely Place, Dublin 2, Ireland.

Received: 1 January 2021 Accepted: 23 April 2021
Published online: 05 May 2021

References

1. Daiger SP. Summaries of genes and loci causing retinal diseases (RetNet). The University of Texas Health Science Center, Houston, Texas, 2020 [June 28, 2020]. https://sph.uth.edu/retnet/. Accessed 6 Sept 2020.
2. Farrar GJ, McWilliam P, Bradley DG, Kenna P, Lawler M, Sharp EM, et al. Autosomal dominant retinitis pigmentosa: linkage to rhodopsin and evidence for genetic heterogeneity. Genomics. 1990;8(1):35–40.
3. McWilliam P, Farrar GJ, Kenna P, Bradley DG, Humphries MM, Sharp EM, et al. Autosomal dominant retinitis pigmentosa (ADRP): localization of an ADRP gene to the long arm of chromosome 3. Genomics. 1989;5(3):619–22.
4. Combs R, Hall G, Payne K, Lowndes J, Devery S, Downes SM, et al. Understanding the expectations of patients with inherited retinal dystrophies. Br J Ophthalmol. 2013;97(8):1057–61.
5. Eden M, Payne K, Jones C, Wright SJ, Hall G, McAllister M, et al. Identifying variation in models of care for the genomic-based diagnosis of inherited retinal dystrophies in the United Kingdom. Eye (Lond). 2016;30(7):966–71.
6. Harrison M, Birch S, Eden M, Ramsden S, Farragher T, Payne K, et al. Variation in healthcare services for specialist genetic testing and implications for planning genetic services: the example of inherited retinal dystrophy in the English NHS. J Community Genet. 2015;6(2):157–65.
7. Black G. In: Keegan D, editor. 2014.
8. Kenna PF, Humphries MM, Kiang AS, Brabet P, Guillou L, Ozaki E, et al. Advanced late-onset retinitis pigmentosa with dominant-acting D477G RPE65 mutation is responsive to oral synthetic retinoid therapy. BMJ Open Ophthalmol. 2020;5(1):e000462.
9. Cukras C, Wiley HE, Jeffrey BG, Sen HN, Turriff A, Zeng Y, et al. Retinal AAV8-RS1 gene therapy for X-linked retinoschisis: initial findings from a phase I/IIa trial by intravitreal delivery. Mol Ther J Am Soc Gene Ther. 2018;26(9):2282–94.
10. Cehajic Kapetanovic J, Patricio MI, MacLaren RE. Progress in the development of novel therapies for choroideremia. Expert Rev Ophthalmol. 2019;14(6):277–85.
11. Cehajic-Kapetanovic J, Xue K, de la Camara CMF, Nanda A, Davies A, Wood LJ, et al. Initial results from a first-in-human gene therapy trial on X-linked retinitis pigmentosa caused by mutations in RPGR. Nat Med. 2020;26(3):354–9.
12. Fischer MD, Michalakis S, Wilhelm B, Zobor D, Muehlfriedel R, Kohl S, et al. Safety and vision outcomes of subretinal gene therapy targeting cone photoreceptors in achromatopsia: a nonrandomized controlled trial. JAMA Ophthalmol. 2020;138:643–51.
13. Liew G, Michaelides M, Bunce C. A comparison of the causes of blindness certifications in England and Wales in working age adults (16–64 years), 1999–2000 with 2009–2010. BMJ Open. 2014;4(2):e004015.
14. Yau JW, Rogers SL, Kawasaki R, Lamoureux EL, Kowalski JW, Bek T, et al. Global prevalence and major risk factors of diabetic retinopathy. Diabetes Care. 2012;35(3):556–64.
15. Akuffo KO, Nolan J, Stack J, Moran R, Feeney J, Kenny RA, et al. Prevalence of age-related macular degeneration in the Republic of Ireland. Br J Ophthalmol. 2015;99(8):1037–44.
16. Galvin O, Chi G, Brady L, Hippert C, Del Valle RM, Daly A, et al. The impact of inherited retinal diseases in the Republic of Ireland (ROI) and the United Kingdom (UK) from a cost-of-illness perspective. Clin Ophthalmol (Auckland, NZ). 2020;14:707–19.
17. Seyedahmadi BJ, Rivolta C, Keene JA, Berson EL, Dryja TP. Comprehensive screening of the USH2A gene in Usher syndrome type II and non-syndromic recessive retinitis pigmentosa. Exp Eye Res. 2004;79(2):167–73.
18. Sutherland JE, Day MA. Advantages and disadvantages of molecular testing in ophthalmology. Expert Rev Ophthalmol. 2011;6(2):221–45.
19. Sergouniotis PI, Maxime E, Leroux D, Olry A, Thompson R, Rath A, et al. An ontological foundation for ocular phenotypes and rare eye diseases. Orphanet J Rare Dis. 2019;14:8.
20. Orphanet. Orphanet rare disease nomenclature 2018 [cited 2018 April 30]. http://www.orpha.net/consor/cgi-bin/index.php?lng=EN. Accessed 9 Sept 2020.
21. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med Off J Am Coll Med Genet. 2015;17(5):405–24.
22. ERN-EYE. Eye diseases (ERN-EYE) 2018 [cited 2018, May 2]. https://www.ern-eye.eu/. Accessed 15 Sept 2020.
23. Russell S, Bennett J, Wellman JA, Chung DC, Yu ZF, Tillman A, et al. Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. Lancet (London, England). 2017;390(10097):849–60.
24. Dockery A, Stephenson K, Keegan D, Wynne N, Silvestri G, Humphries P, et al. Target 5000: Target Capture Sequencing For Inherited Retinal Degenerations. Genes. 2017;8(11):304.
25. Farrar GJ, Carrigan M, Dockery A, Millington-Ward S, Palfi A, Chadderton N, et al. Toward an elucidation of the molecular genetics of inherited retinal degenerations. Hum Mol Genet. 2017;26(R1):R2-r11.
26. Whelan L, Dockery A, Wynne N, Zhu J, Stephenson K, Silvestri G, et al. Findings from a genotyping study of over 1000 people with inherited retinal disorders in Ireland. Genes. 2020;11(1):105.
27. Carrigan M, Duignan E, Malone CP, Stephenson K, Saad T, McDermott C, et al. Panel-based population next-generation sequencing for inherited retinal degenerations. Sci Rep. 2016;6:33248.
28. Khan M, Cornelis SS, Pozo-Valero MD, Whelan L, Runhart EH, Mishra K, et al. Resolving the dark matter of ABCA4 for 1054 Stargardt disease probands through integrated genomics and transcriptomics. Genet Med Off J Am Coll Med Genet. 2020;22:1235–46.
29. Carss KJ, Arno G, Erwood M, Stephens J, Sanchis-Juan A, Hull S, et al. Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. Am J Hum Genet. 2017;100(1):75–90.
30. Duncan JLBPS, Birch DG, Fishman GA, Heon E, Jacobson SG, Pennesi ME, Pierce EA, Scholl HP, Stone EM, Trzupek K, Weleber RG. Recommendations on clinical assessment of patients with inherited retinal degenerations—2016. USA: American Academy of Ophthalmology; 2016. https://www.aao.org/clinical-statement/recommendations-on-clinical-assessment-of-patients. Accessed 3 Nov 2020.
31. Floyd G. ECLO impact tool: UK wide findings 2015–2016: Royal National Institute of Blind People; 2016 [cited 2018 April 8]. http://www.rnib.org.uk/professionals/knowledge-and-research-hub/research-reports/early-reach-research/ECLO-impact-tool. Accessed 17 Sept 2020.
32. Mahroo OAR, Pontikos N, Arno G, Ba-Abbad R, Malka S, Wright G, et al. Genetic basis of inherited retinal disease in a UK cohort of over 2900 families. Investig Ophthalmol Vis Sci. 2019;60(9):401.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix 5

International Journal of
Molecular Sciences

MDPI

Review

Next-Generation Sequencing Applications for Inherited Retinal Diseases

Adrian Dockery *, Laura Whelan, Pete Humphries and G. Jane Farrar

The School of Genetics & Microbiology, Trinity College Dublin, Dublin 2, Ireland; whelanl1@tcd.ie (L.W.); pete.humphries@tcd.ie (P.H.); jane.farrar@tcd.ie (G.J.F.)
* Correspondence: dockerya@tcd.ie

Abstract: Inherited retinal diseases (IRDs) represent a collection of phenotypically and genetically diverse conditions. IRDs phenotype(s) can be isolated to the eye or can involve multiple tissues. These conditions are associated with diverse forms of inheritance, and variants within the same gene often can be associated with multiple distinct phenotypes. Such aspects of the IRDs highlight the difficulty met when establishing a genetic diagnosis in patients. Here we provide an overview of cutting-edge next-generation sequencing techniques and strategies currently in use to maximise the effectivity of IRD gene screening. These techniques have helped researchers globally to find elusive causes of IRDs, including copy number variants, structural variants, new IRD genes and deep intronic variants, among others. Resolving a genetic diagnosis with thorough testing enables a more accurate diagnosis and more informed prognosis and should also provide information on inheritance patterns which may be of particular interest to patients of a child-bearing age. Given that IRDs are heritable conditions, genetic counselling may be offered to help inform family planning, carrier testing and prenatal screening. Additionally, a verified genetic diagnosis may enable access to appropriate clinical trials or approved medications that may be available for the condition.

Keywords: genetic diagnosis; inherited retinal disease; rare disease; retina; sequencing; diagnostics; macula; genomics; variant interpretation; eye

check for updates

Citation: Dockery, A.; Whelan, L.; Humphries, P.; Farrar, G.J. Next-Generation Sequencing Applications for Inherited Retinal Diseases. *Int. J. Mol. Sci.* **2021**, 22, 5684. https://doi.org/10.3390/ijms22115684

Academic Editors: Luca Agnelli and Giovanni Malerba

Received: 26 April 2021
Accepted: 22 May 2021
Published: 26 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

A primary focus in ocular genetics globally is accurate genotyping of patients with rare inherited retinal diseases (IRDs). Next-generation sequencing (NGS) has been a common strategy employed in many countries to achieve this goal for several years [1–16]. This review focuses on the various methods and strategies that are being implemented to elucidate the genetic pathogenesis of IRDs and provides an overview of how these approaches have evolved. IRDs have an estimated prevalence of 1 in 4000 [17]. With a current global population of approximately 7.8 billion [18], it is estimated that approximately 2 million people currently have some form of IRD.

As a global community involved in ocular genetics, the common goal is to achieve a diagnostic success rate of 100% for all IRD patients enrolled in clinical studies. This objective, however, presents several challenges. Firstly, over 270 genes have been associated with the aetiologies of IRDs (RetNet, Retinal Information Network, https://sph.uth.edu/retnet/ accessed on 20 April 2021) [19]. Furthermore, extensive diversity in clinical presentation due to mutations even within a single IRD gene, as well as intersecting clinical phenotypes and phenocopies, is encountered. Mutations in disease genes may affect the retina in isolation, or may have more systemic effects. For example, there are 80 systemic conditions with a retinal phenotype and 200 genes that not only affect retinal health but also the central nervous system, kidneys or heart [20]. Such complexity makes it near-impossible for a diagnosis to be achieved in most instances solely on the basis of disease phenotype [2,21–23]. Furthermore, even a single pathogenic variant can manifest with phenotypic variability [24]. For some IRDs, modifier loci have been identified, somewhat blurring the borders between

Mendelian and polygenic forms of IRD and mirroring similar observations with other disease aetiologies [25,26].

In this review, we aim to provide an overview of the NGS strategies employed globally to maximise the detection of IRD-causing mutations. This includes the use of targeted gene panels for all IRD phenotypes or phenotypic subsets; whole-exome sequencing (WES); whole-gene sequencing, whereby an IRD gene's 5′ and 3′ sequences, exons and introns are interrogated; whole-genome sequencing (WGS); and bespoke methods to compliment other strategies, such as structural variant (SV) detection and copy number variant detection (CNV), among many others. In parallel with the use of the above technologies for the identification of candidate IRD-causing sequence variants, a wide array of methods to explore the functional effects of sequence variants have also been developed, and these are discussed.

While NGS technologies have enabled rapid characterisation of the genetic architecture of IRDs in many disease cohorts with diagnostic rates often approximating 70% [27], much still remains to be optimised. Strategies currently under development to further improve diagnostic rates are reviewed herein, as are the approaches being employed to enable interpretation of novel coding and non-coding candidate variants. Additionally, given the availability of increased numbers of WGS sequences from IRD and control populations, a greater focus is placed on the elucidation of the genetic modifier loci that may influence the effect(s) of the primary disease-causing mutations. An overview of the findings to date is provided.

2. IRDs—Target Panels and Whole Exome Studies

To elucidate the genetic contributions to IRDs, DNA, typically isolated from saliva or peripheral blood sample, is analysed. Optimal processing of the sample will depend on which form NGS is to be employed. In terms of both cost and data generated, NGS methods ranging from low to high involve targeted sequencing (TS), whole-exome sequencing (WES) or whole-genome sequencing (WGS). WES exclusively captures the protein-coding exons, but only accounts for approximately 1% of the genome. It is important to note that exon-based sequencing is likely to also reliably detect intronic variants located close to the targeted exons, such as non-canonical splice site variants, which are known causes of IRDs [28–30]. WGS is significantly more comprehensive including introns, promoters, and intergenic regions; in principle sequencing every nucleotide possible in a sample. TS typically captures the smallest amount of genetic information but does so in a completely customisable manner. For example, some IRD phenotypes are associated with pathogenic variants in a very small number of genes, but some of those genes may also be known to harbour pathogenic deep-intronic variants. In this case, adopting a TS approach would be more fruitful than WES. Arguably, WGS could also be used for this purpose but would generate more off-target data requiring significantly greater levels of analysis and storage and the on-target data would likely be less than a TS approach.

The benefits of TS are that it is an economical method of focusing sequencing capacity in smaller genomic regions including noncoding regions, therefore maximizing the coverage of clinically relevant genes. Enhanced coverage translates to greater sequencing read depth which is valuable, for example, to increase the resolution of detecting genetic variants and to detect smaller levels of heteroplasmy in the mitochondrial genome, or mosaicism in the nuclear genome [31–34]. By reducing the size of the region of genome sequenced per sample, a greater number of samples can be multiplexed together and processed in the same sequencing run. There are other cost-savings elements to TS, smaller file sizes allow for cheaper storage and faster processing. Moreover, targeting specific regions of the genome previously implicated in IRDs, can massively reduce the risk of detecting secondary or incidental findings.

For these reasons, TS strategies have frequently been employed for IRD screening for many years. Shortcomings of a TS strategy are that it often involves multiple gene panels for different conditions and if new IRD genes are identified or new variant associations are made for genes outside of the panel, a panel redesign will be required to include them. It is possible to use TS to detect indicators of large structural variants; however, such genomic breakpoints would likely have to occur within the captured loci reducing the likelihood of identification [1].

The customisable element of TS has become increasingly valuable with the recent detection of several population-enriched rare pathogenic variants likely due to founder effects. For example, a novel *PDE6B* variant was observed in the Māori IRD participant group and is likely to account for 16% of all recessive IRDs in that population [35]. Similarly, *EYS* gene variants were found to be causative in 51% of a RP cohort from Japan [36]. This discovery is not unique, as several other parallel studies have revealed similar founder mutations in their target populations, for example, Belgium, *RAX2* [37]; Costa Rica, *RPE65* [38]; Finland, *CERKL* [39]; Japan, *EYS* [36]; Spain, *RP1* [40] and *ABCA4* [41]; Jewish community in Caucasia, *PDE6B* [42]; Pakistan, *ABCA4* and *NMNAT1* [43]; Guyana, *BBS9* [44]; and Faroe Islands, *MERTK* [45]. The enrichment of these variants, several of which are large structural variants, emphasises the value of population-specific TS panels to target and detect mutations and mutational breakpoints that may be missed by commercial generic gene panel sequencing or even WES.

The use of WES has increased in popularity in recent years (Table 1) compared to previous metadata reported [27] and has many advantages over the TS approach. An effective TS panel design can be optimised with prior knowledge of the spectrum of mutations capable of causing the patient's condition. This includes but is not limited to, knowledge of possible founder mutations in the population, all possible genotype–phenotype associations and breakpoint locations of any large structural variants that may exist. WES is agnostic to these issues. Although WES is not capable of detecting deep intronic mutations without modifying the method, it enables exonic variants to be detected even if their relevance is not entirely elucidated at the time of capture. This provides the potential for future interrogation of WES data as new IRD genes are discovered. Importantly, WES allows for the potential future resolution of a previously unsolved diagnosis.

Table 1. Screening studies of inherited retinal disease (IRD) populations. CRD = cone–rod dystrophy; LCA = Lebers congenital amaurosis; IRD = inherited retinal dystrophy; MD = macular dystrophy; RP = retinitis pigmentosa; TC = target capture; WES = whole-exome sequencing; WGS = whole-genome sequencing.

Country	Author	Year	Pedigrees	Solve Rate	Cohort Details	TC (Genes)	WES	WGS
Australia	Thompson [46]	2017	34	90	LCA	Yes	-	-
Brazil	Motta [47]	2018	559	72	IRD	Yes	Yes	-
China	Liu [48]	2020	800	60	RP	Yes	Yes	-
China	Gao [49]	2019	1243	72	RP	586	-	-
China	Liu [50]	2020	182	48	IRD	Yes	Yes	-
China (Han)	Huang [51]	2017	98	41	RP	-	Yes	-
China	Dan [52]	2020	76	57	IRD	Yes	Yes	-
China	Wang [53]	2018	319	39	IRD	Yes	Yes	-
Finland	Avela [54]	2019	53	77	IRD	Yes	-	-
Germany	Weisschuh [55]	2020	1785	69	IRD	Yes	-	-
Germany	Birtel [56]	2018	251	74	MD/CRD	Yes	-	-
Iran	Tayebi [57]	2019	50	72	IRD	Yes	-	-
Ireland	Whelan [58]	2020	710	70	IRD	Yes	-	-
Israel	Sharon [59]	2020	2420	56	IRD	Yes	Yes	Yes
Japan	Koyanagi [60]	2019	1204	30	RP	Yes	-	-
Japan	Numa [36]	2020	220	45	RP	Yes	-	Yes
Korea	Surl [61]	2020	50	78	LCA	Yes	Yes	-
Korea	Kim [62]	2019	86	44	IRD	Yes	-	-

Table 1. *Cont.*

Country	Author	Year	Pedigrees	Solve Rate	Cohort Details	TC (Genes)	WES	WGS
Mexico	Zenteno [63]	2019	143	66	IRD	Yes	-	-
Norway	Holtan [17]	2020	650	32	IRD	Yes	-	-
Poland	Wawrocka [64]	2018	18	39	CRD	Yes	Yes	-
Polynesian and Māori	Vincent [35]	2017	16	44	IRD	Yes	-	-
Spain	Perea-Romero [41]	2021	3951	53	IRD	Yes	Yes	Yes
Spain	Martin-Merida [65]	2019	877	38	RP	Yes	-	-
Spain	Gonzàlez-Duarte [66]	2019	73	85	IRD	Yes	-	-
Spain	Diñeiro [67]	2020	100	45	IRD	Yes	-	-
Taiwan	Chen [68]	2020	60	53	IRD	Yes	-	-
Tunisia	Habibi [69]	2020	73	68	IRD	-	Yes	-
UAE	Khan [70]	2020	71	100	Pediatric IRD	Yes	Yes	-
UAE	Patel [71]	2018	75	82	IRD	Yes	Yes	-
UK	Jiman [72]	2020	106	49	Syndromic IRD	Yes	-	-
UK	Shah [73]	2020	655	43	IRD	Yes	-	-
UK	Carss [74]	2017	722	56	IRD	-	Yes	Yes
UK	Lenassi [75]	2020	201	64	Pediatric IRD	Yes	-	-
UK	Patel [76]	2019	277	25	Pediatric IRD	Yes	-	-
UK	Taylor [77]	2017	85	79	Pediatric IRD	Yes	-	-
USA and Canada	Goetz [78]	2020	5385	62	IRD	Yes	-	-
USA	Stone [79]	2017	1000	76	IRD	-	Yes	-
USA	Bryant [80]	2018	69	64	IRD	-	Yes	-

Further to this point, many disease phenotype-based gene panels are very specific and therefore typically target only a small number of genes, not allowing for the possibility of new genotype–phenotype correlations or ambiguous phenotypes. Indeed, it was recently established that 23% of cases analysed would not have been resolved if they were sequenced by a commercial panel designed specifically for a patient's phenotype [81]. In the same study, it was also found that for 26% of participants, the cheapest applicable commercial gene panel would have been costlier than performing WES for those patients.

Several large IRD screening studies in recent years have sought to identify the genes responsible for the largest proportions of their cohorts' IRDs. In the UK, over 3000 pedigrees were reviewed, and it was determined that 135 IRD genes contributed to the genetic pathogenesis of the cohort. Interestingly, 70% of resolved cases were deemed to have causative mutations in 20 genes only [82]. Similarly, in over 5000 pedigrees with genetic eye conditions from Canada and the US, 68% of pathogenic or likely pathogenic mutations were identified in just 10 genes [78]. Both of these large studies identified *ABCA4*, *USH2A* and *RPGR* as the top three genes contributing to IRDs.

Although clearly not 100% effective, smaller whole-gene panels may be very effective as a first-tier screening approach. Several gene associations have been recently flagged as unlikely to be as pathogenic as initially reported. Nineteen percent of queried autosomal dominant retinitis pigmentosa (adRP) genes were deemed to harbour variants unlikely to be disease-causing for reasons relating to their respective allele frequencies or variant interpretation at that time [83]. Such variant "false positives" are shortcomings of the diagnostic odyssey, and this implies that, for an initial screening procedure, there may not be the need to screen as many of the genes and variants that are typically included in large gene panel screening studies.

It is important to also note that there have been many reports of the occurrence of multiple IRDs within the same family, or even within a single individual. Although individually IRDs are rare globally, the concurrence of multiple IRDs in a patient or pedigree unfortunately represents another diagnostic challenge. Our team has previously reported a pedigree in which five affected members of the family were broadly categorised as RP phenotypes. After genetic investigation it was revealed that four of these individuals were homozygous for a *FLVCR1* variant, while the remaining affected patient was compound heterozygous for pathogenic variants in *NR2E3* [84]. Similarly, in a US study, involving three IRD pedigrees, each given an initial diagnosis of RP, one with a dominant RP and the other two with a dominant, incompletely penetrant RP, it was found that multiple IRD

genes were responsible for various affected individuals in each of the three families: both *USH2A* and *RP1* was segregating in one family; *PRPH2* and *CRX* in a second family and *PRPH2*, *PRPH8* and *USH2A* in the third family [85]. These studies, however, are trumped in complexity by Birtel et al.'s analysis of a single family with four different IRDs each caused by distinct pathogenic variants and inheritance patterns: father, *RHO*, dominant RP; mother, *ABCA4* and *CACNA1F*, recessive Stargardt and CSNB; first son, *CACNA1F*, CSNB; second son, *MITF*, dominant Waardenburg syndrome [86]. These are some examples of the many that exist, illustrating the complexity of IRD screening and reinforcing the necessity of thorough clinical and genetic investigation prior to genetic counselling [87].

3. Expanding IRD Diagnosis via Whole-Gene or WGS

For any laboratory electing to use NGS, it is essential that the limitations of the NGS approach to be employed are known. Failure to appropriately sequence the target genomic sites clearly will limit the success rate from the outset. Three consistent biases that exist for WES but not WGS are strand bias, evenness of coverage and the proportion of transcripts covered in their entirety. Interestingly, it has also been found that WGS provides a 3% better coverage of exons compared to WES, 98% versus 95% [88]. In another study it was found that the WGS approach offered superior detection of structural variants, variants in regulatory regions and detection of variants in GC-rich regions compared to WES [74]. However, this additional superior detection comes at a significant financial cost. A review of studies that used WES and WGS for clinical practice revealed that the price range for WGS studies was approximately five times higher for WGS on a per sample basis [89].

Costs incurred by WGS include not only the upfront cost of sequencing, but also additional downstream expenses. WGS produces vastly more data, thus immediately requiring additional computational power, people hours and storage to process. Although storage issues may be a limiting factor in the budgets of most research groups, policies regarding data storage can be readily adjusted to meet the needs of the research group, as the needs of two research groups will rarely be the same. This bespoke approach is advisable to avoid issues such as inadequate infrastructure and overspending. Raw sequencing files (such as .fastq files) and output files (variant call files, .vcf) are relatively small in comparison to the alignment files (such as .bam files) that need to be produced as part of the analyses [90–92]. Given this information, it is possible to reduce the capacity required for long term storage of sequencing data by electing to discard the alignment files but keeping the input and output files so that the analysis can be repeated at a later date and outputs can be compared for discrepancies and newer discoveries. However, increased storage will be required again upon reanalysis, as new alignment files will be created as part of the process.

Another viable alternative in reducing the disk footprint of alignment files, is the use of additionally compressed formats, such as .cram files. This format uses reference-based compression, only storing base calls that differ to the reference genome used. This compression can be either lossless or can incur a reduction in base quality scores corresponding to the level of compression. Even the lossless format offers a 40–50% reduction in space required by comparison with BAM [93]. In addition, the use of WGS as a second-tier approach, for cases that remain genetically unresolved following first-tier sequencing will decrease data-storage demands. More research groups are now making the move to cloud-based storage for their NGS data and minimising the amount of data stored has a direct impact on cost [94]. It is important to note that sample processing and analyses are available via cloud-based solutions also, and may be an attractive option for research groups lacking the necessary in-house infrastructure to process NGS data [95].

The additional cost of larger-scale analysis is not the only hazard associated with this data management. Both WES and WGS have an increased likelihood of carrying intrinsic responsibilities regarding the management of incidental or secondary findings (SFs) unrelated to the initial indication for sequencing. For example, Hart et al. (2018) found that a SF is detected in 1.7% of patients who undergo WES [96]. Some IRD studies have employed a nested targeted approach, wherein the entire genome of an individual

is sequenced but only variants in genes relevant to the IRD phenotype are interrogated by use of variant filtering with a virtual gene panel. This still provides benefits over traditional targeted panels, as it also includes sequencing of non-coding regions, as well as the potential for analysis of an expanded panel in the future. For example, Carss et al. (2017) performed WES on 117 individuals, identifying pathogenic variants in 59 cases [74]. Forty-five of the unresolved cases then underwent WGS and positive candidate variants were identified in an additional 14 cases. This approach is likely chosen due to the immense volume of data produced by WGS and WES, and the need to more rapidly analyse the most relevant data available.

This approach also limits the possibility of detecting SFs. In SF v1.0, The American College of Medical Genetics and Genomics (ACMG) recommended analysis of 56 medically actionable gene–phenotype pairs which was then updated to a panel of 59 genes in v2.0 [97,98]. ACMG SF v3.0, recommending the analysis of SFs in 73 gene–phenotype pairs, was very recently released [99,100]. Of particular interest to the ocular genetics community, ACMG SF v3.0 now includes the *RPE65* gene. The *RPE65* gene was included on the basis that an FDA-approved gene therapy now exists for biallelic RPE65-retinopathies and that patients may derive additional benefit from earlier detection and therapeutic intervention. The ACMG recommends application of these SF guidelines in a clinical setting as opposed to a research setting. Nonetheless, as with all genomic testing, it is imperative that the patient's interests are at the forefront. ACMG currently recommend that patients/guardians have the choice to opt-out of SF testing. This highlights the necessity of an appropriate and comprehensive pre-testing consent procedure. This includes and is not limited to information pertaining to what will not be disclosed should the patient/guardian chose to abstain from SF analysis and thorough pre-test and post-test counselling.

Despite additional costs, there may be diagnostic benefits to employing WGS to resolve genotypes. Lionel et al. investigated 103 cases of diverse genetic disorders comparing WGS to targeted panel sequencing. Not only was the solve rate superior when WGS was used, 41% versus 24%, 18 diagnoses were made based on structural variants or intronic variants that were not captured by the TS method [101]. Regardless of the substantial number of genes identified and targeted by TS, as estimated from studies to date, the genetic cause of 43% of all IRDs patients remains unknown and suggests the need for more studies to employ WGS (Table 1). These missing genetic aberrations may reside in introns or intergenic regions, both of which are captured by WGS. There is also the possibly of novel IRD gene discovery that is facilitated by WGS. The superior uniformity of genome coverage enabled by WGS also allows for greater sensitivity when detecting copy number variants (CNVs) that are notoriously difficult to detect by TS and WES.

A cost-effective alternative that retains many of the same benefits as WGS is whole-gene sequencing (GS). GS enables the capture of exonic, intronic and 5′ and 3′ regulatory regions for a target gene of interest but has many of the same limitations as the TS approach, including strand bias and GC-rich impedance to capture. GS has been utilised very successfully for cohorts with phenotypes associated with monogenic or near-monogenic causes. For example, individuals affected with incomplete congenital stationary night blindness (icCSNB) present with a recognisable phenotype. This form of icCSNB is primarily associated with mutations in the *CACNA1F* gene. In a recent large genotyping study of icCSNB–CACNA1F patients (n = 189), 4% of *CACNA1F* causative variants were attributed to intronic and synonymous mutations [102]. It is also probable that there are additional intronic variants yet to be designated as pathogenic in the unresolved portion of this cohort.

Similarly, Khan et al. investigated 1054 unresolved Stargardt cases with a GS approach. Stargardt disease is predominately caused by biallelic variants in the *ABCA4* gene. The authors of the study used a single-molecule molecular inversion probes (smMIPs) approach, which proved reliable and cost-effective. Their study revealed the presence of pathogenic SVs and deep-intronic variants in 25% of biallelic cases [103]. The smMIPs method is gaining in popularity given its superior target capture and low cost compared to

other TS capture methods. In a recent comparative study, 176 IRD patients were analysed with both smMIPs and TS. The smMIPs approach demonstrated enhanced target coverage (97.3% versus 93.9%) and was five times more cost effective when greater than 500 samples were analysed [104].

The GS approach has also been combined with probes for other IRD genes to investigate if this combinatorial strategy could significantly improve diagnostic rates for a range of IRDs when compared to traditional exon-based TS IRD panels. The study design encompassed a second-tier approach for patients who had one previous variant found in *USH2A*, *ABCA4* and *CEP290*. These whole genes, plus exons of 76 additional IRD genes and pathogenic intronic regions of two IRD genes were sequenced in an effort to resolve the "one-hit" patients. An overall diagnostic rate of 58.6% was achieved; two copy number variants were detected in *USH2A* [105]. Although this diagnostic rate was no higher than the average study (Table 1), it does represent significant improvements that can be made to address the large proportion of unresolved patients identified by standard screening studies. The structural variants established in this study would likely not have been detected by use of a more traditional, purely exon-targeting design.

An improved GS study design as outlined above may have additional advantages. The *RPGR* gene is one example of an IRD gene that includes regions that are challenging to sequence comprehensively with traditional TS or WES; sequencing through ORF15 is impeded due to a low-complexity sequence composition [106]. However, it is vital to capture this gene as, for example, it accounts for nearly 40% of X-linked retinopathies in the UK. This makes pathogenic variants in this gene the third most-prevalent cause of IRDs in this population [82]. In an Italian cohort of 48 *RPGR*-related RP cases, approximately half had mutations in ORF15 and presented with a more severe phenotype than the other causative variants in exons 1–14 of *RPGR* [107]. It has been suggested that the sequence coverage of ORF15 could be optimised by modifying NGS library preparation, reducing false-negatives, miscalled variants and false-positives when compared to traditional methods [108]. For this reason, many recent NGS screening studies have adopted bespoke approaches to sequencing *RPGR*, including entirely separate analyses or spiking the NGS libraries with separately generated amplicons for *RPGR* [50,53,56]. In another study improved alignment of sequencing reads mapping to the ORF15 region by using a de novo assembly approach were reported. The accuracy of sequencing can be quickly determined for males when analysing the RPGR gene, as variants called in error will likely not be represented in every sequencing read mapping to this region. Therefore, heterozygous variant calls can be readily identified as errors, since males have only one copy of RPGR. This de novo assembly approach reduced the number of false-heterozygous calls in males and improved the accuracy of indel calls [109].

Another example of genes that benefit from tailored GS design are those encoding the opsins, *OPN1LW* (red cone cells) and *OPN1MW* (green cone cells). These genes encode photopigments in the retina and pose an interesting challenge to sequencing efforts. These two genes are 96% homologous which introduces unique challenges for the IRD gene panel, as short-read sequencing may be unable to determine the best alignment option when mapping back to the genome [110]. A new two-step method from Atilano et al. has demonstrated that long-range PCR can generate specific long amplicons that can be more readily mapped back to the genome [111]. This approach offers a solution that can be analysed separately, by direct sequencing of the amplicons, or alternatively, as part of a larger panel if long-read sequencing (LRseq) is used.

Sequencing the entirety of a gene also facilitates the detection of variants in the upstream regulatory regions which have been implicated in retinal disease previously, for example in Blue-Cone Monochromacy (BCM) [112]. In this condition, a c. −71A>C promoter mutation was initially thought to decrease expression and cause a deutan colour vision deficiency. However, after functional analyses, the mutation was revealed to result in more than double the wild-type expression level of the gene [113]. Other deletions in this area have also been shown to result in BCM phenotypes, suggesting that this gene is

sensitive to alterations in both under and overexpression [114,115]. Similarly, Radziwon et al. used luciferase reporter assays to assess upstream variants detected in the *CHM* gene in patients with choroideremia. Both probands had variants at position c. −98: C>A and C>T. Both mutations led to a reduction in luciferase activity and furthermore, the promoter region for *CHM* was defined as the region encompassing nucleotides c. −119 to c. −76 [116].

Regulatory mutations are often difficult to interpret, particularly for genes associated with recessive forms of inheritance. Previously, consanguineous pedigrees have been useful for identifying homozygous variants in these cases, such as *NMNAT*-related Leber congenital amaurosis (LCA) [117]. Variant interpretation can be further complicated as such variants may not have strong effects on gene expression. In a recent study of promoter variants in *ELOVL4* two variants were found, c. −236 C>T (rs240307) and c. −90 G>C (rs62407622) which resulted in 18% and 14% reduction in expressivity, respectively. However, as the patient in question had the variants in *trans*, a severe phenotype was observed, much more than would have been expected from the modest effect of the two variants analysed separately [118]. This detrimental synergistic effect may emphasise the threshold sensitivity of retinal tissues and cell components to the dosage levels of this protein and its downstream effects.

4. Copy Number and Structural Variants

As discussed above, TS and WES methods are the most universally utilised, yet they are largely incapable of detecting large copy number variants (CNVs), structural variants (SVs) and chromosomal rearrangements. In 2018, an extensive literature-mining endeavour revealed that 1345 copy-number variants (CNVs)—specifically, 317 unique variants—had been reported in 81 distinct IRD genes. When further analysed, the size of the gene correlated with the reported numbers of CNVs associated with that gene. Additionally, many of these large variants affected non-coding and potential *cis*-regulatory elements [119]. The relevance of such variant types is now recognised, and guidelines have been published to assist in the interpretation and classification of them, similarly to those published in recent years for single-nucleotide variants [120,121].

CNVs and SVs can also vary significantly in the complexity of their rearrangements. Gross deletions have previously been detected in many genes, including *BEST1*, *EYS*, *MERTK*, *USH2A* and many more from the aforementioned study alone [119]. Large deletions have also been reported in *RPGR* [122], *CHM* [58], *OPN1LW*/*OPN1MW* [123] and *USH1C* [1], to name but a few. Deletions are likely to be the most detectable CNV type given that most studies employ WES or TS to detect mutations. Homozygous deletions are the most readily detectable from using these methods as the read coverage over the deleted region would be zero, given no template exists for capture or amplification. Heterozygous deletions may be under-reported when using WES or TS, if significant amplification has occurred, which may unintentionally normalise the ultimate read depth aligned to the deleted region. For similar reasons, duplications can be very difficult to detect with these methods. However, such mutations can be more readily detected by WGS due to the superior and more even coverage, or by more specific approaches, such as targeted locus amplification [119]. Some regions of the genome, such as the RP17 locus, have been shown to harbour many complex CNV and SV variants associated with IRDs. These convoluted rearrangements resulted in the interference of the surrounding genome architecture, disrupting enhancer–promoter interactions, and resulting in aberrant gene expression [124].

Genomic rearrangements are more likely to be detected by the presence of broken sequencing reads when aligned back to the reference genome. This occurs when a read, or pair of reads, partially align to one part of the genome and partially to somewhere quite different. This is applicable to translocations and inversions, as unlike CNVs, the read depth is not expected to be altered in these scenarios. Given the significant presence of retrotransposon sequence in the human genome, it is not surprising that several retrotransposons insertions have been reported to disrupt the functionality of IRD genes [125–127]. The *BBS1* gene in particular, has been recently reported to harbour retrotransposons causative of

disease [128,129]. Retrotransposons, much like other large genomic insertions, can be difficult to detect as they are unlikely to disrupt read depth in the genomic region to which they have relocated to, since alignment tools will align these reads to their original positions in the genome. Broken reads may indicate that a rearrangement has occurred. If breakpoints are detected, the region can be directly sequenced to shed light on the nature of the SV. Alternatively, a de novo assembly approach may be used to reconstruct the queried genome [130]. This approach will likely only be beneficial in the case of WGS, since TS or WES will likely not have sequenced the insert because the original genomic region was not an intended target.

In one study, involving an investigation of *PRPF31*-related disease, 45% of probands (10 of 22) tested positive for a CNV. The *PRPF31* gene has no obvious sequence elements that may make it particularly susceptible to genomic rearrangement, such as long interspersed nuclear elements (LINE) and long terminal repeat (LTR) elements [131]. The study emphasises the importance of integrating CNV and SV detection into screening protocols, even for genes that may not appear to be conventionally susceptible to genomic rearrangements. The estimated prevalence of causative SVs in IRD cases is roughly 10% [51,131–133]. This is similar to findings from other rare disease cohorts as 12% of developmental disorders are estimated to be caused by pathogenic CNVs, therefore CNV and SV detection is recommended to be incorporated into first-tier testing for that set of conditions [134]. In a large hearing loss screening study of over 1000 patients, 18% of resolved patients were found to have causative CNVs [135]. CNV detection has also proven very useful in diagnosing atypical syndromic IRD cases resulting in novel genotype–phenotype associations and the refinement of complex phenotypes in multiple cases [136].

Many of the NGS methods discussed so far have revolved around short-read sequencing; however, long-read sequencing is arguably the superior approach for detection of SVs and CNVs. Short-read sequencing is generally preferred to ensure that high-quality data are produced [137]. However, this technology is greatly hindered by features of the genome, such as repetitive elements, which are not only abundant in our genomes, but also known to increase the likelihood of an SV event occurring in IRD genes [123,138]. Long-read sequencing offers superior sequencing of such regions and offers a chance to more accurately recapitulate patients' true genomic sequences through the use of de novo assembly [139–141]. Results from several studies to date have revealed IRD-causing SVs by the use of long-read sequencing, and in some cases, concluded that the complexity of the SV was such that it was likely not possible to fully resolve it by short-read sequencing [44,142].

Another useful application of long-read sequencing is determining the phase of potentially causative recessive variants. Determining the phase of variants is of critical importance, as it may determine whether causative variants have been established, if in *trans*, or not, if in *cis*. This task is challenging for IRD cases primarily for two reasons. Firstly, variants causative of Mendelian IRDs are extremely rare in most cases. This prevents the establishment of known haplotypes or complex alleles in most cases that may otherwise indicate that the two detected variants are likely in *cis*. Secondly, many IRD screening endeavours are still in their infancy. This results in the widest possible age range of patients, since even patients with paediatric onset of their condition, may be elderly when screened. This can often make segregation analysis difficult, as many of their close family members may be immobile or deceased. Long-read sequencing offers the interpreter a greater chance of capturing both variants of interest within the same sequencing molecule and therefore determining phase of the variants without the need for additional family members [140].

5. Modifiers of IRD Phenotypes

There are multiple examples of phenotypes manifesting significantly differently in IRD patients, even when harbouring the same mutation [24]. Such variability may be in overall severity, age of onset or pattern of degeneration. While some instances may be partially explained by variant haplotypes or known interactions with other genes, much of

the source of variation remains unknown [143]. It has also yet to be established how much of this unknown contribution is genetic.

Pathogenic variants in *PRPF31* are the second most common cause of autosomal dominant retinitis pigmentosa [144]. PRPF31 is a universally expressed splicing-factor and has a vital role in the processing of pre-mRNA. However, reduced expression of this ubiquitous splicing-factor results in isolated retinal restricted disease, retinitis pigmentosa. It has been shown in multiple studies that reduced levels of PRPF31 in the retina results in the mis-splicing and reduced expression of several key IRD genes associated with phototransduction and RNA processing [145,146]. This indirect mechanism of action may partially explain the incomplete penetrance and variable severity frequently reported with this IRD gene [144]. It has also been shown that minisatellite repeat elements (MSR1) proximal to the promoter of *PRPF31* can modify the expression of the gene. Alleles with four copies of the MSR1 were shown to correlate with asymptomatic individuals, while alleles with three copies of MSR1 greatly reduced the expression of *PRPF31* [147].

