LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE LIBRARY DUBLIN
Ollscoil Átha Cliath | The University of Dublin

Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

CHARACTERISATION OF POLYMORPHISMS ON THE PROMOTER REGION OF THE HUMAN HISTAMINE 4 RECEPTOR

by

Stephany Micallef

(B.Pharm. (Hons.), M.Sc Pharmaceutical Analysis, MPSI)

A dissertation for the degree of Doctor of Philosophy

at the

University of Dublin, Trinity College Dublin

This research was conducted at the School of Pharmacy and Pharmaceutical Sciences,

Trinity College Dublin, Dublin 2, Ireland

Under the direction and supervision of

Dr.rer.nat. Astrid Sasse



TRINITY COLLEGE DUBLIN
COLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH
THE UNIVERSITY OF DUBLIN

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.

Signature of Author



Stephany Micallef

TRINITY COLLEGE
- 4 MAR 2014
LIBRARY DUBLIN

Thesis 10202

Acknowledgements

First and foremost my sincere thanks go to my supervisor, Dr. Astrid Sasse, for the opportunity to conduct this research and her support throughout. Her contribution, stimulating suggestions and encouragement helped me to complete this project. I especially appreciate her continuous guidance and supervision from the beginning of my work that enabled me to develop a wider understanding of this research. She was not only a supervisor and mentor but also a friend during tough times and I am truly grateful.

I would also like to thank Dr. Derek Morris and Ms. Emma Quinn from the Neuropsychiatric Genetics Research Group, Institute of Molecular Medicine, Trinity Centre for Health Sciences in St. James' Hospital. Their collaborative work and suggestions have been extremely useful in conducting this research and also their generous help throughout this investigation.

My sincere thanks goes to all undergraduate students that contributed to this project, namely Maximilian Schmitt, Jessica Krause, Owen Thorpe, Hazel O'Neill, Caoimhe Cleary, Sarah Hosey, Aurelie Rogeau and Agathe Deniau. They survived the process of cloning into vectors and my catch phrase 'Let's try something different'. I would also like to thank Dr Hans-Georg Koenig from the Department of Physiology and Medical Physics in RCSI for giving me SH-SY5Y cell line and also the group from Dr Luke O'Neill's lab in the School of Biochemistry and Immunology in Trinity College Dublin, especially Moritz Haneklaus who gave me the HEK-293T cells and helped me with the bioluminescence equipment. I am also thankful for the research group in Trinity College Institute of Neuroscience for helping me with the plate-reader for bioluminescence. I also cannot forget all Principal Investigators, staff and technicians in the School of Pharmacy and Pharmaceutical Sciences. Their support and contribution to all postgraduate students is encouraging, helpful and always appreciated.

This research was also supported by EU COST Action BM0806 '*Recent advances in Histamine Receptor H4R Research*', and I am truly grateful for all the opportunities they gave me including attending various COST meetings and also a training school in 'Histamine Bioinformatics' held in Austria.

Furthermore I would like to thank all postgraduates and postdoctoral researchers in the School of Pharmacy and Pharmaceutical Sciences, especially my everyday partners-in-crime; Christine (Chri-Chri), Maria (Marie-Marie), Johanna (Penny), tall Claire, Keith (Arrivederci),

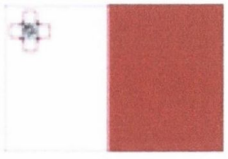
Susan, Sweta, Eva-Maria (Bretzel girl). My day would have been boring without the gossip, nerdy talk, and coffee breaks. Of course, I have to thank Adrian (Coogie), who encouraged me throughout, tried to make me laugh with his daily joke and put up with my singing. I also would like to extend my thanks to all the other postgraduates; Shane, Grainne, Sadhbh, Bo, small Claire, Svenja (Svengi), Elena, Vincent Curtini, Stefano Bianconi, Joanne, Andy (Mandy), Vilmar, Pat, Tadhg, Shu, Conor (Nordy), Jason, Gary, Brian (Money), Cathal, Mary, Ines, Evelyn, Maria P. and postdoctorals Krysztof, Gabor and Jun. I will never forget all the great times we shared, including their crazy characters (which made me feel normal) and great trophy moments in tag rugby. Without the nice friendly environment they create and their moral support, completing my research work would have not been possible.

My thanks also goes to friends I made from working in Boots the chemist for their help, understanding and support. My gratitude also goes out to the staff in Boots in Liffey Valley shopping centre, especially Siobhan, Jackie, Sandra, Christina, Claire, Brida, Jennifer, Andrea and Aileen. Working with them was always a pleasure as they manage to create a great atmosphere in the pharmacy, even during hard times in the lab.

In addition, I would also like to thank my volleyball team mates from Trinity College, especially Jen Fitz, Valeria, Maryline and Jen Six. Their enthusiasm and moral support was immense. Sincere thanks go to my volleyball team mates from home, Paola Hibs, whom always had great faith in me and encouraging. I also cannot forget the great support from my friends back home Delicia, Ann, Romina and Denise, who encouraged me in tough times and were always there for me in time of need. A special thanks also goes to my flatmate Clare Andrews for her daily optimism and feeding me in dire straits. I also cannot forget my good friend Anna with her cheerful character and ongoing (entertaining) life adventures.

My final and deepest thanks go to my family; especially my mum Degna, my dad Joseph and my sister Helena. Their love and support throughout my journey has been exceptional. I especially appreciate their interest in asking questions about my research, and for making me believe they knew what I was talking about. It definitely helped me to build up confidence in explaining my research topic. I would also like to extend my gratitude to my aunty Doris, uncle Joe and my cousins Corinne, Samuel, Nathaniel, Clarissa and Isaac. They were always there for me and my family, and were a great support to all of us. Family is not just an important thing. It is everything.

The research work disclosed in this publication is partially funded by the Strategic Educational Pathways Scholarship (Malta). This Scholarship is part-financed by the European Union – European Social Fund (ESF) under Operational Programme II – Cohesion Policy 2007-2013, "Empowering People for More Jobs and a Better Quality Of Life".



Operational Programme II – Cohesion Policy 2007-2013
Empowering People for More Jobs and a Better Quality of Life
Scholarship part-financed by the European Union
European Social Fund (ESF)
Co-financing rate: 85% EU Funds; 15%National Funds



Investing in your future

To my Dad

ABSTRACT

The human histamine 4 receptor (HRH4) is the recently discovered receptor in the histamine GPCR family. It is mainly expressed in haematopoietic cells but recent publications provide evidence that it is both expressed and functional on neurons in mammalian CNS. This study targets genetic polymorphism in the promoter region of HRH4, for which in preliminary unpublished data it is indicated that a length polymorphism, a variable number of tandem repeats (VNTR), i.e., CAA repeats, is more common in schizophrenia patients. The aim of the project is to test for association for polymorphisms on HRH4 gene using genome-wide association studies (GWAS) from the International Schizophrenia Consortium, genotype HapMap samples and establish linkage disequilibrium (LD) with known SNPs, clone the putative promoter region of HRH4 in a reporter plasmid, and finally analyse these clones in a cell-based assay for functional impact on gene expression. Linkage analyses showed that while there is no strong significance with schizophrenia, the 13 CAA microsatellite repeat has a high LD with rs17797945 (r^2 = 0.956) which is a SNP flagged by the GWAS analysis that had a weak association to schizophrenia (P = 0.007879). Genotyping HapMap samples also indicated that this VNTR region varies between 10 (10R) and 13 (13R) to 19 triplet repeats (19R) with a SNP at the end of the triplet repeat. These genotypes were cloned into a standard vector, sub-cloned into a reporter vector (pGL3-basic) and mutated to obtain the respective allele variants. The sequence cloned for all VNTR polymorphisms was from -2044 bp to -1 of the promoter region as described by Coge *et al*. The 10R was used for further analysis to identify the functional promoter where it was extended at the 5'-end and the 3'-end including exon 1. The -2044 bp 13R reporter construct was also extended on the 3'-end and analysed since this had the highest allele frequency (27%). The promoter constructs were assayed on a Dual-Glo luciferase assay, where the promoter length constructs were transiently transfected into a mammalian cell line (HEK293T) and normalised with the renilla vector pRL-SV40. Gene expression of firefly luciferase was directly quantified as transcriptional activity of the promoter region. The promoter sequence described *in silico* by Coge *et al* ranging from -2044 bp to -1 did not reach statistical significance when compared to the promoter less vector (pGL3-Basic), however the promoter sequence tested *in vitro* ranging from -705 bp to +194 bp was identified as the functional promoter with $P < 0.001$ when compared to pGL3-basic vector. This is the first study that analyses the promoter region of the HRH4 biologically in a cell-based assay.

TABLE OF CONTENTS

Table of Figures ..... **ix**

List of Tables ..... **xv**

List of Abbreviations ..... **xviii**

Legend of clones library ..... **xxi**

1. Introduction **2**

1.1 The histamine receptor family **2**

1.2 Genetic variations in the histamine receptor family **6**

1.2.2 The histaminergic system in the Central Nervous System 6

1.2.2.1 Schizophrenia 7

1.2.2.2 Parkinson's disease 8

1.2.2.3 Alzheimer's disease 8

1.2.2.4 Migraine 9

1.2.3 The histaminergic system in inflammatory conditions 12

1.2.3.1 Allergy and airway inflammation 12

1.2.3.2 Gastric acid secretion 14

1.2.4 The histaminergic system in cancer 18

1.3 The human histamine receptor 4 ..... **20**

1.3.1 The HRH4 and the central nervous system 22

1.3.2 Polymorphisms on the HRH4 gene 24

1.3.3 Schizophrenia and polymorphisms 27

1.4 Aims and objectives 30

2 Materials and Equipment 32

2.1 Reagents and kits 32

2.2 Equipment and Manufacturers 34

2.3 Bacterial Strains 35

2.4 Primers 35

2.4.1 Primers for HRH4 35

2.4.2 Standard primers 37

2.4.3 Primers for Beta-Actin 37

2.5 Specialised Software and Databases 38

2.6 Cell culture 38

2.7 Preparation of solutions, plates and agarose gel 39

2.8 DNA sequences 42

2.8.1 18RCpEX synthesis 42

2.8.2 19RCpEx synthesis 44

2.8.3 XLpUC57 synthesis 46

2.8.4 Mega-primer sequences 48

2.8.4.1 Mega-primer LL 48

2.8.4.2 Mega-primer XL 48

2.8.4.3 Mega-primer Ex1 49

2.8.5 Proposed sequences for alignment 50

2.8.5.1 Proposed sequence for -2044 bp for the 10R VNTR in pGL3 50

2.8.5.2 Proposed sequence for -3329 bp Ex1 for the 10R VNTR in pGL3 51

3. Linkage analysis for HRH4 polymorphism 54

3.1 Methods 55

3.1.1 Associations on the HRH4 Gene Locus 55

3.1.2 Genotyping HapMap Samples 55

3.1.2.1 Polymerase Chain Reaction on HapMap Samples 55

3.1.2.2 Capillary Electrophoresis 57

3.1.2.3 Linkage Disequilibrium Analysis 57

3.2 Results 61

3.2.1 Linkage analysis results 61

3.2.1.1 Association analysis for HRH4 gene locus 61

3.2.1.2 PCR for HapMap samples 62

3.2.1.3 Capillary Electrophoresis 63

3.2.1.4 Linkage Disequilibrium Analysis 65

3.3 Discussion 70

3.4 Conclusion 77

4. Construction of reporter plasmid library 80

4.1 Methods **81**

4.1.1 Overview of methods 81

4.1.1.1 Genomic Human DNA Collection and Extraction 82

4.1.1.2 DNA quantification and qualification 83

4.1.1.3 Amplification of the promoter region 84

4.1.1.4 DNA purification for further analysis 86

4.1.1.5 Ligation 88

4.1.1.6 Transformation into *E.Coli* 90

4.1.1.7 Analysis of transformants - PCR on colonies 92

4.1.1.8 Isolation of plasmids from *E.Coli* 93

4.1.1.9 Sequencing of clones 94

4.1.1.10 Restriction enzyme digest 94

4.1.1.11 Site-directed mutagenesis – point mutation 96

4.1.1.12 Parental plasmid digest after site-directed mutagenesis 99

4.1.1.13 Design and production of mega-primers for mutation 100

4.1.1.14 Site-directed mutagenesis – Extension using mega-primers 102

4.1.2 Construction of clones 104

4.1.2.1 Cloning into pCR-Blunt 104

Summary of clones for section 4.1.2.1 105

4.1.2.2 Sub-cloning into pGL3 reporter vector 106

4.1.2.2.1 Sub-Cloning 10R,13R,14R, 15,16R,17R from pCR-Blunt.....106

4.1.2.2.2 Sub-cloning 18R and 19R from pEX.....107

Summary of clones for section 4.1.2.2.....108

4.1.2.3 Point mutation for the VNTR genotypes.....108

Summary of clones 4.1.2.3.....110

4.1.2.4 Construction of promoter length variations.....111

4.1.2.4.1 Construction of the short length promoter, -705 bp 10RpGL3.....111

4.1.2.4.2 Extension of the 5'-end promoter.....112

4.1.2.4.3 Extension of the 3'-end of the promoter - Addition of exon 1.....113

Summary of clones for section 4.1.2.4.....114

4.2 Results.....115

4.2.1 Cloning the HRH4 promoter in a standard vector.....115

4.2.1.1 Human genomic DNA quantification and qualification.....115

4.2.1.2 Amplification of the promoter region: 10R, 13R, 14R and 15R.....116

4.2.1.3 Ligation into pCR-Blunt vector.....117

4.2.1.4 Analyzing transformants for pCR-Blunt vectors.....118

4.2.1.5 Cloning results for 16R and 17R into pCR-Blunt.....125

4.2.2 Sub-Cloning from pCR-Blunt library into pGL3.....129

4.2.2.1 Amplification of promoter region from clones in pCR-Blunt.....129

4.2.2.2 Ligation into pGL3 reporter plasmid.....131

4.2.1.3 Analysing transformants in reporter plasmid 132

4.2.1.4 Cloning results for 18RC and 19RC using 10RpGL3 as the host vector.. 138

4.2.3 Point mutation results on vector constructs 144

4.2.3.1 Point mutations in pCR-Blunt 144

4.2.3.2 Sub-cloning mutated plasmids into pGL3-basic 152

4.2.3.3 Point mutations in pGL3 reporter plasmid 157

4.2.4 Construction of different promoter lengths 164

4.2.4.1 Constructing the short length promoter: -705bp 10RpGL3 164

4.2.4.1.1 Restriction enzymes digest for extension insert and 10RpGL3 164

4.2.4.1.2 Analysing transformants 165

4.2.4.2 Constructing the long length promoter: -2772bp 10RpGL3 167

4.2.4.2.1 Amplification of the LL mega-primer 168

4.2.4.2.2 Analysis of mutated transformants with LL mega-primers: -2772 bp 10RpGL3 169

4.2.4.3 Construction of the longest length promoter: -3329 bp 10RpGL3..... 171

4.2.4.3.1 Amplification of the XL mega-primer 171

4.2.4.3.2 Analysis of mutated transformants with XL mega-primers: -3329 bp 10RpGL3 172

4.2.4.4 Addition of Exon 1 to promoter length constructs 174

4.2.4.4.1 Amplification of Exon 1 for a mega-primer 174

4.2.4.4.2 Analysing mutated transformats with Exon 1 mega-primers..... 175

4.3 Discussion **180**

4.4 Conclusion **195**

5 Promoter analysis for HRH4 ..... **198**

5.1 Methods **200**

5.1.1 Promoter study for HRH4 200

5.1.1.1 Cell culture 200

5.1.1.2 Preparation of DNA for transfection 201

5.1.1.3 Seeding and transfection 202

5.1.1.4 Dual-Glo luciferase assay 204

5.1.1.5 Statistical analysis 204

5.2 Results **205**

5.2.1 Dual-glo luciferase assay optimisation 205

5.2.2 Gene reporter assay for HRH4 promoter 210

5.2.2.1 Preparation of samples – plasmid purification 210

5.2.2.2 Dual-Glo luciferase assay results 213

5.3 Discussion **221**

5.4 Conclusion **229**

6 Conclusions **232**

References **240**

Appendices

Appendix 1 – Ethical Approval Letter.....250

Appendix 2 – Pedigree analysis and recorded genotypes for HapMap samples.....252

Appendix 3 – Alignment of clones with proposed sequences.....258

Appendix 4 – Presentations, conferences and publications.....274

Table of Figures

Figure 1.1 Key characteristics of the histamine receptor family 3

Figure 1.2 Genomic organisation of the human histamine receptors 4

Figure 1.3 Snake plot of the human histamine H4 receptor [6] 21

Figure 1.4 The histamingeric system in the human brain [86] 22

Figure 1.5 NCBI snapshot image for the VNTR region 24

Figure 2.1 Synthesised sequence for 18RC 42

Figure 2.2 Vector specifications for 18RC 43

Figure 2.3 Synthesised sequence for 19RC 44

Figure 2.4 Vector specifications for 19RC 45

Figure 2.5 Synthesised sequence for XLpUC57 46

Figure 2.6 Vector characteristics for pUC57 47

Figure 2.7 Sequence for amplified mega-primer LL 48

Figure 2.8 Sequence for amplified mega-primer XL 49

Figure 2.9 Sequence for amplified mega-primer Ex1 49

Figure 2.10 Proposed sequence for -2044 bp 10RpGL3 50

Figure 2.11 Proposed sequence for -3229 bp Ex1 10RpGL3 52

Figure 3.1 Gel Electrophoresis for PCR for HapMap samples 62

Figure 3.2 Sequencing results for sample 3, showing the 14 tandem repeat CAA/-AAA .. 63

Figure 3.3 CE of HapMap samples showing 4 different genotypes. 64

Figure 3.4 Location of IMPACT and HRH4 on Chromosome 18 68

Figure 3.5 Formation of microsatellites 71

Figure 3.6 Proposed reporting for the VNTR region of HRH4 gene 74

Figure 4.1 Primers for point mutation site-directed mutagenesis 97

Figure 4.2 PCR qualification for Beta-Actin band 115

Figure 4.3 Agarose gel for PCR purification for 10R and 15R 116

Figure 4.4 Agarose gel for PCR purification for 13R and 14R 116

Figure 4.5 10R insert ligated into pCR-Blunt 118

Figure 4.6 Restriction enzyme digest for plasmids with EcoR1 enzyme 119

Figure 4.7 Sequencing results for 10RpCR-blunt clone with M13Rev (-29) 120

Figure 4.8 Sequencing result for 10RpCR-Blunt clone AS16For primer 120

Figure 4.9 PCR on colonies for 13R clone, using 10R as a positive control 121

Figure 4.10 Sequencing result for 13RApCR-Blunt clone with AS16For 122

Figure 4.11 PCR on colonies for 15RA clone, using 10R as a positive control 122

Figure 4.12 Sequencing result for 15RApCR-Blunt clone with AS16For 123

Figure 4.13 PCR on colonies for 14RC clone, using 10R as a positive control 124

Figure 4.14 Sequencing result for 14RCpCR-Blunt clone with AS16For 125

Figure 4.15 PCR band for 13R/ 16R amplicon 125

Figure 4.16 PCR band for 15R/ 17R amplicon 126

Figure 4.17 PCR on colonies for 16R and 17R clones 127

Figure 4.18 Sequencing result for 16RApCR-Blunt....128
Figure 4.19 Sequencing result for 17RApCR-Blunt....128
Figure 4.20 PCR product for 10R and 13R for pGL3....129
Figure 4.21 PCR product for 14R and 15R for pGL3....129
Figure 4.22 10RpGL3 construct in reporter plasmid....131
Figure 4.23 PCR on colonies for 10RpGL3....132
Figure 4.24 PCR on colonies for 13RApGL3....132
Figure 4.25 PCR on colonies for 14RCpGL3 and 15RApGL3....133
Figure 4.26 Alignment of 10RpGL3 with proposed sequence....134
Figure 4.27 Alignment of 13RApGL3 with proposed sequence....135
Figure 4.28 Alignment of 14RCpGL3 with proposed sequence....136
Figure 4.29 Alignment of 15RApGL3 with proposed sequence....137
Figure 4.30 PCR product for 18RC and 19RC from pEX for insertion into pGL3 vector.138
Figure 4.31 PCR on colonies for 18RC and 19RC in pGL3....140
Figure 4.32 Alignment of 18RCpGL3 with proposed sequence....142
Figure 4.33 Alignment of 19RC with proposed sequence....143
Figure 4.34 Bacterial plates for point mutations after site-directed mutagenesis....145
Figure 4.35 EcoR1 digest for potential point mutations for 13RpCR-Blunt....147
Figure 4.36 EcoR1 digest for potential point mutations for 14RpCR-Blunt....147
Figure 4.37 EcoR1 digest for potential point mutations 15RpCR-Blunt....147