Another key factor in the manifestation of an IRD relates to the naturally occurring expression levels of the implicated genes. In a study by Green et al., IRD genes typically associated with incomplete inheritance significantly correlated with greater variability in levels of expression in a healthy population in several tissue types, including the eye. This implies that *cis* and or *trans* elements may be influencing variability in expression and in turn, contribute to disease states in patients [148]. This may also result in higher-than-expected allele frequencies for pathogenic variants located in these genes in "healthy" control databases, such as gnomAD [149].

Several IRD genes have been found to harbour pathogenic splice mutations. Splice mutations are associated with variable severity as the percentage of wild-type transcript that can still be produced is dependent on the specific mutation in question [150]. It is therefore plausible that some variants may only have marginal disruptive effects on normal splicing and result in a sub-clinical phenotype. It is important that each of these variants is evaluated to determine aberrant effect(s). They can be assessed by a variety of means. The most readily available method to assess variants involves use of plasmid transfection to recapitulate the patient's mutation in mammalian cells. Minigene and Midigene constructs incorporating exons and introns for the IRD gene and including the candidate splice mutation have been hugely successful in providing empirical evidence of the functional effect of candidate splice variants [150–153]. Patient-derived cell and/or 3D retinal organoid models while more laborious to create, remove the need for transient expression of mutations as the mutation is already present in the patient's cells. Organoids provide a more retina-like simulated environment to more accurately evaluate splice processing and thus have been utilised for the interpretation of multiple IRD variants [154–156].

The *RPGR* gene is another example of an IRD gene associated with variable disease severity and age of onset. In the case of *RPGR*, this is particularly true of female carriers of pathogenic variants. Random X-inactivation plays a decisive role on the severity of the IRD manifestation. In a recent analysis, blood and saliva samples were taken from 77 female carriers from 41 RPGR–IRD pedigrees. These samples were analysed for their methylation patterns. It was found that X-inactivation ratios correlated with clinical severity and could be useful indicators for prognostic purposes [157]. Another common feature of *RPGR* pathogenesis is that missense variants often lack an ability to interact with other key IRD gene products which results in the failure of RPGR to correctly localise to the cilia as required [158]. Interestingly, this means that polymorphic variation in the other interacting IRD genes, such as *IQCB1* and *RPGRIP1L*, can further modify the dynamics of this disrupted interaction [159].

Given that many IRDs involve progression over time, age is another modifier of disease. In several studies that have been mentioned previously, it was noted that diagnostic rates are higher in younger patients. In Germany, the mean age of participants with a positive molecular diagnosis was 39.8 years of age, whereas it was 46.3 years of age for unresolved patients [55]. In the UK, patients enrolled with late-onset macular disease

(≥50 years of age) were less likely to receive a genetic diagnosis (18.0%) compared with patients with an onset of less than 50 years of age (24.2%) [73]. In the US, Stone et al. also noted that younger entrants to their screening study were more likely to have their disease-causing mutation identified [79]. These findings suggest that the genetic causes of early onset IRDs are better established and possibly that ageing increases susceptibility to polygenic, or perhaps non-genetic factors, making the older patient cohort more difficult to solve by current genetic screening approaches.

6. Impacts on IRD Diagnosis Outside of NGS Testing

The surge of NGS studies addressing diagnostic challenges in IRDs in the last decade has also encouraged further developments in other areas related to IRD diagnosis. Massive improvements have been made in ophthalmological imaging, not just in resolution, but also in applications and options available to imaging technicians [160]. These techniques help refine phenotypes and can inform whether or not more advanced forms of imagery may be useful for particular patients [161]. Machine learning (ML) is a branch of artificial intelligence that focuses on the concept that systems can learn from training data, such as expert-reviewed fundus images of IRD patients, and thereby learn to identify these patterns independently. This involves expert curation to obtain the training dataset but has the advantage of removing variability due to being a single assessor once training is complete. For example, many studies employ ML to assess fundus and optical coherence tomography images to train and predict outcomes for potential age-related macular degeneration (AMD) patients, including risk of progression to a more severe disease state and response to treatments [162–169]. This approach has enormous potential benefits for improvement of diagnostic and prognostic accuracy. For example, ML has been utilised for the detection of fleck lesions in fundus autofluorescence imaging, a characteristic feature of Stargardt disease. The ML approach could accurately identify and quantify fleck lesions, a potential outcome measure for future clinical trials [170]. Similarly, optic disc photography has been used to systematically detect abnormalities of the optic disc [171]. ML strategies may revolutionise the speed and accuracy at which new patients can be phenotyped in the clinic. More accurate phenotyping may have beneficial implications for NGS screening studies also, particularly those utilising phenotype-based gene panels. Improving the effectiveness of TS is likely to have the most widespread benefits given the relative low cost (less than $20 per sample), making TS the most readily accessible to diagnostic labs, especially those with constrictive budgets [79].

While significant advances in the clinical and genetic diagnosis of IRDs have been made, implementation of IRD screening is under resourced in many countries. A recent study into the cost-of-illness of IRDs in Ireland and the UK demonstrated that lack of access to genetic testing, the absence of an international patient registry and the issues regarding the reimbursement of therapies are common complexities faced by IRD patients in these countries [172]. In the UK, surveyed IRD patients noted that they appreciated pre-test counselling including a discussion of the possibility of an unexpected result [173]. Some countries are actively tackling the issue of integrating genetic services into future planning for their healthcare systems. Portugal and Iran are two of the latest countries to announce the launch of their nationwide registries for IRDs [174,175]. These registries aim to increase accessibility for individuals, while also providing a comprehensive dataset for investigators and clinicians to boost and develop their research. One such example is the French Plan for Genomic Medicine 2025 [176]. This plan aims to improve healthcare by organising and structuring new pathways to care and counselling while also reviewing the current challenges impeding the implementation of genomic testing in France. Despite the many great advances detailed in this review, the successful adaptation of IRD genomics services into national healthcare programs is rare [177]. Given that the technology and expertise is available and accessible to many research teams around the world (Table 1), combining the findings from these studies with effective national healthcare screening

systems may enable better treatment and care through clinical genetic interventions for the patient [177].

7. Conclusions

We are now in the age of genomic medicine, where genetic evidence can often provide precise and robust diagnoses. This evidence may inform prognoses and may aid in directing care plans and therapeutic intervention. This growth has been supported by rapid enhancements in DNA sequencing methodologies, analysis tools and skillsets. The enormous impact of such advances is clearly exemplified in the field of IRDs. The true genetic complexity of IRDs is more fully appreciated with mutations in more than 300 genes implicated in IRDs among many other non-coding and modifying elements. It is clear that many challenges exist for genetic screening of IRDs, as evident from the approximate 40% portion of screened patients whose causative genetic elements are yet to be established. Many of these cases may be resolved in the future as more comprehensive techniques, such as WGS, are more routinely utilised. As such data becomes available, it will undoubtedly also increase our knowledge of genetic elements that have a modifying effect on IRD manifestations and severities. Genetic-screening diagnostic rates are likely to also benefit from advances in technologies related to clinical phenotyping that, where appropriate, may be augmented by machine-learning-based algorithms. Clearly another development that should enhance IRD screening endeavours is the development of national databases and genomics strategies to develop services and enhance clinical genetics collaborations nationally and internationally.

Author Contributions: Conceptualization, G.J.F. and A.D.; data curation, A.D. and L.W.; writing—original draft preparation, A.D., P.H. and G.J.F.; writing—review and editing, A.D., L.W., G.J.F. and P.H.; funding acquisition, G.J.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by grant awards from Fighting Blindness Ireland (FB Irl; FB20DOC, FB16FAR, FB18CRE), the Health Research Board of Ireland (HRB; POR/2010/97) in conjunction with Health Research Charities Ireland (HRCI; MRCG-2013-8, MRCG-2016-14), the Irish Research Council (IRC; GOIPG/2017/1631) and Science Foundation Ireland (SFI; 16/1A/4452).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: This review would not be possible without the approximate 25,000 patients who have been discussed in this review. The authors would like to acknowledge the role that each and every one of these individuals has played in advancing our collective knowledge of inherited retinal conditions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dockery, A.; Stephenson, K.; Keegan, D.; Wynne, N.; Silvestri, G.; Humphries, P.; Kenna, P.F.; Carrigan, M.; Farrar, G.J. Target 5000: Target Capture Sequencing for Inherited Retinal Degenerations. *Genes* **2017**, *8*, 304. [CrossRef]
2. Bernardis, I.; Chiesi, L.; Tenedini, E.; Artuso, L.; Percesepe, A.; Artusi, V.; Simone, M.L.; Manfredini, R.; Camparini, M.; Rinaldi, C.; et al. Unravelling the Complexity of Inherited Retinal Dystrophies Molecular Testing: Added Value of Targeted Next-Generation Sequencing. *BioMed Res. Int.* **2016**, *2016*, 1–14. [CrossRef]
3. Consugar, M.B.; Navarro-Gomez, D.; Place, E.M.; Bujakowska, K.M.; Sousa, M.E.; Fonseca-Kelly, Z.D.; Taub, D.; Janessian, M.; Wang, D.Y.; Au, E.D.; et al. Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing. *Genet. Med.* **2015**, *17*, 253–261. [CrossRef] [PubMed]
4. Ellingford, J.M.; Barton, S.; Bhaskar, S.; O'Sullivan, J.; Williams, S.; Lamb, J.; Panda, B.; Sergouniotis, P.; Gillespie, R.L.; Daiger, S.P.; et al. Molecular findings from 537 individuals with inherited retinal disease. *J. Med. Genet.* **2016**, *53*, 761–767. [CrossRef]

5. Tiwari, A.; Bahr, A.; Bähr, L.; Fleischhauer, J.; Zinkernagel, M.S.; Winkler, N.; Barthelmes, D.; Berger, L.; Gerth-Kahlert, C.; Neidhardt, J.; et al. Next generation sequencing based identification of disease-associated mutations in Swiss patients with retinal dystrophies. *Sci. Rep.* **2016**, *6*, 28755. [CrossRef] [PubMed]
6. Riera, M.; Navarro, R.; Ruiz-Nogales, S.; Méndez, P.; Burés-Jelstrup, A.; Corcóstegui, B.; Pomares, E. Whole exome sequencing using Ion Proton system enables reliable genetic diagnosis of inherited retinal dystrophies. *Sci. Rep.* **2017**, *7*, 42078. [CrossRef]
7. Haer-Wigman, L.; Van Zelst-Stams, W.A.G.; Pfundt, R.; Born, L.I.V.D.; Klaver, C.C.W.; Verheij, J.B.G.M.; Hoyng, C.B.; Breuning, M.H.; Boon, C.J.F.; Kievit, A.J.; et al. Diagnostic exome sequencing in 266 Dutch patients with visual impairment. *Eur. J. Hum. Genet.* **2017**, *25*, 591–599. [CrossRef]
8. Audo, I.; Bujakowska, K.M.; Léveillard, T.; Mohand-Saïd, S.; Lancelot, M.-E.; Germain, A.; Antonio, A.; Michiels, C.; Saraiva, J.-P.; Letexier, M.; et al. Development and application of a next-generation-sequencing (NGS) approach to detect known and novel gene defects underlying retinal diseases. *Orphanet J. Rare Dis.* **2012**, *7*, 8. [CrossRef]
9. De Castro-Miro, M.; Tonda, R.; Escudero-Ferruz, P.; Andrés, R.; Mayor-Lorenzo, A.; Castro, J.; Ciccioli, M.; Hidalgo, D.A.; Rodríguez-Ezcurra, J.J.; Farrando, J.; et al. Novel Candidate Genes and a Wide Spectrum of Structural and Point Mutations Responsible for Inherited Retinal Dystrophies Revealed by Exome Sequencing. *PLoS ONE* **2016**, *11*, e0168966. [CrossRef] [PubMed]
10. Neveling, K.; Collin, R.W.; Gilissen, C.; van Huet, R.A.; Visser, L.; Kwint, M.P.; Gijsen, S.J.; Zonneveld, M.N.; Wieskamp, N.; de Ligt, J.; et al. Next-generation genetic testing for retinitis pigmentosa. *Hum. Mutat.* **2012**, *33*, 963–972. [CrossRef] [PubMed]
11. O'Sullivan, J.; Mullaney, B.G.; Bhaskar, S.S.; Dickerson, J.E.; Hall, G.; O'Grady, A.; Webster, A.; Ramsden, S.C.; Black, G.C. A paradigm shift in the delivery of services for diagnosis of inherited retinal disease. *J. Med. Genet.* **2012**, *49*, 322–326. [CrossRef] [PubMed]
12. Shanks, M.E.; Downes, S.M.; Copley, R.R.; Lise, S.; Broxholme, J.; Hudspith, K.A.; Kwasniewska, A.; Davies, W.I.; Hankins, M.W.; Packham, E.R.; et al. Next-generation sequencing (NGS) as a diagnostic tool for retinal degeneration reveals a much higher detection rate in early-onset disease. *Eur. J. Hum. Genet.* **2012**, *21*, 274–280. [CrossRef] [PubMed]
13. Ge, Z.; Bowles, K.; Goetz, K.; Scholl, H.P.N.; Wang, F.; Wang, X.; Xu, S.; Wang, K.; Wang, H.; Chen, R. NGS-based Molecular diagnosis of 105 eyeGENE® probands with Retinitis Pigmentosa. *Sci. Rep.* **2015**, *5*, 18287. [CrossRef]
14. Zhao, L.; Wang, F.; Wang, H.; Li, Y.; Alexander, S.; Wang, K.; Willoughby, C.; Zaneveld, J.E.; Jiang, L.; Soens, Z.T.; et al. Next-generation sequencing-based molecular diagnosis of 82 retinitis pigmentosa probands from Northern Ireland. *Qual. Life Res.* **2015**, *134*, 217–230. [CrossRef]
15. Perez-Carro, R.; Corton, M.; Sánchez-Navarro, I.; Zurita, O.; Sanchez-Bolivar, N.; Sánchez-Alcudia, R.; Lelieveld, S.H.; Aller, E.; Lopez-Martinez, M.A.; López-Molina, M.I.; et al. Panel-based NGS Reveals Novel Pathogenic Mutations in Autosomal Recessive Retinitis Pigmentosa. *Sci. Rep.* **2016**, *6*, 19531. [CrossRef] [PubMed]
16. Weisschuh, N.; Mayer, A.K.; Strom, T.M.; Kohl, S.; Glöckle, N.; Schubach, M.; Andreasson, S.; Bernd, A.; Birch, D.G.; Hamel, C.P.; et al. Mutation Detection in Patients with Retinal Dystrophies Using Targeted Next Generation Sequencing. *PLoS ONE* **2016**, *11*, e0145951. [CrossRef]
17. Holtan, J.P.; Selmer, K.K.; Heimdal, K.R.; Bragadóttir, R. Inherited retinal disease in Norway—A characterization of current clinical and genetic knowledge. *Acta Ophthalmol.* **2019**, *98*, 286–295. [CrossRef]
18. 2020 World Population Data Sheet Shows Older Populations Growing, Total Fertility Rates Declining—Population Reference Bureau. Available online: https://www.prb.org/2020-world-population-data-sheet/ (accessed on 12 January 2021).
19. RetNet. Retinal Information Network. Available online: https://sph.uth.edu/RetNet/ (accessed on 22 June 2018).
20. Tatour, Y.; Ben-Yosef, T. Syndromic Inherited Retinal Diseases: Genetic, Clinical and Diagnostic Aspects. *Diagnostics* **2020**, *10*, 779. [CrossRef]
21. Carrigan, M.; Duignan, E.; Humphries, P.; Palfi, A.; Kenna, P.F.; Farrar, G.J. A novel homozygous truncatingGNAT1mutation implicated in retinal degeneration. *Br. J. Ophthalmol.* **2015**, *100*, 495–500. [CrossRef]
22. Shankar, S.P.; Hughbanks-Wheaton, D.K.; Birch, D.G.; Sullivan, L.S.; Conneely, K.N.; Bowne, S.J.; Stone, E.M.; Daiger, S.P. Autosomal Dominant Retinal Dystrophies Caused by a Founder Splice Site Mutation, c.828+3A>T, in PRPH2 and Protein Haplotypes in trans as Modifiers. *Investig. Ophthalmol. Vis. Sci.* **2016**, *57*, 349–359. [CrossRef]
23. Tee, J.J.L.; Smith, A.J.; Hardcastle, A.J.; Michaelides, M. RPGR-associated retinopathy: Clinical features, molecular genetics, animal models and therapeutic options. *Br. J. Ophthalmol.* **2016**, *100*, 1022–1027. [CrossRef]
24. Zupan, A.; Fakin, A.; Battelino, S.; Jarc-Vidmar, M.; Hawlina, M.; Bonnet, C.; Petit, C.; Glavač, D. Clinical and Haplotypic Variability of Slovenian USH2A Patients Homozygous for the c. 11864G>A Nonsense Mutation. *Genes* **2019**, *10*, 1015. [CrossRef] [PubMed]
25. Becker, T.; Pich, A.; Tamm, S.; Hedtfeld, S.; Ibrahim, M.; Altmüller, J.; Dalibor, N.; Toliat, M.R.; Janciauskiene, S.; Tümmler, B.; et al. Genetic information from discordant sibling pairs points to ESRP2 as a candidate trans-acting regulator of the CF modifier gene SCNN1B. *Sci. Rep.* **2020**, *10*, 1–19. [CrossRef]
26. Bandres-Ciga, S.; Center, T.A.G.; Saez-Atienzar, S.; Kim, J.J.; Makarious, M.B.; Faghri, F.; Diez-Fairen, M.; Iwaki, H.; Leonard, H.; Botia, J.; et al. Large-scale pathway specific polygenic risk and transcriptomic community network analysis identifies novel functional pathways in Parkinson disease. *Acta Neuropathol.* **2020**, *140*, 341–358. [CrossRef]

27. Farrar, G.J.; Carrigan, M.; Dockery, A.; Millington-Ward, S.; Palfi, A.; Chadderton, N.; Humphries, M.; Kiang, A.S.; Kenna, P.F.; Humphries, P. Toward an elucidation of the molecular genetics of inherited retinal degenerations. *Hum. Mol. Genet.* **2017**, *26*, R2–R11. [CrossRef] [PubMed]
28. Kortüm, F.; Kieninger, S.; Mazzola, P.; Kohl, S.; Wissinger, B.; Prokisch, H.; Stingl, K.; Weisschuh, N. X-Linked Retinitis Pigmentosa Caused by Non-Canonical Splice Site Variants in *RPGR*. *Int. J. Mol. Sci.* **2021**, *22*, 850. [CrossRef] [PubMed]
29. Verbakel, S.K.; Fadaie, Z.; Klevering, B.J.; Van Genderen, M.M.; Feenstra, I.; Cremers, F.P.M.; Hoyng, C.B.; Roosing, S. The identification of a RNA splice variant in TULP1 in two siblings with early-onset photoreceptor dystrophy. *Mol. Genet. Genom. Med.* **2019**, *7*, e660. [CrossRef] [PubMed]
30. Al-Khuzaei, S.; Broadgate, S.; Halford, S.; Jolly, J.K.; Shanks, M.; Clouston, P.; Downes, S.M. Novel Pathogenic Sequence Variants in *NR2E3* and Clinical Findings in Three Patients. *Genes* **2020**, *11*, 1288. [CrossRef]
31. Chen, Z.; Moran, K.; Richards-Yutz, J.; Toorens, E.; Gerhart, D.; Ganguly, T.; Shields, C.L.; Ganguly, A. Enhanced Sensitivity for Detection of Low-Level Germline MosaicRB1Mutations in Sporadic Retinoblastoma Cases Using Deep Semiconductor Sequencing. *Hum. Mutat.* **2014**, *35*, 384–391. [CrossRef]
32. Strubbe, I.; Van Cauwenbergh, C.; De Zaeytijd, J.; De Jaegere, S.; De Bruyne, M.; Rosseel, T.; Van de Sompele, S.; De Baere, E.; Leroy, B.P. Phenocopy of a heterozygous carrier of X-linked retinitis pigmentosa due to mosaicism for a RHO variant. *Sci. Rep.* **2021**, *11*, 1–8. [CrossRef]
33. Dawod, P.G.A.; Jancic, J.; Marjanovic, A.; Brankovic, M.; Jankovic, M.; Samardzic, J.; Potkonjak, D.; Djuric, V.; Mesaros, S.; Novakovic, I.; et al. Whole Mitochondrial Genome Analysis in Serbian Cases of Leber's Hereditary Optic Neuropathy. *Genes* **2020**, *11*, 1037. [CrossRef] [PubMed]
34. Brusa, R.; Mauri, E.; Dell'Arti, L.; Magri, F.; Ronchi, D.; Minorini, V.; Mainetti, C.; Gagliardi, D.; Faravelli, I.; Meneri, M.; et al. Expanding the clinical spectrum of the mitochondrial mutation A13084T in the ND5 gene. *Neurol. Genet.* **2020**, *6*, e511. [CrossRef]
35. Vincent, A.L.; Abeysekera, N.; Van Bysterveldt, K.A.; Oliver, V.F.; Ellingford, J.M.; Barton, S.; Black, G.C. Next-generation sequencing targeted disease panel in rod-cone retinal dystrophies in Māori and Polynesian reveals novel changes and a common founder mutation. *Clin. Exp. Ophthalmol.* **2017**, *45*, 901–910. [CrossRef]
36. Numa, S.; Oishi, A.; Higasa, K.; Oishi, M.; Miyata, M.; Hasegawa, T.; Ikeda, H.O.; Otsuka, Y.; Matsuda, F.; Tsujikawa, A. EYS is a major gene involved in retinitis pigmentosa in Japan: Genetic landscapes revealed by stepwise genetic screening. *Sci. Rep.* **2020**, *10*, 1–9. [CrossRef] [PubMed]
37. Sompele, S.V.D.; Smith, C.; Karali, M.; Corton, M.; Van Schil, K.; Peelman, F.; Cherry, T.; Rosseel, T.; Verdin, H.; Derolez, J.; et al. Biallelic sequence and structural variants in RAX2 are a novel cause for autosomal recessive inherited retinal disease. *Genet. Med.* **2018**, *21*, 1319–1329. [CrossRef]
38. Glen, W.B.; Peterseim, M.M.W.; Badilla, R.; Znoyko, I.; Bourg, A.; Wilson, R.; Hardiman, G.; Wolff, D.; Martinez, J. A high prevalence of biallelic RPE65 mutations in Costa Rican children with Leber congenital amaurosis and early-onset retinal dystrophy. *Ophthalmic Genet.* **2019**, *40*, 110–117. [CrossRef] [PubMed]
39. Avela, K.; Sankila, E.-M.; Seitsonen, S.; Kuuluvainen, L.; Barton, S.; Gillies, S.; Aittomäki, K. A founder mutation inCERKLis a major cause of retinal dystrophy in Finland. *Acta Ophthalmol.* **2018**, *96*, 183–191. [CrossRef] [PubMed]
40. Avila-Fernandez, A.; Cortón, M.; Nishiguchi, K.M.; Muñoz-Sanz, N.; Benavides-Mori, B.; Blanco-Kelly, F.; Riveiro-Alvarez, R.; Garcia-Sandoval, B.; Rivolta, C.; Ayuso, C. Identification of an RP1 Prevalent Founder Mutation and Related Phenotype in Spanish Patients with Early-Onset Autosomal Recessive Retinitis. *Ophthalmology* **2012**, *119*, 2616–2621. [CrossRef] [PubMed]
41. Perea-Romero, I.; Gordo, G.; Iancu, I.F.; Del Pozo-Valero, M.; Almoguera, B.; Blanco-Kelly, F.; Carreño, E.; Jimenez-Rolando, B.; Lopez-Rodriguez, R.; Lorda-Sanchez, I.; et al. Genetic landscape of 6089 inherited retinal dystrophies affected cases in Spain and their therapeutic and extended epidemiological implications. *Sci. Rep.* **2021**, *11*, 1–13. [CrossRef]
42. Tatour, Y.; Tamaiev, J.; Shamaly, S.; Colombo, R.; Bril, E.; Rabinowitz, T.; Yaakobi, A.; Mezer, E.; Leibu, R.; Tiosano, B.; et al. A Novel Intronic Mutation of PDE6B Is a Major Cause of Autosomal Recessive Retinitis Pigmentosa among Caucasus Jews. *Mol. Vis.* **2019**, *25*, 155–164. [PubMed]
43. Rehman, A.U.; Peter, V.G.; Quinodoz, M.; Rashid, A.; Khan, S.A.; Superti-Furga, A.; Rivolta, C. Exploring the Genetic Landscape of Retinal Diseases in North-Western Pakistan Reveals a High Degree of Autozygosity and a Prevalent Founder Mutation in ABCA4. *Genes* **2019**, *11*, 12. [CrossRef] [PubMed]
44. Reiner, J.; Pisani, L.; Qiao, W.; Singh, R.; Yang, Y.; Shi, L.; Khan, W.A.; Sebra, R.; Cohen, N.; Babu, A.; et al. Cytogenomic identification and long-read single molecule real-time (SMRT) sequencing of a Bardet–Biedl Syndrome 9 (BBS9) deletion. *NPJ Genom. Med.* **2018**, *3*, 3. [CrossRef] [PubMed]
45. Ostergaard, E.; Duno, M.; Batbayli, M.; Vilhelmsen, K.; Rosenberg, T. A Novel MERTK Deletion Is a Common Founder Muta-tion in the Faroe Islands and Is Responsible for a High Proportion of Retinitis Pigmentosa Cases. *Mol. Vis.* **2011**, *17*, 1485–1492.
46. Thompson, J.A.; De Roach, J.N.; McLaren, T.L.; Montgomery, H.E.; Hoffmann, L.H.; Campbell, I.R.; Chen, F.K.; Mackey, D.A.; Lamey, T.M. The genetic profile of Leber congenital amaurosis in an Australian cohort. *Mol. Genet. Genom. Med.* **2017**, *5*, 652–667. [CrossRef]
47. Motta, F.L.; Martin, R.P.; Filippelli-Silva, R.; Salles, M.V.; Sallum, J.M.F. Relative frequency of inherited retinal dystrophies in Brazil. *Sci. Rep.* **2018**, *8*, 1–9. [CrossRef]
48. Liu, X.; Tao, T.; Zhao, L.; Li, G.; Yang, L. Molecular diagnosis based on comprehensive genetic testing in 800 Chinese families with non-syndromic inherited retinal dystrophies. *Clin. Exp. Ophthalmol.* **2021**, *49*, 46–59. [CrossRef]

49. Gao, F.-J.; Li, J.-K.; Chen, H.; Hu, F.-Y.; Zhang, S.-H.; Qi, Y.-H.; Xu, P.; Wang, D.-D.; Wang, L.-S.; Chang, Q.; et al. Genetic and Clinical Findings in a Large Cohort of Chinese Patients with Suspected Retinitis Pigmentosa. *Ophthalmology* **2019**, *126*, 1549–1556. [CrossRef]
50. Liu, X.Z.; Li, Y.Y.; Yang, L.P. Comparison study of whole exome sequencing and targeted panel sequencing in molecular diagnosis of inherited retinal dystrophies. *Beijing Da Xue Xue Bao* **2020**, *52*, 836–844.
51. Huang, X.-F.; Mao, J.-Y.; Huang, Z.-Q.; Rao, F.-Q.; Cheng, F.-F.; Li, F.-F.; Wang, Q.-F.; Jin, Z.-B. Genome-Wide Detection of Copy Number Variations in Unsolved Inherited Retinal Disease. *Investig. Ophthalmol. Vis. Sci.* **2017**, *58*, 424–429. [CrossRef] [PubMed]
52. Dan, H.; Huang, X.; Xing, Y.; Shen, Y. Application of targeted panel sequencing and whole exome sequencing for 76 Chinese families with retinitis pigmentosa. *Mol. Genet. Genom. Med.* **2020**, *8*, e1131. [CrossRef] [PubMed]
53. Wang, L.; Zhang, J.; Chen, N.; Wang, L.; Zhang, F.; Ma, Z.; Li, G.; Yang, L. Application of Whole Exome and Targeted Panel Sequencing in the Clinical Molecular Diagnosis of 319 Chinese Families with Inherited Retinal Dystrophy and Comparison Study. *Genes* **2018**, *9*, 360. [CrossRef] [PubMed]
54. Avela, K.; Salonen-Kajander, R.; Laitinen, A.; Ramsden, S.; Barton, S.; Rudanko, S. The genetic aetiology of retinal degeneration in children in Finland—New founder mutations identified. *Acta Ophthalmol.* **2019**, *97*, 805–814. [CrossRef]
55. Weisschuh, N.; Obermaier, C.D.; Battke, F.; Bernd, A.; Kuehlewein, L.; Nasser, F.; Zobor, D.; Zrenner, E.; Weber, E.; Wissinger, B.; et al. Genetic architecture of inherited retinal degeneration in Germany: A large cohort study from a single diagnostic center over a 9-year period. *Hum. Mutat.* **2020**, *41*, 1514–1527. [CrossRef]
56. Birtel, J.; Eisenberger, T.; Gliem, M.; Müller, P.L.; Herrmann, P.; Betz, C.; Zahnleiter, D.; Neuhaus, C.; Lenzner, S.; Holz, F.G.; et al. Clinical and genetic characteristics of 251 consecutive patients with macular and cone/cone-rod dystrophy. *Sci. Rep.* **2018**, *8*, 1–11. [CrossRef] [PubMed]
57. Tayebi, N.; Akinrinade, O.; Khan, M.I.; Hejazifar, A.; Dehghani, A.; Cremers, F.P.; Akhlaghi, M. Targeted next generation sequencing reveals genetic defects underlying inherited retinal disease in Iranian families. *Mol. Vis.* **2019**, *25*, 106–117. [PubMed]
58. Whelan, L.; Dockery, A.; Wynne, N.; Zhu, J.; Stephenson, K.; Silvestri, G.; Turner, J.; O'Byrne, J.J.; Carrigan, M.; Humphries, P.; et al. Findings from a Genotyping Study of over 1000 People with Inherited Retinal Disorders in Ireland. *Genes* **2020**, *11*, 105. [CrossRef]
59. Sharon, D.; Ben-Yosef, T.; Cohen, N.G.; Pras, E.; Gradstein, L.; Soudry, S.; Mezer, E.; Zur, D.; Abbasi, A.H.; Zeitz, C.; et al. A nationwide genetic analysis of inherited retinal diseases in Israel as assessed by the Israeli inherited retinal disease consortium (IIRDC). *Hum. Mutat.* **2020**, *41*, 140–149. [CrossRef] [PubMed]
60. Koyanagi, Y.; Akiyama, M.; Nishiguchi, K.M.; Momozawa, Y.; Kamatani, Y.; Takata, S.; Inai, C.; Iwasaki, Y.; Kumano, M.; Murakami, Y.; et al. Genetic characteristics of retinitis pigmentosa in 1204 Japanese patients. *J. Med. Genet.* **2019**, *56*, 662–670. [CrossRef]
61. Surl, D.; Shin, S.; Lee, S.-T.; Choi, J.R.; Lee, J.; Byeon, S.H.; Han, S.-H.; Lim, H.T.; Han, J. Copy Number Variations and Multiallelic Variants in Korean Patients with Leber Congenital Amaurosis. *Mol. Vis.* **2020**, *26*, 26–35.
62. Kim, M.S.; Joo, K.; Seong, M.-W.; Kim, M.J.; Park, K.H.; Park, S.S.; Woo, S.J. Genetic Mutation Profiles in Korean Patients with Inherited Retinal Diseases. *J. Korean Med. Sci.* **2019**, *34*, e161. [CrossRef] [PubMed]
63. Zenteno, J.C.; García-Montaño, L.A.; Cruz-Aguilar, M.; Ronquillo, J.; Rodas-Serrano, A.; Aguilar-Castul, L.; Matsui, R.; Vencedor-Meraz, C.I.; Arce-González, R.; Graue-Wiechers, F.; et al. Extensive genic and allelic heterogeneity underlying inherited retinal dystrophies in Mexican patients molecularly analyzed by next-generation sequencing. *Mol. Genet. Genom. Med.* **2019**, *8*. [CrossRef]
64. Wawrocka, A.; Skorczyk-Werner, A.; Wicher, K.; Niedziela, Z.; Ploski, R.; Rydzanicz, M.; Sykulski, M.; Kociecki, J.; Weisschuh, N.; Kohl, S.; et al. Novel variants identified with next-generation sequencing in Polish patients with cone-rod dystrophy. *Mol. Vis.* **2018**, *24*, 326–339.
65. Martin-Merida, I.; Avila-Fernandez, A.; Del Pozo-Valero, M.; Blanco-Kelly, F.; Zurita, O.; Perez-Carro, R.; Aguilera-Garcia, D.; Riveiro-Alvarez, R.; Arteche, A.; Trujillo-Tiebas, M.J.; et al. Genomic Landscape of Sporadic Retinitis Pigmentosa. *Ophthalmology* **2019**, *126*, 1181–1188. [CrossRef]
66. Gonzàlez-Duarte, R.; De Castro-Miró, M.; Tuson, M.; Ramírez-Castañeda, V.; Gils, R.V.; Marfany, G. Scaling New Heights in the Genetic Diagnosis of Inherited Retinal Dystrophies. *Adv. Exp. Med. Biol.* **2019**, *1185*, 215–219. [CrossRef] [PubMed]
67. Diñeiro, M.; Capín, R.; Cifuentes, G.Á.; Fernández-Vega, B.; Villota, E.; Otero, A.; Santiago, A.; Pruneda, P.C.; Castillo, D.; Viejo-Díaz, M.; et al. Comprehensive genomic diagnosis of inherited retinal and optical nerve disorders reveals hidden syndromes and personalized therapeutic options. *Acta Ophthalmol.* **2020**, *98*, e1034–e1048. [CrossRef]
68. Chen, Z.-J.; Lin, K.-H.; Lee, S.-H.; Shen, R.-J.; Feng, Z.-K.; Wang, X.-F.; Ms, X.H.; Huang, Z.-Q.; Jin, Z.-B. Mutation spectrum and genotype-phenotype correlation of inherited retinal dystrophy in Taiwan. *Clin. Exp. Ophthalmol.* **2020**, *48*, 486–499. [CrossRef] [PubMed]
69. Habibi, I.; Falfoul, Y.; Turki, A.; Hassairi, A.; El Matri, K.; Chebil, A.; Schorderet, D.F.; El Matri, L. Genetic spectrum of retinal dystrophies in Tunisia. *Sci. Rep.* **2020**, *10*, 10. [CrossRef]
70. Khan, A.O. Phenotype-guided genetic testing of pediatric inherited retinal disease in the United Arab Emirates. *Retina* **2019**, *40*, 1829–1837. [CrossRef]

71. Patel, N.; Alkuraya, H.; Alzahrani, S.S.; Nowailaty, S.; Seidahmed, M.Z.; Alhemidan, A.; Ben-Omran, T.; Ghazi, N.; Al-Aqeel, A.; Al-Owain, M.; et al. Mutations in known disease genes account for the majority of autosomal recessive retinal dystrophies. *Clin. Genet.* **2018**, *94*, 554–563. [CrossRef] [PubMed]
72. Jiman, O.A.; Taylor, R.L.; Lenassi, E.; Smith, J.C.; Douzgou, S.; Ellingford, J.M.; Barton, S.; Hardcastle, C.; Fletcher, T.; Campbell, C.; et al. Diagnostic yield of panel-based genetic testing in syndromic inherited retinal disease. *Eur. J. Hum. Genet.* **2020**, *28*, 576–586. [CrossRef]
73. Shah, M.; Shanks, M.; Packham, E.; Williams, J.; Haysmoore, J.; MacLaren, R.E.; Németh, A.H.; Clouston, P.; Downes, S.M. Next generation sequencing using phenotype-based panels for genetic testing in inherited retinal diseases. *Ophthalmic Genet.* **2020**, *41*, 331–337. [CrossRef]
74. Carss, K.J.; Arno, G.; Erwood, M.; Stephens, J.; Sanchis-Juan, A.; Hull, S.; Megy, K.; Grozeva, D.; Dewhurst, E.; Malka, S.; et al. Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. *Am. J. Hum. Genet.* **2017**, *100*, 75–90. [CrossRef]
75. Lenassi, E.; Clayton-Smith, J.; Douzgou, S.; Ramsden, S.C.; Ingram, S.; Hall, G.; Hardcastle, C.L.; Fletcher, T.A.; Taylor, R.L.; Ellingford, J.M.; et al. Clinical utility of genetic testing in 201 preschool children with inherited eye disorders. *Genet. Med.* **2020**, *22*, 745–751. [CrossRef] [PubMed]
76. Patel, A.; Hayward, J.D.; Tailor, V.; Nyanhete, R.; Ahlfors, H.; Gabriel, C.; Jannini, T.B.; Abbou-Rayyah, Y.; Henderson, R.; Nischal, K.K.; et al. The Oculome Panel Test. *Ophthalmology* **2019**, *126*, 888–907. [CrossRef]
77. Taylor, R.L.; Parry, N.R.; Barton, S.J.; Campbell, C.; Delaney, C.M.; Ellingford, J.M.; Hall, G.; Hardcastle, C.; Morarji, J.; Nichol, E.J.; et al. Panel-Based Clinical Genetic Testing in 85 Children with Inherited Retinal Disease. *Ophthalmology* **2017**, *124*, 985–991. [CrossRef]
78. Goetz, K.E.; Reeves, M.J.; Gagadam, S.; Blain, D.; Bender, C.; Lwin, C.; Naik, A.; Tumminia, S.J.; Hufnagel, R.B. Genetic testing for inherited eye conditions in over 6,000 individuals through the eyeGENE network. *Am. J. Med. Genet. Part C Semin. Med. Genet.* **2020**, *184*, 828–837. [CrossRef]
79. Stone, E.M.; Andorf, J.L.; Whitmore, S.S.; DeLuca, A.P.; Giacalone, J.C.; Streb, L.M.; Braun, T.A.; Mullins, R.F.; Scheetz, T.E.; Sheffield, V.C.; et al. Clinically Focused Molecular Investigation of 1000 Consecutive Families with Inherited Retinal Disease. *Ophthalmology* **2017**, *124*, 1314–1331. [CrossRef] [PubMed]
80. Bryant, L.; Lozynska, O.; Maguire, A.M.; Aleman, T.S.; Bennett, J. Prescreening whole exome sequencing results from patients with retinal degeneration for variants in genes associated with retinal degeneration. *Clin. Ophthalmol.* **2017**, *12*, 49–63. [CrossRef]
81. Dillon, O.J.; Lunke, S.; Stark, Z.; Yeung, A.; Thorne, N.; Gaff, C.; White, S.M.; Tan, T.Y. Exome sequencing has higher diagnostic yield compared to simulated disease-specific panels in children with suspected monogenic disorders. *Eur. J. Hum. Genet.* **2018**, *26*, 644–651. [CrossRef] [PubMed]
82. Pontikos, N.; Arno, G.; Jurkute, N.; Schiff, E.; Ba-Abbad, R.; Malka, S.; Gimenez, A.; Georgiou, M.; Wright, G.; Armengol, M.; et al. Genetic Basis of Inherited Retinal Disease in a Molecularly Characterized Cohort of More Than 3000 Families from the United Kingdom. *Ophthalmology* **2020**, *127*, 1384–1394. [CrossRef]
83. Hanany, M.; Sharon, D. Allele frequency analysis of variants reported to cause autosomal dominant inherited retinal diseases question the involvement of 19% of genes and 10% of reported pathogenic variants. *J. Med. Genet.* **2019**, *56*, 536–542. [CrossRef] [PubMed]
84. Dockery, A.; Carrigan, M.; Wynne, N.; Stephenson, K.; Keegan, D.; Kenna, P.F.; Farrar, G.J. A Novel FLVCR1 Variant Implicated in Retinitis Pigmentosa. *Adv. Exp. Med. Biol.* **2019**, *1185*, 203–207. [CrossRef]
85. Jones, K.D.; Wheaton, D.K.; Bowne, S.J.; Sullivan, L.S.; Birch, D.G.; Chen, R.; Daiger, S.P. Next-generation sequencing to solve complex inherited retinal dystrophy: A case series of multiple genes contributing to disease in extended families. *Mol. Vis.* **2017**, *23*, 470–481.
86. Birtel, J.; Gliem, M.; Hess, K.; Birtel, T.H.; Holz, F.G.; Zechner, U.; Bolz, H.J.; Herrmann, P. Comprehensive Geno- and Phenotyping in a Complex Pedigree Including Four Different Inherited Retinal Dystrophies. *Genes* **2020**, *11*, 137. [CrossRef]
87. Rodríguez-Muñoz, A.; García-Bohórquez, B.; Udaondo, P.; Hervás-Ontiveros, A.; Salom, D.; Aller, E.; Jaijo, T.; García-García, G.; Millán, J.M. Concomitant mutations in inherited retinal dystrophies. *Retina* **2021**. [CrossRef]
88. Lelieveld, S.H.; Spielmann, M.; Mundlos, S.; Veltman, J.; Gilissen, C. Comparison of Exome and Genome Sequencing Technologies for the Complete Capture of Protein-Coding Regions. *Hum. Mutat.* **2015**, *36*, 815–822. [CrossRef]
89. Schwarze, K.; Buchanan, J.; Taylor, J.C.; Wordsworth, S. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. *Genet. Med.* **2018**, *20*, 1122–1130. [CrossRef]
90. MassGenomics. Brace Yourself for Large-Scale Whole Genome Sequencing. Available online: http://massgenomics.org/2014/11/brace-yourself-for-large-scale-whole-genome-sequencing.html (accessed on 20 May 2021).
91. Approximate Sizes of Sequencing Run Output Folders. Available online: https://support.illumina.com/bulletins/2018/01/approximate-sizes-of-sequencing-run-output-folders.html (accessed on 20 May 2021).
92. Strand NGS. Storage and Computation Requirements. Available online: https://www.strand-ngs.com/support/ngs-data-storage-requirements (accessed on 20 May 2021).
93. CRAM Benchmarks. Available online: http://www.htslib.org/benchmarks/CRAM.html (accessed on 20 May 2021).
94. Krumm, N.; Hoffman, N. Practical estimation of cloud storage costs for clinical genomic data. *Pract. Lab. Med.* **2020**, *21*, e00168. [CrossRef]