Figure 4.38 Sequencing result for 13RCpCR-Blunt 148

Figure 4.39 Sequencing comparison for 14RApCR-Blunt point mutation..... 148

Figure 4.40 Sequencing comparison for 15RCpCR-Blunt point mutation 149

Figure 4.41 PCR on colonies for 16RC and 17RC SDM in pCR-Blunt..... 150

Figure 4.42 Sequencing comparison for the 16RCpCR-Blunt point mutation..... 151

Figure 4.43 Sequencing comparison for the 17RCpCR-Blunt point mutation..... 151

Figure 4.44 PCR for 14RA and 15RC for sub-cloning into reporter plasmid 152

Figure 4.45 PCR on colonies for 14RA and 15RC in pGL3 153

Figure 4.46 Alignment of 14RApGL3 with the proposed sequence..... 155

Figure 4.47 Alignment of 15RCpGL3 with a proposed sequence 156

Figure 4.48 PCR on colonies for SDM on 13RCpGL3 157

Figure 4.49 Sequence comparison for point mutation for 13RCpGL3..... 158

Figure 4.50 PCR on colonies for SDM for 18RApGL3 and 19RApGL3..... 159

Figure 4.51 Sequence comparison for point mutation for 18RApGL3 160

Figure 4.52 Sequence comparison for point mutation for 19RApGL3 160

Figure 4.53 Sequencing result for 17RApGL3 161

Figure 4.54 PCR on colonies for SDM for 17RCpGL3 162

Figure 4.55 Sequence comparison for point mutation for 17RCpGL3..... 163

Figure 4.56 Restriction digest for extension insert and 10RpGL3 164

Figure 4.57 PCR on colonies for extension in 10RpGL3 166

Figure 4.58 Alignment of -705 bp 10RpGL3 with proposed sequence 167

Figure 4.59 PCR for LL mega-primer from XLpUC57 168

Figure 4.60 PCR on colonies for -2772 bp10RpGL3.....169

Figure 4.61 Sequencing result for -2772 bp 10RpGL3.....170

Figure 4.62 PCR for mega-primer for -3329 bp 10RpGL3 171

Figure 4.63 PCR on colonies for -3329 bp 10RpGL3.....172

Figure 4.64 Alignment of -3772 bp 10RpGL3 with proposed sequence 173

Figure 4.65 PCR for mega-primer for exon 1 extension.....174

Figure 4.66 PCR on colonies for -705 bp Ex1 10RpGL3.....175

Figure 4.67 PCR on colonies for -2044 bp Ex1 10RpGL3 and -2044 bp Ex1 13RpGL3.176

Figure 4.68 PCR on colonies for -2772 bp Ex1 10RpGL3.....176

Figure 4.69 PCR on colonies with 3329 bp Ex1 10RpGL3.....177

Figure 4.70 Sequence alignments for exon 1 on promoter length constructs.....179

Figure 4.71 Structure and characteristics of pCR-Blunt vector.....181

Figure 4.72 Tag SNP for the 10R genotype.....183

Figure 4.73 Cloning for the potential promoter region into pCR-Blunt.....184

Figure 4.74 Sub-cloning the potential promoter region into a reporter plasmid 185

Figure 4.75 Schematic diagram for the 10R linearised promoter region.....186

Figure 4.76 Site-directed mutagenesis creating a point mutation.....188

Figure 4.77 Virtual restriction enzyme digest for XLpUC57.....192

Figure 4.78 Site-directed mutagenesis mechanism using Mega-primers 193

Figure 5.1 Luciferase activity for optimisation in SH-SY5Y cells 207

Figure 5.2 Optimisation of luciferase assay 209

Figure 5.3 Promoter activity for 10RpGL3 construct without exon 1 215

Figure 5.4 Promoter constructs with for 10RpGL3 with exon 1 with pGL3-SV40 217

Figure 5.5 Promoter activity for 10RpGL3 construct with exon 1 219

Figure 5.6 Luciferase activity for all -2044 bp constructs 220

Figure 5.7 Promoter length constructs assayed for promoter activity 222

Figure 5.8 Luciferase assay for promoter identification 224

Figure 5.9 Proposed in silico promoter sequence by Coge et al [7] 226

Figure 5.10 Proposed in vitro promoter sequence for HRH4 227

List of Tables

Table 1.1 Gene and protein homology for the histamine receptor subtypes 3

Table 1.2 Polymorphisms on HRH affecting central nervous system conditions..... 10

Table 1.3 Polymorphisms on HRH affecting inflammatory conditions..... 15

Table 1.4 Polymoprhisms on HRH potentially affecting cancer 19

Table 1.5 VNTR polymorphism in the promoter region HRH4 25

Table 1.6 VNTR allele frequency in Schizophrenic patients and controls 26

Table 1.7 Genotype frequencies for rs11662595 in Schizophrenia patients 26

Table 1.8 Loci from GWAS in Schizophrenia 28

Table 3.1 Master mix for genotyping HapMap samples..... 56

Table 3.2 Association analysis for HRH4 gene from ISC GWAS data..... 61

Table 3.3 Relabeling alleles according to the VNTR length..... 65

Table 3.4 Allele frequencies for the VNTR polymorphisms 66

Table 3.5 Sub-grouped genotypes and assignment of rs numbers..... 67

Table 3.6 R-squared values from LD plot for 10R and 13R..... 68

Table 3.7 Comparison between studies 73

Table 4.1 Master mix for amplification of the putative promoter region..... 84

Table 4.2 Master mix for PCR on colonies 92

Table 4.3 Master mix for site-directed mutagenesis in point mutation 98

Table 4.4 PCR for mega-primers 101

Table 4.5 Site-directed mutagenesis using mega-primers 102

Table 4.6 Human Genomic DNA quantification 115

Table 4.7 DNA quantification after gel extraction 117

Table 4.8 Plasmid concentrations for 13R colonies after mini-prep 121

Table 4.9 Plasmid concentrations for 15RA colonies after mini-prep 123

Table 4.10 Plasmid concentrations for 14C colonies after mini-prep 124

Table 4.11 Gel extraction for 13R/16R and 15R/17R 126

Table 4.12 Plasmid concentration after miniprep for 16R and 17R in pCR-Blunt 127

Table 4.13 Gel extraction for PCR products into pGL3 130

Table 4.14 Double digest purification for PCR products and pGL3 basic 130

Table 4.15 PCR product for 18RC and 19RC after gel purification 139

Table 4.16 Double digest (Xho1 and Msc1) purification for 18RC, 19RC and 10RpGL3 host vector 139

Table 4.17 Miniprep colonies for 18RC and 19RC 141

Table 4.18 Miniprep for SDM point mutations for 13R, 14R and 15R 146

Table 4.19 Miniprep for SDM on 16RC and 17RC in pCR-Blunt 150

Table 4.20 PCR gel extraction for 14RA and 15RC 152

Table 4.21 Purification for double digested 14RA, 15RC and pGL3-basic 153

Table 4.22 Miniprep for 14RCpGL3 and 15RCpGL3 colonies 154

Table 4.23 Miniprep for 13RCpGL3 clones 157

Table 4.24 Miniprep for 18RApGL3 and 19RApGL3 colonies 159

Table 4.25 Miniprep for 17RCpGL3 colonies 162

Table 4.26 Purification for restriction enzyme digest for XLpUC57 and 10RpGL3 165

Table 4.27 Purified plasmids for short length colonies 166

Table 4.28 Purified LL mega-primer from gel extraction 168

Table 4.29 Miniprep for -2772 bp 10RpGL3 colonies 169

Table 4.30 Purified XL mega-primer after gel extraction 171

Table 4.31 MiniPrep for -3329 bp 10RpGL3 172

Table 4.32 Purification of Exon 1 mega-primer after gel extraction 174

Table 4.33 Miniprep for colonies with exon 1: -705, -2042, -2772, -3329 bp for 10RpGL3 & -2042 bp for 13RpGL3 178

Table 5.1 Endo-free maxi prep for optimisation assays 205

Table 5.2 Luciferase assay optimisation in SH-SY5Y (1) 206

Table 5.3 Luciferase activity for SH-SY5Y optimistation (2) 208

Table 5.4 Luciferase activity for HEK293T optimisation 208

Table 5.5 EndoFree plasmid purification for clones without exon 1 and controls 211

Table 5.6 EndoFree plasmid purification for clones with exon 1 and controls 212

Table 5.7 Luciferase assay activity for constructs without exon 1 213

Table 5.8 Final fold activity for promoter constructs without exon 1 214

Table 5.9 Luciferase assay activity for constructs with exon 1 216

Table 5.10 Final fold activity for promoter constructs with exon 1 218

Table 6.1 Plasmid library 235

List of Abbreviations

AAD	Allergic atopic dermatitis
ABP1	Diamine oxidase
AD	Alzheimer's disease
ADHD	Attention deficit hyperactivity disorder
cAMP	cyclic Adenosine monophosphate
CE	Capillary electrophoresis
CEU	Utah residents with Northern and Western European ancestry from the CEPH collection
CNS	Central nervous system
CNV	Copy number variations
D3R	Dopamine D3 receptor
DAO	Diamine oxidase
dbSNP	Single nucleotide polymorphism database
DD	Double Digest
DNA	Deoxyribonucleic acid
FREC	Faculty research ethics committee
GABA	Gamma- aminobutyric acid
GPCR	G-protein coupled receptor
GTF	General transcription factors
GWAS	Genome-wide association studies

HDC	Histidine decarboxylase
HG	Human genomic
HNMT	Histamine N-methyltransferase
HRH	Human histamine receptor
HRH1	Human histamine H1 receptor
HRH2	Human histamine H2 receptor
HRH3	Human histamine H3 receptor
HRH4	Human histamine H4 receptor
IL16	Interleukin 16
INDEL	Insertion or deletion polymorphism
ISC	International schizophrenia consortium
LD	Linkage disequilibrium
NCBI	National centre for biotechnology information
NMDA	N-methyl D-aspartic acid
NSAID	Non-steroidal anti-inflammatory
OR	Odd ratios
P/S	Penicillin/ streptomycin
PCR	Polymerase chain reaction
PD	Parkinson's disease
PIC	Pre-initation complex

RNA	Ribonuclease acid
RT	Room temperature
SCZ	Schizophrenia
SDM	Site-directed mutagenesis
SNP	Single nucleotide polymorphism
TFIIA	Transcription factor II A
TFIIB	Transcription factor II B
TFIID	Transcription factor II D
TFIIE	Transcription factor II E
TFIIF	Transcription factor II F
TFIIH	Transcription factor II H
TMN	Tuberomammillary nucleus
UTR	Un-translated region
VMAT-2	Vesicular monoamine transporter-2
VNTR	Variable number of tandem repeats
α-INF	Interferon alpha
β-ME	Beta-mercapto ethanol

Legend of clones library

Clones in a standard vector

- ✓ representing HRH4 promoter sequence from -2044 bp to +109 bp

(where -2044 bp to +109 bp is the actual sequence length when 10R triplet repeat is present, and this is kept as the named promoter length sequence for all clones)

Name	Characteristic	Method
10RpCR-Blunt	10 triplets repeat genotype (10R) in pCR-Blunt	Cloning from HapMap sample
13RApCR-Blunt	13 triplets repeat genotype (13R) with an 'A' allele in the last triplet in pCR-Blunt (AAA)	Cloning from extracted human genomic DNA
14RCpCR-Blunt	14 triplet repeats genotype (14R) with an 'C' allele in the last triplet in pCR-Blunt (CAA)	Cloning from extracted human genomic DNA
15RApCR-Blunt	15 triplet repeats genotype (15R) with an 'A' allele in the last triplet in pCR-Blunt (AAA)	Cloning from HapMap sample
16RApCR-Blunt	16 triplet repeats genotype (16R) with an 'A' allele in the last triplet in pCR-Blunt (AAA)	Cloning from extracted human genomic DNA
16RCpCR-Blunt	16 triplet repeats genotype (16R) with an 'A' allele in the last triplet in pCR-Blunt (AAA)	Site-directed mutagenesis from 16RApCR-Blunt
17RApCR-Blunt	17 triplet repeats genotype (17R) with an 'A' allele in the last triplet in pCR-Blunt (AAA)	Cloning from extracted human genomic DNA
17RCpCR-Blunt	17 triplet repeat genotype (17R) with an 'A' allele in the last triplet in pCR-Blunt (CAA)	Site-directed mutagenesis from 17RApCR-Blunt

Clones in a reporter vector

- ✓ representing HRH4 promoter sequence from -2044 bp to -1 bp

(where -2044 bp to -1 bp is the actual sequence length when 10R triplet repeat is present, and this is kept as the named promoter length sequence for all clones)

Name	Characteristic	Method
-2044 bp 10RpGL3	10R in pGL3-basic	Sub-cloning from 10RpCR-blunt
-2044 bp 13RApGL3	13R with 'A' allele in the last triplet repeat (AAA) in pGL3-basic	Sub-cloning from 13RApCR-Blunt
-2044 bp 13RCpGL3	13R with 'C' allele in the last triplet repeat (CAA) in pGL3-basic	Site-directed mutagenesis from 13RApGL3
-2044 bp 14RCpGL3	14R with 'C' allele in the last triplet repeat (CAA) in pGL3-basic	Sub-cloning from 14RCpCR-Blunt
-2044 bp 14RApGL3	14R with 'A' allele in the last triplet repeat (AAA) in pGL3-basic	Site-directed mutagenesis from 14RCpGL3
-2044 bp 15RApGL3	15R with 'A' allele in the last triplet repeat (AAA) in pGL3-basic	Sub-cloning from 15RApCR-Blunt
-2044 bp 15RCpGL3	15R with 'C' allele in the last triplet repeat (CAA) in pGL3-basic	Site-directed mutagenesis from 15RApGL3
-2044 bp 17RApGL3	17R with 'A' allele in the last triplet repeat (AAA) in pGL3-basic	Site-directed mutagenesis from 18RApGl3
-2044 bp 17RCpGL3	17R with 'C' allele in the last triplet repeat (CAA) in pGL3-basic	Site-directed mutagenesis from 17RApGL3
-2044 bp 18RCpGL3	18R with 'C' allele in the last triplet repeat (CAA) in pGL3-basic	Sub-cloning from pEX into 10RpGL3

Name	Characteristic	Method
-2044 bp 18RApGL3	18R with 'A' allele in the last triplet repeat (AAA) in pGL3-basic	Site-directed mutagenesis from 18RCpGL3
-2044 bp 19RCpGL3	19R with 'C' allele in the last triplet repeat (CAA) in pGL3-basic	Sub-cloning from pEX into 10RpGL3
-2044 bp 19RApGL3	19R with 'A' allele in the last triplet repeat (AAA) in pGL3-basic	Site-directed mutagenesis from 19RCpGL3

Promoter length variants in a reporter vector

(where actual length sequences are based on the 10R triplet repeat, and this is kept as the named promoter length sequence for all clones)

Name	Characteristic	Method
-705 bp 10RpGL3	Sequence -705 to -1 for 10R in pGL3-basic	Self-ligation of 10RpGL3 digested with Msc1
-2772 bp 10RpGL3	Sequence -2772 to -1 for 10R in pGL3-basic	Site-directed mutagenesis using mega-primers from -2044 bp 10RpGL3
-3329 bp 10RpGL3	Sequence -3329 to -1 for 10R in pGL3-basic	Site-directed mutagenesis using mega-primers from -2772 bp 10RpGL3
-705 bp Ex1 10RpGL3	Sequence -705 to +197 for 10R in pGL3-basic (where 0 to +197 is exon 1 of HRH4)	Site-directed mutagenesis using mega-primers from -705 bp 10RpGL3
-2044 bp Ex1 10RpGL3	Sequence -2044 to +197 for 10R in pGL3-basic (where 0 to +197 is exon 1 of HRH4)	Site-directed mutagenesis using mega-primers from -2044 bp 10RpGL3
-2044 bp Ex1 13RApGL3	Sequence -2044 to +197 for 13RA in pGL3-basic (where 0 to +197 is exon 1 of HRH4)	Site-directed mutagenesis using mega-primers from -2044 bp 13RApGL3
-2772 bp Ex1 10RpGL3	Sequence -2772 to +197 for 10R in pGL3-basic (where 0 to +197 is exon 1 of HRH4)	Site-directed mutagenesis using mega-primers from -2772 bp 10RpGL3
-3329 bp Ex1 10RpGL3	Sequence -3329 to +197 for 10R in pGL3-basic (where 0 to +197 is exon 1 of HRH4)	Site-directed mutagenesis using mega-primers from -3329 bp 10RpGL3

Chapter 1

Introduction

Parts of this section has been published:

Micallef S, Stark H, Sasse A. Polymorphisms and genetic linkage of histamine receptors. Life Sci 2013; **93**(15): 487-494.

1. Introduction

1.1 The histamine receptor family

Histamine is a biogenic amine that plays an essential role in controlling many physiological functions, both in the central nervous system (CNS) and the peripheral nervous system (PNS). Most of these physiological effects are mediated through interactions with four histamine receptor subtypes, all of which are members of the larger family of rhodopsin-like class A G-protein coupled receptors (GPCRs) [1, 2].

The first two GPCRs from this subtype, namely histamine H1 and histamine H2 receptors were first cloned in 1991 [3]. The human histamine H1 receptor (HRH1) is the largest protein [3] and is expressed primarily in smooth muscle, endothelial cells, the adrenal medulla, the heart, and also in the CNS. The HRH1 is predominantly involved in smooth muscle contraction and increased vascular permeability [4]. The main therapeutic target for HRH1 antagonists are allergic conditions and sleep disorders [5]. The human histamine H2 receptor (HRH2) is mostly expressed in gastric parietal cells, vascular smooth muscle, suppressor T cells, neutrophils, the CNS and the heart. Its main involvement is in stimulation of gastric acid secretion and hence an important therapeutic target for gastric ulcers [4]. The histamine 3 receptor (HRH3), first cloned in 1999 by Johnson & Johnson [6], is a presynaptic autoreceptor on histamine neurons and heteroreceptor on various neurotransmitter neurons, and is expressed in the CNS and peripheral nerves [4]. Homology and phylogenetic analysis of HRH3 suggested it to be markedly different from the previously cloned HRH1 and HRH2 receptors, which suggests that these histamine receptors evolved from different ancestor genes [3]. This discovery led to the identification of a fourth receptor, the human histamine H4 receptor (HRH4) in the year 2000/2001 by several research groups [4, 5, 7-9]. Figure 1.1 highlights the major characteristics on gene, protein and functional level of the histamine receptor family.