95. Tanjo, T.; Kawai, Y.; Tokunaga, K.; Ogasawara, O.; Nagasaki, M. Practical guide for managing large-scale human genome data in research. *J. Hum. Genet.* **2021**, *66*, 39–52. [CrossRef]
96. Hart, M.R.; Biesecker, B.B.; Blout, C.L.; Christensen, K.D.; Amendola, L.M.; Bergstrom, K.L.; Biswas, S.; Bowling, K.M.; Brothers, K.B.; Conlin, L.K.; et al. Secondary findings from clinical genomic sequencing: Prevalence, patient perspectives, family history assessment, and health-care costs from a multisite study. *Genet. Med.* **2019**, *21*, 1100–1110. [CrossRef] [PubMed]
97. Green, R.C.; Berg, J.S.; Grody, W.W.; Kalia, S.S.; Korf, B.R.; Martin, C.L.; JD, A.L.M.; Nussbaum, R.L.; O'Daniel, J.M.; Ormond, K.E.; et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet. Med.* **2013**, *15*, 565–574. [CrossRef]
98. Kalia, S.S.; Adelman, K.; Bale, S.J.; Chung, W.K.; Eng, C.; Evans, J.P.; Herman, G.E.; Hufnagel, S.B.; Klein, T.E.; Korf, B.R.; et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): A policy statement of the American College of Medical Genetics and Genomics. *Genet. Med.* **2017**, *19*, 249–255. [CrossRef] [PubMed]
99. Miller, D.T.; Lee, K.; Gordon, A.S.; Amendola, L.M.; Adelman, K.; Bale, S.J.; Chung, W.K.; Gollob, M.H.; Harrison, S.M.; Herman, G.E.; et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2021 update: A policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* **2021**, 1–8. [CrossRef]
100. Miller, D.T.; Lee, K.; Chung, W.K.; Gordon, A.S.; Herman, G.E.; Klein, T.E.; Stewart, D.R.; Amendola, L.M.; Adelman, K.; Bale, S.J.; et al. ACMG SF v3.0 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* **2021**, 1–10. [CrossRef]
101. Lionel, A.C.; Costain, G.; Monfared, N.; Walker, S.; Reuter, M.S.; Hosseini, S.M.; Thiruvahindrapuram, B.; Merico, D.; Jobling, R.; Nalpathamkalam, T.; et al. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genet. Med.* **2018**, *20*, 435–443. [CrossRef] [PubMed]
102. Zeitz, C.; Michiels, C.; Neuillé, M.; Friedburg, C.; Condroyer, C.; Boyard, F.; Antonio, A.; Bouzidi, N.; Milicevic, D.; Veaux, R.; et al. Where are the missing gene defects in inherited retinal disorders? Intronic and synonymous variants contribute at least to 4% of CACNA1F -mediated inherited retinal disorders. *Hum. Mutat.* **2019**, *40*, 765–787. [CrossRef]
103. Khan, M.; Cornelis, S.S.; Pozo-Valero, M.D.; Whelan, L.; Runhart, E.H.; Mishra, K.; Bults, F.; AlSwaiti, Y.; AlTalbishi, A.; De Baere, E.; et al. Resolving the dark matter of ABCA4 for 1054 Stargardt disease probands through integrated genomics and transcriptomics. *Genet. Med.* **2020**, *22*, 1235–1246. [CrossRef]
104. Almomani, R.; Marchi, M.; Sopacua, M.; Lindsey, P.; Salvi, E.; De Koning, B.; Santoro, S.; Magri, S.; Smeets, H.J.M.; Boneschi, F.M.; et al. Evaluation of molecular inversion probe versus TruSeq® custom methods for targeted next-generation sequencing. *PLoS ONE* **2020**, *15*, e0238467. [CrossRef]
105. Pozo, M.G.-D.; Martín-Sánchez, M.; Bravo-Gil, N.; Méndez-Vidal, C.; Chimenea, Á.; La Rúa, E.R.-D.; Borrego, S.; Antiñolo, G. Searching the second hit in patients with inherited retinal dystrophies and monoallelic variants in ABCA4, USH2A and CEP290 by whole-gene targeted sequencing. *Sci. Rep.* **2018**, *8*, 1–12. [CrossRef]
106. Vervoort, R.; Lennon, A.; Bird, A.C.; Tulloch, B.; Axton, R.; Miano, M.G.; Meindl, A.; Meitinger, T.; Ciccodicola, A.; Wright, A.F. Mutational hot spot within a new RPGR exon in X-linked retinitis pigmentosa. *Nat. Genet.* **2000**, *25*, 462–466. [CrossRef]
107. Di Iorio, V.; Karali, M.; Melillo, P.; Testa, F.; Brunetti-Pierri, R.; Musacchia, F.; Condroyer, C.; Neidhardt, J.; Audo, I.; Zeitz, C.; et al. Spectrum of Disease Severity in Patients with X-Linked Retinitis Pigmentosa Due to RPGR Mutations. *Investig. Ophthalmol. Vis. Sci.* **2020**, *61*, 36. [CrossRef]
108. Chiang, J.P.W.; Lamey, T.M.; Wang, N.K.; Duan, J.; Zhou, W.; McLaren, T.L.; Thompson, J.A.; Ruddle, J.; De Roach, J.N. Development of High-Throughput Clinical Testing ofRPGRORF15 Using a Large Inherited Retinal Dystrophy Cohort. *Investig. Ophthalmol. Vis. Sci.* **2018**, *59*, 4434–4440. [CrossRef] [PubMed]
109. Maggi, J.; Roberts, L.; Koller, S.; Rebello, G.; Berger, W.; Ramesar, R. De Novo Assembly-Based Analysis of *RPGR* Exon ORF15 in an Indigenous African Cohort Overcomes Limitations of a Standard Next-Generation Sequencing (NGS) Data Analysis Pipeline. *Genes* **2020**, *11*, 800. [CrossRef]
110. Deeb, S. The molecular basis of variation in human color vision. *Clin. Genet.* **2005**, *67*, 369–377. [CrossRef]
111. Atilano, S.R.; Kenney, M.C.; Briscoe, A.D.; Jameson, K.A. A two-step method for identifying photopigment opsin and rhodopsin gene sequences underlying human color vision phenotypes. *Mol. Vis.* **2020**, *26*, 158–172.
112. Ueyama, H.; Li, Y.-H.; Fu, G.-L.; Lertrit, P.; Atchaneeyasakul, L.-O.; Oda, S.; Tanabe, S.; Nishida, Y.; Yamade, S.; Ohkubo, I. An A-71C substitution in a green gene at the second position in the red/green visual-pigment gene array is associated with deutan color-vision deficiency. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3357–3362. [CrossRef]
113. Katagiri, S.; Iwasa, M.; Hayashi, T.; Hosono, K.; Yamashita, T.; Kuniyoshi, K.; Ueno, S.; Kondo, M.; Ueyama, H.; Ogita, H.; et al. Genotype determination of the OPN1LW/OPN1MW genes: Novel disease-causing mechanisms in Japanese patients h blue cone monochromacy. *Sci. Rep.* **2018**, *8*, 1–10. [CrossRef]
114. Yatsenko, S.; Bakos, H.; Vitullo, K.; Kedrov, M.; Kishore, A.; Jennings, B.; Surti, U.; Wood-Trageser, M.; Cercone, S.; Yatsenko, A.; et al. High-resolution microarray analysis unravels complex Xq28 aberrations in patients and carriers affected by X-linked blue cone monochromacy. *Clin. Genet.* **2015**, *89*, 82–87. [CrossRef]
115. Wang, C.; Hosono, K.; Kachi, S.; Suto, K.; Nakamura, M.; Terasaki, H.; Miyake, Y.; Hotta, Y.; Minoshima, S. Novel OPN1LW/OPN1MW deletion mutations in 2 Japanese families with blue cone monochromacy. *Hum. Genome Var.* **2016**, *3*, 16011. [CrossRef]

116. Radziwon, A.; Arno, G.; Wheaton, D.K.; McDonagh, E.M.; Baple, E.L.; Webb-Jones, K.; Birch, D.G.; Webster, A.R.; Macdonald, I.M. Single-base substitutions in the CHM promoter as a cause of choroideremia. *Hum. Mutat.* **2017**, *38*, 704–715. [CrossRef]
117. Coppieters, F.; Todeschini, A.L.; Fujimaki, T.; Baert, A.; De Bruyne, M.; Van Cauwenbergh, C.; Verdin, H.; Bauwens, M.; Ongenaert, M.; Kondo, M.; et al. Hidden Genetic Variation in LCA9-Associated Congenital Blindness Explained by 5′UTR Mutations and Copy-Number Variations of NMNAT1. *Hum. Mutat.* **2015**, *36*, 1188–1196. [CrossRef]
118. Donato, L.; Scimone, C.; Rinaldi, C.; Aragona, P.; Briuglia, S.; D'Ascola, A.; D'Angelo, R.; Sidoti, A. Stargardt Phenotype Associated with Two ELOVL4 Promoter Variants and ELOVL4 Downregulation: New Possible Perspective to Etiopathogenesis? *Investig. Ophthalmol. Vis. Sci.* **2018**, *59*, 843–857. [CrossRef]
119. Van Schil, K.; Naessens, S.; Van De Sompele, S.; Carron, M.; Aslanidis, A.; Van Cauwenbergh, C.; Mayer, A.K.; Van Heetvelde, M.; Bauwens, M.; Verdin, H.; et al. Mapping the genomic landscape of inherited retinal disease genes prioritizes genes prone to coding and noncoding copy-number variations. *Genet. Med.* **2017**, *20*, 202–213. [CrossRef]
120. Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody, W.W.; Hegde, M.; Lyon, E.; Spector, E.; et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **2015**, *17*, 405–423. [CrossRef]
121. Brandt, T.; Sack, L.M.; Arjona, D.; Tan, D.; Mei, H.; Cui, H.; Gao, H.; Bean, L.J.H.; Ankala, A.; Del Gaudio, D.; et al. Adapting ACMG/AMP sequence variant classification guidelines for single-gene copy-number variants. *Genet. Med.* **2019**, *22*, 670–671, Correction in **2020**, *22*, 336–344. [CrossRef] [PubMed]
122. Li, H.-P.; Yuan, S.-Q.; Wang, X.-G.; Sheng, X.-L.; Li, X.-R. Myopia with X-linked retinitis pigmentosa results from a novel gross deletion of RPGR gene. *Int. J. Ophthalmol.* **2020**, *13*, 1306–1311. [CrossRef] [PubMed]
123. Buena-Atienza, E.; Nasser, F.; Kohl, S.; Wissinger, B. A 73,128 bp de novo deletion encompassing the OPN1LW/OPN1MW gene cluster in sporadic Blue Cone Monochromacy: A case report. *BMC Med. Genet.* **2018**, *19*, 107. [CrossRef]
124. De Bruijn, S.E.; Fiorentino, A.; Ottaviani, D.; Fanucchi, S.; Melo, U.S.; Corral-Serrano, J.C.; Mulders, T.; Georgiou, M.; Rivolta, C.; Pontikos, N.; et al. Structural Variants Create New Topological-Associated Domains and Ectopic Retinal Enhancer-Gene Contact in Dominant Retinitis Pigmentosa. *Am. J. Hum. Genet.* **2020**, *107*, 802–814. [CrossRef]
125. Jones, K.D.; Radziwon, A.; Birch, D.G.; MacDonald, I.M. A novel SVA retrotransposon insertion in the CHM gene results in loss of REP-1 causing choroideremia. *Ophthalmic Genet.* **2020**, *41*, 341–344. [CrossRef]
126. Hurk, J.A.J.M.V.D.; Van De Pol, D.J.R.; Wissinger, B.; Van Driel, M.A.; Hoefsloot, L.H.; De Wijs, I.J.; Born, L.I.V.D.; Heckenlively, J.R.; Brunner, H.G.; Zrenner, E.; et al. Novel types of mutation in the choroideremia (CHM) gene: A full-length L1 insertion and an intronic mutation activating a cryptic exon. *Qual. Life Res.* **2003**, *113*, 268–275. [CrossRef]
127. Hitti-Malin, R.J.; Burmeister, L.M.; Ricketts, S.L.; Lewis, T.W.; Pettitt, L.; Boursnell, M.; Schofield, E.C.; Sargan, D.; Mellersh, C.S. A LINE-1 insertion situated in the promoter of IMPG2 is associated with autosomal recessive progressive retinal atrophy in Lhasa Apso dogs. *BMC Genet.* **2020**, *21*, 1–17. [CrossRef]
128. Tavares, E.; Tang, C.Y.; Vig, A.; Li, S.; Billingsley, G.; Sung, W.; Vincent, A.; Thiruvahindrapuram, B.; Héon, E. Retrotransposon insertion as a novel mutational event in Bardet-Biedl syndrome. *Mol. Genet. Genom. Med.* **2018**, *7*, e00521. [CrossRef]
129. Delvallée, C.; Nicaise, S.; Antin, M.; Leuvrey, A.-S.; Nourisson, E.; Leitch, C.C.; Kellaris, G.; Stoetzel, C.; Geoffroy, V.; Scheidecker, S.; et al. A BBS1 SVA F retrotransposon insertion is a frequent cause of Bardet-Biedl syndrome. *Clin. Genet.* **2020**, 13878. [CrossRef]
130. Ma, X.; Fan, J.; Wu, Y.; Zhao, S.; Zheng, X.; Sun, C.; Tan, L. Whole-genome de novo assemblies reveal extensive structural variations and dynamic organelle-to-nucleus DNA transfers in African and Asian rice. *Plant J.* **2020**, *104*, 596–612. [CrossRef]
131. Zampaglione, E.; Kinde, B.; Place, E.M.; Navarro-Gomez, D.; Ma, C.F.; Jamshidi, F.; Nassiri, S.; Mazzone, J.A.; Finn, C.; Schlegel, D.; et al. Copy-number variation contributes 9% of pathogenicity in the inherited retinal degenerations. *Genet. Med.* **2020**, *22*, 1079–1087. [CrossRef]
132. Bujakowska, K.M.; Fernandez-Godino, R.; Place, E.; Consugar, M.; Navarro-Gomez, D.; White, J.; Bedoukian, E.C.; Zhu, X.; Xie, H.M.; Gai, X.; et al. Copy-number variation is an important contributor to the genetic causality of inherited retinal degenerations. *Genet. Med.* **2016**, *19*, 643–651. [CrossRef]
133. Ellingford, J.M.; Horn, B.; Campbell, C.; Arno, G.; Barton, S.; Tate, C.; Bhaskar, S.; Sergouniotis, P.I.; Taylor, R.L.; Carss, K.J.; et al. Assessment of the incorporation of CNV surveillance into gene panel next-generation sequencing testing for inherited retinal diseases. *J. Med. Genet.* **2017**, *55*, 114–121. [CrossRef]
134. Dong, X.; Liu, B.; Yang, L.; Wang, H.; Wu, B.; Liu, R.; Chen, H.; Chen, X.; Yu, S.; Chen, B.; et al. Clinical exome sequencing as the first-tier test for diagnosing developmental disorders covering both CNV and SNV: A Chinese cohort. *J. Med. Genet.* **2020**, *57*, 558–566. [CrossRef]
135. Sloan-Heggen, C.M.; Bierer, A.O.; Shearer, A.E.; Kolbe, D.L.; Nishimura, C.J.; Frees, K.L.; Ephraim, S.S.; Shibata, S.B.; Booth, K.T.; Campbell, C.; et al. Comprehensive genetic testing in the clinical evaluation of 1119 patients with hearing loss. *Qual. Life Res.* **2016**, *135*, 441–450. [CrossRef]
136. Sanchez-Navarro, I.; Da Silva, L.R.J.; Blanco-Kelly, F.; Zurita, O.; Sanchez-Bolivar, N.; Villaverde, C.; Lopez-Molina, M.I.; Garcia-Sandoval, B.; Swafiri, S.T.S.; Minguez, P.; et al. Combining targeted panel-based resequencing and copy-number variation analysis for the diagnosis of inherited syndromic retinopathies and associated ciliopathies. *Sci. Rep.* **2018**, *8*, 1–11. [CrossRef] [PubMed]
137. Goodwin, S.; McPherson, J.D.; McCombie, W.R. Coming of age: Ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **2016**, *17*, 333–351. [CrossRef]

138. De Koning, A.P.J.; Gu, W.; Castoe, T.A.; Batzer, M.A.; Pollock, D.D. Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. *PLoS Genet.* **2011**, *7*, e1002384. [CrossRef]
139. Chaisson, M.J.P.; Wilson, R.K.; Eichler, E.E. Genetic variation and the de novo assembly of human genomes. *Nat. Rev. Genet.* **2015**, *16*, 627–640. [CrossRef]
140. Mantere, T.; Kersten, S.; Hoischen, A. Long-Read Sequencing Emerging in Medical Genetics. *Front. Genet.* **2019**, *10*, 426. [CrossRef]
141. Botton, M.R.; Yang, Y.; Scott, E.R.; Desnick, R.J.; Scott, S.A. Phased Haplotype Resolution of the *SLC6A4* Promoter Using Long-Read Single Molecule Real-Time (SMRT) Sequencing. *Genes* **2020**, *11*, 1333. [CrossRef] [PubMed]
142. Sanchis-Juan, A.; Stephens, J.; French, C.E.; Gleadall, N.; Mégy, K.; Penkett, C.; Shamardina, O.; Stirrups, K.; Delon, I.; Dewhurst, E.; et al. Complex structural variants in Mendelian disorders: Identification and breakpoint resolution using short- and long-read genome sequencing. *Genome Med.* **2018**, *10*, 1–10. [CrossRef]
143. Beales, P.L.; Badano, J.L.; Ross, A.J.; Ansley, S.J.; Hoskins, B.E.; Kirsten, B.; Mein, C.A.; Froguel, P.; Scambler, P.J.; Lewis, R.A.; et al. Genetic Interaction of BBS1 Mutations with Alleles at Other BBS Loci Can Result in Non-Mendelian Bardet-Biedl Syndrome. *Am. J. Hum. Genet.* **2003**, *72*, 1187–1199. [CrossRef] [PubMed]
144. Wheway, G.; Douglas, A.; Baralle, D.; Guillot, E. Mutation spectrum of PRPF31, genotype-phenotype correlation in retinitis pigmentosa, and opportunities for therapy. *Exp. Eye Res.* **2020**, *192*, 107950. [CrossRef] [PubMed]
145. Pormehr, L.A.; Ahmadian, S.; Daftarian, N.; Mousavi, S.A.; Shafiezadeh, M. PRPF31 reduction causes mis-splicing of the phototransduction genes in human organotypic retinal culture. *Eur. J. Hum. Genet.* **2019**, *28*, 491–498. [CrossRef]
146. Yang, D.; Yao, Q.; Li, Y.; Xu, Y.; Wang, J.; Zhao, H.; Liu, F.; Zhang, Z.; Liu, Y.; Bie, X.; et al. A c.544_618del75bp mutation in the splicing factor gene PRPF31 is involved in non-syndromic retinitis pigmentosa by reducing the level of mRNA expression. *Ophthalmic Physiol. Opt.* **2020**, *40*, 289–299. [CrossRef]
147. Rose, A.M.; Shah, A.Z.; Venturini, G.; Krishna, A.; Chakravarti, A.; Rivolta, C.; Bhattacharya, S.S. Transcriptional regulation of PRPF31 gene expression by MSR1 repeat elements causes incomplete penetrance in retinitis pigmentosa. *Sci. Rep.* **2016**, *6*, 19450. [CrossRef] [PubMed]
148. Green, D.J.; Sallah, S.R.; Ellingford, J.M.; Lovell, S.C.; Sergouniotis, P.I. Variability in Gene Expression is Associated with Incomplete Penetrance in Inherited Eye Disorders. *Genes* **2020**, *11*, 179. [CrossRef] [PubMed]
149. The Genome Aggregation Database. Available online: https://gnomad.broadinstitute.org/ (accessed on 14 November 2019).
150. Sangermano, R.; Khan, M.; Cornelis, S.S.; Richelle, V.; Albert, S.; Garanto, A.; Elmelik, D.; Qamar, R.; Lugtenberg, D.; Born, L.I.V.D.; et al. ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome Res.* **2017**, *28*, 100–110. [CrossRef] [PubMed]
151. Di Scipio, M.; Tavares, E.; Deshmukh, S.; Audo, I.; Green-Sanderson, K.; Zubak, Y.; Zine-Eddine, F.; Pearson, A.; Vig, A.; Tang, C.Y.; et al. Phenotype Driven Analysis of Whole Genome Sequencing Identifies Deep Intronic Variants that Cause Retinal Dystrophies by Aberrant Exonization. *Investig. Ophthalmol. Vis. Sci.* **2020**, *61*, 36. [CrossRef]
152. Fadaie, Z.; Khan, M.; Del Pozo-Valero, M.; Cornelis, S.S.; Ayuso, C.; Cremers, F.P.M.; Roosing, S. Identification of splice defects due to noncanonical splice site or deep-intronic variants in ABCA4. *Hum. Mutat.* **2019**, *40*, 2365–2376. [CrossRef] [PubMed]
153. Jonsson, F.; Westin, I.M.; Österman, L.; Sandgren, O.; Burstedt, M.; Holmberg, M.; Golovleva, I. ATP-binding cassette subfamily A, member 4 intronic variants c.4773+3A>G and c.5461-10T>C cause Stargardt disease due to defective splicing. *Acta Ophthalmol.* **2018**, *96*, 737–743. [CrossRef]
154. Vig, A.; Poulter, J.A.; Poulter, J.A.; Ottaviani, D.; Tavares, E.; Toropova, K.; Tracewska, A.M.; Mollica, A.; Kang, J.; Kehelwathugoda, O.; et al. DYNC2H1 hypomorphic or retina-predominant variants cause nonsyndromic retinal degeneration. *Genet. Med.* **2020**, *22*, 2041–2051. [CrossRef]
155. Brooks, M.J.; Chen, H.Y.; Kelley, R.A.; Mondal, A.K.; Nagashima, K.; De Val, N.; Li, T.; Chaitankar, V.; Swaroop, A. Improved Retinal Organoid Differentiation by Modulating Signaling Pathways Revealed by Comparative Transcriptome Analyses with Development In Vivo. *Stem Cell Rep.* **2019**, *13*, 891–905. [CrossRef]
156. Cowan, A.S.; Renner, M.; De Gennaro, M.; Roma, G.; Nigsch, F.; Roska, B.; Cowan, C.S.; Gross-Scherf, B.; Goldblum, D.; Hou, Y.; et al. Cell Types of the Human Retina and Its Organoids at Single-Cell Resolution. *Cell* **2020**, *182*. [CrossRef]
157. Fahim, A.T.; Sullivan, L.S.; Bowne, S.J.; Jones, K.D.; Wheaton, D.K.; Khan, N.W.; Heckenlively, J.R.; Jayasundera, K.T.; Branham, K.H.; Andrews, C.A.; et al. X-Chromosome Inactivation Is a Biomarker of Clinical Severity in Female Carriers of RPGR-Associated X-Linked Retinitis Pigmentosa. *Ophthalmol. Retin.* **2020**, *4*, 510–520. [CrossRef] [PubMed]
158. Zhang, Q.; Giacalone, J.C.; Searby, C.; Stone, E.M.; Tucker, B.A.; Sheffield, V.C. Disruption of RPGR protein interaction network is the common feature of RPGR missense variations that cause XLRP. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 1353–1360. [CrossRef] [PubMed]
159. Fahim, A.T.; Bowne, S.J.; Sullivan, L.S.; Webb, K.D.; Williams, J.T.; Wheaton, D.K.; Birch, D.G.; Daiger, S.P. Polymorphic variation of RPGRIP1L and IQCB1 as modifiers of X-linked retinitis pigmentosa caused by mutations in RPGR. *Single Mol. Single Cell Seq.* **2011**, *723*, 313–320. [CrossRef]
160. Li, Y.; Xia, X.; Paulus, Y.M. Advances in Retinal Optical Imaging. *Photonics* **2018**, *5*, 9. [CrossRef]
161. Litts, K.M.; Woertz, E.N.; Georgiou, M.; Patterson, E.J.; Lam, B.L.; Fishman, G.A.; Pennesi, M.E.; Kay, C.N.; Hauswirth, W.W.; Michaelides, M.; et al. Optical Coherence Tomography Artifacts Are Associated with Adaptive Optics Scanning Light Ophthalmoscopy Success in Achromatopsia. *Transl. Vis. Sci. Technol.* **2021**, *10*, 11. [CrossRef]

162. Russakoff, D.B.; Lamin, A.; Oakley, J.D.; Dubis, A.M.; Sivaprasad, S. Deep Learning for Prediction of AMD Progression: A Pilot Study. *Investig. Ophthalmol. Vis. Sci.* **2019**, *60*, 712–722. [CrossRef] [PubMed]
163. Schmidt-Erfurth, U.; Waldstein, S.M.; Klimscha, S.; Sadeghipour, A.; Hu, X.; Gerendas, B.S.; Osborne, A.; Bogunovic, H. Prediction of Individual Disease Conversion in Early AMD Using Artificial Intelligence. *Investig. Ophthalmol. Vis. Sci.* **2018**, *59*, 3199–3208. [CrossRef]
164. Schmidt-Erfurth, U.; Bogunovic, H.; Sadeghipour, A.; Schlegl, T.; Langs, G.; Gerendas, B.S.; Osborne, A.; Waldstein, S.M. Machine Learning to Analyze the Prognostic Value of Current Imaging Biomarkers in Neovascular Age-Related Macular Degeneration. *Ophthalmol. Retin.* **2018**, *2*, 24–30. [CrossRef]
165. Burlina, P.M.; Joshi, N.; Pacheco, K.D.; Freund, D.E.; Kong, J.; Bressler, N.M. Use of Deep Learning for Detailed Severity Characterization and Estimation of 5-Year Risk Among Patients with Age-Related Macular Degeneration. *JAMA Ophthalmol.* **2018**, *136*, 1359–1366. [CrossRef]
166. Burlina, P.; Joshi, N.; Pacheco, K.D.; Freund, D.E.; Kong, J.; Bressler, N.M. Utility of Deep Learning Methods for Referability Classification of Age-Related Macular Degeneration. *JAMA Ophthalmol.* **2018**, *136*, 1305–1307. [CrossRef]
167. Grassmann, F.; Mengelkamp, J.; Brandl, C.; Harsch, S.; Zimmermann, M.E.; Linkohr, B.; Peters, A.; Heid, I.M.; Palm, C.; Weber, B.H. A Deep Learning Algorithm for Prediction of Age-Related Eye Disease Study Severity Scale for Age-Related Macular Degeneration from Color Fundus Photography. *Ophthalmology* **2018**, *125*, 1410–1420. [CrossRef] [PubMed]
168. Goldhagen, B.E.; Al-Khersan, H. Diving Deep into Deep Learning: An Update on Artificial Intelligence in Retina. *Curr. Ophthalmol. Rep.* **2020**, *8*, 1–8. [CrossRef]
169. Peng, Y.; Dharssi, S.; Chen, Q.; Keenan, T.D.; Agrón, E.; Wong, W.T.; Chew, E.Y.; Lu, Z. DeepSeeNet: A Deep Learning Model for Automated Classification of Patient-based Age-related Macular Degeneration Severity from Color Fundus Photographs. *Ophthalmology* **2019**, *126*, 565–575. [CrossRef] [PubMed]
170. Charng, J.; Xiao, D.; Mehdizadeh, M.; Attia, M.S.; Arunachalam, S.; Lamey, T.M.; Thompson, J.A.; McLaren, T.L.; De Roach, J.N.; Mackey, D.A.; et al. Deep learning segmentation of hyperautofluorescent fleck lesions in Stargardt disease. *Sci. Rep.* **2020**, *10*, 1–13. [CrossRef] [PubMed]
171. Ha, A.; Sun, S.; Kim, Y.K.; Lee, J.; Jeoung, J.W.; Kim, H.C.; Park, K.H. Deep-learning-based enhanced optic-disc photography. *PLoS ONE* **2020**, *15*, e0239913. [CrossRef]
172. Galvin, O.; Chi, G.; Brady, L.; Hippert, C.; Rubido, M.D.V.; Daly, A.; Michaelides, M. The Impact of Inherited Retinal Diseases in the Republic of Ireland (ROI) and the United Kingdom (UK) from a Cost-of-Illness Perspective. *Clin. Ophthalmol.* **2020**, *14*, 707–719. [CrossRef]
173. McVeigh, E.; Jones, H.; Black, G.; Hall, G. The psychosocial and service delivery impact of genomic testing for inherited retinal dystrophies. *J. Community Genet.* **2019**, *10*, 425–434. [CrossRef]
174. Sabbaghi, H.; Daftarian, N.; Suri, F.; Mirrahimi, M.; Madani, S.; Sheikhtaheri, A.; Khorrami, F.; Saviz, P.; Nejad, M.Z.; Tivay, A.; et al. The First Inherited Retinal Disease Registry in Iran: Research Protocol and Results of a Pilot Study. *Arch. Iran. Med.* **2020**, *23*, 445–454. [CrossRef]
175. Marques, J.P.; Carvalho, A.L.; Henriques, J.; Murta, J.N.; Saraiva, J.; Silva, R. Design, development and deployment of a web-based interoperable registry for inherited retinal dystrophies in Portugal: The IRD-PT. *Orphanet J. Rare Dis.* **2020**, *15*, 1–13. [CrossRef] [PubMed]
176. Lejeune, C.; Amado, I.F. Valuing genetic and genomic testing in France: Current challenges and latest evidence. *J. Community Genet.* **2021**, 1–9. [CrossRef]
177. Sergouniotis, P.I. Inherited Retinal Disorders: Using Evidence as a Driver for Implementation. *Ophthalmology* **2019**, *242*, 187–194. [CrossRef]

Appendix 6

npj | Genomic Medicine

www.nature.com/npjgenmed

ARTICLE OPEN

Check for updates

Whole genome sequencing and in vitro splice assays reveal genetic causes for inherited retinal diseases

Zeinab Fadaie[1,2,12], Laura Whelan[3,12], Tamar Ben-Yosef[4], Adrian Dockery[3], Zelia Corradi[1,2], Christian Gilissen[1,5], Lonneke Haer-Wigman[1], Jordi Corominas[1,5], Galuh D. N. Astuti[1,6], Laura de Rooij[1], L. Ingeborgh van den Born[7], Caroline C. W. Klaver[8,9,10], Carel B. Hoyng[2,10], Niamh Wynne[11], Emma S. Duignan[11], Paul F. Kenna[3,11], Frans P. M. Cremers[1,2], G. Jane Farrar[3] and Susanne Roosing[1,2]✉

Inherited retinal diseases (IRDs) are a major cause of visual impairment. These clinically heterogeneous disorders are caused by pathogenic variants in more than 270 genes. As 30–40% of cases remain genetically unexplained following conventional genetic testing, we aimed to obtain a genetic diagnosis in an IRD cohort in which the genetic cause was not found using whole-exome sequencing or targeted capture sequencing. We performed whole-genome sequencing (WGS) to identify causative variants in 100 unresolved cases. After initial prioritization, we performed an in-depth interrogation of all noncoding and structural variants in genes when one candidate variant was detected. In addition, functional analysis of putative splice-altering variants was performed using in vitro splice assays. We identified the genetic cause of the disease in 24 patients. Causative coding variants were observed in genes such as *ATXN7*, *CEP78*, *EYS*, *FAM161A*, and *HGSNAT*. Gene disrupting structural variants were also detected in *ATXN7*, *PRPF31*, and *RPGRIP1*. In 14 monoallelic cases, we prioritized candidate noncanonical splice sites or deep-intronic variants that were predicted to disrupt the splicing process based on in silico analyses. Of these, seven cases were resolved as they carried pathogenic splice defects. WGS is a powerful tool to identify causative variants residing outside coding regions or heterozygous structural variants. This approach was most efficient in cases with a distinct clinical diagnosis. In addition, in vitro splice assays provide important evidence of the pathogenicity of rare variants.

npj Genomic Medicine (2021)6:97 ; https://doi.org/10.1038/s41525-021-00261-1

INTRODUCTION

Inherited retinal diseases (IRDs) are a group of conditions showing dysfunction and/or degeneration of the neural retina or retinal pigment epithelium, resulting in visual impairment. They affect more than two million people worldwide and display a very high degree of clinical and genetic heterogeneity[1,2]. Therefore, defining the precise underlying genetic cause of the disease has a profound impact on the potential to diagnose, counsel, and provide accurate family risk assessment. Furthermore, genetic diagnoses can enable the development of new therapeutic approaches and access to these treatments[2–4].

High-throughput DNA sequencing technologies are widely used to investigate IRD pathogenesis as they display a high degree of heterogeneity, with more than 270 genes and loci associated with IRDs, encompassing all Mendelian inheritance patterns and digenic inheritance (https://sph.uth.edu/retnet/). Next-generation sequencing (NGS) facilitates the analysis of genetic variation in multiple regions of the genome in a single experiment[5]. For many years, the application of NGS in IRD molecular diagnostics had been limited to target-capture sequencing (TCS) and whole-exome sequencing (WES). TCS has been shown to be an effective preliminary diagnostic technique, boasting cost effectiveness, rapid data generation, and lower volumes of data interpretation as major advantages[6–8]. Sequence analysis of only previously IRD-associated genes is one of the major disadvantages of this technique[9,10]. On the contrary, WES is not limited to IRD-associated genes, as this approach also enables the analysis of genes not yet associated with IRD. It is estimated that 85% of pathogenic variants are located within protein-coding regions, the primary focus of WES[11,12]. However, this may vary depending on the particular disease entity. It is evident that sequencing noncoding regions of the genome is required to also identify causal deep-intronic variants and to improve the identification of structural variants (SVs) and their breakpoints[13–16].

An accurate splicing process requires highly conserved sequences in the canonical splice donor site (SDS) and splice acceptor site (SAS). All SASs contain AG at the canonical −1 and −2 positions and 98.7% of SDSs contain GT in the canonical +1 and +2 positions. In addition, the first and last three nucleotides of an exon as well as the −3 to −14 nucleotides from the SAS and +3 to +6 of the SDS are conserved and ensure a precise splicing process[17]. However, variants in intronic regions of the genome can introduce or strengthen a cryptic splice site, which resembles a canonical splice site, and in the presence of a nearby SAS or SDS, may lead to pseudoexon insertion and potentially disrupt the reading frame causing protein truncation[17,18]. In both TCS and WES approaches, the analysis is restricted to genetic variations in protein-coding and flanking splice site regions of the genome.

[1]Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands. [2]Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center, Nijmegen, The Netherlands. [3]The School of Genetics and Microbiology, Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland. [4]Rappaport Faculty of Medicine, Technion-Israel Institute of Technology, Haifa, Israel. [5]Radboud Institute of Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands. [6]Division of Human Genetics, Center for Biomedical Research (CEBIOR), Faculty of Medicine, Diponegoro University, Semarang, Indonesia. [7]The Rotterdam Eye Hospital, Rotterdam, The Netherlands. [8]Department of Ophthalmology, Erasmus Medical Center, Rotterdam, The Netherlands. [9]Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands. [10]Department of Ophthalmology, Radboud University Medical Center, Nijmegen, The Netherlands. [11]Department of Ophthalmology, Royal Victoria Eye and Ear Hospital, Dublin, Ireland. [12]These authors contributed equally: Zeinab Fadaie, Laura Whelan. ✉email: Susanne.Roosing@radboudumc.nl

Published in partnership with CEGMR, King Abdulaziz University

npj nature partner journals

Furthermore, these techniques are limited in their ability to detect SVs such as deletions, duplications, or inversions when breakpoints lie in intronic or intergenic regions[19].

These limitations are less evident in WGS, which enables the identification of most types of variation across the entire genome[2,20]. TCS and WES are the most commonly employed techniques for the genetic diagnosis of IRDs. However, because of decreasing costs and increasing data interpretability, WGS is becoming an attractive alternative albeit predominantly for research purposes, as diagnostic laboratories are able to interpret variants only in previously associated IRD genes with confidence using the American College of Medical Genetics and Genomics (ACMG) guidelines[19]. In addition, IRD cases are ideal candidates for WGS analysis as they are largely presumed to be monogenic diseases with a high degree of genetic heterogeneity.

In this study, we aimed to investigate the sensitivity and accuracy of WGS in a cohort of 100 individuals with different IRDs. All individuals were previously tested by TCS or WES, but a genetic cause of disease remained undetermined. The study provides insights into the increased resolution of IRD cases that can be achieved by employing WGS.

RESULTS

Patient characterization

We performed WGS on 100 IRD cases with a suspected autosomal recessive inheritance recruited in Ireland, Israel, and The Netherlands. Among these, TCS was previously performed on 38 Irish and eight Israeli cases, which did not result in a complete genetic diagnosis. The remaining 54 cases (43 Dutch and 11 Israeli) were previously examined using WES and also remained genetically unresolved. In 56/100 cases, a first plausible pathogenic variant was selected from prior WES or TCS data based on a minor allele frequency of <1% in the ExAC database[21] and 5% in dbSNP as well as <1% in an in-house database of 15,576 mostly Caucasian individuals and <2% in an in-house database of 454 mostly Asian individuals[22]. In the prior prioritization of monoallelic cases, all coding and splice site variants (nucleotides −17 to −1 and +1 to +6) were manually classified according to a five-class system based on the ACMG-Association for Molecular Pathology (AMP) classification system[23], followed by evaluating the prioritized variants using practice guidelines by Wallis et al.[24]. Therefore, these were included as monoallelic cases in this study (Supplementary Table 1).

A broad range of IRDs were included in this cohort (Table 1 and Fig. 1). Retinitis pigmentosa (RP, OMIM: 268000) is the most frequently listed phenotype (53%), followed by cone–rod dystrophy (9%, OMIM: 601777), macular dystrophy (8%, OMIM: 616152), and Stargardt disease (STGD1, 6%, OMIM: 248200). Rare forms of IRD such as Leber congenital amaurosis (LCA, OMIM: 204000), Bardet–Biedl syndrome (OMIM: 209900), and Senior–Løken syndrome (OMIM: 266900) were also included.

Identification of coding and noncoding candidate variants after WGS data analysis

Following comprehensive WGS analysis and in vitro functional assays, we identified a genetic cause of disease in 24 probands (24%). Pathogenic variants were confirmed by Sanger sequencing and segregation analysis when family members were available (9/24) (Table 2 and Supplementary Table 2). Among these, 15 probands (55.5%) carried previously undetected pathogenic variants in coding regions and three individuals (11.1%) carried SVs in *RPGRIP1* and *PRPF31*, or a triplet tandem repeat expansion in *ATXN7*. In proband Pt-4, presenting with cone–rod dystrophy

Table 1. Phenotypic diversity of patients and the number of cases referred for TCS, WES, and WGS.

Clinical diagnosis	No. of cases previously tested by WES	No. of cases previously tested by TCS	No. of cases tested by WGS
Retinitis pigmentosa	21	32	53
Cone dystrophy	2	1	3
Rod–cone dystrophy	3	0	3
Cone–rod dystrophy	7	2	9
Alström syndrome	0	1	1
Bardet–Biedl syndrome	1	0	1
Leber congenital amaurosis	2	0	2
Macular dystrophy	8	1	9
Oguchi disease	1	0	1
Central areolar choroidal dystrophy	1	0	1
Fundus albipunctatus	1	0	1
Usher syndrome type III	1	0	1
Senior–Løken syndrome	1	0	1
Nystagmus	1	0	1
Stargardt disease	0	7	7
Clumped pigmentary retinal dystrophy	1	0	1
Cone–rod dystrophy with progressive neurodegeneration	1	0	1
Retinitis pigmentosa with hypogonadism	1	0	1
Bietti crystalline corneoretina dystrofie	1	0	1
Congenital stationary night blindness	0	1	1
Achromatopsia	0	1	1
Total number	54	46	100

WES whole-exome sequencing, *TCS* target-capture sequencing, *WGS* whole-genome sequencing.

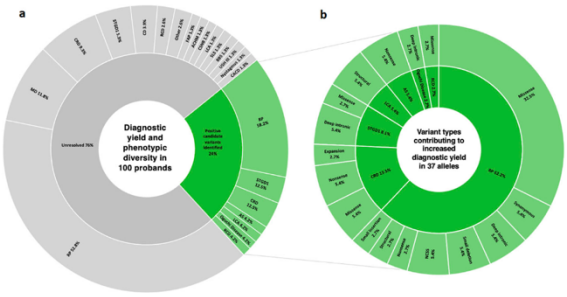


Fig. 1 Sunburst diagrams illustrating diagnostic yield, phenotypic diversity and variants contributing to increased diagnostic yield in this study. a Diagnostic yield and phenotypic diversity. The inner rings represent solved (green) and unresolved (gray) cases following whole-genome sequencing. The outer ring encompasses the phenotypic presentation of patients in these groups. **b** Variant types contributing to increased diagnostic yield. The types of variants contributing to increased diagnostic yield are illustrated on the outer ring, with their corresponding clinical phenotypes on the inner ring. ACHM achromatopsia, AS Alström syndrome, BBS Bardet–Biedl syndrome, CACD central areolar choroidal dystrophy, CD cone dystrophy, CRD cone–rod dystrophy, CSNB congenital stationary night blindness, FAP fundus albipunctatus, LCA Leber congenital amaurosis, MD macular dystrophy, NCSS noncanonical splice site, RCD rod–cone dystrophy, RP retinitis pigmentosa, SLS Senior–Loken syndrome, STGD1 Stargardt disease type 1, USH III Usher syndrome type III. Small insertions include insertions <4-bp in length. Small deletions include deletions <4-bp in length.

and progressive neurodegeneration (OMIM: 164500), we identified a pathogenic CAG repeat expansion in exon 3 of *ATXN7* (NG_008227.1, p.(Gln30[70])) (Table 2 and Supplementary Fig. 1A). While the wild-type allele carries 3–19 CAG repeats in *ATXN7*, the pathogenic allele can be between 37 and 460 repeats[25].

The second proband (Pt-22) illustrates the strength of WGS for SV detection compared to WES. The patient was diagnosed with LCA (MIM: 605446), with one pathogenic allele in *CEP290* (p.(Arg360Gln)), *PDE6A* (p.(Val685Met)), and *RPE65* (p.(Val226Ile)) (Supplementary Table 1). However, WGS data revealed—and segregation analysis confirmed—a homozygous *RPGRIP1* (NG_008933.1) exon 20 deletion (p.(Asp1080Glyfs*6)), missed previously due to incapability of WES analysis tools to detect copy number variants (CNVs) using the SureSelect XT Human All Exon V4 (SOLiD 5500xl sequencer in 2012). While we cannot rule out that a more recent WES strategy would have detected this single exon deletion, this approach would not have revealed the breakpoints. A 12-bp microhomology region was identified in the breakpoint boundaries (Table 2 and Supplementary Fig. 1D). Completing the variant analysis, the noncoding regions of *CEP290*, *RPE65*, and *PDE6A* did not contain additional potential pathogenic variants. The third proband (Pt-23) had a 26.68-kb heterozygous deletion on chromosome 19, encompassing *PRPF31*, *TFPT*, and the promoter region of *NDUFA3*. This large deletion was not detected previously using TCS assessment (Table 2 and Supplementary Fig. 1B).

Furthermore, we prioritized ten deep-intronic variants and three noncanonical splice site (NCSS) variants in 14 individuals using criteria in Alamut and SpliceAI (Supplementary Tables 3 and 4) outlined in the variant frequency and pathogenicity prediction parameters section of the methods. For the indel variants detected in *ABCA4* (NG_009073.1, in Pt-28, Pt-29, and Pt-30), *CYP4V2* (NG_007965.1, in Pt-52), and *RLBP1* (NG_008116.1, in Pt-65), no SpliceAI scores were provided due to the nature of these variants. Therefore, inclusion of these three variants for in vitro splice assays was based on Alamut prediction scores only. The pathogenicity of noncoding variants was determined in seven cases (46.6%) after in vitro splice assays (see below) (Table 2 and Fig. 1). While employing WGS, significant progress was made in defining the genetic pathogenesis of IRDs in this prescreened IRD cohort; 76% of cases remained genetically unexplained.

Splice defects due to NCSS variants in *CDHR1*, *HGSNAT*, and *USH2A*

In Pt-7, we identified a homozygous synonymous variant in the last nucleotide of *CDHR1* exon 8, c.783G>A (NG_028034.1). In 2017, Stingl et al. reported that a patient was homozygous for the same variant. A midigene splice assay encompassing exons 7, 8, and 9 demonstrated that this variant causes in-frame skipping of exon 8 (p.Asp214_Pro261del)[26]. Likewise, a similar defect was identified in messenger RNA (mRNA) from the retina of an individual heterozygous for this variant. The authors suggest that this variant is associated with a relatively mild form of IRD, based on clinical examination of a patient homozygous for this variant[26]. Similarly, our patient displays a mild phenotype and late age-at-onset as previously described[26].

In Pt-17, a pathogenic missense variant (c.1843G>A) was identified in *HGSNAT* (NG_009552.1) as the first allele. Through WGS analysis, a NCSS variant, c.493+5G>A, was identified in the same gene, resulting in exon 4 skipping (p.Arg124Serfs*25) as evaluated using a midigene splice assay spanning exons 3, 4, and 5 (Fig. 2a). The variant was classified as severe due to the absence of remaining wild-type mRNA when testing the mutant construct.

Individual Pt-24 with autosomal recessive RP carried a heterozygous variant, c.4758+3A>G in *USH2A* (NG_009497.2) together with a deep-intronic variant, c.784+14389G>T, in the same gene. The NCSS variant was found to be pathogenic after performing a midigene splice assay showing exon elongation (p.Gln1586_Gly1587ins5*) (Fig. 2b). Two fragments were observed upon transfection of the mutant midigene in human embryonic kidney 293 T (HEK293T) cells. After quantification, we observed 78.9% of a larger fragment (332 nt) corresponding to 107-bp elongation of exon 22 and 21.1% of the wild-type fragment (225 nt) enabling classification of this variant as severe (Supplementary Table 5).

Splice defects due to deep-intronic variants in *ABCA4*, *EYS*, *GRM6*, and *USH2A*

In the aforementioned Pt-24, after confirming the pathogenicity of the c.4758+3A>G NCSS variant in *USH2A*, an extensive analysis of the intronic regions of this gene was performed and a deep-intronic variant, c.784+14389G>T, was detected in *trans* in *USH2A*

Table 2. Overview of pathogenic variants in genetically solved cases.

ID	Ethnicity	Phenotype	Gene	DNA allele 1	Protein allele 1	DNA allele 2	Protein allele 2	Previous genetic testing method	Segregation analysis, # of family tested members
Pt-1	Irish	STGD1	*ABCA4*	c.6079C>T[b]	p.(Leu2027Phe)	c.5196+1137G>A	p.[=,Met1733Glufs*78]	TCS	No
Pt-2	Irish	STGD1	*ABCA4*	c.6743T>C[b]	p.(Phe2248Ser)	c.4539+2028C>T	p.[=,Arg1514Leufs*36]	TCS	No
Pt-3	Irish	STGD1	*ABCA4*	c.5603A>T[b]	p.(Asn1868Ile)	c.286A>C	p.(Asn96His)	TCS	No
Pt-4	Dutch	CRD[a]	*ATXN7*	c.88_117CAG[70]	p.(Gln30[70])	–	–	WES	Yes, 2
Pt-5	Irish	RP	*C21ORF2*	c.218G>C	p.(Arg73Pro)	c.76T>C	p.(Trp26Arg)	TCS	No
Pt-6	Irish	RP	*C21ORF2*	c.218G>C	p.(Arg73Pro)	c.218G>C	p.(Arg73Pro)	TCS	No
Pt-7	Irish	RP	*CDHR1*	c.783G>A	p.Asp214_Pro261del	c.783G>A	p.Asp214_Pro261del	TCS	No
Pt-8	Israeli	AS	*CEP78*	c.1033G>T	p.(Gly345*)	c.1033G>T	p.(Gly345*)	TCS	No
Pt-9	Dutch	RP	*EYS*	c.[3906C>A;9468T>A][b]	p.[Tyr3156*;His1302Gln]	c.5644+70912A>G	p.Asp1882Glyfs*18	WES	Yes, 1
Pt-10	Israeli	RP	*EYS*	c.403_423delinsCTTTT[b]	p.(Thr135Argfs*26)	c.539C>A	p.(Ser180*)	TCS	Yes, 1
Pt-11	Israeli	CRD	*EYS*	c.9468T>A	p.(Tyr3156*)	c.9468T>A	p.(Tyr3156*)	WES	No
Pt-12[c]	Dutch	RCD	*EYS*	c.1660T>A[b]	p.(Cys554Ser)	c.5044G>T	p.(Asp1682Tyr)	WES	No
Pt-13	Israeli	RP	*FAM161A*	c.1753_1756del	p.(Lys585Alafs*6)	c.1753_1756del	p.(Lys585Alafs*6)	TCS	No
Pt-14	Dutch	Oguchi disease	*GRM6*	c.1732C>T[b]	p.(Arg578Cys)	c.1355-587dup	p.[Asn451_Ala452ins25, Gly385Alafs*42,=]	WES	No
Pt-15	Dutch	RP	*HGSNAT*	c.1843G>A	p.Ala615Thr	c.1843G>A	p.Ala615Thr	WES	No
Pt-16	Dutch	RP	*HGSNAT*	c.1030C>T	p.(Arg344Cys)	c.1843G>A	p.Ala615Thr	WES	No
Pt-17	Dutch	RP	*HGSNAT*	c.1843G>A[b]	p.Ala615Thr	c.493+5G>A	p.Arg124Serfs*25	WES	No
Pt-18	Dutch	RP	*HGSNAT*	c.1843G>A	p.Ala615Thr	c.1843G>A	p.Ala615Thr	WES	No
Pt-19	Irish	RP	*HGSNAT*	c.1622C>T	p.(Ser541Leu)	c.1843G>A	p.Ala615Thr	TCS	Yes, 3
Pt-20	Irish	RP	*PCARE*	c.3604C>T[b]	p.(Arg1202*)	c.3099_3100insCAGG	p.(Val1034Glnfs*74)	TCS	Yes, 5
Pt-21	Israeli	CRD	*RGS9BP*	c.583T>G	p.(Ser195Ala)	c.583T>G	p.(Ser195Ala)	WES	No
Pt-22	Dutch	LCA	*RPGRIP1*	c.3238+1118_3339+323del	p.(Asp1080Glyfs*6)	c.3238+1118_3339+323del	p.Asp1080Glyfs*6	WES	Yes, 2
Pt-23	Irish	RP	*PRPF31*	c.-396_*287(0) (g.54106454-54133135del)	p.(0)	–	–	TCS	Yes, 4
Pt-2	Israeli	RP	*USH2A*	c.4758+3A>G	p.Gln1586_Gly1587ins*5	c.784+14389G>T	p.Gly262Aspfs*26	TCS	No

ABCA4: NG_009073.1; *ATXN7*: NG_008227.1; *C21ORF2*: NG_032952.1; *CDHR1*: NG_028034.1; *CEP78*: NG_053171.1; *EYS*: NG_023443.2; *FAM161A*: NG_028125.1; *GRM6*: NG_008105.1; *HGSNAT*: NG_009552.1; *PCARE*: NG_021427.1; *RGS9BP*: NG_016751.1; *RPGRIP1*: NG_008933.1; *PRPF31*: NG_009759.1; *USH2A*: NG_009497.2.
AS Alström syndrome, *CD* cone dystrophy, *CRD* cone–rod dystrophy, *LCA* Leber congenital amaurosis, *RCD* rod–cone dystrophy, *RP* retinitis pigmentosa, *TCS* target-capture sequencing, *WES* whole-exome sequencing.
[a]With neurodegeneration.
[b]A first pathogenic allele was identified in previous genetic testing.
[c]This patient is assigned as likely solved.

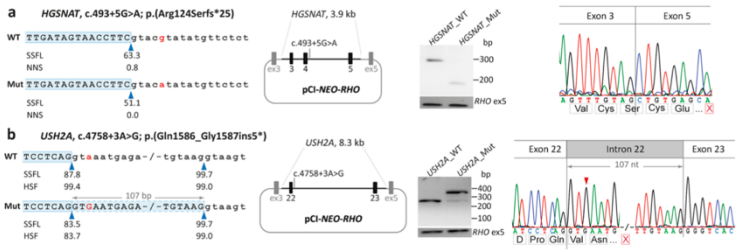


Fig. 2 Splice defects due to noncanonical splice site variants. For *HGSNAT* c.493+5G > A (**a**) and *USH2A*, c.4758+3A>G (**b**), the left panels are schematic representations of the gene segments of interest containing the corresponding variants in the wild-type (WT) and mutant (Mut) sequences. The SpliceSiteFinder-like (SSFL, range 0–100), and Human Splicing Finder (HSF, range 0–100) or NNSPLICE (NNS, range 0–1) scores for the splice sites are indicated below each sequence. The blue triangles indicate the position of splice donor sites in both wild-type and mutant sequences. The red highlighted nucleotides indicate the variants in the mutant sequence. The second panels show the mutant pCI-*NEO-RHO* vectors. They contain the variants and flanking exons 3, 4, and 5 in *HGSNAT* and exons 22 and 23 in *USH2A* flanked by *RHO* exons 3 and 5. Wild-type and mutant constructs were used to transfect HEK293T cells. The next panels show the gel images of RT-PCR products in wild-type and mutant midigenes. RT-PCR analysis of *RHO* exon 5 was performed as a control for transfection efficiencies. In the last panels, the Sanger sequence analysis of the RT-PCR fragments confirmed the exon 4 skipping in *HGSNAT*, c.493+5G>A (**a**) and 107-nt exon elongation in *USH2A* c.4758+3A>G (**b**). The red "X" in a red box indicates the stop codon in each schematic chromatogram. For each variant, both WT and mutant variants were examined in parallel in the same experiment and processed in parallel.

that led to a pseudoexon insertion (p.Gly262Aspfs*26) (Fig. 3c). Using the mutant construct for midigene analysis, 77.1% of the mutant fragment (371 nt) corresponding to the presence of a pseudoexon in mature mRNA was observed and 22.9% of the wild-type fragment (274 nt). Therefore, this allele was classified as severe (Supplemental File 1: Table 5). This patient also carried a frameshift variant in *PCARE* (p.(Lys919Thrfs*2)), which was detected previously by TCS. However, WGS analysis did not reveal any plausible second candidate variant in intronic regions of the *PCARE* gene (Supplemental File 1: Table 1).

Two STGD1 cases (Pt-1 and Pt-2) carried c.6079C>T and c.674T>C missense variants, individually, as pathogenic alleles (Supplementary Table 1). Upon WGS analysis, we identified previously published pathogenic deep-intronic *ABCA4* variants in these probands, c.4539+2028C>T and c.5196+1137G>A, respectively[27]. The c.4539+2028C>T variant did not show a splice defect in fibroblast cells, but revealed a retina-specific 345-nt pseudoexon insertion in a proportion of the mRNA (p.[=,Arg1514Leufs*36]) extracted from patient-derived photoreceptor precursor cells[4]. The variant c.5196+1137G>A showed a 73-nt pseudoexon insertion in a small fraction of mRNA in a HEK293T cell–midigene-based splice assay. This effect was stronger in patient-derived photoreceptor precursor cells (p.[=,Met1733Glufs*78])[28]. Based on genotype–phenotype correlations, variants c.4539+2028C>T and c.5196+1137G>A were deemed moderately severe[27,28].

EYS c.5644+70912A>G (NG_023443.2) was identified in Pt-9, who carried c.[3906C>A;9468T>A] (p.[His1302Gln;Tyr3156*], NM_001292009.1) in *EYS* as the first pathogenic allele for which the truncating variant is mostly reported as c.9405T>A (p.(Tyr3135*), NM_001142800.1) (Supplementary Table 1). Due to the large size of intron 26 (150.8 kb), c.5644+70912A>G was tested in a minigene with a small genomic DNA insert (0.97 kb; Fig. 3a). Reverse transcription-polymerase chain reaction (RT-PCR) analysis of the mRNA resulting from the wild-type *EYS* minigene revealed the expected wild-type fragment of 274 nt (76.0% of the total product) and a larger fragment of 378 nt (24.0% of the total product) corresponding to the presence of a 104-nt pseudoexon, which leads to frameshift variant after 18 amino acids (p.Asp1882Glyfs*18). mRNA products from the mutant *EYS* c.5644+70912A>G minigene only contained the larger fragment (378 nt), with no remaining wild-type mRNA, and therefore the allele was classified as severe (Fig. 3a and Supplementary Table 5).

In Pt-14, who was monoallelic for *GRM6*, c.1732C>T (NG_008105.1; Supplementary Table 1), *GRM6* c.1355-587dup was identified, which resulted in the insertion of an in-frame pseudoexon encoding 25 amino acids (p.Asn451_Ala452ins25) (Fig. 3b). Surprisingly, a 97-nt partial deletion of exon 6 (359 nt) in both wild-type and mutant midigene assays was observed, which was calculated to be 54.4% and 37.0% of total RNA molecules, respectively. In the *GRM6* mutant midigene assay, the 567-nt fragment corresponding to the presence of a pseudoexon was quantified as 21.0% of total mRNA molecules. The percentage of full-length wild-type fragments (492 nt) for wild-type and mutant *GRM6* midigene assays were 45.6% and 42.0%, respectively. Therefore, we normalized the percentage of wild-type fragment in the assays using the mutant construct (42.0%) to the wild-type construct (46.0%). *GRM6* c.1355-587dup was classified as a mild allele due to the remaining 91.0% of the wild-type fragment (42/46 × 100) (Supplementary Table 5).

Finally, five noncoding variants in seven cases did not show any splice defects after midi- or midigene in vitro splice assays and we therefore classify these as likely benign (Supplementary Fig. 2A–D): *ABCA4*, c.6148-89G>A in Pt-27; *ABCA4*, c.5460+1315_5460+1317delinsTA in three cases (Pt-28–30); *CYP4V2*, c.214+879_214+882delinsG in Pt-52; *PDE6B* c. 469-776C>G in Pt-61; *RLBP1* c.525+425_525+433delinsATA in Pt-65. These patients were included in this study as monoallelic cases having the first pathogenic allele in *ABCA4*, *CYP4V2*, *PDE6B*, and *RLBP1*, respectively (Supplementary Table 1). However, all seven probands remained genetically unresolved given that a potential pathogenic second allele was not identified.

DISCUSSION

The sophistication of sequencing methods employed for the detection of disease-causing variants in the human genome has increased rapidly over the past decade, from targeted approaches such as panel sequencing to more comprehensive diagnostic methods such as WES and WGS. TCS and WES have undoubtedly played a valuable role in the genetic diagnosis of IRDs. For genetically undiagnosed cases following TCS and WES, it has become clear that more comprehensive sequencing procedures

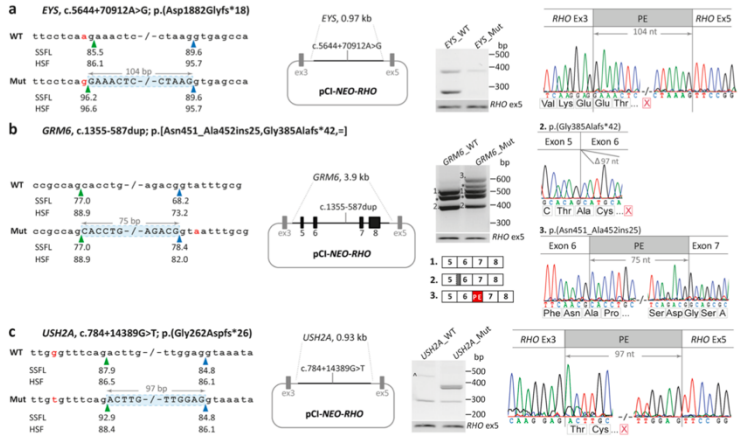


Fig. 3 Splice defects due to deep-intronic variants. For each variant, i.e., **a** *EYS*, c.5644+70912A>G, **b** *GRM6*, c.1355-587dup, and **c** *USH2A*, c.784+14389G>T, the left panels are schematic representations of the gene segments of interest containing the corresponding variants in the wild-type (WT) and mutant (Mut) sequences. The SpliceSiteFinder-like (SSFL, range 0–100), and Human Splicing Finder (HSF, range 0–100) scores for the splice sites are indicated below each sequence. The green and blue triangles indicate the position of splice acceptor sites and splice donor sites in both wild-type and mutant sequences. The red highlighted nucleotides indicate the variants in the mutant sequence. The second panels show the mutant pCI-*NEO-RHO* vectors. They contain the variants and flanking exons (except for *EYS*, as c.5644+70912A>G is in a minigene construct) flanked by *RHO* exons 3 and 5. Wild-type and mutant constructs were used to transfect HEK293T cells. The next panels show the gel images of RT-PCR products in wild-type and mutant midi- or minigenes. RT-PCR analysis of *RHO* exon 5 was performed as a control for transfection efficiencies. For the *GRM6* c.1355-587dup variant, the asterisks in the gel image indicate the heteroduplex fragments. For the *USH2A* c.784+14389G>T variant, the "^" in the wild-type construct indicates an artifact product. In the last panels, the Sanger sequence analysis of the RT-PCR fragments confirmed the predicted splice defects for each variant. For *GRM6* c.1355-587dup, the sequence of fragment 2 in both midigenes showed the 97-bp deletion of exon 6, which led to the frameshift variant. The third fragment was only observed in the mutant construct and corresponded to the insertion of an in-frame pseudoexon in intron 6. The red "X" in a red box indicates the stop codon in each schematic chromatogram. For each variant, both WT and mutant variants were examined in parallel in the same experiment and processed in parallel.

are required. Importantly, WGS facilitates the identification of noncoding variants, now widely considered a prominent cause of a wide spectrum of IRDs[18,19,29–33]. In order to investigate these possible advantages, we performed WGS for 100 probands with clinical indications of IRD who had received inconclusive results from TCS and WES previously.

This study cohort consisted of 56 monoallelic individuals who carried one candidate pathogenic variant and 44 individuals with no candidate pathogenic variants previously identified in an IRD-associated gene via WES or TCS (Supplementary Table 1). Following extensive WGS data analysis and employment of in vitro splice assays, we identified the genetic cause of disease in 24 cases. These findings demonstrate that WGS is a more comprehensive sequencing method than TCS and WES for IRDs. As previously described in the literature, we have also observed a higher detection rate of pathogenic variants in WGS compared to WES and TCS in the coding and NCSS regions[2,20]. For instance, in seven of 24 solved cases (29.2%), one or both pathogenic alleles in the coding regions of *EYS* and *HGSNAT* were not detected or remained unrecognized as pathogenic in the previous genetic testing methods. This is due to incomplete gene panels (previous absence of *HGSNAT*), flaws in the annotation pipeline, or a lack of evidence or knowledge regarding the implication of these variants as IRD-associated at the time of WES or TCS data analysis. In one case (Pt-12), we have identified a missense variant in *EYS*, c.5044G>T, which was assigned previously as "variant with unknown significance"[34]. However, due the strong in silico predictions for this missense variant and the highly conserved amino acid and nucleotide, we categorized this patient as likely solved. We have identified the hypomorphic allele p.Ala615Thr in *HGSNAT* in two homozygous (Pt-15 and Pt-18) and three heterozygous (Pt-16, Pt-17, and Pt-19) patients. After extensive analysis of IRD- and ciliopathies-associated genes as well as the complete sequence of *HGSNAT*, no other candidate variant was identified in these probands. Recently, Schiff et al. investigated the enzymatic activity of the HGSNAT protein and showed the decreased enzyme activity and urinary GAG/creatinine ratio in patients homozygous for p.Ala615Thr. They suggested that this allele is associated with non-syndromic retinal disease with the influence of transacting genetic and/or environmental modifiers on the retina[35]. Currently, our cases have non-syndromic RP; however, we suggest monitoring as a protracted syndromic phenotype may arise.

In addition, WGS facilitated the discovery of pathogenic noncoding variants that were not previously captured by TCS and WES. Two NCSS variants, i.e. *HGSNAT*, c.493+5G>A and *USH2A*, c.4758+3A>G, were detected using WGS. These noncoding variants were overlooked in earlier testing methods due to the lack of evidence for pathogenicity at the time of sequencing. After in vitro splice assays of these variants, they were both classified as severe due to the presence of <25% of the wild-type fragment in the total mRNA[18] (Fig. 2 and Supplementary Table 5).

In Pt-7, we identified *CDHR1* c.783G>A in a homozygous state, which previously was determined to lead to in-frame skipping of

exon 8 (p.Asp214_Pro261del)[26]. No other potentially causative variants were identified as part of our analysis. The allele frequency of this variant (total allele frequency of 0.003052 and allele frequency of 0.004903 in non-Finnish Europeans; Supplementary Table 6) indicates that this variant may not be fully penetrant or acts as a hypomorphic allele. The splice defect is in-frame and may confer a subtle change in protein function. Additionally, Stingl et al. showed reduced expression of the mutant allele by comparison with the wild type[26]. In line with this, Pt-7 displays a mild phenotype consistent with previously described *CDHR1*-related disease and late age at disease onset[26]. It is also possible that other genetic factors affect the penetrance of this variant, potentially explaining the relatively high allele frequency[35].

Importantly, WGS identified pathogenic coding variants in IRD-associated genes that did not reside in the TCS panels. Genes such as *FAM161A* (NG_028125.1), *RGS9BP* (NG_016751.1), *C21ORF2* (NG_032952.1), and *CDHR1* had not been included in the TCS panels at the time of sequencing (Table 2). These findings prompt redesign of TCS panels to include these genes where WES and WGS remain too costly to apply on a wider scale.

In seven out of 24 (29.2%) resolved cases, we identified five incidents of deep-intronic variants that could not be detected with previous sequencing techniques. The tested deep-intronic variants in this study led to frameshift variants, except for *GRM6*, c.1355-587dup, and were classified as severe alleles with <25% of remaining wild-type mRNA[18]. We observed the insertion of a 75-nt pseudoexon in *GRM6* midigene assays using the mutant construct (21%), which encodes 25 amino acids. In addition, a partial deletion of exon 6 in wild-type and mutant midigene assays was identified, which was more abundant in products from the wild-type midigene compared to those from the mutant midigene (54% compared to 37% in the mutant) (Fig. 3b). Of note, the partial exon 6 deletion in *GRM6* was not reported previously, and further research needs to be undertaken to investigate this phenomenon in a retina-specific cell model. The deletion of part of exon 30 in *ABCA4* has previously been reported in normal retinal mRNA[4,36]. In addition, we classified *GRM6* c.1355-587dup as a mild allele in this study, although a stronger effect on splicing in retina-specific cells cannot be discarded as a similar effect was previously observed for the *ABCA4* c.4539+2028C>T and c.5196+1137G>A variants. These findings highlight the importance of functional analyses in proving intronic variant pathogenicity, the contribution of such variants to IRD pathogenesis, and the diagnostic imperative imposed by them.

No splice defects were observed as a result of *ABCA4*, c.6148-89G>A (Pt-27); *ABCA4*, c.5460+1315_5460+1317delinsTA (Pt-28-30); *CYP4V2*, c.214+879_214+882delinsG (Pt-52); *PDE6B* c. 469-776C>G (Pt-61); and *RLBP1* c.525+425_525+433delinsATA (Pt-65). For the *ABCA4*, *CYP4V2*, and *RLBP1* indel variants, only Alamut prediction scores were used for their inclusion in the in vitro functional analysis as the SpliceAI algorithm cannot provide prediction scores for indel variants. All three variants create a cryptic donor site and have not been reported previously in the gnomAD database (Supplementary Fig. 2A, C, and E). The *ABCA4*, c.6148-89G>A variant strengthens the cryptic SDS in intron 44, from 67.6 to 79.1 in SpliceSiteFinder-like (range 0–100), and also activates an exonic splice enhancer (SF2/ASF) motif in the presence of c.6148-89G>A (Supplementary Fig. 2B). Similarly, the SpliceAI algorithm predicts a gain of a nearby SDS with a delta score of 0.12 (Supplementary Table 4). Surprisingly, we did not observe a splice defect in the presence of variant c.6148-89G>A in *ABCA4*. Possibly, the functional assay in HEK293T cells is not a representative model for this variant due to differences in the presence of splice enhancer and silencer proteins in kidney and retina-specific cells. These five variants and in particular the *ABCA4*, c.6148-89G>A variant, may still be proven pathogenic when investigated in retina-like cells. Such findings have been reported previously, e.g., splice defects resulting from *ABCA4* c.4539+2001G>A and c.4539+2028C>T were only detected in photoreceptor precursor cells derived from patient fibroblasts[4]. Therefore, for variants that do not display a splice defect, exploring variant pathogenicity in models that better recapitulate the retinal cell identity should be considered. Likewise, we have established that a minority of wild-type mRNA remains for a number of variant alleles, while analysis of variants in photoreceptor precursor cells may determine an alternative allele severity compared to the result from in vitro splice assays in HEK293T cells. Furthermore, Pt-52 was included as a suspected Bietti crystalline corneoretinal dystrophy case, which has a strong genotype–phenotype association to pathogenic variants in *CYP4V2*. This may support our hypothesis that providing pathogenic evidence for deep-intronic variants possibly failed due to the disadvantages of the HEK293T midigene system. Alternatively, the disadvantage lies within the flaws of short-read sequencing to detect the actual pathogenic variants or the initial pathogenic allele may represent a chance-finding.

WGS is the optimal sequencing method for the detection of SVs by comparison with WES and TCS, as indicated by previous studies[2,19,37]. In line with this, we have detected a homozygous 3.5 kb SV in *RPGRIP1* in an individual with LCA (Pt-22), which remained undetected by WES. The SV led to exon 20 deletion in the mRNA, altering the open-reading frame and resulting in a stop codon after six amino acids (p.Asp1080Glyfs*6) (Supplementary Fig. 1D). The assessment of WES was set to only detect deletions when spanning 2 exons or more and therefore could not detect this SV. Previously, heterozygous pathogenic variants in *CEP290*, *PDE6A*, and *RPE65* were reported as the potential first allele, with all three genes strongly associated with the given phenotype, which may have distracted from assessing other clinically relevant genes such as *RPGRIP1* (Supplementary Table 1). However, a comprehensive analysis of the entirety of these genes did not identify additional pathogenic variants. This finding emphasizes the importance of analysis of all IRD-associated genes to identify the true genetic cause of disease.

In addition, a 26.86 kb heterozygous deletion on chromosome 19 was detected in proband Pt-23 (Supplementary Fig. 2B). This deletion encompasses the entire *PRPF31* and *TFPT* genes, and the promoter of *NDUFA3*. This large deletion was not detected using TCS as the breakpoints of the variant lay within noncoding regions not covered by the TCS panel. In general, it is difficult to detect read depth reduction in heterozygous SVs when there is no read coverage in the surrounding region for comparison. In addition, this specific individual was thought to have an autosomal recessive disease given the lack of family history. Segregation analysis confirmed maternal inheritance of the SV. However, the proband's mother was phenotypically unaffected, due to incomplete penetrance associated with variants in *PRPF31*, noted in other studies[38–40]. *TFPT* is a molecular partner of *TCF3* and is associated with childhood acute lymphoblastic leukemia (OMIM: 613065)[41]. *NDUFA3* is required for the formation of the extramembrane arm of human mitochondrial complex 1[42]. No additional disease features could be attributed to the disruption of these genes at this time. In addition, there are no reported pathogenic variants in either *TFPT* or *NDUFA3* in the ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/). A similar deletion encompassing *PRPF31*, *TFPT*, and *NDUFA3* on chromosome 19 has been reported previously[43].

In this study, 76 (76%) cases remain genetically unresolved following WGS analysis, indicating the presence of pathogenic variants that eluded our data analysis approach. The pipeline employed in this study was designed to be stringent, comprehensive, and time effective, given the immense volume of data created by WGS. Unresolved cases may harbor pathogenic variants that were called, but not individually interpreted because they did not pass the prioritization criteria employed. Our data

analysis approach relied on the presence of one coding or NCSS pathogenic variant before interrogation of intronic regions of the same gene. Therefore, it is possible that some unresolved probands carry homozygous or compound heterozygous pathogenic intronic variants. Alternatively, some may carry pathogenic variants in genes not yet associated with IRD and therefore were not included in the WGS data analysis.

Furthermore, it is likely that some individuals carry pathogenic variants located within regulatory regions of IRD genes, as reported in other studies[44,45]. However, the identification and interpretation of such variants remains complex. These variants are possibly not prioritized by our current strategy as they likely have their own interpretation criteria that are yet to be established. In addition, it is important to note that the line between Mendelian and complex disease may be blurred by modifier loci and environmental influences[46]. Thus, we should consider that some of the remaining unresolved IRD cases in this study may have a more complex etiology than previously thought as shown by the presence of hypomorphic alleles in cases of our cohort.

Lastly, it is possible that some variants may not have been called by variant calling software as they may be located in repetitive or low complexity regions with poor coverage, e.g., *RPGR* open-reading frame 15 (*ORF15*). Pathogenic variants in *RPGR ORF15* are responsible for 50–90% of X-linked RP (MIM: 312610)[47]. While in many instances male cases are analyzed for pathogenic variants in *RPGR ORF15* prior to WES, in our WGS cohort, 35/76 (46%) of unresolved cases are male individuals with an RP phenotype and it is possible that a proportion carries pathogenic variants in *RPGR ORF15*. As data analysis tools for WGS and knowledge surrounding variant interpretation improve, it is anticipated that further cases will receive a genetic diagnosis in the future.

We performed WGS on a cohort of 100 IRD probands that remained unresolved following the employment of TCS and WES as preliminary sequencing measures. A genetic cause of the disease was determined in 24 (24%) cases through detection of variants in genes that were not included in TCS panels or by identification and functional analyses of pathogenic intronic variants as well as detection of both small and large SVs not previously identified or recognized as causative. While a 24% yield may be modest, the number of "obvious" genetic causes detected in the coding regions of established IRD-associated genes is surprising, especially, considering the extensively previously examined cohort we assessed. Reasons for this may include an incomplete gene panel, poor read quality in previous testing methods, or lack of evidence surrounding the implication of some IRD genes at the time of initial sequencing. While the coverage and quality of reads in WGS are higher than prescreening techniques enabling better variant detection in this cohort, our findings emphasize the necessity of improvements in variant interpretation and bioinformatics pipelines as well as performing in vitro functional assays to obtain a higher diagnostic yield not only by WGS but also by less comprehensive sequencing techniques such as WES and TCS.

The findings of this study demonstrate that WGS can be an optimal sequencing approach for individuals diagnosed with an IRD. In future studies, a combined approach of WGS and optical genome mapping or long-read sequencing can be beneficial to overcome the difficulties of SV detection in short-read sequencing and provide a higher genetic diagnostic yield. It is a diagnostic imperative to continue to analyze individuals who have received an inconclusive genetic testing result following less comprehensive sequencing methods. Identification of the underlying genetic pathogenesis of a disease is becoming increasingly important in the advent of gene-based therapeutics, where a known genetic cause is essential to access clinical trials and approved treatments. With decreasing costs, increasing data interpretability, and continued functional analysis, we foresee that WGS may become the sequencing methodology of choice to provide IRD patients with an accurate genetic diagnosis.

METHODS

Ethical considerations

The study adhered to the tenets of the Declaration of Helsinki and was approved by the local ethics committees of the Radboud University Medical Center (Nijmegen, The Netherlands), the Rotterdam Eye Hospital (Rotterdam, The Netherlands) (MEC-2010-359; OZR protocol no. 2009-32), The Department of Ophthalmology, The Royal Victoria Eye and Ear Hospital (Dublin, Ireland) (13-06-2011: HRA-POR201097), Ramabam Health Care Campus (Haifa, Israel), and HaEmek Medical Center (Afula, Israel). Written informed consent was obtained from patients prior to DNA analysis and inclusion in this study.

Patient cohort

100 IRD probands were subjected to WGS, of which 43 were recruited from the Radboud University Medical Center, Nijmegen and The Rotterdam Eye Hospital (both in The Netherlands), 19 from the Technion-Israel Institute of Technology (Haifa, Israel), and 38 from the Research Foundation of the Royal Victoria Eye and Ear Hospital (Dublin, Ireland). A suspected recessive inheritance pattern and unresolved status following prior preliminary genetic testing using TCS or WES were prerequisites for participation in this study.

DNA acquisition and preliminary genetic analysis

Participants provided either a blood or saliva sample for analysis. DNA from Irish participants was isolated using the Qiagen DNA Blood Maxi Kit (Hilden, Germany) or the DNA Genotek Oragene-DNA Kit (Ontario, Canada). In the remaining individuals, the genomic DNA was extracted according to the standard protocol[48]. TCS and subsequent data analysis of single-nucleotide variants (SNVs), SVs, and CNVs were carried out as described previously in the Irish cohort[6,9] and eight Israeli cases[49,50] (Supplementary Note 1). All Dutch (43 cases) and remaining Israeli (11 cases) participants had undergone WES (Supplementary Note 1) and the subsequent data were analyzed based on the identification of known pathogenic variants as well as likely pathogenic variants or variants of unknown significance, i.e., assessment of all type of variant including SNVs, SVs, and CNVs. Variant prioritization in WES data was based on a strong correlation between genotype and phenotype in individuals, i.e., a first pathogenic allele was identified in a gene previously found to be mutated in the IRD cases with a similar phenotype. Thirty-three out of 43 Dutch, 14 out of 38 Irish, and nine out of 19 Israeli cases were monoallelic, carrying one reported pathogenic or candidate allele in IRD-associated genes prior to WGS.

WGS was performed by BGI (Hongkong, China) on a BGISeq500 using either 2 × 100 or 2 × 150 bp paired-end reads, with a 30-fold minimal median coverage per genome. Burrows–Wheeler Aligner was utilized to map the data to the human genome (hg19). The quality of the WGS data was based on the insert size, percentage mapped reads, percentage duplicated mapped reads, coverage, bases with >20× coverage and error rate, which were evaluated using Qualimap V.2.2.1[51]. Variant calling was performed by xAtlas V.0.1[52] and variants were annotated using the Variant Effect Predictor (VEP V.91) and Gencode V.34lift37 basic gene annotations. It consists of various features such as chromosomal and nucleotide position, gene name, and its component (e.g., intron, exon, splice site, 5′-untranslated region (5′-UTR), 3′-UTR, intragenic). In addition, we used an in-house developed pipeline that provides further information including different population frequency databases (e.g., gnomAD[22] (https://gnomad.broadinstitute.org), GoNL[53], Wellderly[54], 1000genomes[55], in-house variant frequencies database containing WES data of 15,576 individuals), various in silico prediction scores (e.g., Grantham, PhyloP, CADD_PHRED, SpliceAI)[56–59], predicted protein effect, gene and disease OMIM description, and gene regulation and expression data. Runs of homozygosity were detected using plink V.1.07[60] with the following parameters: homozyg-window-het=3, homozyg-snp=50 and homozyg-kb=300.

SVs were detected based on paired-end and split-read evidence using Manta Structural Variant Caller V.1.1.0 (Illumina) with default parameters[61]. CNVs were called using Control-FREEC, which detects copy number changes and allelic imbalances based on the read depth[62]. SVs and CNVs were annotated using an in-house developed pipeline[63], i.e., annotations

for chromosomal location and position, gene(s) and their component (e.g., intronic, exonic), type of SV (gain/loss), percentage overlap and frequency of various population frequency databases (e.g., gnomAD-SV[22], GoNL[53], Decipher[64], Wellderly[54], 1000genomes[55]), and disease OMIM description. Short tandem repeats (STRs) for 33 known pathogenic sites were detected using Expansion Hunter V.3.1.2. using default settings[65].

Variant prioritization approach

Sequencing data from a customized panel of 446 IRD and ocular defect-associated genes and 163 ciliopathy associated genes were extracted from the WGS data obtained from each patient. To facilitate the identification of pathogenic variants, a two-step protocol was designed for variant prioritization, in which the first step was the utilization of an automated prioritization pipeline composed in R studio[66]. The details of this pipeline can be found in Supplementary Note 2. The second step was a manual review of the remaining variants identified as follows below.

The manual variant prioritization approach differed depending on the findings of previous genetic studies. In cases where no first candidate variant was identified during previous genetic testing efforts, CNVs and SVs were evaluated first followed by assessment of SNVs. Subsequently, SNVs were manually prioritized according to their predicted pathogenic effect (i.e., nonsense, frameshift, canonical splice site variants, NCSS variants, in-frame deletions/insertions, missense variants, and synonymous variants). Coding or noncoding SNVs with a minor allele frequency of >1% in the gnomAD database were not considered causative[22]. Once a candidate pathogenic variant was identified, the predicted effects of rare intronic variants (with allele frequency <1%) in the gene of interest were interrogated in silico.