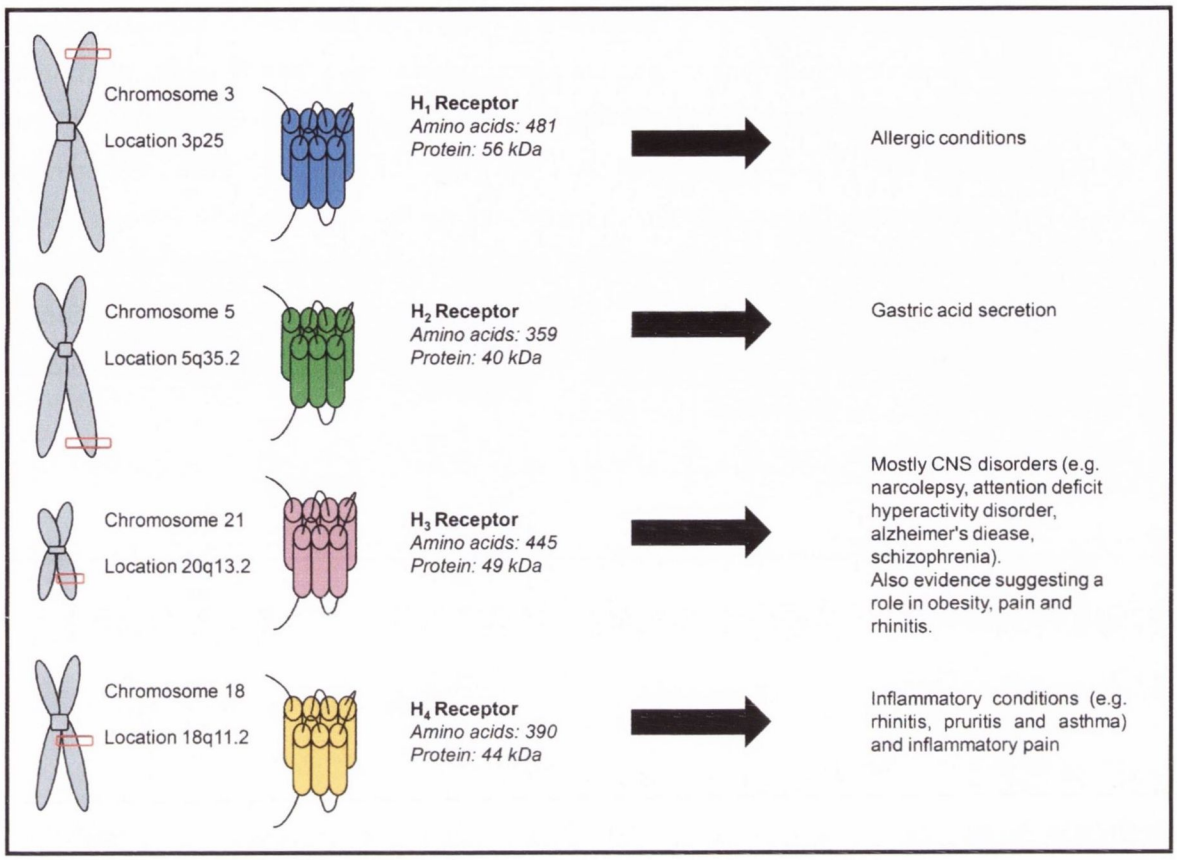


Figure 1.1 Key characteristics of the histamine receptor family

On comparing these four receptors, it becomes evident that HRH3 and HRH4 share a similar genomic organisation, consisting of three exons and two introns (Figure 1.2). Using a bioinformatics tool, GenomeNet multiple sequence alignment CLUSTALW, homology for genes can be estimated. Comparing the DNA sequence between all HRH genes the homology appears to be in the same range (18-20%) for all four HRHs (Table 1.1).

Table 1.1 Gene and protein homology for the histamine receptor subtypes

Homology for HRHs		Gene %			
		HRH1	**HRH2**	**HRH3**	**HRH4**
Protein %	***HRH1***		18.60	20.20	20.60
	HRH2	*21.50*		20.70	18.95
	HRH3	*25.70*	*19.50*		18.30
	HRH4	*23.00*	*22.00*	*37.00*	

Looking into the translated sequence in amino acids for each protein gives a different picture. An amino acid sequence resemblance can be seen between two GPCR family members, namely HRH3 and HRH4. The histamine 3 receptor displays a 37% homology to HRH4 but when compared to HRH1 and HRH2, the similarity drops to 23% and 22%, respectively (Table 1.1). Looking at the transmembrane domains, the similarity between HRH3 and HRH4 increases up to 54% and if the physiochemical properties of the amino acids is taken into account this similarity increases even more to 68% [10]. This similarity between the H_3 receptor and H_4 receptor also led to the development of imidazole-containing ligands that have a dual action on histamine H_3 and histamine H_4 receptors [11].

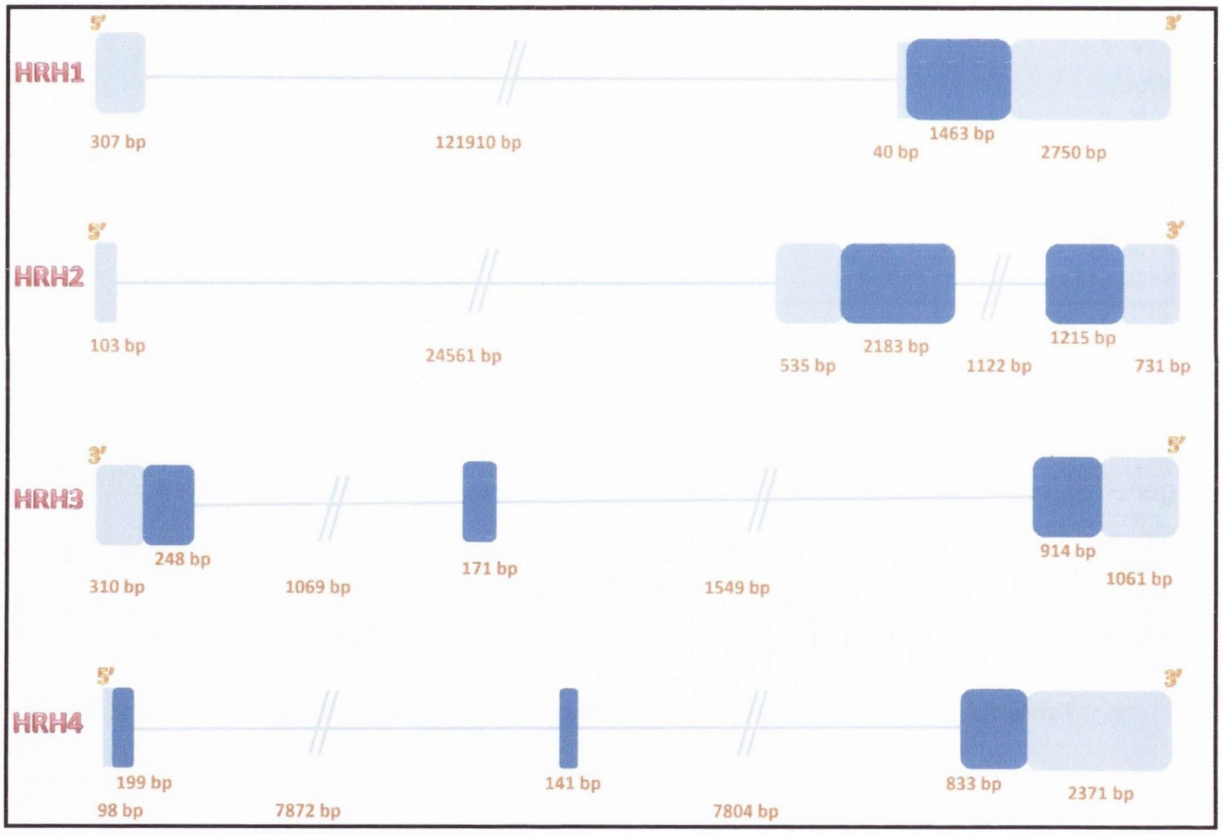


Figure 1.2 Genomic organisation of the human histamine receptors as per GRCH37.p10 Primary Assembly

(where the dark blue boxes represent the exons, while the light blue boxes represent the UTRs at both ends of the genes)

Initial histamine research focused on the role of histamine in allergic diseases, which resulted in the development of several potent antihistamine agents such as mepyramine. The first generation of antihistamines developed to treat allergic conditions were HRH1 antagonists, but unfortunately these were shown to ultimately induce marked side effects such as sedation. These ligands crossed the blood-brain-barrier hence newly generated antihistamines were made with the appropriate structural modifications to prevent this physiological effect [12]. These modifications resulted in new non-sedating HRH1 antagonists that are now widely used against common allergic reactions such as cetirizine, fexofenadine and loratidine.

As the actions of histamine continued to be studied, it was observed that antihistamines failed to antagonize the full spectrum of histamine-induced effects. This subsequently led to the discovery that histamine has a prominent physiological role in controlling gastric acid secretion, later confirmed with the discovery of HRH2 antagonists. The first ligands identified to target this receptor, by Sir James Black and his research group, were burimamide and cimetidine [13], proved critical in the treatment of peptic ulcers and gastro–oesophageal reflux disease. Following the success of cimetidine, further compounds such as ranitidine were developed, which proved even more potent, and unlike cimetidine they did not affect cytochrome *P450* in the liver; a major drawback for the ligands initially synthesised by Black *et al* [12].

Histamine receptor HRH1 and HRH2 subtypes have proven to be excellent drug targets, however, the more recently discovered histamine receptors HRH3 and HRH4 remain under investigation. Ligands for HRH3 are currently in clinical studies. Some companies have HRH3 antagonists for Phase 1 and Phase 2 clinical trials under review which could provide potential treatment for Alzheimer's disease, ADHD, schizophrenia, epilepsy, narcolepsy, obesity, neuropathic pain and allergic rhinitis [1]. The HRH3 antagonist pitolisant is now in Phase 3 clinical trial for treatment of narcolepsy and excessive daytime sleepiness associated with Parkinson's disease. The recently discovered H_4 receptor is still undergoing intense preclinical research, however novel H_4 receptor antagonists are currently undergoing clinical trials such as; UR-63325 which recently completed Phase 2 clinical trial for allergic rhinitis [14], JNJ-39758979 completing Phase 2 clinical trial for patients suffering from asthma and PF-3893787 which is in Phase 1 clinical trial for asthmatic patients subjected to allergen change [15].