For cases where a first candidate variant in an IRD or ciliopathy associated gene was previously identified, the assessment of CNVs, SVs, and subsequently all rare intronic SNVs in the gene of interest was performed to identify a second causal variant. If a second causal variant remained elusive, an analysis of all CNVs, SVs, and SNVs was implemented in the same manner as for cases in which no first candidate variant was previously established. In addition, all IRD cases were analyzed for putative pathogenic STR expansions in IRD-associated genes.

Variant frequency and pathogenicity prediction parameters

All CNVs and SVs were assessed based on their potential effects to disrupt the reading frame or regulatory regions in 5′- and 3′-UTRs. SVs were compared to previously identified SVs reported in gnomAD-SV and the Database of Genomic Variants. In the presence of pathogenic SVs, the breakpoint regions were manually assessed for the presence of microhomology or repetitive elements.

For the coding SNVs, the pathogenicity of missense variants was measured using threshold scores of in silico prediction tools; PhyloP (range 14.1–6.4; predicted pathogenic ≥2.7)[56], CADD-PHRED (range 1–99; predicted pathogenic ≥15)[57], and Grantham (range 0–215; predicted pathogenic ≥80)[58]. Missense variants that only passed one of the thresholds were not considered causative.

Noncoding variants were selected for in vitro splice assay analysis based on criteria that were defined previously[18]. In short, the predicted effect of noncoding variants on splicing was evaluated using the algorithms SpliceSiteFinder-like, MaxEntScan, GeneSplicer, and Human Splicing Finder embedded in the Alamut Visual software version 2.10 (Interactive Biosoftware, Rouen, France; http://www.interactive-biosoftware.com)[67–71]. In addition, the SpliceAI algorithm[59] was utilized to select candidate noncoding variants. Other than the default for SpliceAI, we selected variants with at least one delta score above the threshold of 0.02 for SDS or SAS gain or loss. For each variant, 500 bp upstream and downstream were included as the input sequences for SpliceAI analysis.

In vitro splice assays

After completing the data analysis of 100 cases, 10 deep-intronic and 3 NCSS candidate variants in 14 individuals were identified, which adhered to the stringent criteria defined above for noncoding variants. Subsequently, a midigene or minigene splice assay was employed as described previously to assess potential splicing defects in the presence of variants[63,72]. In short, the regions of interest of a genomic DNA sample were amplified by primers that contain attB1 and attB2 tags at their 5′ end to facilitate the Gateway cloning. After generating the wild-type constructs, they served as templates to generate mutant constructs by PCR-based site-directed mutagenesis. The *ABCA4* variant c.6148-89G>A was introduced into a vector containing the wild-type fragment of *ABCA4* from exons 43 to 47, as described previously[15]. For *PDE6B* c. 469-776C>G in Pt-61 and *RLBP1* c.525 +425_525+433delinsATA in Pt-65, mutant and wild-type constructs were both generated from the patient's DNA as the mutagenesis PCR technique failed due to the complexity of the regions surrounding these variants. Subsequently, the wild-type and mutant constructs were separately incorporated into the pCI-*NEO-RHO* or pcDNA3native/DEST Gateway-adapted vector to generate wild-type and mutant midi(mini)genes. The pcDNA3native/DEST vector was utilized when the region of interest contained the first exon of the gene. Generated wild-type and mutant midigenes and minigenes were independently transfected into HEK293T cells. After 48 h of incubation, mRNAs were isolated and utilized for transcript analysis by RT-PCR with primers in flanking exons for midigenes or primers in *RHO* exon 3 and 5 for minigenes. All primers designed for splice assays are listed in Supplementary Table 7. Electrophoresis gel images are derived from raw images that are provided in Supplementary Fig. 3. The Fiji software was utilized for quantification analysis in the presence of multiple mRNA fragments after gel electrophoresis. The variants with <25% of the wild-type fragments present in mutant constructs were classified as severe alleles. This variant classification refers only to the severity of the mRNA defect observed in HEK293T cells.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

All data relevant to the study are included in the article or uploaded as Online Supplementary data. The pathogenic variant data are submitted to Leiden Open Variation Database (LOVD). The whole-genome sequencing data are not publicly available as these could compromise research participant privacy. Whole-genome sequencing data may become available upon a data transfer agreement approved by local (Irish, Israeli and/or Dutch) ethical committees. Patient sample identifiers from this study can be released upon reasonable request from "Pt-1 to Pt-100" to the corresponding local "DNA-number." Specific variant requests or other data are available from the corresponding author (S.R.) upon reasonable request.

Received: 20 May 2021; Accepted: 21 October 2021;
Published online: 18 November 2021

REFERENCES

1. Berger, W., Kloeckener-Gruissem, B. & Neidhardt, J. The molecular basis of human retinal and vitreoretinal diseases. *Prog. Retin. Eye Res.* **29**, 335–375 (2010).
2. Ellingford, J. M. et al. Whole genome sequencing increases molecular diagnostic yield compared with current diagnostic testing for inherited retinal disease. *Ophthalmology* **123**, 1143–1150 (2016).
3. Ellingford, J. M. et al. Pinpointing clinical diagnosis through whole exome sequencing to direct patient care: a case of Senior-Loken syndrome. *Lancet* **385**, 1916 (2015).
4. Albert, S. et al. Identification and rescue of splice defects caused by two neighboring deep-intronic ABCA4 mutations underlying Stargardt disease. *Am. J. Hum. Genet.* **102**, 517–527 (2018).
5. Buermans, H. P. & den Dunnen, J. T. Next generation sequencing technology: advances and applications. *Biochim. Biophys. Acta* **1842**, 1932–1941 (2014).
6. Whelan, L. et al. Findings from a genotyping study of over 1000 people with inherited retinal disorders in Ireland. *Genes* **11**, https://doi.org/10.3390/genes11010105 (2020).
7. Consugar, M. B. et al. Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing. *Genet. Med.* **17**, 253–261 (2015).
8. Patel, A. et al. The Oculome Panel Test: next-generation sequencing to diagnose a diverse range of genetic developmental eye disorders. *Ophthalmology* **126**, 888–907 (2019).
9. Dockery, A. et al. Target 5000: target capture sequencing for inherited retinal degenerations. *Genes* **8**, https://doi.org/10.3390/genes8110304 (2017).
10. Lin, X. et al. Applications of targeted gene capture and next-generation sequencing technologies in studies of human deafness and other genetic disabilities. *Hear. Res.* **288**, 67–76 (2012).
11. Tucker, T., Marra, M. & Friedman, J. M. Massively parallel sequencing: the next big thing in genetic medicine. *Am. J. Hum. Genet.* **85**, 142–154 (2009).

12. Choi, M. et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl Acad. Sci. USA* **106**, 19096–19101 (2009).
13. Lewis, C. A. et al. Tubby-like protein 1 homozygous splice-site mutation causes early-onset severe retinal degeneration. *Invest. Ophthalmol. Vis. Sci.* **40**, 2106–2114 (1999).
14. Stojic, J., Stöhr, H. & Weber, B. H. Three novel ABCC5 splice variants in human retina and their role as regulators of ABCC5 gene expression. *BMC Mol. Biol.* **8**, 42 (2007).
15. Khan, M. et al. Resolving the dark matter of ABCA4 for 1054 Stargardt disease probands through integrated genomics and transcriptomics. *Genet. Med.* **22**, 1235–1246 (2020).
16. Haer-Wigman, L. et al. Diagnostic exome sequencing in 266 Dutch patients with visual impairment. *Eur. J. Hum. Genet.* **25**, 591–599 (2017).
17. Anna, A. & Monika, G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J. Appl. Genet.* **59**, 253–268 (2018).
18. Fadaie, Z. et al. Identification of splice defects due to noncanonical splice site or deep-intronic variants in ABCA4. *Hum. Mutat.* **40**, 2365–2376 (2019).
19. Carss, K. J. et al. Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. *Am. J. Hum. Genet.* **100**, 75–90 (2017).
20. Belkadi, A. et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc. Natl Acad. Sci. USA* **112**, 5473–5478 (2015).
21. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
22. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
23. Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
24. Wallis, Y. et al. *Practice Guidelines for the Evaluation of Pathogenicity and the Reporting of Sequence Variants in Clinical Molecular Genetics* 1–16 (Association for Clinical Genetic Science & Dutch Society of Clinical Genetic Laboratory Specialists, 2013).
25. Gardiner, S. L. et al. Large normal-range TBP and ATXN7 CAG repeat lengths are associated with increased lifetime risk of depression. *Transl. Psychiatry* **7**, e1143 (2017).
26. Stingl, K. et al. CDHR1 mutations in retinal dystrophies. *Sci. Rep.* **7**, 6992 (2017).
27. Khan, M. et al. Detailed phenotyping and therapeutic strategies for intronic ABCA4 variants in Stargardt disease. *Mol. Ther. Nucleic Acids* **21**, 412–427 (2020).
28. Braun, T. A. et al. Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. *Hum. Mol. Genet.* **22**, 5136–5145 (2013).
29. Cremers, F. P. et al. Autosomal recessive retinitis pigmentosa and cone-rod dystrophy caused by splice site mutations in the Stargardt's disease gene ABCR. *Hum. Mol. Genet.* **7**, 355–362 (1998).
30. Sangermano, R. et al. Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. *Genet. Med.* **21**, 1751–1760 (2019).
31. Weisschuh, N. et al. Deep-intronic variants in CNGB3 cause achromatopsia by pseudoexon activation. *Hum. Mutat.* **41**, 255–264 (2020).
32. Weisschuh, N., Buena-Atienza, E. & Wissinger, B. Splicing mutations in inherited retinal diseases. *Prog. Retin. Eye Res.* **80**, 100874 (2021).
33. Di Scipio, M. et al. Phenotype driven analysis of whole genome sequencing identifies deep intronic variants that cause retinal dystrophies by aberrant exonization. *Invest. Ophthalmol. Vis. Sci.* **61**, 36 (2020).
34. Di, Y. et al. Whole-exome sequencing analysis identifies mutations in the EYS gene in retinitis pigmentosa in the Indian population. *Sci. Rep.* **6**, 19432 (2016).
35. Schiff, E. R. et al. A genetic and clinical study of individuals with nonsyndromic retinopathy consequent upon sequence variants in HGSNAT, the gene associated with Sanfilippo C mucopolysaccharidosis. *Am. J. Med. Genet. C* **184**, 631–643 (2020).
36. Gerber, S. et al. Complete exon-intron structure of the retina-specific ATP binding transporter gene (ABCR) allows the identification of novel mutations underlying Stargardt disease. *Genomics* **48**, 139–142 (1998).
37. González-Del Pozo, M. et al. Unmasking retinitis pigmentosa complex cases by a whole genome sequencing algorithm based on open-access tools: hidden recessive inheritance and potential oligogenic variants. *J. Transl. Med.* **18**, 73 (2020).
38. Bhatia, S., Goyal, S., Singh, I. R., Singh, D. & Vanita, V. A novel mutation in the PRPF31 in a North Indian adRP family with incomplete penetrance. *Doc. Ophthalmol.* **137**, 103–119 (2018).
39. Rose, A. M. & Bhattacharya, S. S. Variant haploinsufficiency and phenotypic non-penetrance in PRPF31-associated retinitis pigmentosa. *Clin. Genet.* **90**, 118–126 (2016).
40. Vithana, E. N. et al. Expression of PRPF31 mRNA in patients with autosomal dominant retinitis pigmentosa: a molecular clue for incomplete penetrance? *Invest. Ophthalmol. Vis. Sci.* **44**, 4204–4209 (2003).
41. Brambillasca, F. et al. Promoter analysis of TFPT (FB1), a molecular partner of TCF3 (E2A) in childhood acute lymphoblastic leukemia. *Biochem. Biophys. Res. Commun.* **288**, 1250–1257 (2001).
42. Rak, M. & Rustin, P. Supernumerary subunits NDUFA3, NDUFA5 and NDUFA12 are required for the formation of the extramembrane arm of human mitochondrial complex I. *FEBS Lett.* **588**, 1832–1838 (2014).
43. Rose, A. M., Mukhopadhyay, R., Webster, A. R., Bhattacharya, S. S. & Waseem, N. H. A 112 kb deletion in chromosome 19q13.42 leads to retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.* **52**, 6597–6603 (2011).
44. Radziwon, A. et al. Single-base substitutions in the CHM promoter as a cause of choroideremia. *Hum. Mutat.* **38**, 704–715 (2017).
45. Bauwens, M. et al. ABCA4-associated disease as a model for missing heritability in autosomal recessive disorders: novel noncoding splice, cis-regulatory, structural, and recurrent hypomorphic variants. *Genet. Med.* **21**, 1761–1771 (2019).
46. Runhart, E. H. et al. Association of sex with frequent and mild ABCA4 alleles in Stargardt disease. *JAMA Ophthalmol.* **138**, 1035–1042 (2020).
47. Maggi, J. et al. De novo assembly-based analysis of RPGR exon ORF15 in an Indigenous African Cohort overcomes limitations of a standard next-generation sequencing (NGS) data analysis pipeline. *Genes* **11**, https://doi.org/10.3390/genes11070800 (2020).
48. Diekstra, A. et al. Translating sanger-based routine DNA diagnostics into generic massive parallel ion semiconductor sequencing. *Clin. Chem.* **61**, 154–162 (2015).
49. Weisschuh, N. et al. Molecular and clinical analysis of 27 German patients with Leber congenital amaurosis. *PLoS ONE* **13**, e0205380 (2018).
50. Sharon, D. et al. A nationwide genetic analysis of inherited retinal diseases in Israel as assessed by the Israeli inherited retinal disease consortium (IIRDC). *Hum. Mutat.* **41**, 140–149 (2020).
51. Okonechnikov, K., Conesa, A. & García-Alcalde, F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292–294 (2016).
52. McLaren, W. et al. The Ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
53. The Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat. Genet.* **46**, 818–825 (2014).
54. Erikson, G. A. et al. Whole-genome sequencing of a healthy aging cohort. *Cell* **165**, 1002–1011 (2016).
55. Abecasis, G. R. et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
56. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
57. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
58. Grantham, R. Amino acid difference formula to help explain protein evolution. *Science* **185**, 862–864 (1974).
59. Jaganathan, K. et al. Predicting splicing from primary sequence with deep learning. *Cell* **176**, 535–548.e524 (2019).
60. Farek, J. et al. xAtlas: scalable small variant calling across heterogeneous next-generation sequencing experiments. Preprint at *bioRxiv* https://doi.org/10.1101/295071 (2018).
61. Chen, X. et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).
62. Boeva, V. et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **28**, 423–425 (2012).
63. Verbakel, S. K. et al. The identification of a RNA splice variant in TULP1 in two siblings with early-onset photoreceptor dystrophy. *Mol. Genet. Genom. Med.* **7**, e660 (2019).
64. Firth, H. V. et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am. J. Hum. Genet.* **84**, 524–533 (2009).
65. Dolzhenko, E. et al. Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome Res.* **27**, 1895–1903 (2017).
66. Allaire, J. RStudio: integrated development environment for R. *RStudio* **770**, 394 (2012).
67. Shapiro, M. B. & Senapathy, P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15**, 7155–7174 (1987).
68. Reese, M. G., Eeckman, F. H., Kulp, D. & Haussler, D. Improved splice site detection in Genie. *J. Comput. Biol.* **4**, 311–323 (1997).
69. Pertea, M., Lin, X. & Salzberg, S. L. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res.* **29**, 1185–1190 (2001).
70. Yeo, G. & Burge, C. B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **11**, 377–394 (2004).

71. Desmet, F. O. et al. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **37**, e67 (2009).
72. Sangermano, R. et al. ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome Res.* **28**, 100–110 (2018).

ACKNOWLEDGEMENTS

We thank the affected individuals and their families for participating in this study and our funding bodies. We thank B. van Gestel for his expert bioinformatic input. The work of Z.F. is funded by the Foundation Fighting Blindness USA Project Program Award, grant no. PPA-0517-0717-RAD (to F.P.M.C., S.R., and C.B.H.). The research was supported by the European Union's Horizon 2020 research and innovation program under the EJP RD COFUND-EJP No. 825575 (to F.P.M.C. and S.R.), the Algemene Nederlandse Vereniging ter Voorkoming van Blindheid, Oogfonds, Landelijke Stichting voor Blinden en Slechtzienden; Rotterdamse Stichting Blindenbelangen, Stichting Blindenhulp, Stichting tot Verbetering van het Lot der Blinden, and Stichting Blinden-Penning (to S.R. and F.P.M.C.). The work of L.W. and A.D. was supported by grant awards from Fighting Blindness Ireland (FB Irl; FB16FAR, FB18CRE, FB20DOC) (to F.P.M.C., S.R., and G.J.F.), The Health Research Board of Ireland (HRB; POR/2010/97) (to G.J.F.) in conjunction with Health Research Charities Ireland (HRCI; MRCG-2013-8, MRCG-2016-14) (to G.J.F.), the Irish Research Council (IRC; GOIPG/2017/1631) (to G.J.F.), and Science Foundation Ireland (SFI; 16/1A/4452) (to G.J.F.).

AUTHOR CONTRIBUTIONS

Z.F. and L.W. are co-first authors. Z.F., L.W., and S.R. performed the data analysis and experiments. C.G., J.C., and G.D.N.A. provided infrastructure and bioinformatic expert input. T.B.-Y., L.I.v.d.B., C.B.H., N.W., E.S.D., and P.F.K. collected clinical cases and performed clinical examinations of patients. Z.C. and L.d.R. contributed to the experiments. S.R., F.P.M.C., T.B.-Y., A.D., and G.J.F. contributed significantly to the design of the study, provided infrastructure, and strategic support. Z.F., L.W., F.P.M.C., and S.R. wrote the manuscript. All authors read and approved the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41525-021-00261-1.

Correspondence and requests for materials should be addressed to Susanne Roosing.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021

Appendix 7

International Journal of *Molecular Sciences*

MDPI

Article

Clinical and Genetic Re-Evaluation of Inherited Retinal Degeneration Pedigrees following Initial Negative Findings on Panel-Based Next Generation Sequencing

Kirk A. J. Stephenson [1,*,†], Julia Zhu [1,*,†], Adrian Dockery [2], Laura Whelan [3], Tomás Burke [1], Jacqueline Turner [1], James J. O'Byrne [1], G. Jane Farrar [3] and David J. Keegan [1]

[1] Mater Clinical Ophthalmic Genetics Unit, The Mater Misericordiae University Hospital, D07 R2WY Dublin, Ireland; tomasburke@mater.ie (T.B.); jturner@mater.ie (J.T.); jamesobyrne@mater.ie (J.J.O.); dkeegan@mater.ie (D.J.K.)
[2] Next Generation Sequencing Laboratory, Pathology Department, The Mater Misericordiae University Hospital, D07 R2WY Dublin, Ireland; AdrianDockery@mater.ie
[3] The School of Genetics & Microbiology, Trinity College Dublin, D02 PN40 Dublin, Ireland; whelan11@tcd.ie (L.W.); jane.farrar@tcd.ie (G.J.F.)
* Correspondence: kirkstephenson@hotmail.com (K.A.J.S.); T5000@mater.ie (J.Z.)
† These authors contributed equally to this work.

Abstract: Although rare, inherited retinal degenerations (IRDs) are the most common reason for blind registration in the working age population. They are highly genetically heterogeneous (>300 known genetic loci), and confirmation of a molecular diagnosis is a prerequisite for many therapeutic clinical trials and approved treatments. First-tier genetic testing of IRDs with panel-based next-generation sequencing (pNGS) has a diagnostic yield of ≈70–80%, leaving the remaining more challenging cases to be resolved by second-tier testing methods. This study describes the phenotypic reassessment of patients with a negative result from first-tier pNGS and the rationale, outcomes, and cost of second-tier genetic testing approaches. Removing non-IRD cases from consideration and utilizing case-appropriate second-tier genetic testing techniques, we genetically resolved 56% of previously unresolved pedigrees, bringing the overall resolve rate to 92% (388/423). At present, pNGS remains the most cost-effective first-tier approach for the molecular assessment of diverse IRD populations Second-tier genetic testing should be guided by clinical (i.e., reassessment, multimodal imaging, electrophysiology), and genetic (i.e., single alleles in autosomal recessive disease) indications to achieve a genetic diagnosis in the most cost-effective manner.

Keywords: inherited retinal degenerations; retinal dystrophy; genetic testing; next generation sequencing; whole exome sequencing; single gene sequencing; unresolved inherited retinal degenerations

check for updates

Citation: Stephenson, K.A.J.; Zhu, J.; Dockery, A.; Whelan, L.; Burke, T.; Turner, J.; O'Byrne, J.J.; Farrar, G.J.; Keegan, D.J. Clinical and Genetic Re-Evaluation of Inherited Retinal Degeneration Pedigrees following Initial Negative Findings on Panel-Based Next Generation Sequencing. *Int. J. Mol. Sci.* **2022**, *23*, 995. https://doi.org/10.3390/ijms23020995

Academic Editors: Christina Zeitz and Isabelle Audo

Received: 11 December 2021
Accepted: 12 January 2022
Published: 17 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Inherited retinal degenerations (IRDs) are rare genetic disorders associated with pathogenic variation in over 300 known genetic loci, manifesting with variably progressive visual dysfunction [1]. IRDs represent the main cause of visual loss in the working age group in many Western nations [2,3]. The clinically and genetically heterogeneous nature of IRDs make accurate molecular diagnosis challenging. First-tier genetic testing of IRDs with panel-based next-generation sequencing (pNGS) has a diagnostic yield of ≈70–80% [4–6]. Whole exome/genome sequencing (WES/WGS) may detect further variants, resolving up to 79% of pedigrees [7,8]; however, this increased scope of sequencing comes with increased cost and requires resources to process and store data as well as effectively manage significant secondary findings [9,10].

Meticulous phenotyping may guide the choice of most appropriate genetic testing modality [11]. Novel therapeutic options including gene and stem cell therapies are burgeoning, and confirmation of genetic etiology is often a prerequisite for access to clinical trials and approved treatments [12–15]. Confirmation of genotype is important also for family risk determination so that families with genetically determined IRDs can be evaluated, tested, and counseled appropriately with regard to reproductive options (e.g., prenatal diagnosis, pre-implantation genetic testing). Thus, it is critical to maximise the genetic resolve rate for IRDs.

Herein, we describe the process of reassessing the phenotype of patients with an initial negative genetic result from pNGS and the rationale of further clinical care and genetic testing strategies.

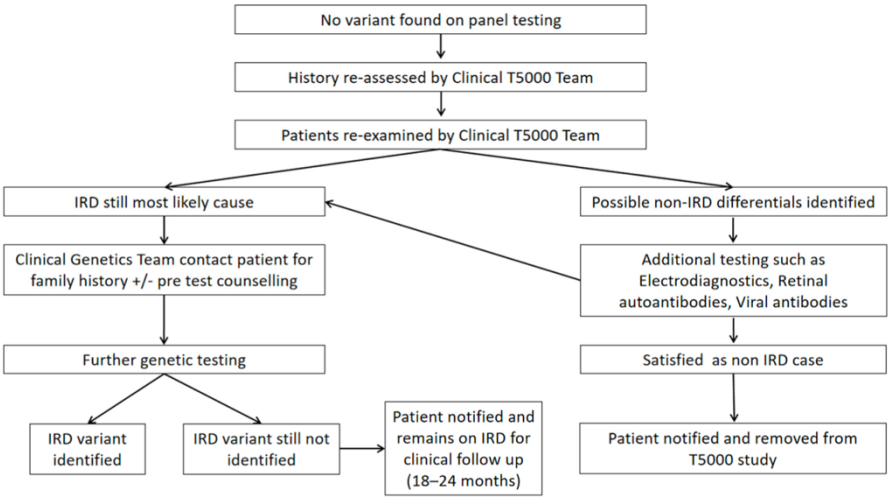
2. Methods

Patients enrolled on the Mater Clinical Ophthalmic Genetics arm of the Irish national inherited retinal degeneration registry (Target 5000) were assessed for a genetic cause of their ophthalmic ± syndromic phenotype. All patients had undergone comprehensive clinical assessment including LogMAR visual acuity (VA, Optos plc, Dunfermline, Scotland, UK), formal visual fields (VF, Humphrey Field Analyser, Carl Zeiss MediTec, Dublin, CA, USA), colour vision assessment (Lanthony D15, Gulden Ophthalmics, Elkins Park, PA, USA), ocular motility/nystagmus assessment, and dilated slit lamp biomicroscopy with Goldmann applanation tonometry (Haag-Streit UK Ltd, Harlow, UK). Multimodal imaging included colour fundus photography and autofluorescence (Optos 'California,' Optos plc, Dunfermline, Scotland, UK) and spectral domain optical coherence tomography (OCT, Cirrus 5000, Carl Zeiss MediTec, Dublin, CA, USA). Visual electrophysiology (ERG, Metrovision, Perenchies, France) was assessed where diagnostically relevant. All clinically assessed patients had undergone a research grade pNGS of 250 IRD-implicated genes at the Ocular Genetics Unit, Trinity College Dublin [4,5,16].

All patients with a negative result from an initial pNGS approach (exonic regions of 250 IRD-implicated genes) were reassessed by reviewing their existing records (clinical examinations, multimodal imaging, visual fields, and electrophysiology). This was performed by three clinicians in a masked fashion (KS, TB, DK). A fourth investigator (JZ) assessed these reports for agreement, with consensus cases progressing to the clinical genetics multidisciplinary team (MDT) and disputed cases brought back for in-person clinical reassessment. In-person reassessment included functional (VA, VF, electrophysiology) and structural assessment (multimodal imaging) as required for diagnostic clarification. See Figure 1A.

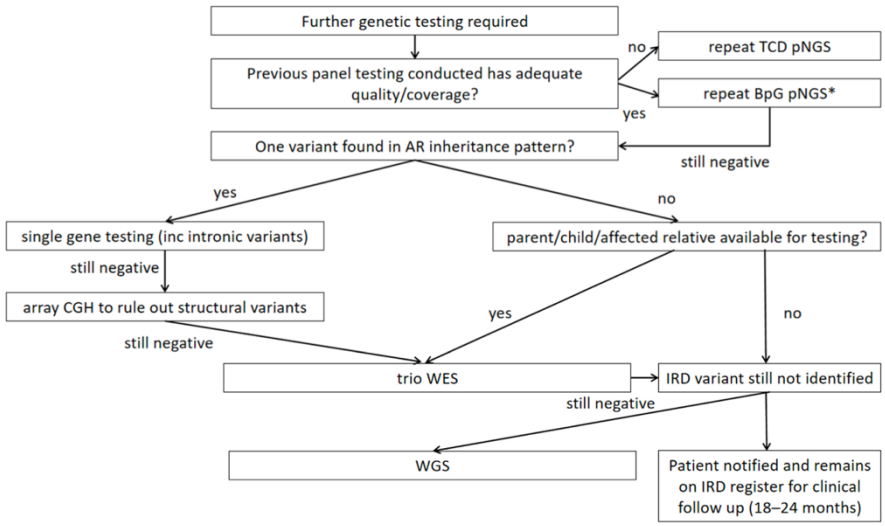
Unresolved pedigrees that retained a phenotype consistent with IRD were referred to the clinical genetics MDT for discussion of the most appropriate further genetic testing modalities based on their clinical findings, family history, and any findings from 1st-tier pNGS (e.g., single allele in autosomal recessive disease). The second-tier genetic testing modalities employed in this study are as follows. For cases in which gene coverage from the initial pNGS run was felt to be inadequate, a repeat pNGS sequencing (250 gene panel) run and/or direct manual inspection of Binary Alignment Map (BAM) files was performed (pedigrees 19, 20, 23). For cases where primary pNGS coverage of genes of interest was adequate, a larger pNGS of 351 genes was applied to extracted DNA by a commercial laboratory (Blueprint Genetics, Helsinki, Finland) (pedigrees 24, 26) [17]. Gene panels can be found in Supplementary Table S1. Single gene testing was applied to cases with a single candidate variant previously detected on 1 allele or to cases with a classic phenotype e.g., Stargardt Disease (OMIM#248200). This approach included sequencing of both exons and introns of the gene of interest and flanking regions, in particular *ABCA4* (OMIM*601691, pedigrees 21 and 22), *ADGRV1* (OMIM*602851, pedigree 31), *BBS1* (OMIM*209901, pedigree 32), *CDH23* (OMIM*601067, pedigree 29), *CNNM4* (OMIM*607805, pedigree 28), *EYS* (OMIM*612424, pedigree 25), *PEX7* (OMIM*601757, pedigree 18), and *TRIM32* (OMIM*602290, pedigree 30). Trio WES (sequencing of all exonic DNA and flanking regions) was performed by a

commercial laboratory (Blueprint Genetics) for 4 pedigrees (proband and their parents) where no candidate variants had been identified on 1st-tier pNGS. Three pedigrees undergoing trio WES had a non-syndromic retinitis pigmentosa (RP) phenotype (one each had X-linked, autosomal dominant, and autosomal recessive modes of inheritance, pedigrees 27, 33, and 34, respectively) with the remaining pedigree having an autosomal dominant vitreoretinopathy phenotype (pedigree 35). All detected genetic variants were reported with HGNC nomenclature, confirmed with bidirectional Sanger sequencing (to rule out incorrect reads, etc.), and compared with the reference genome (GRCH37/HG19) (see Figure 1B).



(A)

Figure 1. *Cont.*



(**B**)

Figure 1. Proposed algorithm for clinical and genetic reassessment of 'gene-negative' cases. (**A**) Algorithm for clinical reassessment of 'gene-negative' cases. (**B**) Algorithm for selecting the most appropriate further genetic testing modalities. * https://blueprintgenetics.com/tests/panels/ophthalmology/retinal-dystrophy-panel/ (accessed on 8 November 2021).

3. Results

Of 441 patients (331 pedigrees) in the Mater Clinical Ophthalmic Genetics Unit IRD cohort that underwent genetic testing before 2019, 69 patients (16%) from 52 pedigrees were genetically unresolved after 1st-tier pNGS. Following reassessment of their phenotypes, 74% of patients (n = 51) retained a clinical diagnosis of IRD, and 26% (n = 18) were deemed non-IRD acquired disease (Table 1 and Figure 2). Mean age was 58.06 (SD ± 16.97) years and 50.57 (SD ± 16.12) years, and 61% and 53% were female for the acquired and IRD groups, respectively. Of the 51 patients (35 pedigrees) who remained clinically consistent with IRD, 34 patients (67%) from 18 pedigrees were available for further genetic testing (Table 2). A total of 17 clinically IRD patients were unavailable for further investigation (deceased (n = 2), no family member available for trio WES (n = 4), patient preference due to SARS-CoV-2 pandemic (n = 11)).

Table 1. Non-IRD diagnosis categories and demographics.

Group	Mean Age, Years (SD)	Female (%)	n = (%)
Posterior Uveitis	51 (16.39)	57%	7 (39%)
AMD	76 (8.39)	25%	4 (22%)
Myopic/Pachychoroid Degeneration	45.5 (3.54)	50%	2 (11%)
ION	50	100%	1 (6%)
Normal	49 (12.92)	75%	4 (22%)

AMD—age-related macular degeneration. ION—inherited optic neuropathy. SD—standard deviation.

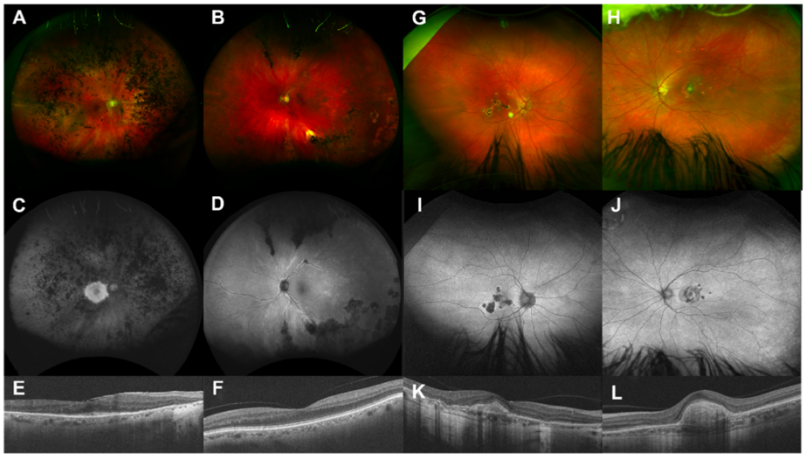


Figure 2. Non-IRD ('gene negative') examples with markedly asymmetric disease. Pedigree #2: Colour fundus photographs a 70-year-old man with markedly asymmetrical retinal pigmentation. The right eye (**A**) has more classic features of retinitis pigmentosa (RP) while the left eye (**B**) has less pronounced paraarteriolar intraretinal pigment migration only. Fundus autofluorescence delineates a small central island of residual RPE in the right eye (**C**) while the left eye (**D**) shows subtle hyperautofluorescent paravascular changes not in keeping with RP. Optical coherence tomography confirms asymmetric disease with relative sparing of the central macular outer retina in the right eye (**E**) with entirely normal retinal lamination in the left eye (**F**). No relevant genetic variants were detected on pNGS. This patient had a history of childhood meningitis with no family history, stable visual acuity (6/12 right, 6/6 left), and no progression of visual symptoms. Clinical consensus opinion was reached of asymmetric post-inflammatory pigmentary retinal changes, and no further genetic testing was indicated. Pedigree #3: Colour fundus photographs (**G**,**H**) of a 47-year-old woman with asymmetrical macular atrophy. Autofluorescence (**I**,**J**) shows hypoautofluorescent areas of macular atrophy with surrounding/intervening areas of hyperautofluorescence. Retinal vasculature and periphery are otherwise normal. OCT (**K**,**L**) shows outer retinal atrophy, subretinal fibrosis, and focal choroidal attenuation. The clinical phenotype was reassessed, and a diagnosis of punctate inner choroidopathy was made, with the decision to not pursue further genetic testing. The patient was referred to the uveitis clinic.

Clinical reassessment has allowed a revision of the clinical diagnosis (i.e., non-IRD) in 18 patients and, with the genetic investigation techniques outlined in Table 2, a further 16 patients from 10 IRD pedigrees have been resolved bringing the total genetic resolution rate for IRDs in this cohort to 92% (388/423).

Repeated pNGS testing led to genetic resolution in 100% of patients assessed in this manner (n = 5, pedigrees 19, 20, 23, 24, 26). If coverage with the 1st-tier pNGS (250 gene panel) approach was adequate, the larger 351 gene panel was used for 2nd-tier genetic assessment; however, if coverage of genes deemed to be relevant to the clinical phenotype was inadequate (e.g., *BBS10*), repeat of the initial 250 gene panel was performed.

Table 2. Summary of results obtained after additional genetic testing. Pedigrees 18–27 are resolved, and pedigrees 28–36 remained unresolved after additional testing. Variants in plain text were identified from the primary pNGS run, and variants in **bold** were detected with further testing modalities as outlined in the 'Method' column. All variants reported here were ACMG class 5 (pathogenic).

Pedigree	N =	Phenotype	Inheritance	Gene	Variant 1	Variant 2	Method	1st-Tier Problem
1–17	18	Non-IRD	-	-	-	-	-	-
18	1	Refsum Disease	AR	*PEX7*	c.875T > A, p.Leu292 *	**c.40A > C, p.Thr14Pro**	Single gene testing	Limited coverage
19	2	EOSRD	AR	*CFAP410*	c.218G > C, p.Arg73Pro	**c.218G > C, p.Arg73Pro**	Repeat pNGS (R)	Misaligned reads, index hopping
20	1	sRP	AR	*FLVCR1*	**c.1022A > G, p.Tyr341Cys**	**c.1307 + 5G > T †**	Repeat pNGS (R)	Additional phenotype information
21	1	STGD	AR	*ABCA4*	c.752del, p.Phe251Serfs*11 †	**c.5461 − 10T > C p.Thr1821Aspfs*6, Thr1821Valfs*13**	Single gene testing	Intronic variant
22	1	STGD	AR	*ABCA4*	c.4363T > C, p.Cys1455Arg	**c.4253 + 43G > A p.Ile1377Hisfs*3**	Single gene testing	Intronic variant
23	1	BBS	AR	*BBS10*	c.2119_2120del, p.Val707 *	**c.687del, p.Val230Phefs*7**	Repeat pNGS (R)	Poor coverage
24	3	nsRP	AD	*RP1*	**c.2321_2322insAlu**	-	Repeat pNGS (A)	Complex structural variant
25	1	nsRP	AR	*EYS*	c.2620C > T, p.Gln874 *	**c.(?_-538-1)_(2023+1_2024-1)del †**	Single gene testing	Copy number variants
26	3	nsRP	XL	*RPGR*	**c.2777_2778del, p.Glu926Glyfs*152 †**	-	Repeat pNGS (A)	Low complexity ORF15 region
27	2	nsRP	XL	*RPGR*	**c.2571_2572del, p.Glu859Glyfs*219**	-	Trio WES	Low complexity ORF15 region
28	1	sMD	AR	*CNNM4*	c.1660G > T, p.Ala554Ser	Unresolved	Single gene testing	-
29	1	USH	AR	*CDH23*	c.289-1G > A, p.Arg964Gln	Unresolved	Single gene testing	-
30	1	nsRP	AR	*TRIM32*	c.691del, p.Ala231Glnfs*21	Unresolved	Single gene testing	-
31	1	USH	AR	*ADGRV1*	c.18025C > T, p.Arg6009 *	Unresolved	Single gene testing	-
32	1	BBS	AR	*BBS1*	c.478C > T, p.Arg160Trp	Unresolved	Single gene testing	-
33	3	nsRP	AD *	-	Unresolved	-	Trio WES	-
34	1	nsRP	AR *	-	Unresolved	Unresolved	Trio WES	-
35	9	VRO	AD *	-	Unresolved	-	Trio WES	-
36–52	16	Still consistent with IRDs	-	-	Unresolved	Unresolved	-	Retest delayed due to SARS-CoV-2 pandemic

AD—Autosomal Dominant. AR—Autosomal Recessive. XL—X-linked. BBS—Bardet–Biedl Syndrome. EOSRD—Early-onset severe retinal dystrophy. sRP—Syndromic Retinitis Pigmentosa. nsRP—Non-syndromic Retinitis Pigmentosa. sMD—Syndromic Macular Dystrophy. STGD—Stargardt Disease. USH—Usher Syndrome. VRO—Vitreoretinopathy. * Presumed inheritance pattern based on available family history. (A)—repeat pNGS at accredited laboratory. (R)—repeat pNGS at research laboratory. † novel variant.

Single gene testing (introns and exons) was used for cases of clinically autosomal recessive IRDs where one pathogenic allele had been previously identified on pNGS. This approach was used for nine cases, with 44% detecting a second pathogenic allele (pedigrees 18, 21, 22, and 25).

A total of 15 patients (four pedigrees) with no candidate genetic variants from 1st-tier pNGS underwent WES, but three out of four pedigrees remain unresolved (pedigrees 34, 35, and 36). In total, four novel variants in *ABCA4*, *EYS*, *FLVCR1*, and *RPGR* were

detected (Table 2). Pedigrees that remain unresolved following 2nd-tier testing will undergo array comparative genomic hybridisation and/or WGS techniques to assess for structural variants or copy number variations.

The 18 patients (17 pedigrees) deemed non-IRD on clinical re-evaluation have all been referred to the appropriate ophthalmic sub-specialties (e.g., uveitis, neuro-ophthalmology, medical retina) and subsequently discharged from the IRD clinic.

4. Discussion

For 69 patients (16%) previously unresolved by 1st-tier pNGS, clinical reassessment guided the most appropriate course of action: further genetic testing modalities in 74% and reclassification as acquired disease in 26%. Nearly half (47%) of patients (16 of 34) or 56% of pedigrees (10 of 18) available for 2nd-tier genetic reassessment were genetically resolved after further investigation, including four novel variants (Table 2). Following these additional tests, 92% (388/423) of this total IRD cohort received a molecular diagnosis for their retinal/syndromic phenotype.

4.1. Clinical Reassessment

Useful factors for confirmation of IRD were (1) symptom onset before 40 years, (2) symmetrical disease, (3) positive family history, (4) signs of progression, and (5) associated ocular/systemic features (e.g., juvenile onset posterior subcapsular cataract, sensorineural hearing loss, post-axial polydactyly, etc.) [18,19]. Typical retinal phenotypes suggestive of IRD can be seen in Figure 3. Conversely, non-IRD diagnoses were associated with unilateral or asymmetrical disease (e.g., uveitis, trauma, Figure 2) and onset after 60 years (e.g., age-related macular degeneration). These factors are not individually/independently diagnostic of IRD, but considering the weight of the total evidence can be suggestive. Many acquired retinal conditions with advanced manifestations may mimic IRDs with features such as arteriolar attenuation (e.g., retinal vasculitis, retinal arteriolar occlusion), disc pallor (e.g., anterior ischemic optic neuropathy, glaucoma) and intraretinal RPE migration (e.g., retinal pigment epitheliitis, late-stage multifocal choroiditis) [20]. In this cohort, significant interocular asymmetry was found in 0% (0/16) of resolved IRD cases versus 33.33% (6/18) of non-inherited cases. Thus, this finding should prompt thorough clinical and biochemical analysis for an acquired cause before considering IRD genotyping. Subjective (e.g., past medical history, family history, cadence of progression) and objective (e.g., clinical aspect, multimodal imaging, electrophysiology) clues help to clarify the likelihood of a genetic etiology for each case of retinal degeneration. Classic mimickers of IRD include autoimmune retinopathy, infectious and non-infections posterior uveitides, and drug-induced retinal toxicity (e.g., hydroxychloroquine (simulating bullseye maculopathy/STGD1), deferoxamine). In the initial stages of this research collaboration, difficult cases with some IRD features were progressed to genetic testing in the hope of useful insights from genetic data. This approach has now changed with the benefit of MDT experience showing that the addition of potentially unrelated genetic findings may obfuscate the true etiology, adding unnecessary delay and anxiety for the affected patient and their wider family. A negative result from appropriate genetic screening for IRDs (e.g., pNGS) should prompt clinical reassessment for an acquired cause. This may enhance the percentage of resolved pedigrees in all cohorts as reasonable exclusion of a primary genetic etiology has been performed with a minimum of overlooked pedigrees (i.e., acquired cases are removed from the total number in cohort).