Considering the inherent biological importance of histamine in various diseases, it is evident that looking into genetic polymorphisms for these receptors could turn out to be of clinical significance to the disease and its therapy. Histamine-related enzymes for anabolic or catabolic effects seem to be a major focus when it comes to polymorphisms and histamine, such as histamine *N*-methyltransferase (HMNT), histamine decarboxylase (HDC) and diamine oxidase (ABP1). Such examples include the identification of the allele variant Thr105Ile in HNMT that reduces thermal stability and a decrease of activity in the enzyme [16] and a recent identified polymorphism on ABP1 (rs10156191), which potentially impacts the development of hypersensitivity to non-steroidal anti-inflammatory drugs (NSAIDs) [17]. Such enzymes have proven to be critical in other diseases as well, such as HMNT and DAO, where genetic polymorphisms have been found in colonic adenoma patients [18]. In another study, both of these enzymes' activity was found to be significantly lower in colonic adenoma when compared to mucosa of healthy individuals [19]. However, further investigations need to confirm this linkage. Keeping that in mind, identifying a linkage between polymorphisms on histamine receptor genes and diseases is an important aspect since they are a major or potential drug target.

1.2 Genetic variations in the histamine receptor family

1.2.2 The histaminergic system in the Central Nervous System

The brain is continually processing information through aminergic systems present in the mammalian CNS such as the serotonergic, dopaminergic, noradrenergic and histaminergic systems which function by projecting signals throughout the brain and spinal cord [20]. Histamine has become progressively more apparent as one of the major neurotransmitters and is involved in a wide range of physiological functions such as cognition, emotion, feeding behaviour and sleep-wake cycle [12, 21]. Lately, *post mortem* studies have revealed alterations in the histaminergic systems in neurological and psychiatric diseases. Brain histamine levels are noticeably reduced in Alzheimer's disease patients whereas higher histamine concentrations are found in the brain of Parkinson's disease and schizophrenic patients. Low histamine levels are associated with convulsions and seizures. Neuronal histamine is also involved in pain perception, and ligands that cause an increase in the brain and spinal histamine concentrations have antinociceptive properties [22]. The histaminergic system seems to also contribute in the etiology of tourette syndrome, where a recent study highlights the

involvement of histamine pathways (mainly HRH1 and HRH2) to copy number variants identified in pedigree tourettes and autism spectrum disorder [23]. Due to the importance of maintaining healthy levels of histamine for normal cognitive function, genetic variations of histamine receptors have been extensively studied, Table 1.2 summarizes these findings.

1.2.2.1 Schizophrenia

An important objective in genetic research in schizophrenia (SCZ) is to identify target loci that can contribute or improve treatment of the disease. Taking into account the importance of histamine and its receptors in many biological processes, studies looking into genetic variations are of high interest as it provides an appealing approach to the investigation of the disease. Schizophrenia is a complex disorder and has a clear absence of a target gene or a biological marker. Various attempts have been made in identifying polymorphisms of HRHs in patients diagnosed with schizophrenia. The histaminergic system proved a good angle in the quest to identify genetic targets for SCZ after it has been found that levels of N-tele-methylhistamine, a major brain histamine metabolite, was elevated in the cerebrospinal fluid of schizophrenia patients and HRH1 binding sites decreased [24-26]. Evidence for a potential role in the disease is also seen by the different binding affinity that therapeutic drugs have for SCZ. Clozapine, a common drug used in SCZ treatment, is an atypical antipsychotic which has a diverse binding profile to all four aminergic systems (serotoninergic, dopaminergic, histaminergic and noradrenergic) [27]. Clozapine and amitriptyline were found to bind to HRH4 in low micromolar range where clozapine showed partial agonism to HRH4 and a high binding affinity to HRH2, both in a clinically relevant concentration. Other antipsychotics were also studied on all four receptors, where almost all showed a high bind affinity to HRH1 [28, 29]. Studies focusing on HRH2 and schizophrenia so far have been controversial. An allele variant identified to cause an amino acid exchange of Asn217Asp was suggested to be more common in schizophrenia patients [30, 31], however this could not be replicated in other studies [32]. To date no strong association has been linked between HRH2 and the disease, yet famotidine, a selective HRH2 antagonist, has been successfully used to treat the negative symptoms of schizophrenia [33]. Low brain distribution of available HRH2 antagonists/inverse agonists most probably contribute to the missing significance of data. The interactions between antipsychotics and the four histamine receptor subtypes generate an interesting perspective for the involvement of HRHs and the disease.

Genetic variations were also found to play a role in depression in bipolar disorder, where the SNPs in the dopamine 3 receptor (D3R) and HRH1 indicated to a significant association to the improvement of the condition following olanzapine and fluoxetine treatment [34].

1.2.2.2 Parkinson's disease

Parkinson's disease (PD) is a neurodegenerative disorder resulting in dopamine depletion. The histamingeric system has been studied in relation to dopamineric related diseases such as PD and it was observed that PD patients have an increased density of histaminergic fibers, and considerable interaction between histaminergic fibers and dopaminergic neurons in the substantia nigra [35]. To date HRH1 and HRH2 polymorphisms have been studied for association studies to the disorder as shown in Table 1.2. In a study focusing on SNPs for histamine related enzymes to PD, both HNMT and ABP1 were looked into. The nonsynonymous polymorphism for ABP1, His645Asp, showed no association with the disease, however the SNP on HNMT showed some evidence of associations with this neurological disorder. The genetic polymorphism located on exon 4 of the gene in HNMT causing a functional amino acid exchange Thr105Ile proved to have an association with Parkinson's disease since Ile105 leads to a lower activity of the enzyme [36, 37], yet in another study the same allele showed to have no direct association to PD and tremor [35]. It is pertinent to note that in the first two studies showing this association, patient samples were of mainly European descent, whereas the other study showing no significant difference were from North American patients; thus any discrepancy in their results could be based on the grounds of variation of ethnic background and genetic diversity between populations.

1.2.2.3 Alzheimer's disease

Alzheimer's disease (AD) is a progressive neurodegenerative disorder where cognitive deficits and memory loss is observed in patients. Studies have shown that histamine levels are significantly reduced in hypothalamus, hippocampus and temporal cortex, and it is suggested that there is a degradation of histaminergic neurons in the pathogenesis of AD [38, 39]. Imbalance in the histamine levels triggered studies on the only enzyme that metabolises histamine in the brain, HNMT. The polymorphism found on HNMT, which was associated with PD [36, 37], Thr105Ile, was also looked into for biological significance in AD, but there was

no correlation between this functional polymorphism and the disorder, unlike in PD and attention deficit hyperactivity disorder [40] studies. As yet, there is no direct relation between genetic polymorphisms in AD and the HRHs but newly synthesised drugs, such as H3 antagonists, are being investigated in clinical trials to treat cognitive disorders in AD since it was shown to improve attention, vigilance, and wakefulness [41].

1.2.2.4 Migraine

A polymorphism which was linked to disorders of the CNS was an amino acid substitution on the HRH3 third intracellular loop, A280V, which was found to be a risk factor for migraine [42]. Migraine is a neurovascular disorder and histamine is thought to participate in the pathophysiology of the disease [43]. The V allele for this SNP was found to be higher in patients when compared to controls, and especially the genotype VV or VA was found to be 12.92% when compared to controls with 3.22% (P = 0.001). This polymorphism is essentially the same variation as reported in the Shy-Drager syndrome study [44] (Table 1.2). In a recent publication a functional implication of this amino acid exchange was suggested as it reduces signalling efficacy in HRH3 when expressed in CHO-K1 cell line [45]. Therefore, this variation can indicate a possible contribution to diseases affected by this SNP.

Migraine tends to be more frequent in patients with allergic conditions and it was shown that histamine levels tend to increase in at-risk patients [42, 46, 47]. With the increased interest in the HMNT mutation, Thr105Ile, known to decrease enzyme activity, was also investigated for a possible contribution to migraine. However, the overall finding was that there was no clear evidence that Thr105Ile can be associated with the risk of migraine [48]. Even though the pathway and role of mediators in migraine are not fully understood overall, these findings suggest a possible role for HRH3 in this condition, especially due to its location in the CNS.

Table 1.2 Polymorphisms on HRH affecting central nervous system conditions

Histamine receptor	rs number	SNP	Location on gene	Condition	Functional consequence	PubMed ID
HRH1	NA[†]	G/C	Promoter (intron)	Schizophrenia	Significant LD ($P< 10^{-6}$) found in the excess of C allele but no strong association with SCZ in multiple testing. However, the G/C allele may represent a marker for more functional polymorphisms located further upstream.	12429384 [49]
HRH1	rs346074 rs346070	A/G C/T	Intron 1 3' UTR	Schizophrenia	Significant association between patients using antipsychotics with high affinity for HRH1 (clozapine, olanzapine and quetiapine) in obesity and BMI; where haplotype for rs346074-rs346070 A-T in obesity $P = 0.025$ and haplotype A-T in BMI $P = 0.005$.	21336576 [50]
HRH1	rs2067467	G/T	Exon 1	Schizophrenia	No association for antipsychotic induced weight gain in a Chinese population diagnosed with SCZ.	21937795 [51]
HRH1	rs2067470	C/T	Exon 1	Parkinson's disease	Results in a Leu449Ser amino acid substitution might be related to PD	18366640 [52]
HRH2	rs2067474	G/A	Promoter	Schizophrenia	Significant LD ($P< 10^{-6}$) found in patients' response to clozapine but this relationship was not observed in multiple testing possibly due to small patient sample. Could have a potential role in patient outcome to clozapine treatment upon further studies.	12429384 [49]

Histamine receptor	rs number	SNP	Location on gene	Condition	Functional consequence	PubMed ID
HRH2	rs2067474	G/A	Promoter	Parkinson's disease	Not a major risk factor for PD	18366640 [52]
HRH3	rs3787430 rs3787429	C/T C/T	Exon 3 Exon 3	Schizophrenia	Strong association between rs3787430(P = 0.024, 0.010) and a positive response to risperidone treatment in a Chinese Han population after 4 and 8 weeks of treatment respectively. This polymorphism could be a potential genetic marker for predicting treatment outcome or response. Association for rs3787429 was only noted after 4 weeks (P = 0.013)	21652606 [53]
HRH3	NA*	C/T	Exon 3	Shy-Drager syndrome	A polymorphism C839T causing a Ala280Val exchange was observed in one patient which was associated to the disease state.	11956964 [44]
HRH4	rs4483927	G/T	Intron 1	Schizophrenia	Significant association for risperidone efficacy where the TT genotype predicts poor response to treatment both on the positive, negative, and general subscales and on the total scale of PANSS scores (P = 0.017, 0.019, 0.021, 0.002 respectively). The SNP could be a biological marker for therapeutic response.	23422377 [54]

NA denotes rs number was not assigned and NA† denotes rs number could not be identified*

1.2.3 The histaminergic system in inflammatory conditions

Besides the central role histamine plays in allergies and autoimmune diseases, it was shown by various research studies that histamine can strongly influence in immune response [55-57]. Histamine is stored in specialised granules in mast cells and basophils, and the level of histamine is controlled by two enzymes; diamine oxidase (DAO), and histamine *N*-methyltransferase (HNMT) [58]. Different receptor affinities to histamine and different cell types expressing the receptor trigger distinct biological responses leading to numerous inflammatory disorders.