Figure 3. Multimodal imaging (MMI) of cases consistent with IRD. Pedigree #18: (**A**,**B**): Colour fundus photographs demonstrating symmetrical predominantly midperipheral pigmentary changes. (**C**,**D**): Autofluorescence showing patchy midperipheral and focal posterior pole hypoautofluorescence consistent with retinal pigment epithelium (RPE) atrophy. (**E**,**F**): OCT demonstrating predominant preservation of photoreceptor/RPE laminae with a focal nasal defect of photoreceptor inner segments (IS) in E and multiple (nasal and temporal) IS/RPE defects in F. A second pathogenic *PEX7* variant (OMIM*601757, c.40A > C, p.Thr14Pro) was detected via single gene sequencing, confirming a genetic diagnosis of autosomal recessive (AR) Refsum disease (OMIM#614879). This patient also had systemic features of this condition including ataxia. Pedigree #22: (**G**,**H**): Colour fundus photographs showing macular atrophy with surrounding subretinal flecks with sparing outside the vascular arcades. (**I**,**J**): Autofluorescence confirms foveal hypoautofluorescence surrounded by hyperautofluorescent flecks mostly confined to the macula with some flecks nasal to the optic nerve head in J. (**K**,**L**): OCT showing foveal outer retinal atrophy. These multimodal imaging features are in keeping with AR Stargardt disease (OMIM#248200). Single gene (i.e., *ABCA4*) testing allowed the detection of a second pathogenic allele (c.4253 + 43G > A p.[=, Ile1377Hisfs*3]), genetically resolving this case.

4.2. Second-Tier Genetic Testing Approaches and Costs

The panel-based NGS approach in a research-based academic laboratory with validation in an accredited laboratory adopted by Target 5000, as outlined in previous publications, accrued substantial cost savings (Table 3) [5,16]. Using pNGS compared to WES as a first-pass approach for 'gene hunting' is favorable, as it reduces bioinformatic demand, identifies the molecular diagnosis for the majority of IRD patients, and frees resources for 2nd-tier testing of more difficult cases as necessary (Figure 1B). Similar approaches have been adopted by other centers with resolution of approximately one-third of partially resolved cases on 2nd-tier testing (i.e., single pathogenic allele detected in autosomal recessive IRD) [6,21].

Table 3. Indicative total cost of genetic testing approaches used in this study (Trinity College Dublin, Ireland and Blueprint Genetics, Finland) and resolve rate for each testing modality. pNGS—panel-based Next-Generation Sequencing. WES—Whole Exome Sequencing.

Test Type	Cost	N = Per Test Type	Total Cost	Resolve Rate, % (*n*=)
pNGS (resolved at research laboratory + accredited laboratory validation)	€600 (€250 + €350)	441	€240,450 *	84.4% (372/441)
pNGS (negative research laboratory result + expanded accredited laboratory panel + validation at an accredited laboratory for other affected family members)	€1120 (€250 + €870 ± €350)	5 + 3 variant confirmation	€6350	100% (5/5)
Single Gene Testing (negative research laboratory result + accredited single gene test + validation at an accredited laboratory for other affected family members)	€700 (€250 + €450 ± €350)	9	€6300	44% (4/9)
WES/trio WES (negative research laboratory result + accredited trio WES + validation at an accredited laboratory for other affected family members)	€2550 (€250 + €2300 ± €350)	4 + 1 variant confirmation	€10,550	25% (1/4) **

* Being unresolved, the 69 unresolved cases discussed in this paper only underwent initial research-grade pNGS prior to this study, and thus, this total represents (372 × €600) + (69 × €250). ** The case identified with WES could be solved using expanded pNGS (351 genes).

The cost and resolve rate of the various genetic testing modalities utilised in this study are outlined in Table 3. In total, 44% (n = 9) and 25% (n = 4) of re-tested patients had a genetically diagnostic result from single gene testing and WES, respectively. Thus, the most cost-effective initial genetic testing approach remains pNGS with the use of more costly and bioinformatic resource-consuming modalities reserved for genetically unresolved cases applied after reassessment of phenotype. The relatively low resolve rate for single gene sequencing may represent spurious 1st-tier variant detection, skewing focus to a single gene when a broader re-screening approach (e.g., WGS, array comparative genomic hybridisation) may yield an increased likelihood of resolution [22,23]. Closer scrutiny of single variants detected by pNGS (such as ACMG class and in silico functional analysis) will be undertaken prior to committing to 2nd-tier testing choice. A low resolve rate with WES may represent the majority of exonic variants being already screened with pNGS; in this case, WGS may be a more suitable approach to interrogate the deep intronic variants or structural variants of genes of interest. Primary application of these broader techniques to an untested IRD population may resolve significantly higher proportions of cases but at greater relative cost than pNGS [21,24].

Ongoing revision and upgrading of the NGS panel design allows the reapplication of this technique to existing DNA samples, which is the most cost-effective approach, as the primary cost of pNGS is preparation, sequencing, and panel design [25]. The total cost incurred to reassess the 34 patients was €23,200 (mean €1450 per resolved case) on top of the original pNGS costs and an additional €10,600 for estimated cost of clinicians and genetic counsellors).

4.3. Examples of Resolution Problems and Their Solutions

In research sequencing endeavors, there is a need to maximise the number of patients that can be analysed on a single sequencing run in order to optimise cost effectiveness. This aggressive pooling approach during sample preparation can result in some patients falling below the acceptable sequencing coverage thresholds. In certain cases, a specific phenotype may allow for the manual review of plausible candidate genes in the sequencing data (i.e., BAM files). This was the scenario for pedigree #23. Re-phenotyping of patient #170 revealed additional systemic features including diabetes mellitus and polydactyly in addition to teenage-onset RP in keeping with Bardet–Biedl syndrome (BBS, OMIM#209900). This narrowed the scope of the genetic search from all genes associated with autosomal recessive RP (i.e., 30 non-syndromic and > 27 syndromic) [26,27] to BBS-associated genes (n = 16), of which two (i.e., *BBS1* and *BBS10*) explain 45% of BBS cases [28]. Despite sub-optimal sequencing coverage, this refined phenotype allowed for the targeted direct manual inspection of BBS-associated genes, revealing two clearly pathogenic frameshift variants in the *BBS10* gene: c.2119_2120del, p.(Val707*) and c.687del, p.(Val230Phefs*7). These variants were then confirmed by direct sequencing.

Panel-based sequencing typically relies heavily on amplification methods to efficiently capture gene panels and incorporate necessary components (e.g., indexing tags for multiplexing DNA) for parallel sequencing. As a result, regions of DNA that are difficult to amplify may sequence poorly. A prime example in the IRD gene panel is the ORF15 region of the *RPGR* gene (OMIM*312610), which includes highly repetitive regions with sequences of low nucleotide diversity (e.g., purine-rich). Cloning allowing bidirectional sequence reads of such regions has been demonstrated as a cost-effective way to overcome this issue [29,30]. The *RPGR* ORF15 region is a crucial locus to assess X-linked RP (XLRP) pedigrees, as ≈60% of variants causing XLRP may be found in this mutational 'hot spot' [31]. Pedigrees that were resolved by the resequencing of *RPGR* include pedigree #26 (resolved by an expanded pNGS (351 genes) at an accredited laboratory) and pedigree #27 (resolved using trio WES) [32]. In retrospect, the variant identified in pedigree #27 was identifiable with the expanded pNGS (351 genes, commercial laboratory); their techniques to achieve better amplification/coverage of the ORF15 region are proprietary (commercial laboratory) but likely reflect superior capture or hybridisation methods than those used in the research laboratory and lends further credence to pNGS use as 1st-tier investigation for IRDs.

In rare circumstances, a patient's zygosity of a variant may be misdetermined by alignment methods. This can result in a patient appearing to have a single variant in a gene associated with a recessive mode of inheritance, when in actuality, the patient is homozygous for that variant. In the case of pedigree #19 (*CFAP410* OMIM*603191), this was attributed to index hopping in the sequencing run, which is a sequencing phenomenon known to occur when multiplexing samples [33]. It occurs when the sequencing platform erroneously attempts to demultiplex the pool of samples and misassigns sequencing reads to the wrong sample of origin. In the case of pedigree #19, this was sufficient to result in a homozygous variant to be classified as heterozygous.

Panel-based sequencing frequently places a primary focus on exonic DNA regions. Canonical splice site variants can be routinely detected by any exon-focused approach given proximity to the exon, while the opposite is true for most deep intronic variants. However, variants in close proximity to exons may be inconsistently detected by exon-focused methods. The detection of variants located near exons depends on several variables such as capture probe design, efficiency of probe hybridisation, sequence coverage depth, as well as the browser extensible data file used for analysis. To ensure consistency and to simplify the variant interpretation process, it is possible to establish bioinformatic cut-offs based on purely exonic variants and canonical splice variants (±1–2 nucleotides from the exon). For this reason, some pathogenic intronic variants may be filtered out during the variant filtering process. Additionally, sequence coverage depth may significantly drop off outside of the exon target, resulting in near-exonic variants falling below detection thresholds: for example, near-exon aberrant RNA (NEAR) variants and deep-intronic

variants [34]. A relevant example here is pedigree #22, whom had one variant identified on the initial pNGS run (*ABCA4* c.4363T > C, p.Cys1455Arg), and through retesting (single gene sequencing of the *ABCA4* gene), a second pathogenic variant in *ABCA4* c.4253 + 43G > A, p.Ile1377Hisfs*3 has been identified.

For clinically recessive cases, the initial detection of one likely causative allele in an appropriate IRD-associated gene adds extra evidence supportive of a clinical/genetic diagnosis of IRD. Single gene testing is appropriate in this situation, and 2nd-tier testing of intronic regions within the candidate gene identified a second pathogenic variant in 44% (n = 4/9) of the cohort tested in this manner (Table 2). Deep intronic variants resulting in activation of cryptic splice sites account for up to 5% of all known pathogenic *ABCA4* alleles [35,36], and in our cohort, both cases identified a second variant in the *ABCA4* gene (pedigree #21 novel non-canonical splice site variant/NEAR; pedigree #22 NEAR intronic variant). For 56% of partially resolved cases that were not resolved by 2nd-tier single gene sequencing, perhaps the originally 1st-tier pNGS-detected heterozygous variants (single alleles) were spurious, leading to further assessment of an unrelated gene. In fact, the single alleles detected by 1st-tier pNGS were ACMG class 5 (i.e., pathogenic) for pedigrees resolved by 2nd-tier single gene sequencing, while the five unresolved pedigrees were class 3 (variant of unknown significance) or 4 (likely pathogenic). In light of this, greater caution will be taken prior to this 2nd-tier approach, with only single variants with ACMG class 5 progressing to single gene sequencing and rigorous in silico analyses performed on < class 5 or novel variants prior to a decision regarding 2nd-tier genetic testing modalities (i.e., further evaluate the gene in question or reassess a wider range of IRD-implicated genes). If the initial variant cannot be upgraded to ACMG class 5, then broader genetic screening techniques, such as WES or WGS, are more appropriate. Although WGS is a powerful technique for resolving such issues, currently, the resource demand (financial and infrastructure) is prohibitive, as 1st- or 2nd-tier screening unless other more affordable yet highly effective options have been exhausted.

WES is gaining popularity as a primary method for the genetic assessment for IRDs [31]. Benefits include the ability to store entire exomes and apply a 'virtual panel' for genes of interest in specific areas of medicine [37]. This allows for the expansion of gene panels for subsequent re-analysis in cases where new genotype–phenotype associations have been established in the time since the original interpretation without the need for re-sequencing. Alternate gene panels (e.g., for other body systems, e.g., cardiac, neurological) can be applied to the pre-existing WES data at a later date if new phenotypic detail becomes apparent as well as for use in establishing population-specific variant frequency. Costs are upfront in terms of DNA sequencing, and data can be securely stored for future research or clinical applications. Published series report a 49–63% genetic resolve rate for their IRD cohorts using WES [22,31,38,39]; thus, the financial and efficiency benefit remains with pNGS as in our pathway [4,5]. Although a panel-based method eliminates the possibility of re-analysis with a broader virtual gene panel, this is less likely to be an issue for conditions whereby the gene associations have been largely already established. For IRDs, this is evident from the relatively small number of new gene associations in recent years [1].

Approximately 70–80% of IRDs are resolved by analysis of exons (e.g., by pNGS), while deep intronic mutations may account for 1.4–25%, and copy number variations account for up to 9% of IRDs and are unlikely to be detected by short read sequencing such as NGS [5,35,40,41]. The use of WGS as a primary genetic investigation is not practical at this time due to the high costs, large bioinformatic workload, and limited addition of genetic resolve rate over 1st-tier pNGS [30]. This broader sequencing approach may obscure the true genotype by detecting rare non-pathogenic polymorphisms, causing delay in reaching an accurate harmonised molecular/clinical diagnosis. As per WES, digital storage of WGS data from unresolved cases/pedigrees may allow virtual panel tests to be performed in the future and for new gene associations to be made. The added risk in 1st-tier WGS approaches is the workload required to process the output data, which may require artificial intelligence algorithms to create actionable clinical outcomes within

a reasonable timeframe [42], and detection of significant secondary findings relevant to other body systems (e.g., cancer risk genes) [10,43]. However, recent data from the UK 100,000 genomes project showed >40% genetic diagnosis rate using 2nd-tier WGS for heritable ophthalmic disease with previously negative 1st-tier genetic testing (e.g., pNGS, WES); WGS addressed issues including non-coding/structural variants, mitochondrial DNA variants, and poor coverage by exon-based approaches [8,44].

The 'cost-effectiveness' of genetic testing for IRDs includes not only the direct costs of genetic sequencing but also the surrounding infrastructure (e.g., genetic counsellor, MDT meeting, molecular geneticist/scientist time/training, clinic use). The broader the scope of a test (e.g., WGS), the greater the yield of non-diagnostic (potentially misleading) findings, including secondary findings, which must be reported [10,16,43]. The expenditure of limited resources on the use and interpretation of 1st-tier WGS in a health service of limited size reduces the number of patients/pedigrees that can be served. Thus, pNGS maximises the genetic/diagnostic resolution rate for IRD in small/medium-sized countries with the reservation of broader tests for the minority of unresolved cases. Large collaborative groups involved in such multicenter WGS projects will likely identify condition-specific algorithms for optimal genetic resolution yield while having strategies for addressing secondary findings from these broad-scope techniques. WGS may supersede pNGS as 1st-tier sequencing for IRDs on a global scale once further cost reduction and better interpretation/infrastructure guidelines are in place.

4.4. Relevance to Gene Therapies

The need to achieve an accurate genetic diagnosis for IRD patients is increasingly important as novel gene therapies are in clinical trials for a growing number of etiologies (Table 4). Few (15.6% (n = 5)) of the patients in this study are eligible for upcoming gene therapy clinical trials for *RPGR* gene mutations (NCT03316560, NCT03252847, and NCT03116113) for which specific genotype parameters are an inclusion criteria prerequisite. A validated genetic diagnosis also allows accurate genetic counselling and family planning, as many patients with IRD are in the reproductive age (e.g., prenatal diagnosis and preimplantation genetic diagnosis) [2,3].

Table 4. List of ongoing IRD gene therapy clinical trials.

Gene	NCT Number	Technique	Phase	Status
Rod-Cone Dystrophies				
MERTK	NCT01482195	AAV	1/2	Completed
PDE6B	NCT03328130	AAV	1/2	Recruiting
RHO	NCT04123626	AON	1/2	Recruiting
RPGR	NCT03252847	AAV	1/2	Completed
	NCT03116113	AAV	1/2	Completed
	NCT03316560	AAV	1/2	Recruiting
RLBP1	NCT03374657	AAV	1/2	Recruiting
USH2A	NCT03780257	AON	1/2	Not recruiting
MYO7A	NCT01505062	LV	1/2	Terminated

Table 4. *Cont.*

Gene	NCT Number	Technique	Phase	Status
Macular/Cone Dystrophies or Cone Dysfunction Syndromes				
RS1	NCT02416622	AAV	1/2	Terminated
	NCT02317887	AAV	1/2	Recruiting
ABCA4	NCT01367444	LV	1/2	Terminated
CNGB3	NCT03001310	AAV	1/2	Completed
	NCT02599922	AAV	1/2	Recruiting
CNGA3	NCT03758404	AAV	1/2	Completed
	NCT02935517	AAV	1/2	Recruiting
	NCT02610582	AAV	1/2	Recruiting
Leber Congenital Amaurosis				
RPE65	NCT02781480 NCT02946879	AAV	1/2	Recruiting
	NCT00643747	AAV	1/2	Completed
	NCT01496040	AAV	1/2	Completed
	NCT00821340	AAV	1	Completed
	NCT00749957	AAV	1/2	Completed
	NCT00481546	AAV	1	Completed
GUCY2D	NCT03920007	AAV	1/2	Recruiting
CEP290	NCT03913143	AON	3	Not recruiting
	NCT03872479	Gene editing	1/2	Recruiting
Choroidal Dystrophies				
CHM	NCT02341807	AAV	1/2	Completed
	NCT02671539	AAV	2	Completed
	NCT01461213	AAV	1/2	Completed
	NCT02077361	AAV	1/2	Not recruiting
	NCT02553135	AAV	1/2	Completed
	NCT03507686 NCT03496012	AAV	2 3	Completed Completed
	NCT02407678	AAV	2	Completed
	NCT04483440	AAV	1	Recruiting

AAV = adeno-associated virus. AON = antisense oligonucleotide. LV = lentiviral vector NCT = Reference number for study on clinicaltrials.gov.

4.5. Limitations

This study was carried out during the global SARS-CoV-2 pandemic, and thus, some patients were unwilling to travel for further in-person clinical assessment, potentially reducing the total power of the investigations in terms of final genetic resolution rate. Two patients were deceased by time of reassessment, and thus, further genetic testing was not carried out. Trio WES was not possible on five patients, as relevant family members were not available (e.g., deceased, out of country, unwilling to attend/participate during SARS-CoV-2 pandemic). IRDs are rare by definition; thus, despite being a national effort including over 1000 patients, statistical analyses to make definitive conclusions regarding the most appropriate genetic testing methodology for each gene/variant cannot be reliably drawn from the data included herein. The most powerful evidence for initial pNGS is the cost savings accrued and high genetic resolve rate, which then frees funding for more

expensive 2nd-tier genomic techniques to be used in a targeted manner, as outlined above once the phenotype has been refined.

5. Conclusions

This study showed that 2nd-tier genetic testing resolved 56% of previously unresolved pedigrees (pathogenic variants in IRD-implicated genes), leading to an overall resolve rate of 92% (388/423). Second-tier genetic testing should be guided by detailed clinical (i.e., reassessment, multimodal imaging, electrophysiology) and genetic (i.e., single alleles in AR disease) indications to achieve a molecular diagnosis of IRD in the most cost-effective manner. Detailed pedigree phenotyping can help to reclassify disease and guide non-IRD patients (e.g., AMD, uveitis) into more appropriate care pathways while preventing further unnecessary genetic testing. The glass ceiling of sequencing resolution in international data may represent a patient population including a proportion of non-inherited pathology. Diagnostic refinement may improve genetic resolution rates for true cases of IRD.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms23020995/s1.

Author Contributions: Conceptualisation, D.J.K., K.A.J.S., J.Z., J.J.O.; methodology, K.A.J.S., A.D., L.W., D.J.K.; software, J.Z., L.W., A.D.; validation, K.A.J.S., J.Z., A.D., T.B., D.J.K., G.J.F.; formal analysis, J.Z., L.W., A.D.; investigation, K.A.J.S., J.Z.; resources, D.J.K.; data curation, J.Z., A.D.; writing—original draft preparation, J.Z. and K.A.J.S.; writing—review and editing, K.A.J.S., J.Z., A.D., L.W., T.B., J.T., J.J.O., G.J.F., D.J.K.; visualization, D.J.K.; supervision, K.A.J.S., D.J.K.; project administration, J.Z.; funding acquisition, D.J.K., G.J.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by grant awards from Fighting Blindness Ireland (FB Irl; FB16FAR, FB18CRE, FB20DOC), The Health Research Board of Ireland (HRB; POR/2010/97) in conjunction with Health Research Charities Ireland (HRCI; MRCG-2013-8, MRCG-2016-14), the Irish Research Council (IRC; GOIPG/2017/1631), and Science Foundation Ireland (SFI; 16/1A/4452).

Institutional Review Board Statement: This study was carried out at the Mater Clinical Ophthalmic Genetics Unit at the Mater Misericordiae University Hospital, Dublin Ireland. All patients signed informed consent for participation in this study and for publication of their data in anonymised fashion. This study complies with General Data Protection Regulation guidelines (articles 6 and 9) and abides by the tenets of the Declaration of Helsinki.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Anonymised source data are available upon reasonable request.

Acknowledgments: Fighting Blindness Ireland. We thank the patients and families for their involvement in this research endeavor.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Daiger, S.P. Summaries of Genes and Loci Causing Retinal Diseases (RetNet). Available online: https://sph.uth.edu/retnet/ (accessed on 5 October 2021).
2. Heath Jeffery, R.C.; Mukhtar, S.A.; McAllister, I.L.; Morgan, W.H.; Mackey, D.A.; Chen, F.K. Inherited retinal diseases are the most common cause of blindness in the working-age population in Australia. *Ophthalmic Genet.* **2021**, *42*, 431–439. [CrossRef] [PubMed]
3. Liew, G.; Michaelides, M.; Bunce, C. A comparison of the causes of blindness certifications in England and Wales in working age adults (16–64 years), 1999–2000 with 2009–2010. *BMJ Open* **2014**, *4*, e004015. [CrossRef]
4. Dockery, A.; Stephenson, K.; Keegan, D.; Wynne, N.; Silvestri, G.; Humphries, P.; Kenna, P.F.; Carrigan, M.; Farrar, G.J. Target 5000: Target capture sequencing for inherited retinal degenerations. *Genes* **2017**, *8*, 304. [CrossRef]
5. Whelan, L.; Dockery, A.; Wynne, N.; Zhu, J.; Stephenson, K.; Silvestri, G.; Turner, J.; O'Byrne, J.J.; Carrigan, M.; Humphries, P.; et al. Findings from a genotyping study of over 1000 people with inherited retinal disorders in Ireland. *Genes* **2020**, *11*, 105. [CrossRef]

6. Shah, M.; Shanks, M.; Packham, E.; Williams, J.; Haysmoore, J.; MacLaren, R.E.; Németh, A.H.; Clouston, P.; Downes, S.M. Next generation sequencing using phenotype-based panels for genetic testing in inherited retinal diseases. *Ophthalmic Genet.* **2020**, *41*, 331–337. [CrossRef] [PubMed]
7. Ellingford, J.M.; Barton, S.; Bhaskar, S.; Williams, S.G.; Sergouniotis, P.I.; O'Sullivan, J.; Lamb, J.A.; Perveen, R.; Hall, G.; Newman, W.G.; et al. Whole genome sequencing increases molecular diagnostic yield compared with current diagnostic testing for inherited retinal disease. *Ophthalmology* **2016**, *123*, 1143–1150. [CrossRef]
8. Smedley, D.; Smith, K.R.; Martin, A.; Thomas, E.A.; McDonagh, E.M.; Cipriani, V.; Ellingford, J.M.; Arno, G.; Tucci, A.; Vandrovcova, J.; et al. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care—Preliminary Report. *N. Engl. J. Med.* **2021**, *385*, 1868–1880.
9. Hart, M.R.; Biesecker, B.B.; Blout, C.L.; Christensen, K.D.; Amendola, L.M.; Bergstrom, K.L.; Biswas, S.; Bowling, K.M.; Brothers, K.B.; Conlin, L.K.; et al. Secondary findings from clinical genomic sequencing: Prevalence, patient perspectives, family history assessment, and health-care costs from a multisite study. *Genet. Med.* **2019**, *21*, 1100–1110. [CrossRef]
10. Zhu, J.; Stephenson, K.A.; Farrar, G.J.; Turner, J.; O'Byrne, J.J.; Keegan, D. Management of significant secondary genetic findings in an ophthalmic genetics clinic. *Eye* **2021**, *3*, 1–3. [CrossRef]
11. Stone, E.M.; Andorf, J.L.; Whitmore, S.S.; DeLuca, A.P.; Giacalone, J.C.; Streb, L.M.; Braun, T.A.; Mullins, R.F.; Scheetz, T.E.; Sheffield, V.C.; et al. Clinically focused molecular investigation of 1000 consecutive families with inherited retinal disease. *Ophthalmology* **2017**, *124*, 1314–1331. [CrossRef]
12. Bush, R.A.; Zeng, Y.; Colosi, P.; Kjellstrom, S.; Hiriyanna, S.; Vijayasarathy, C.; Santos, M.; Li, J.; Wu, Z.; Sieving, P.A. Preclinical dose-escalation study of intravitreal AAV-RS1 gene therapy in a mouse model of X-linked retinoschisis: Dose-dependent expression and improved retinal structure and function. *Hum. Gene Ther.* **2016**, *27*, 376–389. [CrossRef]
13. Cukras, C.; Wiley, H.E.; Jeffrey, B.G.; Sen, H.N.; Turriff, A.; Zeng, Y.; Vijayasarathy, C.; Marangoni, D.; Ziccardi, L.; Kjellstrom, S.; et al. Retinal AAV8-RS1 gene therapy for X-linked retinoschisis: Initial findings from a phase I/IIa trial by intravitreal delivery. *Mol. Ther.* **2018**, *26*, 2282–2294. [CrossRef] [PubMed]
14. Cehajic-Kapetanovic, J.; Xue, K.; de la Camara, C.M.; Nanda, A.; Davies, A.; Wood, L.J.; Salvetti, A.P.; Fischer, M.D.; Aylward, J.W.; Barnard, A.R.; et al. Initial results from a first-in-human gene therapy trial on X-linked retinitis pigmentosa caused by mutations in RPGR. *Nat. Med.* **2020**, *26*, 354–359. [CrossRef]
15. Méjécase, C.; Malka, S.; Guan, Z.; Slater, A.; Arno, G.; Moosajee, M. Practical guide to genetic screening for inherited eye diseases. *Ther. Adv. Ophthalmol.* **2020**, *12*, 2515841420954592. [CrossRef] [PubMed]
16. Stephenson, K.A.; Zhu, J.; Wynne, N.; Dockery, A.; Cairns, R.M.; Duignan, E.; Whelan, L.; Malone, C.P.; Dempsey, H.; Collins, K.; et al. Target 5000: A standardized all-Ireland pathway for the diagnosis and management of inherited retinal degenerations. *Orph. J. Rare Dis.* **2021**, *16*, 1–8. [CrossRef] [PubMed]
17. Genetics, B. Retinal Dystrophy Panel. Available online: https://blueprintgenetics.com/tests/panels/ophthalmology/retinal-dystrophy-panel/ (accessed on 8 November 2021).
18. Koenig, R. Bardet-Biedl syndrome and Usher syndrome. *Dev. Ophthalmol.* **2003**, *37*, 126–140.
19. Prokofyeva, E.; Troeger, E.; Wilke, R.; Zrenner, E. Early visual symptom patterns in inherited retinal dystrophies. *Ophthalmologica* **2011**, *226*, 151–156. [CrossRef]
20. Sevgi, D.D.; Davoudi, S.; Comander, J.; Sobrin, L. Retinal pigmentary changes in chronic uveitis mimicking retinitis pigmentosa. *Graefe's Arch. Clin. Exp. Ophthalmol.* **2017**, *255*, 1801–1810. [CrossRef]
21. Haer-Wigman, L.; van Zelst-Stams, W.A.; Pfundt, R.; van den Born, L.I.; Klaver, C.C.; Verheij, J.B.; Hoyng, C.B.; Breuning, M.H.; Boon, C.J.; Kievit, A.J.; et al. Diagnostic exome sequencing in 266 Dutch patients with visual impairment. *Eur. J. Hum. Genet.* **2017**, *25*, 591–599. [CrossRef] [PubMed]
22. Van Nimwegen, K.J.; van Soest, R.A.; Veltman, J.A.; Nelen, M.R.; van der Wilt, G.J.; Vissers, L.E.; Grutters, J.P. Is the $1000 genome as near as we think? A cost analysis of next-generation sequencing. *Clin. Chem.* **2016**, *62*, 1458–1464. [CrossRef]
23. Zampaglione, E.; Kinde, B.; Place, E.M.; Navarro-Gomez, D.; Maher, M.; Jamshidi, F.; Nassiri, S.; Mazzone, J.A.; Finn, C.; Schlegel, D.; et al. Copy-number variation contributes 9% of pathogenicity in the inherited retinal degenerations. *Genet. Med.* **2020**, *22*, 1079–1087. [CrossRef] [PubMed]
24. Van Cauwenbergh, C.; Van Schil, K.; Cannoodt, R.; Bauwens, M.; Van Laethem, T.; De Jaegere, S.; Steyaert, W.; Sante, T.; Menten, B.; Leroy, B.P.; et al. arrEYE: A customized platform for high-resolution copy number analysis of coding and noncoding regions of known and candidate retinal dystrophy genes and retinal noncoding RNAs. *Genet. Med.* **2017**, *19*, 457–466. [CrossRef] [PubMed]
25. Wetterstrand, K. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Available online: www.genome.gov/sequencingcostsdata (accessed on 8 November 2021).
26. Hims, M.M.; Diager, S.P.; Inglehearn, C.F. Retinitis pigmentosa: Genes, proteins and prospects. *Dev. Ophthalmol.* **2003**, *37*, 109–125.
27. Ferrari, S.; Di Iorio, E.; Barbaro, V.; Ponzin, D.; SSorrentino, F.; Parmeggiani, F. Retinitis pigmentosa: Genes and disease mechanisms. *Curr. Genom.* **2011**, *12*, 238–249.
28. Gerth, C.; Zawadzki, R.J.; Werner, J.S.; Héon, E. Retinal morphology in patients with BBS1 and BBS10 related Bardet–Biedl Syndrome evaluated by Fourier-domain optical coherence tomography. *Vis. Res.* **2008**, *48*, 392–399. [CrossRef] [PubMed]
29. Parmeggiani, F.; Barbaro, V.; De Nadai, K.; Lavezzo, E.; Toppo, S.; Chizzolini, M.; Palù, G.; Parolin, C.; Di Iorio, E. Identification of novel X-linked gain-of-function RPGR-ORF15 mutation in Italian family with retinitis pigmentosa and pathologic myopia. *Sci. Rep.* **2016**, *6*, 1–8. [CrossRef] [PubMed]

30. Dockery, A.; Whelan, L.; Humphries, P.; Farrar, G.J. Next-Generation Sequencing Applications for Inherited Retinal Diseases. *Int. J. Mol. Sci.* **2021**, *22*, 5684. [CrossRef] [PubMed]
31. Cho, A.; de Carvalho, J.R.; Tanaka, A.J.; Jauregui, R.; Levi, S.R.; Bassuk, A.G.; Mahajan, V.B.; Tsang, S.H. Fundoscopy-directed genetic testing to re-evaluate negative whole exome sequencing results. *Orphanet J. Rare Dis.* **2020**, *15*, 1–11. [CrossRef]
32. Ebenezer, N.D.; Michaelides, M.; Jenkins, S.A.; Audo, I.; Webster, A.R.; Cheetham, M.E.; Stockman, A.; Maher, E.R.; Ainsworth, J.R.; Yates, J.R.; et al. Identification of novel RPGR ORF15 mutations in X-linked progressive cone-rod dystrophy (XLCORD) families. *Investig. Ophthalmol. Vis. Sci.* **2005**, *46*, 1891–1898. [CrossRef]
33. Farouni, R.; Djambazian, H.; Ferri, L.E.; Ragoussis, J.; Najafabadi, H.S. Model-based analysis of sample index hopping reveals its widespread artifacts in multiplexed single-cell RNA-sequencing. *Nat. Commun.* **2020**, *11*, 1–8. [CrossRef]
34. Verbakel, S.K.; Fadaie, Z.; Klevering, B.J.; van Genderen, M.M.; Feenstra, I.; Cremers, F.P.; Hoyng, C.B.; Roosing, S. The identification of a RNA splice variant in TULP1 in two siblings with early-onset photoreceptor dystrophy. *Mol. Genet. Genom. Med.* **2019**, *7*, e660. [CrossRef]
35. Khan, M.; Cornelis, S.S.; del Pozo-Valero, M.; Whelan, L.; Runhart, E.H.; Mishra, K.; Bults, F.; AlSwaiti, Y.; AlTalbishi, A.; De Baere, E.; et al. Resolving the dark matter of ABCA4 for 1054 Stargardt disease probands through integrated genomics and transcriptomics. *Genet. Med.* **2020**, *22*, 1235–1246. [CrossRef]
36. Whelan, L.; Dockery, A.; Stephenson, K.A.J.; Zhu, J.; Kopcic, E.; Khan, M.; Corradi, Z.; Post, I.J.M.; Wynne, N.; Duignan, E.; et al. Enrichment of ABCA4 c.4539+2028C>T in Stargardt disease patients in Ireland—A detailed geno-type-phenotype analysis. Hum. *Mutat. Current* reference number humu-2021-0469. *Hum. Mutat.*.
37. Molina-Ramírez, L.P.; Kyle, C.; Ellingford, J.M.; Wright, R.; Taylor, A.; Bhaskar, S.S.; Campbell, C.; Jackson, H.; Fairclough, A.; Rousseau, A.; et al. Personalised virtual gene panels reduce interpretation workload and maintain diagnostic rates of proband-only clinical exome sequencing for rare disorders. *J. Med. Genet.* **2021**. [CrossRef] [PubMed]
38. Beryozkin, A.; Shevah, E.; Kimchi, A.; Mizrahi-Meissonnier, L.; Khateb, S.; Ratnapriya, R.; Lazar, C.H.; Blumenfeld, A.; Ben-Yosef, T.; Hemo, Y.; et al. Whole exome sequencing reveals mutations in known retinal disease genes in 33 out of 68 Israeli families with inherited retinopathies. *Sci. Rep.* **2015**, *5*, 1–11. [CrossRef] [PubMed]
39. Ma, D.J.; Lee, H.S.; Kim, K.; Choi, S.; Jang, I.; Cho, S.H.; Yoon, C.K.; Lee, E.K.; Yu, H.G. Whole-exome sequencing in 168 Korean patients with inherited retinal degeneration. *BMC Med. Genom.* **2021**, *14*, 1–12. [CrossRef] [PubMed]
40. Siemiatkowska, A.M.; Collin, R.W.; den Hollander, A.I.; Cremers, F.P. Genomic approaches for the discovery of genes mutated in inherited retinal degeneration. *Cold Spring Harb. Perspect. Med.* **2014**, *4*, a017137. [CrossRef] [PubMed]
41. Qian, X.; Wang, J.; Wang, M.; Igelman, A.D.; Jones, K.D.; Li, Y.; Wang, K.; Goetz, K.E.; Birch, D.G.; Yang, P.; et al. Identification of Deep-Intronic Splice Mutations in a Large Cohort of Patients With Inherited Retinal Diseases. *Front. Genet.* **2021**, *12*. [CrossRef] [PubMed]
42. Yang, A.; Zhang, W.; Wang, J.; Yang, K.; Han, Y.; Zhang, L. Review on the Application of Machine Learning Algorithms in the Sequence Data Mining of DNA. *Front. Bioeng. Biotechnol.* **2020**, *8*, 1032. [CrossRef] [PubMed]
43. Hehir-Kwa, J.Y.; Claustres, M.; Hastings, R.J.; Van Ravenswaaij-Arts, C.; Christenhusz, G.; Genuardi, M.; Melegh, B.; Cambon-Thomsen, A.; Patsalis, P.; Vermeesch, J.; et al. Towards a European consensus for reporting incidental findings during clinical NGS testing. *Eur. J. Hum. Genet.* **2015**, *23*, 1601–1606. [CrossRef]
44. Fadaie, Z.; Whelan, L.; Ben-Yosef, T.; Dockery, A.; Corradi, Z.; Gilissen, C.; Haer-Wigman, L.; Corominas, J.; Astuti, G.D.; de Rooij, L.; et al. Whole genome sequencing and in vitro splice assays reveal genetic causes for inherited retinal diseases. *NPJ Genom. Med.* **2021**, *6*, 1–11. [CrossRef]

Appendix 8

Human Genetics
https://doi.org/10.1007/s00439-022-02441-0

ORIGINAL INVESTIGATION

Usher syndrome type IV: clinically and molecularly confirmed by novel *ARSG* variants

Hedwig M. Velde[1,2] · Janine Reurink[2,3] · Sebastian Held[4] · Catherina H. Z. Li[2,5] · Suzanne Yzer[2,5] · Jaap Oostrik[1,2] · Jack Weeda[5] · Lonneke Haer-Wigman[2,3] · Helger G. Yntema[2,3] · Susanne Roosing[2,3] · Laurenz Pauleikhoff[6] · Clemens Lange[6] · Laura Whelan[7] · Adrian Dockery[7,8] · Julia Zhu[9] · David J. Keegan[9] · G. Jane Farrar[7] · Hannie Kremer[1,2,3] · Cornelis P. Lanting[1,2] · Markus Damme[4] · Ronald J. E. Pennings[1,2]

Received: 23 November 2021 / Accepted: 10 February 2022
© The Author(s) 2022

Abstract
Usher syndrome (USH) is an autosomal recessively inherited disease characterized by sensorineural hearing loss (SNHL) and retinitis pigmentosa (RP) with or without vestibular dysfunction. It is highly heterogeneous both clinically and genetically. Recently, variants in the arylsulfatase G (*ARSG*) gene have been reported to underlie USH type IV. This distinct type of USH is characterized by late-onset RP with predominantly pericentral and macular changes, and late onset SNHL without vestibular dysfunction. In this study, we describe the USH type IV phenotype in three unrelated subjects. We identified three novel pathogenic variants, two novel likely pathogenic variants, and one previously described pathogenic variant in *ARSG*. Functional experiments indicated a loss of sulfatase activity of the mutant proteins. Our findings confirm that *ARSG* variants cause the newly defined USH type IV and support the proposed extension of the phenotypic USH classification.

Abbreviations

ARSG	Arylsulfatase G
bp	Base pair
dB HL	Decibel hearing level
ERG	Electroretinography
ER	Endoplasmic reticulum
HL	Hearing loss
KO	Knockout
OCT	Optical coherence tomography
pNCS	P-nitrocatechol sulfate
PTA	Pure-tone average
OD/OS	Right/left eye
RP	Retinitis pigmentosa
SNHL	Sensorineural hearing loss
USH	Usher syndrome
WT	Wild type

Introduction

Usher syndrome (USH) is an autosomal recessively inherited disease characterized by sensorineural hearing loss (SNHL) and retinitis pigmentosa (RP) with or without vestibular dysfunction. It is the leading cause of inherited deaf-blindness with an estimated prevalence of 3–17 per 100,000 individuals (Boughman et al. 1983; Grondahl 1987; Hallgren 1959; Hope et al. 1997; Kimberling et al. 2010; Rosenberg et al. 1997; Spandau and Rohrschneider 2002). USH is highly heterogeneous from both a clinical and a genetic perspective. Understanding the different genotypes and associated phenotypes is essential for genetic counseling and informing patients about their prognosis. It is also necessary to evaluate the efficacy of future therapies and to select patients that are eligible for such therapies.

Previously, three clinical types of USH were distinguished based on the age of onset, severity, and progression of symptoms (Davenport and Omenn 1977; Smith et al. 1994). USH type I is characterized by severe to profound congenital SNHL, the onset of RP in the first decade of life, and vestibular areflexia (Nisenbaum et al. 2021). USH type II is the most common type of the disease, with moderate to severe congenital SNHL and onset of RP in the second

Cris P. Lanting, Hannie Kremer, and Markus Damme contributed equally.