Histamine levels have been studied in various inflammatory diseases suggesting an association of these conditions to histamine receptors. Histamine levels are found to be elevated in bronchoalveolar fluids extracted from asthma patients, in the skin and plasma of patients with atopic dermatitis, in chronic urticaria biopsies, in both plasma and synovial fluid of patients with rheumatoid arthritis and in the plasma of patients with psoriatic arthritis [59]. The major receptors shown to be linked to inflammation and immune response are HRH1, HRH2 and HRH4 [60]. HRH1 activation increases histamine release and the release of other mediators, while also augmenting the pro-inflammatory activity of the immune system by promoting migration to the area of inflammation. On the other hand, HRH2 works by suppressing inflammation by decreasing eosinophil and neutrophil chemotaxis; decreasing IL-12 by dendritic cells; increasing IL-10 and inducing the development of Th2 or tolerance-inducing dendritic cells. HRH4 promotes the accumulation of inflammatory cells at sites of allergic inflammation by increasing calcium flux in human eosinophils; increasing eosinophil chemotaxis and increasing IL-16 production [58, 61].

1.2.3.1 Allergy and airway inflammation

To date HRH1 has been the major target in the treatment of allergic conditions. HRH1 antagonists have been used for decades to treat allergies, and although they provide symptomatic relief for most allergic conditions such as hay fever, they are not as effective for allergic asthma [62]. Recent goals in the histamine research include establishing a clearer role for the new receptor HRH4, with inflammatory disorders being one of the main targets. Only few genetic variations on this gene have been associated so far with asthma and underlying

conditions, as shown in Table 1.3. These SNPs (rs17187619, rs527790, rs487202, rs1421125, rs615283) do not give a missense mutation since they are not located in an exon leading to an amino acid exchange, yet they still seem to contribute to the condition.

Other pharmacological approaches focusing on the HRH4 ligands have shown that it is a promising target for the treatment of allergic conditions such as asthma and allergic rhinitis. The HRH4 antagonist, UR-63325, which recently completed phase 2 clinical trials [14], has been shown to be a good candidate for allergic asthma and rhinitis. Overall, this is a promising step forward for HRH4s' contribution to airway inflammation.

A recent publication has identified polymorphisms on the HRH4 that are associated to allergic atopic dermatitis (AAD). Atopic dermatitis is a complex disorder, which is caused by airborn and food allergens and is characterized by eczematous skin lesions and pruritus. Studies have shown that the HRH4 is expressed in dendritic epithelial cells and can play a role in regulating histamine levels. It was suggested that HRH4 agonists can be beneficial in downregulating inflammation in AAD [63]. One of the SNPs identified in this linkage study was rs77485247, a nucleotide exchange found in the 5'UTR which is located 51 bp upstream to the gene [60], representing the potential promoter region of the gene. A variation in this part of the gene could be responsible for changes in gene expression [64], but further investigations need to be carried out for confirmation. Two other polymorphisms are located in exon 3. The nonsense mutation of SNP rs74604924 leads to a stop codon, whereas rs77041280 results in a missense mutation leading to an amino acid exchange. These SNPs were also looked into for a genetic association to systemic lupus erythematosus but no significant association was found, however copy number variants found on the HRH4 were identified as a significant link to the disease [65].

Knowing the potential linkage of SNPs to AAD, it becomes natural to try and identify ligands that can treat or prevent this condition. Known HRH4 antagonists, JNJ7777120 and JNJ28307474, were studied for their potential use in canine models in comparison to HRH1 antagonists, cetirizine and hydroxyzine but it was shown that there was no impact on the skin lesion prevention for AAD [66]. Looking into the homology for the histamine 4 receptor, it only shows 71% homology between canine animals and humans [67], thus the polymorphisms outlined here might be human specific. Hence, these flagged polymorphisms could result in a possible target for HRH4 ligands leading to a diagnostic tool or potentially personalised medicine.

1.2.3.2 Gastric acid secretion

A polymorphism found in the promoter region of the HRH2 was identified to be associated with gastric mucosal atrophy [68], which is interesting considering that HRH2 blockers are the major therapeutic strategy used against gastric acid disorders. Atrophic gastritis is a chronic inflammation in the muscosal lining and together with *H. pylori* infections and other underlining conditions, can even result in gastric cancer if left untreated [69, 70]. The GG genotype of rs2607474 proved to give a significant increase in the risk of gastric mucosal atrophy, especially in *H. pylori* infected patients as shown in Table 1.3. The minor allele frequency for the GA genotype in gastric atrophy patients compared to controls was 8.02% and 13.3% respectively ($P = 0.057$), with the GG genotype being significantly increased in the patient group ($P = 0.055$) [68]. As a result, this polymorphism in particular seems to provide a biological marker for this condition and can be used as a tool to identify the risk and severity of gastric mucosal atrophy in combination with *H. pylori* infection.

Table 1.3 ***Polymorphisms on HRH affecting inflammatory conditions***

Histamine receptor	rs number	SNP	Location on gene	Condition	Functional consequence	PubMed ID
HRH1	NA*	C/T	Promoter	Aspirin-induced urticaria/ angioedema	No significant difference	15953854 [71]
	NA*	G/A	Exon3			
HRH2	NA*	G/A	Exon 1	Aspirin-induced urticaria/ angioedema	No significant difference	15953854 [71]
	NA*	C/T	Exon 1			
HRH2	rs2607474	G/A	Promoter	Gastric mucosal atrophy	Suggested as a risk factor for gastric muscular atrophy under the influence of *H. pylori* infections; where in GG homozygote had a significant risk factor in gastric atrophy (P = 0.042) and an enhanced effect in patients aged above 60 years (P = 0.022), and a more significant association in *H. pylori* infections (P = 0.023).	22720301 [68]
HRH3	rs6062144	C/G	Upstream 5'UTR	Asthma	From 8 SNPs on HRH3 tested for association, this SNP located 8 kbp upstream to the gene was the most associated (P=0.036), but there was no overall evidence for association.	19824886 [72]

Histamine receptor	rs number	SNP	Location on gene	Condition	Functional consequence	PubMed ID
HRH4	rs17187619 rs527790 rs487202 rs1421125 rs615283	T/C G/A C/G C/A G/A	Intron 2 After H4 After H4 3'UTR promoter	Asthma	rs17187619, rs527790, rs487202 - Significant difference in patients with or without infection-induced asthma (P = 0.002, P = 0.0002, P = 0.00007 respectively) rs1421125, rs487202 & rs527790 – Associated with asthma patients with or without allergic rhinitis (P = 0.005, P = 0.006, P = 0.006). rs615283 - Associated with conjunctivitis in asthmatics in GG genotype (P = 0.008). rs487202 & rs527790 –Where the CC allele combination for these SNPs was found to be associated with infection-induced asthma (P = 0.0009) rs487202 & rs574913 – Where CA haplotype was more common in infection-induced asthma patients (P = 0.0006)	22653292 [73]
HRH4	rs77485247 rs74604924 rs77041280	A/T A/T A/T	5'UTR Exon 3 Exon 3	Atopic dermatitis	SNPs suggest a genetic linkage to Atopic Dermatitis where rs74604924 gives a functional amino acid exchange Lys376Ter[*] and rs77041280 gives an amino acid exchange Ile376Lys (P = 0.002, P< 0.001, P< 0.001, respectively)	20199554 [60]

Histamine receptor	rs number	SNP	Location on gene	Condition	**Functional consequence**	**PubMed ID**
HRH4	rs77485247	A/T	5'UTR	Lupus erythematosus	No significant association for these SNPs	20618322 [65]
	rs74604924	A/T	Exon 3		(P <0.694, P <0.439, P <0.561 respectively)	
	rs77041280	A/T	Exon 3			

NA denotes rs number could not be identified from publication*

1.2.4 The histaminergic system in cancer

Histamine has also been shown to be a contributing factor to cancer. Histamine seems to play a role in cell proliferation in malignant cells, where high histamine biosynthesis was reported in different human neoplasias [74, 75]. It has been proven that HRH1 and HRH2 are expressed in normal and malignant cells lines, and that HRH3 and HRH4 are expressed also in cell lines of human mammary gland [75]. Studies focusing on the latter pair of receptors also demonstrated that the presence in human mammary tissue with 67% for HRH3 being detected in benign lesions and in all carcinomas, while 14% for HRH4 being detected in benign lesions and 44% in carcinomas [76]. The same research group also investigated the effect HRH4 agonists have on triple negative human breast cancer, and they showed that the agonists can significantly increase survival rate and tumour cell apoptosis [77]. These findings suggest HRH3 and HRH4 could be potential molecular therapeutic targets.