✉ Ronald J. E. Pennings
Ronald.Pennings@radboudumc.nl

Published online: 28 February 2022

decade of life or later (Nisenbaum et al. 2021; Tsilou et al. 2002). Vestibular function was previously considered unaffected. However, recent studies have also identified vestibular defects in USH II patients (Magliulo et al. 2017; Stemerdink et al. 2021; Wafa et al. 2021). USH type III displays variable progressive SNHL with the age of onset usually in the first decade of life, variable onset of RP, typically by the second decade of life, and variable vestibular function (Geng et al. 2017; Nisenbaum et al. 2021). Nine USH genes have so far been identified: *MYO7A* (OMIM 276900 (Weil et al. 1995)), *USH1C* (OMIM 276904 (Verpy et al. 2000)), *CDH23* (OMIM 601067 (Bolz et al. 2001; Bork et al. 2001)), *PCDH15* (OMIM 602083 (Ahmed et al. 2001; Alagramam et al. 2001)), *USH1G* (OMIM 606943 (Weil et al. 2003)) (type I); *USH2A* (OMIM 276901 (Eudy et al. 1998; Van Wijk et al. 2004)), *ADGRV1* (OMIM 605472 (Weston et al. 2004)), *WHRN* (OMIM 611383 (Ebermann et al. 2007)) (type II); and *CLRN1* (OMIM 276902 (Sankila et al. 1995)) (type III). *CIB2* (OMIM 605564) was described to be associated with USH type I but this was later refuted (Booth et al. 2018; Riazuddin et al. 2012). Finally, *PDZD7* (OMIM 612971) has been suggested to be a modifier gene of retinal disease and a contributor to digenic USH (Ebermann et al. 2010) and was shown to be causal to non-syndromic hearing loss type DFNB57 (OMIM 618003 (Booth et al. 2015; Vona et al. 2016)).

In 2018, a homozygous missense variant (Supplemental Table 1) in the arylsulfatase G (*ARSG*) gene (OMIM 610008) was associated with an atypical USH phenotype characterized by late-onset RP with a distinctive retinal phenotype of ring-shaped atrophy along the arcades, late-onset progressive SNHL, and no vestibular involvement (Khateb et al. 2018). Later, when a second homozygous *ARSG* missense variant was described that caused a similar remarkably late onset of visual and hearing impairment, this phenotype was defined as USH type IV (OMIM 618144) (Abad-Morales et al. 2020). The classification as a fourth, distinct USH type was confirmed by the identification of three additional variants associated with similar phenotypic characteristics (Peter et al. 2020). Five other novel *ARSG* variants were subsequently reported to underlie this newly described phenotype (Fowler et al. 2021; Igelman et al. 2021).

ARSG encodes the sulfatase ARSG, which is transported to lysosomes after biosynthesis in the endoplasmic reticulum (ER) (Ferrante et al. 2002; Kowalewski et al. 2014). ARSG has been shown to hydrolyze pseudosubstrates such as *p*-nitrocatechol sulfate (pNCS) and 4-methylumbelliferyl sulfate (Frese et al. 2008). Moreover, the physiological substrate of ARSG has been determined to be 3-*O*-sulfated glucosamine in heparan sulfate and ARSG was shown to be critical for the complete lysosomal degradation of 3-*O*-glucosamine modified heparan sulfate (Kowalewski et al. 2014). The ARSG sulfatase is ubiquitously expressed in mice, including the kidney, testis, liver, spleen, brain, retina, and cochlea (Girotto et al. 2014; Kowalewski et al. 2014; Kowalewski 2015; Kruszewski et al. 2016; Ratzka 2010).

This report confirms both clinically and molecularly that *ARSG* variants are causal to USH type IV. We describe three novel pathogenic *ARSG* variants, two novel likely pathogenic variants, and one previously described pathogenic variant in three unrelated individuals. All subjects share a phenotype of late-onset SNHL and RP, characterized by predominantly pericentral and macular changes.

Materials and methods

For this retrospective case study, clinical and genetic data of included subjects were obtained from the Radboud university medical center (Nijmegen, the Netherlands), the University Medical Center Freiburg (Germany) and the Mater Misercordiae University Hospital (Dublin, Ireland). All included subjects were previously seen at the out-patient clinic of these hospitals.

Clinical evaluation of recruited subjects

To obtain a complete description of the subjects' phenotypes, we retrospectively collected demographic data and general medical history, ophthalmic data, audiological data, and data on vestibular function. Demographics and medical history included age, descent, comorbidities, medication use, and medical family history. For ophthalmic phenotyping, we collected ocular history, age of onset and progression of ophthalmic symptoms, best corrected visual acuity, refractive error, slit lamp examination findings, and multimodal imaging including color fundus photography, optical coherence tomography (OCT), visual field tests, electroretinography (ERG), and fundus autofluorescence. Audiological data included age of onset, type and progression of audiological symptoms, medical history of the ear, otoscopic assessment, pure tone and speech audiometry, and speech-in-noise testing. The medical history of the ear included recurrent ear infections, previous noise exposure, the use of ototoxic medication, meningitis, severe head injuries, and ear surgeries. Finally, we collected data on vestibular function by identifying whether vestibular symptoms were present and whether vestibular function had been examined. Incomplete data were supplemented with additional diagnostics if possible.

Auditory phenotype analyses

RStudio (version 1.4.1106) was used to analyze and plot the thresholds for pure-tone audiometry. Air conduction thresholds were used to analyze the observed severity and degree of symmetry of hearing loss and audiogram configuration. In case of

an air–bone gap, the bone conduction thresholds of all audiograms of the subject in question were used instead. In addition, age-related typical audiograms (ARTA) were derived from linear regression analysis of all available audiometric data of included and previously reported subjects as described to provide insight into the rate of progression of hearing loss (Huygen et al. 2003). In addition to assessing potential differences in the progression of hearing loss, the age of onset of hearing loss was determined. The annual threshold deterioration was calculated using cross-sectional non-linear regression analyses of hearing thresholds as a function of age. First, the pure-tone average (PTA) across 0.5–4.0 kHz ($PTA_{0.5–4\,kHz}$) for each subject was determined. A two-parameter logistic function was used to fit the data, i.e. $PTA_{0.5–4kHz} \sim \frac{130}{1+e^{(-scale*(age-age_{mid}))}}$, similar to what has been described previously (Pauw et al. 2011). The parameter age_{mid} describes the age at which the hearing thresholds are halfway between 0 dB and the asymptotic value (fixed at 130 dB). The parameter *scale* represents the slope of the function at this midpoint, which was used as an estimate for the annual threshold deterioration. The calculated age of onset was defined as when the calculated hearing threshold exceeds 25 dB, defined by the WHO as abnormal hearing, and subjects could benefit from amplification (Informal Working Group on Prevention of Deafness and Hearing Impairment Programme Planning & World Health Organization 1991). The calculated age of onset was determined by interpolating the function fit using the parameters age_{mid} and *scale*. For comparison, the age of onset of SNHL for USH IIa was calculated the same way based on previously published data (Hartel et al. 2016).

Genetic analyses

For subject N, exome sequencing with targeted gene panels for vision disorders (version DG-2.13, consisting of 408 genes) and hearing impairment (version DG-2.15 and DG-2.18, consisting of 173 and 208 genes) was performed in the ISO15189 accredited Genome Diagnostic Laboratory of the Radboud university medical center (Nijmegen, the Netherlands) according to routine diagnostic procedures (Haer-Wigman et al. 2017). Variants in genes from these panels with a frequency below 5% in the dbSNP database and below 1% in an in-house database (consisting of exome sequencing data of 24,488 individuals mainly of Dutch origin) and within the exon or within the intronic position of −8 to + 3 were interpreted and classified based on existing variant classification guidelines established by the Association for Clinical Genetic Science (Plon et al. 2008; Wallis et al. 2013).

Subject D and F were pre-screened to exclude biallelic variants in other USH-associated genes in Trinity College Dublin (Ireland) as part of the Target 5000 study (Whelan et al. 2020) (subject D) and in the University Medical Center Freiburg (Germany) with the 2016 CeGat panel for USH (CeGaT GmbH, Tübingen, Germany) (subject F). The CeGat panel for subject F was later expanded to include *ABCA4* (OMIM 601691), *CNGA1* (OMIM 123825), *EYS* (OMIM 612424), *PDE6A* (OMIM 180071), *PDE6B* (OMIM 180072), and *RPE65* (OMIM 180069).

Genome sequencing was performed for subjects D and F as described before, with 2 × 150nt paired-end reads and a minimal median coverage of 30× (Fadaie et al. 2021). In the analysis of genome sequencing data, single nucleotide variants were first prioritized based on the in-house frequency (< 1%, database consisting of genome sequencing data of 702 individuals mainly of Dutch origin) and the Genome Aggregation Database (gnomAD) minor allele frequency (≤ 1%) (Karczewski et al. 2020). All exonic variants, variants at exon–intron boundaries, and deep-intronic variants in genes associated with USH were considered and ranked based on predicted pathogenicity. All stop-gain variants were considered pathogenic. Missense variants were evaluated using the following prediction tools and threshold scores: CADD_PHRED (≥ 15) (Kircher et al. 2014), PhyloP (≥ 2.7) (Pollard et al. 2010), Grantham (≥ 80) (Grantham 1974), SIFT (Ng and Henikoff 2001), and MutationTaster (Schwarz et al. 2014). Potential effects on splicing were evaluated by employing SpliceAI (≥ 0.1/1) (Jaganathan et al. 2019). Copy number variants (detected with Control-FREEC Copy number and allelic content caller (Boeva et al. 2012)) and structural variants (detected with Manta Structural Variant Caller (Chen et al. 2016)) were prioritized based on whether they passed our in-house quality filter and whether they affected genes known to be associated with USH. All variants were visualized with Integrative Genomics Viewer (version 2.4) to confirm their presence.

When DNA of relatives was available or variants were in close proximity to each other, the presence of variants *in trans* was determined by PCR and subsequent Sanger sequencing according to standard protocols.

Functional analyses of identified variants

Immunoblot and sulfatase activity assays were performed as previously described (Khateb et al. 2018; Peter et al. 2020). Therefore, cDNA expression vectors coding for C-terminally 3xFLAG-tagged wildtype (WT) and mutated ARSG were generated by PCR and cloning into the pcDNA4/TO-C-3xFlag vector with *Eco*R1 and *Xho*I followed by transfection and selection of stable cell lines (primer sequences are provided in Supplemental Table 2). Both patient-derived missense variants were introduced by site-directed mutagenesis as previously described in the WT *ARSG* (Zheng et al. 2004) (primer sequences are provided in Supplemental Table 2). All constructs were sequence-validated by Sanger

sequencing. Subsequently, HT1080 cells were transfected as previously described (Peter et al. 2020). HT1080 cells were treated with 600 µg/ml Zeocin (Invivogen, San Diego, CA, USA) three days after transfection, and stable cell lines were used for immunoblotting and activity assays after reaching confluence and passaging in Zeocin-containing selection medium after two passages. Immunoblotting of transfected HT1080 cells was performed according to standard procedures with a nitrocellulose membrane and antibodies against the FLAG-tag (mouse monoclonal, clone M2; Sigma Aldrich, St. Louis, MO, USA), as well as glyceraldehyde 3-phosphate dehydrogenase (rabbit polyclonal, SC-335; Santa Cruz Biotechnology, Dallas, TX, USA). Sulfatase activity was measured photometrically by the turnover of pNCS from protein lysates of stable HT1080 cell lines with the WT *ARSG* construct or the construct harboring an *ARSG* variant.

Minigene splice assays were performed for variants with a predicted effect on splicing. The genomic regions of interest, including the exon and 500nt intron sequence at both ends, were amplified from patient DNA with PCR (primer sequences are provided in Supplemental Table 2). The PCR product was then inserted in a pCI-NEO vector, adapted to the Gateway® cloning system as previously described (Sangermano et al. 2016). The vector inserts were verified by Sanger sequencing. WT and mutant constructs (500 ng) were subsequently transfected individually into HEK293T cells using polyethylenimine (100 µg/ml in 150 mM NaCl). Approximately 24 h post-transfection, RNA was isolated by employing the NucleoSpin RNA Clean-up Kit (Macherey-Nagel, Düren, Germany). 500 ng of RNA was used for cDNA synthesis with the iScript cDNA synthesis kit (Bio-Rad, Hercules, CA, USA), following the manufacturer's instructions. The cDNA derived from the vector insert was amplified with primers for RHO exons 3 and 5 and amplicons were screened with Sanger sequencing. A PCR for *GAPDH* was performed as cell lysis control.

Results

Clinical evaluation of recruited subjects

Subject N is a 51-year-old Dutch man of Turkish descent seen at the Radboud university medical center in Nijmegen, the Netherlands. His medical history was uneventful, besides hypothyroidism and occupational noise exposure for many years. In particular, there was no relevant ophthalmic history nor any family history related to hearing or vision loss. His parents were not known to be consanguineous. He experienced peripheral vision loss and night blindness since the age of 46 and was diagnosed with RP aged 49 (Supplemental Table 3). His visual acuity was relatively spared (OD (right eye) 20/25 and OS (left eye) 20/32 Snellen). Ophthalmoscopy revealed waxy pallor of the optic disc, mild attenuation of the vessels, and a preserved fovea with perifoveal retinal pigment epithelial mottling. Atrophy was present around the vascular arcades; adjacent to the atrophy, predominantly located nasally and superior to the optic disc, intraretinal bone spicules were noted (Supplemental Fig. 1). The far periphery appeared relatively preserved. OCTs showed foveal sparing with only paracentral loss of the outer retina, corresponding to the relatively mild loss of visual acuity. On fundus autofluorescence, a ring of hyperautofluorescence was visible in the posterior pole at the transition zone. The mid-periphery showed confluent areas of hypoautofluorescence with preserved autofluorescence in the far periphery. ERG revealed lower scotopic than photopic responses; both were diminished. Goldmann perimetry showed an absolute ring scotoma with intact central vision and a relatively preserved peripheral field in accordance with a normal V-4e Isopter. At the age of 42, before his RP diagnosis, subject N was diagnosed with bilateral SNHL for which he is rehabilitated with bilateral conventional hearing aids (Supplemental Table 4). In addition, he experienced tinnitus. Physical examination and otoscopy showed no abnormalities. Pure tone audiometry revealed bilateral moderate to severe SNHL, with air and bone conduction PTAs (0.5–4.0 kHz) of 49 and 46 dB hearing level (dB HL) (right) and 55 and 53 dB HL (left) at age 46. At age 49, there was asymmetry in the air conduction thresholds, mainly in the high frequencies to the detriment of the right. An air–bone gap was observed primarily in the right ear, which could not be explained by otoscopy or CT. The latter revealed no congenital temporal bone abnormalities. At that time, air and bone conduction PTAs were 70 and 51 dB HL (right) and 64 and 55 dB HL (left). Based on the bone conduction PTAs, the progression in three years was 5 (right) and 3 dB HL (left). The subject did not report any vestibular problems.

Subject F is a 72-year-old woman of Caucasian origin, recruited at the Eye Center of the University of Freiburg, Germany. She was diagnosed with hypothyroidism earlier in life. There was no relevant medical or family history related to hearing, vision, or vestibular function. Her parents were not known to be consanguineous. This individual experienced central vision loss from the age of 40 (Supplemental Table 3). Visual acuity deteriorated to hand movement vision only. Ophthalmoscopy revealed a pale optic disc, mild attenuation of the vessels, and atrophy around the arcades with intraretinal bone spicules predominantly located nasally and superior to the atrophy. The far periphery appeared intact. OCT showed complete loss of the outer retinal layers. Fundus autofluorescence showed extensive hypoautofluorescence in the periphery that extended to the inner part of the vascular arcades, leaving a small ring of relative normal autofluorescence in the posterior pole. ERG

revealed absence of both cone and rod responses. As the disease progressed, concentric restriction of the peripheral visual field and a central scotoma developed. Subject F also experienced progressive, bilateral SNHL since the age of 20 for which she has been using bilateral conventional hearing aids since age 31–40 (Supplemental Table 4). At the age of 60 and 72, pure-tone audiometry showed moderate to severe SNHL with a flat audiogram configuration and PTAs of 73 dB HL (both ears) and 90 (right) and 85 dB HL (left). There were no additional problems, notably no reported vestibular symptoms. Bilateral bithermal vestibular caloric stimulation revealed a symmetrical response of the lateral semicircular canals.

Subject D is an 87-year-old man of Caucasian origin seen at the Mater Misercordiae University Hospital in Dublin, Ireland. He had pulmonary tuberculosis when he was 19 years old. Later in life, he developed gastro-esophageal reflux disease, atrial fibrillation, and benign prostatic hyperplasia. This subject was diagnosed with exotropia of the left eye in his early youth. When he was over 60 years old, a posterior subcapsular cataract was observed bilaterally. Medical and family history was uneventful, his parents were not known to be consanguineous. At the age of 51–60, he experienced night blindness and photophobia (Supplemental Table 3). Over time, visual acuity deteriorated to light perception in the right eye and no-light perception in the left eye at age 86. Retinal imaging revealed excavation of the nerve heads, mild attenuation of the vessels, and central atrophy connecting to atrophy around the vascular arcades. This was surrounded by a ring of intraretinal bone-spicule hyperpigmentation, nasally and superior to the disc. Fundus autofluorescence showed central hypoautofluorescence and large confluent areas of hypoautofluorescence adjacent and posterior to the vascular arcades. OCT revealed complete loss of outer retinal layers. At the age of 31–40, subject D developed bilateral SNHL for which he was rehabilitated by bilateral conventional hearing aids (Supplemental Table 4). No other audiological data were available or could be obtained. He reported no vestibular symptoms.

Audiograms of the right and left ears of included and previously reported subjects showed a predominantly bilateral moderate to severe high-frequency SNHL (Supplemental Fig. 2). The ARTA displayed an average progression of hearing loss of 1.0–1.5 dB HL per year (Fig. 1). This corresponds with the mean progression of 1.25 dB HL per year of both included subjects with more than one audiogram available (subjects N and F). The calculated age of SNHL onset was 17 years (Fig. 1).

Identification of *ARSG* variants

The genetic analyses of the three described cases revealed six heterozygous *ARSG* variants (Table 1, Fig. 2).

In subject N, exome sequencing targeting gene panels for vision disorders and hearing impairment initially identified a heterozygous *SPP2* (OMIM 602637, isolated RP (Liu et al. 2015)) variant (NM_006944.3:c.439G > A NP_008875.1:p.(Glu147Lys)) and a possibly explanatory heterozygous *MCM2* (OMIM 116945, isolated SNHL OMIM 616968 (Gao et al. 2015)) variant (NM_004526.4:c.547C > T NP_004517.2:p.(Arg183Cys)). However, these variants were also identified in unaffected relatives (Supplemental Fig. 3). Re-analysis of the exome sequencing data with an updated gene panel for hearing impairment identified two variants in *ARSG*: NM_014960.5:c.1212 + 1G > A NP_055775.2:p.(?); NM_014960.5:c.275T > C NP_055775.2:p.(Leu92Pro). Segregation analysis showed that these variants were present in a compound heterozygous state (Supplemental Fig. 3) as there were no unaffected relatives with both *ARSG* variants. The c.275T > C p.(Leu92Pro) variant is a missense variant that was predicted to be pathogenic by all used *in-silico* prediction tools (Table 1). Skipping of exon 10 (out of frame) was predicted to be a consequence of the c.1212 + 1G > A p.(?) variant (SpliceAI: Donor splice site loss at −1nt (0.99/1) and acceptor splice site loss at −121nt (0.78/1)).

In subject F, a heterozygous variant (NM_206933.4:c.6524G > A NP_996816.3:p.(Arg2175His)) was identified in *USH2A* (OMIM 608400, USH and isolated RP OMIM 613809 (Rivolta et al. 2000)) and genome sequencing was performed to identify a second causal variant in this gene. However, no second *USH2A* variant could be identified and therefore variants in other USH-associated genes were addressed. This revealed two *ARSG* variants: NM_014960.5:c.1326del NP_055775.2:p.(Ser443Alafs*12); NM_014960.5:c.1024C > T NP_055775.2:p.(Arg342Trp). The c.1326del p.(Ser443Alafs*12) variant is a previously described pathogenic stop-gain variant (Peter et al. 2020). The c.1024C > T p.(Arg342Trp) variant is a missense variant that was predicted to be pathogenic by all used *in-silico* prediction tools (Table 1). In addition, this variant was predicted to result in skipping of exon 9 (out of frame) (SpliceAI: Donor splice site loss at + 67nt (0.15/1) and acceptor splice site loss at −41nt (0.25/1)). Segregation analysis could not be performed because relatives were not available.

Genome sequencing was initiated in subject D because no variants could be identified in genes associated with USH. Initially, four heterozygous variants in USH-associated genes were identified: a stop-gain variant in *ARSG* (NM_014960.5:c.588C > A NP_055775.2:p.(Tyr196*)), a non-canonical splice site variant in *ADGRV1* (OMIM 602,851, USH, NM_032119.4:c.5525-7C > T NP_115495.3:p.(?)), a missense variant in *MYO7A* (OMIM 276903, USH and isolated SNHL OMIM 601317 (Liu et al. 1997a, Liu et al. 1997b) and 600060 (Liu et al. 1997a, 1997b; Weil et al. 1997), NM_000260.4:c.905G > A NP_000251.3:p.(Arg302His)), and a deep-intronic variant in *PCDH15* (OMIM 605514, USH and

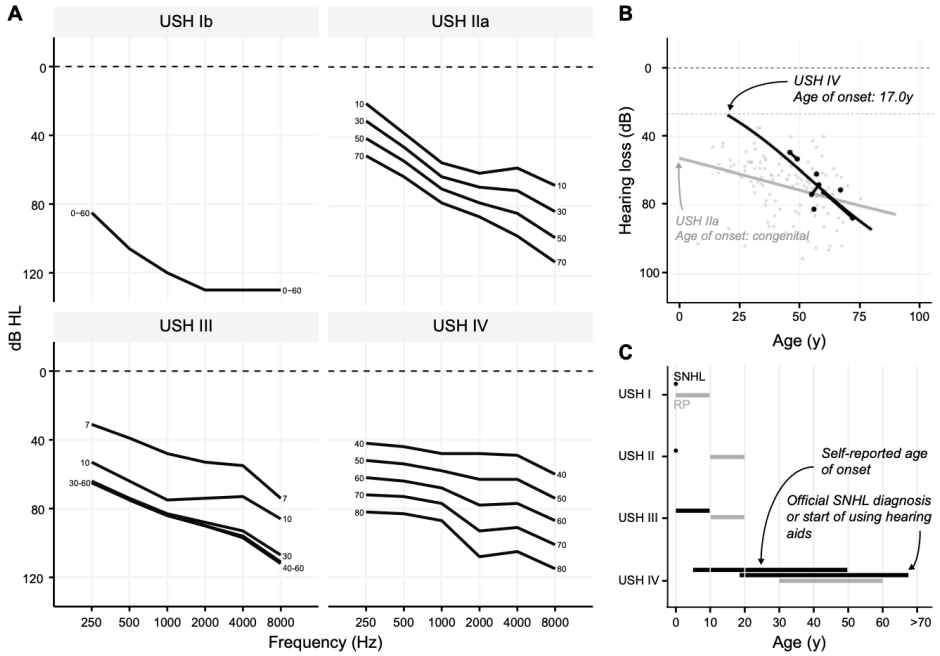


Fig. 1 Audiologic features of USH IV compared to other USH types. **A** ARTA of USH IV compared to USH Ib (Wagenaar et al. 1999), USH IIa (Hartel et al. 2016), and USH III (Plantinga et al. 2004). Age (in years) is indicated at both ends of all lines. **B** Calculated age of onset of SNHL for USH IV based on individual $PTA_{0.5–4\ kHz}$ and compared to USH IIa (Hartel et al. 2016). Black dots represent USH IV $PTA_{0.5–4\ kHz}$ data; gray dots represent USH IIa $PTA_{0.5–4\ kHz}$ data, connected dots represent longitudinal data when more than one audiogram was available for one subject. **C** Visualization of age of onset of RP (in gray) and SNHL (in black) of USH IV compared to published data for other USH types (Geng et al. 2017; Millan et al. 2011; Nisenbaum et al. 2021; Tsilou et al. 2002) (Table 2). *dB HL* decibel hearing level; *RP* retinitis pigmentosa; *SNHL* sensorineural hearing loss; *y* years

isolated SNHL OMIM 609533 (Ahmed et al. 2003), NM_001142764.2:c.1098+2354G>A NP_001136236.1:p.(?)). The *in-silico* predictions for the missense variant in *MYO7A* were moderate (CADD_PHRED: 24.5 (threshold: 15), PhyloP: 2.5 (threshold: 2.7), Grantham: 29 (threshold: 80), MutationTaster: 'Benign' and SIFT: 'Tolerated') and the SpliceAI predictions for the *ADGRV1* and *PCDH15* variants were low (*ADGRV1*: acceptor splice site loss at+11nt (0.01/1); *PCDH15:* acceptor splice site gain at+124nt (0.24/1), acceptor splice site loss at −54nt (0.02/1), donor splice site gain at+3nt (0.19/1)). Also, SpliceAI predicted no pseudo-exon inclusion for the *PCDH15* variant. Therefore, the *ARSG* variant was considered the strongest candidate in a gene associated with USH. Manual inspection of the sequencing reads (CRAM files) revealed indications for a deletion spanning exons 7 and 8 of *ARSG*. PCR and Sanger sequencing confirmed the deletion and breakpoints: NM_014960.5:c.705-3940_982+2952del NP_055775.2:p.(Ser235Argfs*29). PCR with a reverse primer designed to bind nucleotides deleted in the c.705-3940_982+2952del allele demonstrated that the c.588C>A p.(Tyr196*) variant and the deletion are in trans.

The identified *ARSG* variants were either not detected or detected at low frequencies (<0.01%) in gnomAD (Table 1). The previously described variant (c.1326del p.(Ser443Alafs*12)) was classified as 'pathogenic' in the ClinVar database (Landrum et al. 2020) and according to the American College of Medical Genetics and Genomics (ACMG) Standards and Guidelines and its specifications for hearing loss (Table 1) (Abou Tayoun et al. 2018; Oza et al. 2018; Richards et al. 2015). Three novel variants were classified as 'pathogenic' and two as 'likely pathogenic' according to these guidelines. The identified missense variants affect amino acids within protein domains and that have been highly conserved throughout evolution (Fig. 2) (Mistry et al. 2021).

Table 1 Overview of *ARSG* variants of recruited subjects

Variant GRCh37 (hg19) NM_014960.5	Effect on RNA	Effect on protein NP_055775.2	GnomAD minor allele frequency (%)	Variant classification[a]	CADD_ PHRED	Grantham score	Mutation taster	PhyloP	SIFT
Subject N									
Chr17:g.66391335G > A c.1212 + 1G > A	r.1212_1213ins[a;1212 + 2_ 1212 + 13]	p.(Val405Ilefs*41)	–	Pathogenic (PVS1-S, PS3-M, PM2, PM3, PP3, PP4)	34	N/a	N/a	6.163	N/a
Chr17:g.66339801T > C c.275T > C	r.(275u > c)	p.(Leu92Pro)	0.000398 / 0.008700 (Southern Europe)	Likely pathogenic (PS3-M, PM2, PM3, PP3, PP4)	27.5	98	Disease causing	6.704	Deleterious
Subject F									
Chr17:g.66416352del c.1326del	r.(1326del)	p.(Ser443Alafs*12)	0.007781 / 0.025860 (Southern Europe)	Pathogenic (PVS1, PS3-M, PM2-Supp, PM3-Supp, PP4)	24.4	N/a	N/a	N/a	N/a
Chr17:g.66381246C > T c.1024C > T	r.1024c > u	p.(Arg342Trp)	0.000796 / 0.009930 (African/ African American XX)	Likely pathogenic (PS3-M, PM2, PM3-Supp, PP3, PP4)	32	101	Disease causing	1.654	Deleterious
Subject D									
Chr17:g.66352829C > A c.588C > A	r.(588c > a)	p.(Tyr196*)	0.000396 / 0.002370 (Northwestern Europe)	Pathogenic (PVS1-S, PS3-M, PM2, PM3, PP4)	36	N/a	N/a	1.532	N/a
Chr17:g.66360749_66369617del c.705-3940_ 982 + 2952del	r.(705_982del)	p.(Ser235Argfs*29)	-	Pathogenic (PVS1-S, PS3-M, PM2, PM3, PP4)	N/a	N/a	N/a	N/a	N/a

[a]Variant classification was performed using the American College of Medical Genetics and Genomics Standards and Guidelines (Abou Tayoun et al. 2018; Oza et al. 2018; Richards et al. 2015); CADD_PHRED, Combined Annotation Dependent Depletion, threshold value ≥ 15 (Kircher et al. 2014); ClinVar database (Landrum et al. 2020); GnomAD, Genome Aggregation Database, total population exome frequency/maximum exome frequency (corresponding population); Grantham score, threshold value ≥ 80 (Grantham 1974); MutationTaster (Schwarz et al. 2014); PhyloP, threshold value ≥ 2.7 (Pollard et al. 2010); SIFT, Sorting Intolerant From Tolerant (Ng and Henikoff 2001); N/a, not applicable; -, not present in database

Functional analyses of identified variants

The minigene splice assay for variant c.1212 + 1G > A (subject N), showed that the original splice donor site was lost, and that an alternative splice donor site at + 13nt in intron 10 was used. This resulted in an out of frame elongation of exon 10 and a premature stop codon (r.1212_1213ins[a;1212 + 2_1212 + 13] p.(Val405Ilefs*41) (Fig. 3). Stable expression in HT1080 fibrosarcoma cells of a construct bearing the second *ARSG* variant of subject N (c.275T > C p.(Leu92Pro)), followed by immunoblot analysis, revealed a modestly increased ARSG expression level compared to the WT construct (Fig. 4). An additional ARSG-specific band was detected for the p.(Leu92Pro) construct migrating at a higher apparent molecular weight. ARSG is extensively post-translationally modified by N-glycosylation and limited proteolysis (Kowalewski et al. 2014) and it could be speculated that the posttranslational modifications of this polypeptide differ, e.g., due retention in the endoplasmic reticulum. Notably, a minor band detected for the WT, most likely representing a proteolytically processed form migrating below the full-length protein, was nearly absent in the lane for the p.(Leu92Pro) construct. Sulfatase activity assays revealed a significant activity increase for the artificial substrate pNCS in cells stably expressing WT ARSG compared to untransfected

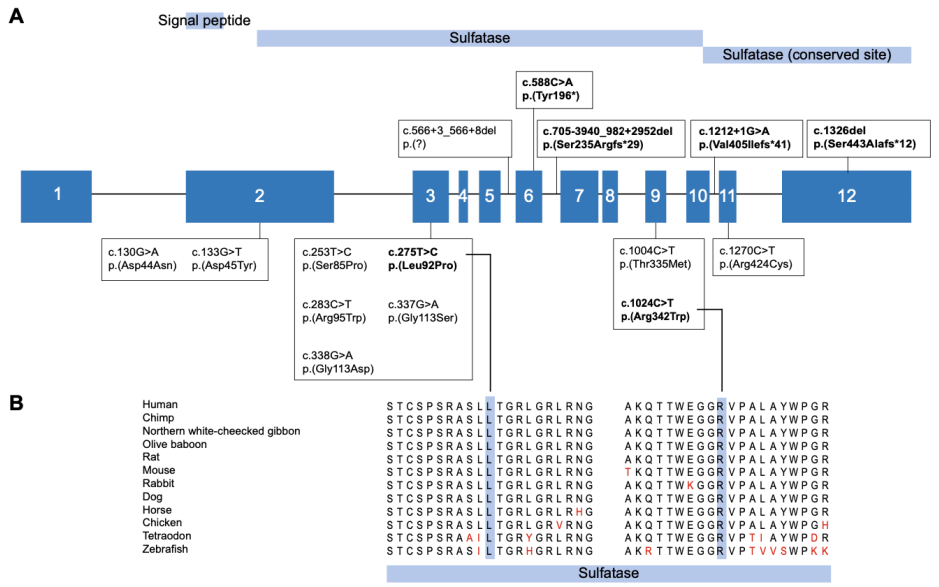


Fig. 2 *ARSG* variants. **A** Schematic representation of the variability in location and type of novel and previously reported *ARSG* variants. The twelve *ARSG* exons are shown in dark blue squares. Protein domains are shown in light blue bars (Mistry et al. 2021). Variants that were identified in the present study are shown in bold. Truncating variants are presented above the exons, missense variants below. **B** Multiple sequence alignments of the sequence regions containing the two novel missense variants in 12 orthologs (Alamut Visual v.2.13). One-letter amino acid abbreviations are presented. Amino acids that do not correspond with the most conserved amino acid at their position are shown in red. The position of the variants are highlighted in blue. Both variants are located within the sulfatase protein domain

cells. There was also a minimal increase in total sulfatase activity observed in cells expressing the p.(Leu92Pro) construct variant compared to untransfected cells, which, however, can likely be explained by the slightly higher expression (Fig. 4). This analysis indicates that the amino acid substitution presumably results in a nearly complete loss of enzyme activity.

Constructs encoding the truncated c.1326del p.(Ser443Alafs*12) ARSG (subject F) previously revealed similar ARSG expression levels as found in constructs encoding WT ARSG, but with a slightly lower molecular weight, in line with a truncated protein. Also, a significantly lower sulfatase activity was observed than for WT ARSG (Peter et al. 2020). These findings are confirmed in the present study in which we also found similar expression levels of the mutant construct as compared to WT ARSG with a slightly lower molecular weight, as well as significantly lower sulfatase activity in cells expressing the p.(Ser443Alafs*12) variant than in cells expressing WT ARSG (Fig. 4). Immunoblot analysis of ARSG with the amino acid substitution c.1024C > T p.(Arg342Trp) revealed that the protein is expressed at similar levels when compared to WT ARSG. Also, sulfatase activity showed no increase in the cells expressing the mutant ARSG compared to untransfected cells, indicating a complete loss of enzyme activity of the mutant ARSG. The minigene splice assay for this latter variant (c.1024C > T) revealed no effect on splicing (r.1024c > u p.(Arg342Trp)) (Fig. 3).

Finally, immunoblot analysis of the two truncated ARSG proteins c.588C > A p.(Tyr196*) and c.705-3940_982 + 2952del p.(Ser235Argfs*29) identified in subject D confirmed the truncated proteins to be of lower molecular weight than the WT ARSG, within their respective expected ranges (Fig. 4). Slightly lower signal intensities were observed for both mutant constructs compared to the WT construct, indicating lower expression levels of the mutant ARSG proteins. High-molecular weight bands were observed for both mutant constructs, indicating possibly SDS-resistant dimers of the truncated proteins. Determination of the sulfatase activity of cells expressing either mutant ARSG showed no increase of pNCS turnover compared to untransfected cells, indicating a complete loss of enzyme activity.

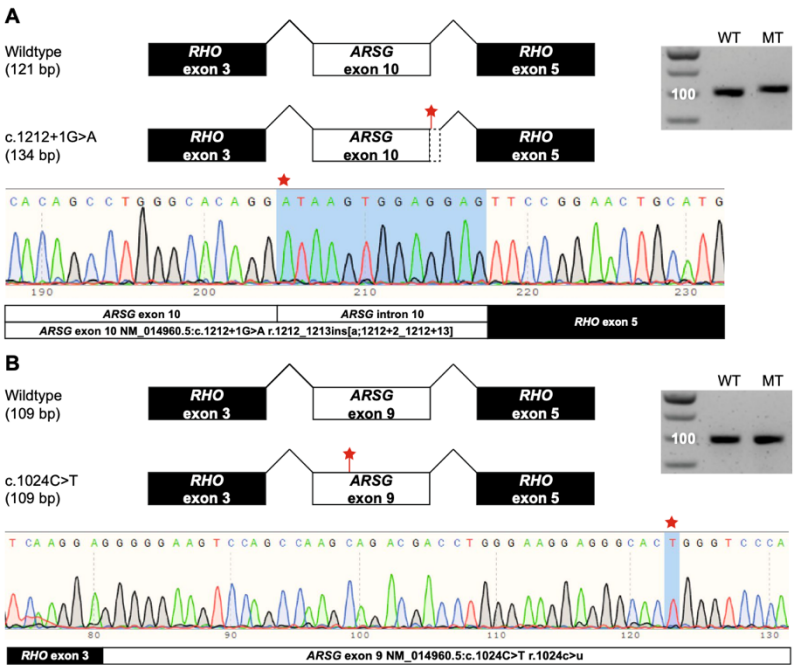


Fig. 3 Results of minigene splice assays for *ARSG* variants c.1212+1G>A and c.1024C>T. In vitro splice assays were performed in HEK293T cells to validate the predicted splice defects. **A** A canonical splice site variant (c.1212+1G>A) was detected in *ARSG* (subject N) and was predicted by SpliceAI to cause skipping of exon 10 due to a donor splice site loss at −1nt (0.99/1) and acceptor splice site loss at −121nt (0.78/1). A minigene splice assay revealed the use of an alternative splice donor site at+13nt in intron 10 which leads to an out of frame elongation of exon 10 (NM_014960.5:c.1212+1G>A r.1212_1213ins[a;1212+2_1212+13] NP_055775.2:p.(Val405Ilefs*41)). **B** A missense variant (c.1024C>T) was predicted by SpliceAI to cause skipping of exon 9 due to a donor splice site loss at+67nt (0.15/1) and acceptor splice site loss at −41nt (0.25/1). A minigene splice assay did not confirm this and demonstrated no effect on splicing (NM_014960.5:c.1024C>T r.1024c>u NP_055775.2:p.(Arg342Trp)). *Bp* base pair; *WT* wildtype; *MT* mutant

Discussion

The present study confirms the recent introduction of USH type IV as being distinctive from the other three types of USH. We describe three subjects with an USH type IV phenotype caused by *ARSG* variants. This phenotype consists of late-onset RP and late-onset progressive SNHL. The retinal phenotype is characterized by a specific distribution of atrophy around the arcades, peripherally surrounded by intraretinal bone spicules best appreciated nasally and superior to the disc. There is progressive macular involvement eventually resulting in complete loss of the outer retina. There are no signs of vestibular dysfunction, although accurate testing to entirely exclude vestibular involvement could not be performed. In the three studied subjects, we identified six *ARSG* variants: one previously reported pathogenic single nucleotide deletion, three novel pathogenic variants, including the first reported structural *ARSG* variant, and two novel likely pathogenic missense variants (Table 1). Five identified variants were shown to cause loss of enzyme activity, in line with previous research (Khateb et al. 2018; Peter et al. 2020). For one variant, sulfatase was not tested, but this variant (c.1212+1G>A p.(Val405Ilefs*41)) is likely to result in loss of sulfatase activity since it leads to premature protein truncation. This is supported by the observation that the c.1212+1G>A p.(Val405Ilefs*41) variant occurs 38 amino acids N-terminal of the c.1326del p.(Ser443Alafs*12) variant, for which significantly reduced sulfatase activity was found in the current study (Fig. 4).

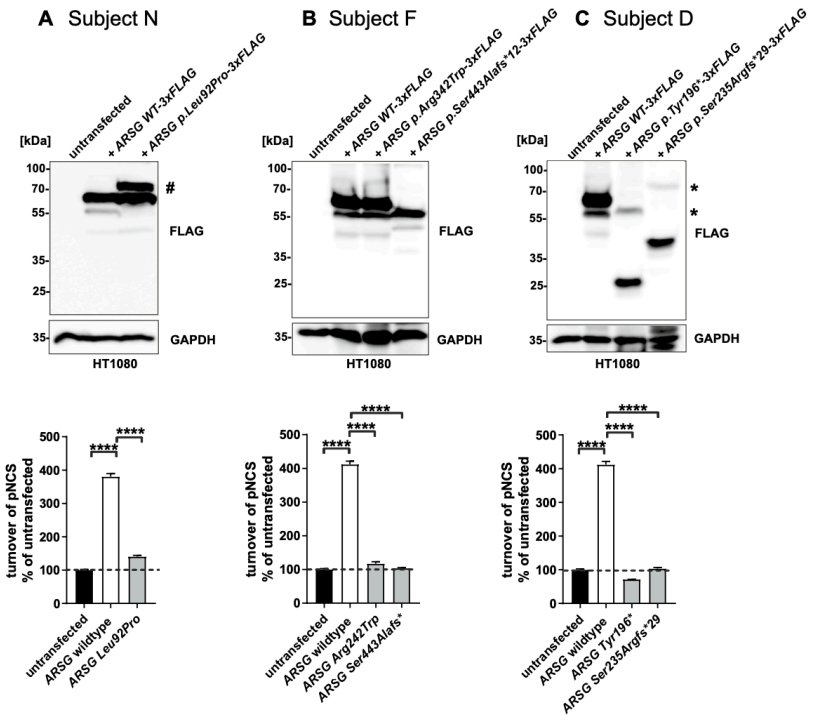


Fig. 4 Functional analysis of *ARSG* variants identified in subjects N, F, and D. **A** Subject N: Immunoblot analysis of untransfected or stably transfected HT1080 cells with plasmids coding for 3xFLAG-tagged wildtype (WT) ARSG, along with the ARSG variant p.(Leu92Pro). An additional ARSG p.(Leu92Pro)-specific band is labeled with an #. Other bands in the figure are likely to be caused by limited proteolytic cleavage of ARSG. Total sulfatase activity of cell lysates from untransfected cells, or transfected cells expressing WT ARSG or ARSG with the pathogenic variant p.(Leu92Pro) against the artificial substrate P-nitrocatechol sulfate (pNCS). $N=3$ replicates. Unpaired, two-tailed *t*-test, Mean ± SEM of the replicates. **B** Subject F: Immunoblot analysis of untransfected or stably transfected HT1080 cells with plasmids coding for 3xFLAG-tagged WT ARSG, along with the pathogenic variants p.(Arg342trp) and p.(Ser443Alafs*12). Other bands in the figure are likely to be caused by limited proteolytic cleavage of ARSG. Total sulfatase activity of cell lysates from untransfected cells, or from cells expressing WT ARSG or the pathogenic variants p.(Arg342Trp) and p.(Ser443Alafs*12) against the artificial substrate pNCS. $N=3$ replicates. Unpaired, two-tailed *t*-test, Mean ± SEM of the replicates. **C** Subject D: Immunoblot analysis of untransfected or stably transfected HT1080 cells with plasmids coding for 3xFLAG-tagged WT ARSG, along with the pathogenic variants p.(Tyr196*) and p.(Ser235Argfs*29). High-molecular weight bands indicating possibly SDS-resistant dimers of the truncated proteins are labeled with an *. Other bands in the figure are likely to be caused by limited proteolytic cleavage of ARSG. Total sulfatase activity of cell lysates from untransfected cells, or cells expressing WT ARSG or the pathogenic variants p.(Tyr196*) and p.(Ser235Argfs*29) against the artificial substrate pNCS. $N=3$ replicates. Unpaired, two-tailed *t*-test, Mean ± SEM of the replicates. Antibodies in all panels are against FLAG and GAPDH, used as a loading control. pNCS, P-nitrocatechol sulfate; SEM, standard error of the mean

The lysosomal sulfatase ARSG is ubiquitously expressed in mice and has been detected the retina and the cochlea (Girotto et al. 2014; Kruszewski et al. 2016). An *Arsg* knockout (KO) mouse was shown to accumulate heparan sulfate in enlarged lysosomes and develop typical mucopolysaccharidosis-like features (Kowalewski et al. 2012). An early feature of the *Arsg* KO model is progressive photoreceptor degeneration in the retina and dysregulation of various lysosomal proteins, however, with no clear signs of lysosomal storage defects in retinal cells. Generally, these *Arsg* KO mice present with a milder phenotype than other mouse models of disorders resulting from impaired lysosomal degradation

of heparan sulfate (Kruszewski et al. 2016). Immunohistochemistry of murine retina revealed ARSG expression in the retinal pigment epithelium (Kruszewski et al. 2016). A homozygous missense loss of function *ARSG* variant in a canine animal model resulted in sulfatase deficiency and neuronal ceroid lipofuscinosis, however, without affecting the retina (Abitbol et al. 2010). In mouse cochlea, ARSG was demonstrated to be located at the apical side of sensory hair cells in the organ of Corti (Girotto et al. 2014). In human, *ARSG* mRNA expression levels were the highest in the liver, kidney, pancreas and brain. Retina and cochlea were not included in these analyses (Ferrante et al. 2002; Frese et al. 2008). ARSG is ubiquitously expressed with the highest expression levels in the brain (Carithers et al. 2015; DeLuca et al. 2012; GTEx Consortium 2013, 2020; Melé et al. 2015). A major open question is how variants in *ARSG* cause USH type IV, and in particular how the impaired degradation of 3-*O*-modified heparan sulfate is causing cellular dysfunction and cell death despite no apparent lysosomal enlargement. Moreover, it is surprising that none of the patients reported so far show any signs of lysosomal storage defects in peripheral organs or central nervous system (CNS) involvement, as would be expected from the phenotype of the KO mice and ARSG-deficient dogs. Remarkably, a similar situation has been described for other genes causative for lysosomal storage diseases. While variants in genes associated with neuronal ceroid lipofuscinosis (*CLN3* (OMIM 607042), *CLN5* (OMIM 608102), and *MFSD8*/*CLN7* (OMIM611124)) typically lead to severe neuronal dysfunction and lysosomal storage in neurons and the peripheral tissues, some variants present with isolated retina disease (Bauwens et al. 2020; Khan et al. 2017; Ku et al. 2017; Magliyah et al. 2021). Similarly, variants in *HGSNAT* (OMIM 610453) are associated with mucopolysaccharidosis in the vast majority of patients, but isolated retina degeneration has also been described (Haer-Wigman et al. 2015). One could hypothesize that *ARSG* variants may also exhibit a broad spectrum of disease as well as isolated retina and cochlear defects. However, the variants we identified are all loss of function variants and lead to phenotypes of RP and SNHL, without neurological symptoms or signs of generalized clinical manifestations of lysosomal storage disease. Therefore, it is not expected that other variants would cause a more broad phenotype.

Comparing the twelve previously described cases supplemented with the presently identified subjects, we observed a fairly similar phenotype (Supplemental Table 3, 4) (Abad-Morales et al. 2020; Fowler et al. 2021; Igelman et al. 2021; Khateb et al. 2018; Peter et al. 2020). All individuals were diagnosed with RP with a midlife age of onset (35–60 years), which is later than generally seen in USH type I (first decade) (Tsilou et al. 2002; Van Camp and Smith 2020), type II (first or second decade) (Millan et al. 2011; Tsilou et al. 2002), or type III (variable but usually begins in the second decade) (Nisenbaum et al. 2021) (Fig. 1, Table 2). RP in the identified subjects is characterized by onset at the age of 40–60 years with unique but similar funduscopic changes. Funduscopic abnormalities were primarily seen around the arcades on initial examination. Atrophy was seen mainly around the arcades and extended into the mid-periphery with intraretinal bone-spicule hyperpigmentation adjacent to the atrophic border, most evident nasally and superior to the disc. The far periphery appeared normal. Fundus autofluorescence was consistent with the clinically visible changes and showed a typical pattern of hypoautofluorescence in the mid-periphery but with normal fundus autofluorescence in the far periphery. With age, the macula became more involved, resulting in complete loss of the outer retina in the fovea, as seen on OCT, leading to light perception vision. ERGs revealed lower scotopic than photopic responses, but eventually showed complete absence of both types of responses. All subjects identified in the present study and all but one previously described individuals suffered from progressive SNHL with a self-reported onset ranging from childhood to 50 years of age and an official diagnosis or the start of hearing aid use ranging from 18 to 67 years. Based on the PTA data, we calculated an age of onset of 17 years. Presumably, patients do not experience symptoms of their hearing loss until a more significant loss than the WHO cut-off value of 25 dB HL and thus have their hearing tested at a later age (Informal Working Group on Prevention of Deafness and Hearing Impairment Programme Planning & World Health Organization 1991). One individual reported no symptoms of SNHL. However, since no audiometric examination had been performed at his last visit at the age of 48, it could be that he was not yet aware of the possibly increased hearing thresholds or that this will manifest later in life (Igelman et al. 2021). The reported and calculated age of onset is obviously later than for USH types I and II (congenital), but also later than for type III with an age of onset of SNHL usually in the first decade of life (Fig. 1, Table 2) (Geng et al. 2017). Although, due to the limited age range of the subjects, we could only determine the age of onset of SNHL for USH type IV by interpolating the data. When comparing the USH IV ARTA with that of USH type Ib (Wagenaar et al. 1999), IIa (Hartel et al. 2016), and III (Plantinga et al. 2004) (Fig. 1), the difference in the severity of SNHL with USH Ib and the higher rate of progression compared to USH IIa are also notable.

It is noteworthy that two subjects (N and F) had a medical history of hypothyroidism, which was not reported in previous *ARSG* cases. Identification of larger numbers of USH type IV cases may answer the question on whether hypothyroidism is associated with *ARSG* defects.

Recent USH type IV reports expand the information on the variability of USH phenotypes. This aids in accurate

Table 2 Suggested extension of clinical classification of Usher syndrome

	Hearing impairment	Visual impairment	Vestibular impairment
USH I	Congenital (Nisenbaum et al. 2021) Severe to profound (Nisenbaum et al. 2021) Stable	RP onset in first decade of life (Nisenbaum et al. 2021) Mean age of perceived night blindness: 10y (SD: 6.6) (Tsilou et al. 2002)	Severe (Nisenbaum et al. 2021)
USH II	Congenital (Nisenbaum et al. 2021) Moderate to severe (Nisenbaum et al. 2021) Variable progression (Stemerdink et al. 2021)	RP onset in second decade of life or later (Millan et al. 2011) Mean age of perceived night blindness: 15y (SD: 6.9) (Tsilou et al. 2002)	Uncertain (Stemerdink et al. 2021)
USH III	Variable age of onset, usually in first decade of life, range: childhood-35y (Geng et al. 2017) Progressive (Nisenbaum et al. 2021)	Variable RP onset, typically begin by second decade of life (Nisenbaum et al. 2021)	Variable (Nisenbaum et al. 2021)
USH IV	Range of self-reported age of onset: childhood-50y Range of official SNHL diagnosis or start of using hearing aids: 18–67y Moderate to severe	Range of RP onset: 30–60y	Uncertain (no patient-reported symptoms)

Classification of USH type I–III based on the original classification (Smith et al. 1994; Davenport and Omenn 1977), modified and expanded in response to recent findings. *y* years

patient counseling, a recurring challenge when patients are diagnosed with newly found or rarely described genotypes. In the literature, this considerable variability is evident in several reports of atypical USH phenotypes. First, some genes may be causative for both USH and isolated deafness or RP, for example *CDH23* (OMIM 605516, USH and isolated SNHL OMIM 601386 (Bork et al. 2001)). For that matter, there is no evidence of *ARSG* to be associated with non-syndromic RP or SNHL, neither in our in-house database (containing exome sequencing data of 2086 individuals analyzed for vision disorders and 2300 individuals analyzed for hearing impairment) nor in the literature to the best of our knowledge. Second, there are cases with variants in known USH-associated genes and phenotypes that differ from the corresponding type of USH, such as an USH II phenotype caused by *CDH23* variants or a phenotype with vestibular dysfunction caused by *USH2A* variants (Liu et al. 1999; Valero et al. 2019).

In conclusion, our observations confirm the newly described USH type IV caused by *ARSG* variants (Table 2). This type differs from the other USH types in the late onset of RP and SNHL. The retinal phenotype is characterized by a specific distribution of atrophy around the arcades, peripherally surrounded by intraretinal bone spicules best appreciated nasally and superior to the optic disc. There is progressive macular involvement, eventually resulting in complete loss of the central outer retina. Contrary to atypical USH, which we believe is an aberrant phenotype that results from variants in one of the previously identified USH genes, the *ARSG* phenotype is consistent and distinct from the original USH types I, II, and III. This study contributes to the determination of genotype–phenotype associations that are important not only for patient counseling, but also given the development of therapeutic strategies.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00439-022-02441-0.

Acknowledgements We acknowledge Dr. Jordi Corominas Galbany, Dr. Galuh DN Astuti, and Prof. Dr. Christian Gilissen for the provision of infrastructure and bioinformatic expert input. Also, we thank Prof. Dr. Caroline Klaver for her contribution to the review and description of the ophthalmic data.

Author contributions HMV, HK, CPL, MD, and RJEP contributed to the conception and design of the study. HMV, JR, CHZL, JW, LP, CL, LW, AD, JZ, DJK, and GJF were involved in acquisition of data. Analysis and interpretation of the data was performed by HMV, JR, SH, CHZL, SY, JO, JW, LHW, HGY, SR, HK, CPL, MD, and RJEP. All authors were involved in either drafting the manuscript (HMV, JR, CHZL, JO, MD) or in its critical revision.

Funding This study was supported by a grant from the Heinsius-Houbolt foundation (to HK and RJEP), the VELUX Stiftung (to HK and SR), and Fighting Blindness Ireland, the Health Research Board of Ireland, and Health Research Charities Ireland (to GJF).

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information. The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Ethics approval Written informed consent for participation and publication was received from all individuals adherent to the tenets of the Declaration of Helsinki, and the study was approved by the local ethics committee of each participating center.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abad-Morales V, Navarro R, Bures-Jelstrup A, Pomares E (2020) Identification of a novel homozygous ARSG mutation as the second cause of Usher syndrome type 4. Am J Ophthalmol Case Rep. 19:100736

Abitbol M, Thibaud JL, Olby NJ, Hitte C, Puech JP, Maurer M et al (2010) A canine arylsulfatase G (ARSG) mutation leading to a sulfatase deficiency is associated with neuronal ceroid lipofuscinosis. Proc Natl Acad Sci USA 107(33):14775–14780

Abou Tayoun AN, Pesaran T, DiStefano MT, Oza A, Rehm HL, Biesecker LG et al (2018) Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. Hum Mutat 39(11):1517–1524

Ahmed ZM, Riazuddin S, Bernstein SL, Ahmed Z, Khan S, Griffith AJ et al (2001) Mutations of the protocadherin gene PCDH15 cause Usher syndrome type 1F. Am J Hum Genet 69(1):25–34

Ahmed ZM, Riazuddin S, Ahmad J, Bernstein SL, Guo Y, Sabar MF et al (2003) PCDH15 is expressed in the neurosensory epithelium of the eye and ear and mutant alleles are responsible for both USH1F and DFNB23. Hum Mol Genet 12(24):3215–3223

Alagramam KN, Yuan H, Kuehn MH, Murcia CL, Wayne S, Srisailpathy CR et al (2001) Mutations in the novel protocadherin PCDH15 cause Usher syndrome type 1F. Hum Mol Genet 10(16):1709–1718

Bauwens M, Storch S, Weisschuh N, Ceuterick-de Groote C, De Rycke R, Guillemyn B et al (2020) Functional characterization of novel MFSD8 pathogenic variants anticipates neurological involvement in juvenile isolated maculopathy. Clin Genet 97(3):426–436

Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G et al (2012) Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics 28(3):423–425

Bolz H, von Brederlow B, Ramírez A, Bryda EC, Kutsche K, Nothwang HG et al (2001) Mutation of CDH23, encoding a new member of the cadherin gene family, causes Usher syndrome type 1D. Nat Genet 27(1):108–112

Booth KT, Azaiez H, Kahrizi K, Simpson AC, Tollefson WT, Sloan CM et al (2015) PDZD7 and hearing loss: more than just a modifier. Am J Med Genet A 167(12):2957–2965

Booth KT, Kahrizi K, Babanejad M, Daghagh H, Bademci G, Arzhangi S et al (2018) Variants in CIB2 cause DFNB48 and not USH1J. Clin Genet 93(4):812–821

Bork JM, Peters LM, Riazuddin S, Bernstein SL, Ahmed ZM, Ness SL et al (2001) Usher syndrome 1D and nonsyndromic autosomal recessive deafness DFNB12 are caused by allelic mutations of the novel cadherin-like gene CDH23. Am J Hum Genet 68(1):26–37

Boughman JA, Vernon M, Shaver KA (1983) Usher syndrome: definition and estimate of prevalence from two high-risk population. J Chronische Dis 36(8):595–603

Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA et al (2015) A novel approach to high-quality postmortem tissue procurement: the GTEx project. Biopreserv Biobank 13(5):311–319

Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M et al (2016) Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 32(8):1220–1222

Davenport SHL, Omenn GS (1977) The heterogeneity of Usher syndrome. In: Fifth international conference on birth defects. Excerpta Medica, Amsterdam, pp 87–88

DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C et al (2012) RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics 28(11):1530–1532

Ebermann I, Scholl HP, Charbel Issa P, Becirovic E, Lamprecht J, Jurklies B et al (2007) A novel gene for Usher syndrome type 2: mutations in the long isoform of whirlin are associated with retinitis pigmentosa and sensorineural hearing loss. Hum Genet 121(2):203–211

Ebermann I, Phillips JB, Liebau MC, Koenekoop RK, Schermer B, Lopez I et al (2010) PDZD7 is a modifier of retinal disease and a contributor to digenic Usher syndrome. J Clin Invest 120(6):1812–1823

Eudy JD, Weston MD, Yao S, Hoover DM, Rehm HL, Ma-Edmonds M et al (1998) Mutation of a gene encoding a protein with extracellular matrix motifs in Usher syndrome type IIa. Science 280(5370):1753–1757

Fadaie Z, Whelan L, Dockery A, Li CHZ, van den Born LI, Hoyng CB, et al (2021) BBS1 branchpoint variant is associated with non-syndromic retinitis pigmentosa. J Med Genet

Ferrante P, Messali S, Meroni G, Ballabio A (2002) Molecular and biochemical characterisation of a novel sulphatase gene: Arylsulfatase G (ARSG). Eur J Hum Genet 10(12):813–818

Fowler NH, El-Rashedy MI, Chishti EA, Vander Kooi CW, Maldonado RS (2021) Multimodal imaging and genetic findings in a case of ARSG-related atypical Usher syndrome. Ophthalmic Genet 42(3):338–343

Frese MA, Schulz S, Dierks T (2008) Arylsulfatase G, a novel lysosomal sulfatase. J Biol Chem 283(17):11388–11395

Gao J, Wang Q, Dong C, Chen S, Qi Y, Liu Y (2015) Whole exome sequencing identified MCM2 as a novel causative gene for autosomal dominant nonsyndromic deafness in a Chinese family. PLoS ONE 10(7):e0133522

Geng R, Omar A, Gopal SR, Chen DH, Stepanyan R, Basch ML et al (2017) Modeling and preventing progressive hearing loss in Usher Syndrome III. Sci Rep 7(1):13480

Girotto G, Vuckovic D, Buniello A, Lorente-Canovas B, Lewis M, Gasparini P et al (2014) Expression and replication studies to identify new candidate genes involved in normal hearing function. PLoS ONE 9(1):e85352

Grantham R (1974) Amino acid difference formula to help explain protein evolution. Science 185(4154):862–864

Grondahl J (1987) Estimation of prognosis and prevalence of retinitis pigmentosa and Usher syndrome in Norway. Clin Genet 31(4):255–264

GTEx Consortium (2013) The genotype-tissue expression (GTEx) project. Nat Genet 45(6):580–585

GTEx Consortium (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369(6509):1318–1330

Haer-Wigman L, Newman H, Leibu R, Bax NM, Baris HN, Rizel L et al (2015) Non-syndromic retinitis pigmentosa due to mutations in the mucopolysaccharidosis type IIIC gene, heparan-alpha-glucosaminide N-acetyltransferase (HGSNAT). Hum Mol Genet 24(13):3742–3751

Haer-Wigman L, van Zelst-Stams WA, Pfundt R, van den Born LI, Klaver CC, Verheij JB et al (2017) Diagnostic exome sequencing in 266 Dutch patients with visual impairment. Eur J Hum Genet 25(5):591–599

Hallgren B (1959) Retinitis pigmentosa combined with congenital deafness: with vestibulo-cerebellar ataxia and mental abnormality in a proportion of cases. Acta Psychiatr Scand Suppl 34(138):1–101

Hartel BP, Löfgren M, Huygen PL, Guchelaar I, Lo ANKN, Sadeghi AM et al (2016) A combination of two truncating mutations in USH2A causes more severe and progressive hearing impairment in Usher syndrome type IIa. Hear Res 339:60–68

Hope CI, Bundey S, Proops D, Fielder AR (1997) Usher syndrome in the city of Birmingham-prevalence and clinical classification. Br J Opthalmol 81(1):46–53

Huygen PL, Pennings RJ, Cremers CW (2003) Characterizing and distinguishing progressive phenotypes in nonsyndromic autosomal dominant hearing impairment. Audiological Med 1(1):37–46

Igelman AD, Ku C, da Palma MM, Georgiou M, Schiff ER, Lam BL et al (2021) Expanding the clinical phenotype in patients with disease causing variants associated with atypical Usher syndrome. Ophthalmic Genet 42(6):664–673

Informal Working Group on Prevention of Deafness and Hearing Impairment Programme Planning & World Health Organization (1991) Report of the Informal Working Group on prevention of deafness and hearing impairment programme planning. World Health Organization, Geneva. Available from: https://apps.who.int/iris/handle/10665/58839

Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Yi Li et al (2019) Predicting splicing from primary sequence with deep learning. Cell 176(3):535–548

Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q et al (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581(7809):434–443

Khan KN, El-Asrag ME, Ku CA, Holder GE, McKibbin M, Arno G et al (2017) Specific alleles of CLN7/MFSD8, a protein that localizes to photoreceptor synaptic terminals, cause a spectrum of nonsyndromic retinal dystrophy. Invest Ophthalmol vis Sci 58(7):2906–2914

Khateb S, Kowalewski B, Bedoni N, Damme M, Pollack N, Saada A et al (2018) A homozygous founder missense variant in arylsulfatase G abolishes its enzymatic activity causing atypical Usher syndrome in humans. Genet Med 20(9):1004–1012

Kimberling WJ, Hildebrand MS, Shearer AE, Jensen ML, Halder JA, Trzupek K et al (2010) Frequency of Usher syndrome in two pediatric populations: Implications for genetic screening of deaf and hard of hearing children. Genet Med 12(8):512–516

Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J (2014) A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 46(3):310–315

Kowalewski B, Lamanna WC, Lawrence R, Damme M, Stroobants S, Padva M et al (2012) Arylsulfatase G inactivation causes loss of heparan sulfate 3-O-sulfatase activity and mucopolysaccharidosis in mice. Proc Natl Acad Sci USA 109(26):10310–10315

Kowalewski B, Lubke T, Kollmann K, Braulke T, Reinheckel T, Dierks T et al (2014) Molecular characterization of arylsulfatase G: expression, processing, glycosylation, transport, and activity. J Biol Chem 289(40):27992–28005

Kruszewski K, Lullmann-Rauch R, Dierks T, Bartsch U, Damme M (2016) Degeneration of photoreceptor cells in arylsulfatase G-deficient mice. Invest Ophthalmol vis Sci 57(3):1120–1131

Ku CA, Hull S, Arno G, Vincent A, Carss K, Kayton R et al (2017) Detailed clinical phenotype and molecular genetic findings in CLN3-associated isolated retinal degeneration. JAMA Ophthalmol 135(7):749–760

Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J et al (2020) ClinVar: improvements to accessing data. Nucleic Acids Res 48(D1):D835–D844

Liu XZ, Walsh J, Mburu P, Kendrick-Jones J, Cope MJ, Steel KP et al (1997a) Mutations in the myosin VIIA gene cause non-syndromic recessive deafness. Nat Genet 16(2):188–190

Liu XZ, Walsh J, Tamagawa Y, Kitamura K, Nishizawa M, Steel KP et al (1997b) Autosomal dominant non-syndromic deafness caused by a mutation in the myosin VIIA gene. Nat Genet 17(3):268–269

Liu XZ, Hope C, Liang CY, Zou JM, Xu LR, Cole T et al (1999) A mutation (2314delG) in the Usher syndrome type IIA gene: high prevalence and phenotypic variation. Am J Hum Genet 64(4):1221–1225

Liu Y, Chen X, Xu Q, Gao X, Tam PO, Zhao K et al (2015) SPP2 mutations cause autosomal dominant retinitis pigmentosa. Sci Rep 5:14867

Magliulo G, Iannella G, Gagliardi S, Iozzo N, Plateroti R, Mariottini A et al (2017) Usher's syndrome type II: a comparative study of genetic mutations and vestibular system evaluation. Otolaryngol Head Neck Surg 157(5):853–860

Magliyah MS, Geuer S, Alsalamah AK, Lenzner S, Drasdo M, Schatz P (2021) Association of the recurrent rare variant c415T>C pPhe139Leu in CLN5 with a recessively inherited macular dystrophy. JAMA Ophthalmol 139(3):339–343

Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M et al (2015) Human genomics. The human transcriptome across tissues and individuals. Science 348(6235):660–665

Millan JM, Aller E, Jaijo T, Blanco-Kelly F, Gimenez-Pardo A, Ayuso C (2011) An update on the genetics of usher syndrome. J Ophthalmol 2011:417217

Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL et al (2021) Pfam: the protein families database in 2021. Nucleic Acids Res 49(D1):D412–D419

Ng PC, Henikoff S (2001) Predicting deleterious amino acid substitutions. Genome Res 11(5):863–874

Nisenbaum E, Thielhelm TP, Nourbakhsh A, Yan D, Blanton SH, Shu Y et al (2021) Review of Genotype-Phenotype correlations in Usher syndrome. Ear Hear 43:1–8

Oza AM, DiStefano MT, Hemphill SE, Cushman BJ, Grant AR, Siegert RK et al (2018) Expert specification of the ACMG/AMP variant interpretation guidelines for genetic hearing loss. Hum Mutat 39(11):1593–1613

Pauw RJ, Huygen PL, Colditz GM, Cremers CW (2011) Phenotype analysis of an Australian DFNA9 family with the 1109N COCH mutation. Ann Otol Rhinol Laryngol 120(6):414–421

Peter VG, Quinodoz M, Sadio S, Held S, Rodrigues M, Soares M et al (2020) New clinical and molecular evidence linking mutations in ARSG to Usher syndrome type IV. Hum Mutat 42(3):261–271

Plantinga RF, Kleemola L, Huygen PL, Joensuu T, Sankila EM, Pennings RJ et al (2004) Serial audiometry and speech recognition findings in Finnish Usher syndrome type III patients. Audiol Neurootol 10(2):79–89

Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS et al (2008) Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat 29(11):1282–1291

Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A (2010) Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res 20(1):110–121

Riazuddin S, Belyantseva IA, Giese AP, Lee K, Indzhykulian AA, Nandamuri SP et al (2012) Alterations of the CIB2 calcium- and

integrin-binding protein cause Usher syndrome type 1J and non-syndromic deafness DFNB48. Nat Genet 44(11):1265–1271
Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J et al (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17(5):405–424
Rivolta C, Sweklo EA, Berson EL, Dryja TP (2000) Missense mutation in the USH2A gene: association with recessive retinitis pigmentosa without hearing loss. Am J Hum Genet 66(6):1975–1978
Rosenberg T, Haim M, Hauch AM, Parving A (1997) The prevalence of Usher syndrome and other retinal dystrophy-hearing impairment associations. Clin Genet 51(5):314–321
Sangermano R, Bax NM, Bauwens M, van den Born LI, De Baere E, Garanto A et al (2016) Photoreceptor progenitor mRNA analysis reveals exon skipping resulting from the ABCA4 c.5461–10T→C mutation in Stargardt disease. Ophthalmology 123(6):1375–1385
Sankila EM, Pakarinen L, Kääriäinen H, Aittomäki K, Karjalainen S, Sistonen P et al (1995) Assignment of an Usher syndrome type III (USH3) gene to chromosome 3q. Hum Mol Genet 4(1):93–98
Schwarz JM, Cooper DN, Schuelke M, Seelow D (2014) MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods 11(4):361–362
Smith RJ, Berlin CI, Hejtmancik JF, Keats BJ, Kimberling WJ, Lewis RA et al (1994) Clinical diagnosis of the Usher syndromes. Usher Syndrome Consortium. Am J Med Genet 50(1):32–38
Spandau UH, Rohrschneider K (2002) Prevalence and geographical distribution of Usher syndrome in Germany. Graefes Arch Clin Exp Ophthalmol 240(6):495–498
Stemerdink M, Garcia-Bohorquez B, Schellens R, Garcia-Garcia G, Van Wijk E, Millan JM (2021) Genetics, pathogenesis and therapeutic developments for Usher syndrome type 2. Hum Genet
Tsilou ET, Rubin BI, Caruso RC, Reed GF, Pikus A, Hejtmancik JF et al (2002) Usher syndrome clinical types I and II: could ocular symptoms and signs differentiate between the two types? Acta Ophthalmol Scand 80(2):196–201
Valero R, de Castro-Miro M, Jimenez-Ochoa S, Rodriguez-Ezcurra JJ, Marfany G, Gonzalez-Duarte R (2019) Aberrant splicing events associated to CDH23 noncanonical splice site mutations in a proband with atypical Usher syndrome 1. Genes (basel) 10(10):732
Van Camp G, Smith RJH (202) Hereditary hearing loss homepage. https://hereditaryhearingloss.org [updated 2020 August 12; cited 2020 November 16]
van Wijk E, Pennings RJ, te Brinke H, Claassen A, Yntema HG, Hoefsloot LH et al (2004) Identification of 51 novel exons of the Usher syndrome type 2A (USH2A) gene that encode multiple conserved functional domains and that are mutated in patients with Usher syndrome type II. Am J Hum Genet 74(4):738–744
Verpy E, Leibovici M, Zwaenepoel I, Liu XZ, Gal A, Salem N et al (2000) A defect in harmonin, a PDZ domain-containing protein expressed in the inner ear sensory hair cells, underlies Usher syndrome type 1C. Nat Genet 26(1):51–55
Vona B, Lechno S, Hofrichter MA, Hopf S, Läig AK, Haaf T et al (2016) Confirmation of PDZD7 as a nonsyndromic hearing loss gene. Ear Hear 37(4):e238–e246
Wafa TT, Faridi R, King KA, Zalewski C, Yousaf R, Schultz JM et al (2021) Vestibular phenotype-genotype correlation in a cohort of 90 patients with Usher syndrome. Clin Genet 99(2):226–235
Wagenaar M, van Aarem A, Huygen P, Pieke-Dahl S, Kimberling W, Cremers C (1999) Hearing impairment related to age in Usher syndrome types 1B and 2A. Arch Otolaryngol Head Neck Surg 125(4):441–445
Wallis Y, Payne S, McAnulty C, Bodmer D, Sister-mans E, Robertson K, et al. (eds) (2013) Practice guidelines for the evaluation of pathogenicity and the reporting of sequence variants in Clinical molecular genetics
Weil D, Blanchard S, Kaplan J, Guilford P, Gibsont F, Walsht J et al (1995) Defective myosin VIIA gene responsible for Usher syndrome type 1B. Nature 374(6517):60–61
Weil D, Küssel P, Blanchard S, Lévy G, Levi-Acobas F, Drira M et al (1997) The autosomal recessive isolated deafness, DFNB2, and the Usher 1B syndrome are allelic defects of the myosin-VIIA gene. Nat Genet 16(2):191–193
Weil D, El-Amraoui A, Masmoudi S, Mustapha M, Kikkawa Y, Laine S et al (2003) Usher syndrome type I G (USH1G) is caused by mutations in the gene encoding SANS, a protein that associates with the USH1C protein, harmonin. Hum Mol Genet 12(5):463–471
Weston MD, Luijendijk MW, Humphrey KD, Moller C, Kimberling WJ (2004) Mutations in the VLGR1 gene implicate G-protein signaling in the pathogenesis of Usher syndrome type II. Am J Hum Genet 74(2):357–366
Whelan L, Dockery A, Wynne N, Zhu J, Stephenson K, Silvestri G et al (2020) Findings from a genotyping study of over 1000 people with inherited retinal disorders in Ireland. Genes (basel) 11(1):105
Zheng L, Baumann U, Reymond JL (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. Nucleic Acids Res 32(14):e115

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix 9

Automated prioritization pipeline employed in section 5.

```
# Load the required packages library(stringr)
library(data.table)
# Define the IRD WGS filtering function IRD_WGS_filtering <- function(sample, path){ # Perform filtering for sv file


svfile <- fread(file = paste0(path,
sample, "/00_raw.unfiltered/", "/", sample,
sep = '\t', fill = TRUE)
svfile_filtered <- svfile[
(str_detect(svfile$DISEASE, "BLIND") &
(svfile$SVTYPE != "BND") &
str_detect(svfile$filter, "PASS"))]
write.csv(svfile_filtered,
file = paste0(path,
sample, "/01_filtered/", sample,
".sv.segments_filtered.txt"))
# Perform filtering for cnv file
cnvfile <- fread(file = paste0(path,
sample, "/00_raw.unfiltered/", "/", sample,
sep = '\t', fill = TRUE)
cnvfile_filtered <- cnvfile[
(str_detect(cnvfile$DISEASE, "BLIND"))]
write.csv(cnvfile_filtered,
file = paste0(path,
sample, "/01_filtered/", sample,
".cnv.segments_filtered.txt"))
# Perform filtering for snv.IRDextract file
snvIRD <- fread(file = paste0(path,
sample, "/00_raw.unfiltered/", "/",
".WGS.IRDextract.hcdiffs.txt"),
sep = '\t', fill = TRUE)
snvIRD_filtered <- snvIRD[
((snvIRD$`NonCausative - Frequency`
(snvIRD$`gnomAD-E AF` <= 5.0) &
(snvIRD$`gnomAD-G AF` <= 5.0) &
(str_detect(snvIRD$`Protein Effect`, paste0("stop\\_gained|frameshift\\_variant|stop\\_lost|",
"inframe\\_deletion|inframe\\_insertion|stop\\_gained|", "splice\\_region\\_variant")) |
(str_detect(snvIRD$`Protein Effect`,"missense\\_variant") & (snvIRD$phyloP >= 2.7)) |
(str_detect(snvIRD$`Protein Effect`,"missense\\_variant") & (snvIRD$CADD_PHRED >= 15)) |
(str_detect(snvIRD$`Protein Effect`,"missense\\_variant") & (snvIRD$`Grantham Score` >= 80)) |
(str_detect(snvIRD$`Protein Effect`,"synonymous\\_variant") & (snvIRD$phyloP >= 2.7) &
(snvIRD$CADD_PHRED >= 15)) |
str_detect(snvIRD$`Gene component`, paste0("INTRON\\_REGION|SA\\_SITE\\_CANONICAL|SD\\_SITE\\_CANONICAL|",
"CODING\\_SPLICE\\_SITE_REGION|NONCODING\\_SPLICE\\_SITE\\_REGION"))))]
write.csv(snvIRD_filtered,
file = paste0(path,
sample, "/01_filtered/", sample, ".WGS.IRDextract.hcdiffs_filtered.txt"))
# Perform filtering for snv file
snv <- fread(file = paste0(path,
sample, "/00_raw.unfiltered/", "/", sample, ".snv.hcdiffs.txt"),
sep = '\t', fill = TRUE) sample, <= 5.0) & ".sv.segments.txt"),".cnv.segments.txt"),snv_filtered <- snv[
((snv$`NonCausative - Frequency` <= 5.0) &
(snv$`gnomAD-E AF` <= 5.0) &
(snv$`gnomAD-G AF` <= 5.0) &
(str_detect(snv$`Protein Effect`, paste0("stop\\_gained|frameshift\\_variant|stop\\_lost|",
"inframe\\_deletion|inframe\\_insertion|stop\\_gained|", "splice\\_region\\_variant")) |
(str_detect(snv$`Protein Effect`,"missense\\_variant") & (snv$phyloP >= 2.7)) |
(str_detect(snv$`Protein Effect`,"missense\\_variant") & (snv$CADD_PHRED >= 15)) |
(str_detect(snv$`Protein Effect`,"missense\\_variant") & (snv$`Grantham Score` >= 80)) |
(str_detect(snv$`Protein Effect`,"synonymous\\_variant") & (snv$phyloP >= 2.7) &
(snv$CADD_PHRED >= 15)) |
str_detect(snv$`Gene component`, paste0("SA\\_SITE\\_CANONICAL|SD\\_SITE\\_CANONICAL|",
"CODING\\_SPLICE\\_SITE_REGION|NONCODING\\_SPLICE\\_SITE\\_REGION" ))))]
write.csv(snv_filtered,
file = paste0(path,
sample, "/01_filtered/", sample,
".snv.hcdiffs_filtered.txt"))}
```

Appendix 10

ACMG-AMP Classification Tables

Criteria for Classifying Pathogenic Variants

Very strong evidence of pathogenicity

PVS1 Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion) in a gene where loss of function (LOF) is a known mechanism of disease

Caveats:

- Beware of genes where LOF is not a known disease mechanism (e.g. *GFAP, MYH7*)
- Use caution interpreting LOF variants at the extreme 3' end of a gene
- Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact
- Use caution in the presence of multiple transcripts

Strong evidence of pathogenicity

PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change

Example: Val->Leu caused by either G>C or G>T in the same codon

Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level

PS2 *De novo* (both maternity and paternity confirmed) in a patient with the disease and no family history

Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, *etc*. can contribute to non-maternity

PS3 Well-established *in vitro* or *in vivo* functional studies supportive of a damaging effect on the gene or gene product

Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well-established

PS4 The prevalence of the variant in affected individuals is significantly increased compared to the prevalence in controls

Note 1: Relative risk (RR) or odds ratio (OR), as obtained from case-control studies, is >5.0 and the confidence interval around the estimate of RR or OR does not include 1.0. See manuscript for detailed guidance.

Note 2: In instances of very rare variants where case-control studies may not reach statistical significance, the prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in controls, may be used as moderate level of evidence.

Moderate evidence of pathogenicity

PM1 Located in a mutational hot spot and/or critical and well-established functional domain (*e.g*. active site of an enzyme) without benign variation

PM2 Absent from controls (or at extremely low frequency if recessive) (see Table 6) in Exome Sequencing Project, 1000 Genomes or ExAC

Caveat: Population data for indels may be poorly called by next generation sequencing

PM3 For recessive disorders, detected in *trans* with a pathogenic variant

Note: This requires testing of parents (or offspring) to determine phase

PM4 Protein length changes due to in-frame deletions/insertions in a non-repeat region or stop-loss variants

PM5 Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before

Example: Arg156His is pathogenic; now you observe Arg156Cys

Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level

PM6 Assumed *de novo*, but without confirmation of paternity and maternity

Supporting evidence of pathogenicity

PP1 Co-segregation with disease in multiple affected family members in a gene definitively known to cause the disease

Note: May be used as stronger evidence with increasing segregation data

PP2 Missense variant in a gene that has a low rate of benign missense variation and where missense variants are a common mechanism of disease

PP3 Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc)

Caveat: As many *in silico* algorithms use the same or very similar input for their predictions, each algorithm should not be counted as an independent criterion. PP3 can be used only once in any evaluation of a variant.

PP4 Patient's phenotype or family history is highly specific for a disease with a single genetic etiology

PP5 Reputable source recently reports variant as pathogenic but the evidence is not available to the laboratory to perform an independent evaluation

Criteria for Classifying Benign Variants

Stand-Alone evidence of benign impact	
BA1	Allele frequency is above 5% in Exome Sequencing Project, 1000 Genomes, or ExAC
Strong evidence of benign impact	
BS1	Allele frequency is greater than expected for disorder (see table 6)
BS2	Observed in a healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder with full penetrance expected at an early age
BS3	Well-established *in vitro* or *in vivo* functional studies shows no damaging effect on protein function or splicing
BS4	Lack of segregation in affected members of a family Caveat: The presence of phenocopies for common phenotypes (*i.e.* cancer, epilepsy) can mimic lack of segregation among affected individuals. Also, families may have more than one pathogenic variant contributing to an autosomal dominant disorder, further confounding an apparent lack of segregation.
Supporting evidence of benign impact	
BP1	Missense variant in a gene for which primarily truncating variants are known to cause disease
BP2	Observed in *trans* with a pathogenic variant for a fully penetrant dominant gene/disorder; or observed in *cis* with a pathogenic variant in any inheritance pattern
BP3	In-frame deletions/insertions in a repetitive region without a known function
BP4	Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc) Caveat: As many *in silico* algorithms use the same or very similar input for their predictions, each algorithm cannot be counted as an independent criterion. BP4 can be used only once in any evaluation of a variant.
BP5	Variant found in a case with an alternate molecular basis for disease
BP6	Reputable source recently reports variant as benign but the evidence is not available to the laboratory to perform an independent evaluation
BP7	A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site AND the nucleotide is not highly conserved

Table 5

Rules for Combining Criteria to Classify Sequence Variants

Pathogenic

1 1 Very Strong (PVS1) *AND*
- **a.** ≥1 Strong (PS1–PS4) *OR*
- **b.** ≥2 Moderate (PM1–PM6) *OR*
- **c.** 1 Moderate (PM1–PM6) and 1 Supporting (PP1–PP5) *OR*
- **d.** ≥2 Supporting (PP1–PP5)

2 ≥2 Strong (PS1–PS4) *OR*

3 1 Strong (PS1–PS4) *AND*
- **a.** ≥3 Moderate (PM1–PM6) *OR*
- **b.** 2 Moderate (PM1–PM6) *AND* ≥2 Supporting (PP1–PP5) *OR*
- **c.** 1 Moderate (PM1–PM6) *AND* ≥4 Supporting (PP1–PP5)

Likely Pathogenic

1 1 Very Strong (PVS1) *AND* 1 Moderate (PM1–PM6) *OR*

2 1 Strong (PS1–PS4) *AND* 1–2 Moderate (PM1–PM6) *OR*

3 1 Strong (PS1–PS4) *AND* ≥2 Supporting (PP1–PP5) *OR*

4 ≥3 Moderate (PM1–PM6) *OR*

5 2 Moderate (PM1–PM6) *AND* ≥2 Supporting (PP1–PP5) *OR*

6 1 Moderate (PM1–PM6) *AND* ≥4 Supporting (PP1–PP5)

Benign

1 1 Stand-Alone (BA1) *OR*

2 ≥2 Strong (BS1–BS4)

Likely Benign

1 1 Strong (BS1–BS4) and 1 Supporting (BP1–BP7) *OR*

2 ≥2 Supporting (BP1–BP7)

* Variants should be classified as Uncertain Significance if other criteria are unmet or the criteria for benign and pathogenic are contradictory.