Some SNPs have been identified for HRH in cancer (Table 1.4). Looking into breast cancer research, HRH4 polymorphisms have recently been associated with the malignancy of the disease. This was evidenced upon the analysis of genotypes and haplotypes of a Chinese Han population; this study showed that SNPs rs623590, rs11662595, rs1421125 had a significant association with the risk and malignancy of breast cancer [78]. Further investigations need to be undertaken in order to demonstrate if these results are replicable in other populations and understand if this linkage is ethnicity-based or not. Studies related to gastric carcinoma and HRH4 have also been carried out, where no significant difference was found between copy number variations on HRH4 and patients with gastric carcinoma, but there was a substantial correlation between copy number deletions and HRH4 down regulation. This suggests that abnormalities in HRH4 have a potential role in histamine-mediated regulation of tumour growth in gastric cancer [79]. Analysis carried out on the HRH2 in gastric carcinoma identified promoter polymorphism rs2607474 to be associated with the disease, especially the GG homozygous genotype. This polymorphism was identified by the same group to be associated with gastric mucosal atrophy as previously mentioned. This association with gastric cancer was even higher at an advanced age where the atrophy and metaplasia scores were found to be higher [80]. Due to the location of this SNP in an important region for gene regulation, it seems like it has an important implication in gene transcription, but further analyses are needed to prove this biological significance.

Table 1.4 ***Polymoprhisms on HRH potentially affecting cancer***

Histamine receptor	rs number	SNP	Location on gene	Condition	Functional consequence	PubMed ID
HRH1	rs2606731 rs346076 rs346070 rs901865	G/T A/G C/T G/A	Intron 1 Intron 1 3'UTR 5'UTR	Opioid use in cancer pain	Opioid efficacy in cancer patients for pain is not genetically associated with these SNPs.	21398039 [81] 21570824 [82]
HRH2	rs1800689	G/A	Exon 1	Non-melanoma skin cancer	SNP is not associated with UV-induced immunosuppression and risk of non-melanoma skin cancer	21760883 [83]
HRH2	rs2607474	G/A	Promoter	Gastric cancer	Homozygous GG genotype was found to be associated with gastric mucosal atrophy (P = 0.0052) with a subsequent development of gastric cancer (mostly intestinal, P = 0.0047) especially at an advanced age.	22615049 [80]
HRH4	rs623590 rs16940762 rs11662595 rs1421125	A/C/T A/G A/G A/C	Intron 1 Intron 1 Exon 3 3'UTR	Breast cancer	T allele in rs623590 had a decreased risk of breast cancer (P = 0.012), while A allele of rs1421125 had an increased risk of cancer (P= 0.008) The haplotype C-A-A of rs623590 - rs11662595 - rs1421125 was found to be more common in breast cancer patients (P= 0.003)..	23481304 [78]

With the completion of the Human Genome Project in April 2003, numerous studies have emerged which help to understand the genetic basis of many common phenotypes of biochemical importance. Several diseases have been studied and are now known to be caused by a combination of multiple genetic and environmental factors. Deeper understanding of these factors can prove to be a useful tool in leading to better markers and targets to treat human diseases [84, 85]. The knowledge to date on HRH polymorphisms generates even more interest in identifying novel genetic variations that contribute to a disease, especially with the importance histamine has in biological processes.

The studies outlined here show a positive step forward in the quest for personalised care to patients. However, more genomic research in the context of on biogenic amines, here histamine, is needed to establish a direct association between diseases, biological markers and drug treatments in this new era of personalized medicine.

1.3 The human histamine receptor 4

Several research groups cloned and identified the new HRH member, the human histamine receptor 4 [2-5]. As previously mentioned in Table 1.1, its homology and sequence is more comparable to H3 receptor (31% at protein level, and 54% in the transmembrane domains) than to H1 and H2 receptors. Cloning HRH4 in mouse, rat, guinea-pig and pig revealed limited homology with the human receptor [6], which reduces the accuracy of using them as animal models. To date, only two isoforms have been identified for the HRH4 (full length is represented as H_4R_{390}) [6]; the H_4R_{67} isoform contains only 67 N-terminal residues and the H_4R_{302} has deletion of amino acids 68-155. Both of these isoforms (H_4R_{67} and H_4R_{302}) are non-functional. However when compared to the other member HRH3, 20 natural occurring isoforms have been identified so far, where some display differential expression patterns in various brain areas and functionalities compared with its respective full-length isoforms [1].

Similar to the other HRH family members, the HRH4 possesses all consensus motifs identified for Class A rhodopsin-like GPCRs, that is $N1.50^{33}$, $D2.50^{61}$, $R3.50^{112}$, $W4.50^{140}$, $P5.50^{186}$, $P6.50^{318}$ and $P7.50^{355}$ (represented in black circles in Figure 1.3), The receptor also has two potential glycosylation sites, Asn^5 and Asn^9 (marked with arrows in Figure 1.3) and a

potential palmitoylation site (Cys^{374}) at the C-terminal tail is suggested to be close to the membrane [6].

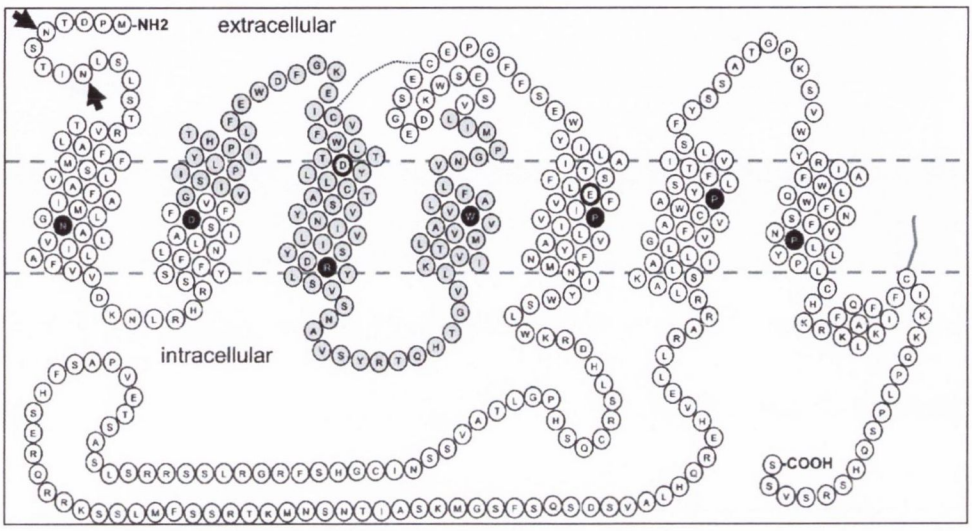


Figure 1.3 Snake plot of the human histamine H4 receptor [6]

The receptor consists of 390 amino acids, has three extracellular loops and three intracellular loops, with an extracellular N-terminal and an intracellular C-terminal peptide. The conserved residues in the family A of G protein-coupled receptors are depicted in black circles, while the putative glycosylation sites (Asn^5 and Asn^9) are indicated with arrows.

Investigations of potential therapeutic use for this receptor are strongly related to the localisation and expression of the receptor. The HRH4 is preferentially expressed in various cells of the immune system and mast cells. It also induces chemotaxis of eosinophils, mast cells and monocyte-derived dendritic cells. Other involvement of HRH4 was also reported in the control of interleukin (IL)-16 release from human lymphocytes and it has been speculated that a H4 selective antagonist might be useful in the treatment of asthma [2, 12]. Various research groups suggest that this receptor can also be a potential therapeutic target to other inflammatory diseases, such as chronic allergy, atopic dermatitis and inflammatory bowel disease. Other possible roles for HRH4 are rheumatoid arthritis and colorectal cancer since this receptor has been detected in primary synovial culture of these tissues [2].

1.3.1 The HRH4 and the central nervous system

Histaminergic neurons are expressed in the tuberomammillary nucleus (TMN) of the hypothalamus which project to all major areas of the brain and are involved in many functions including regulation of sleep/wakefulness, feeding and memory process [12]. The TMN neurons provide a widespread input to all areas of the CNS (Figure 1.4) but the most densely innervated target of histaminergic varicose fibres is the hypothalamus. Histamine is synthesised in the brain from L-histidine by the enzyme L-histidine decarboxylase (HDC). This enzyme acts as an active homodimer and is stored in synaptic vesicles via the vesicular monamine transporter-2 (VMAT-2), and released by exocytosis. The termination of its action in the brain requires its catabolism to *tele*-methylhistamine by the enzyme histamine N-methyltransferase (HNMT) [20].

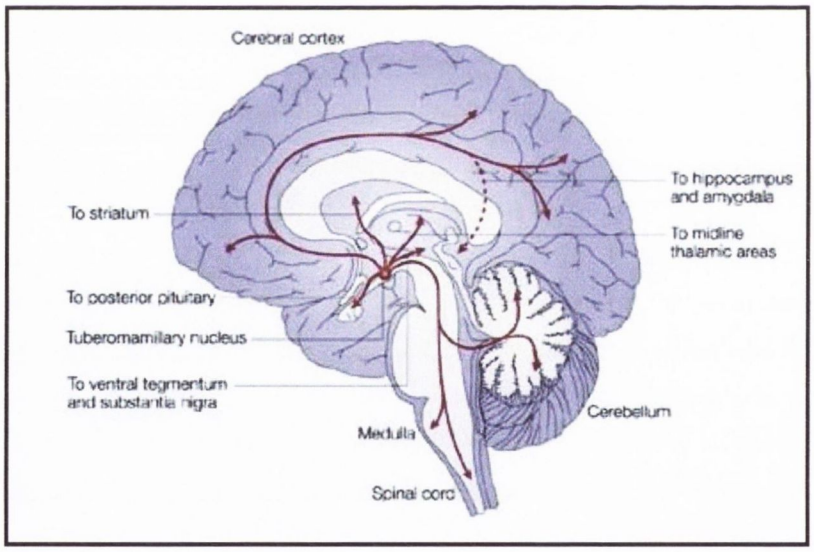


Figure 1.4 The histamingeric system in the human brain [86]

The histaminergic fibers emanating from the tuberomamillary nucleus project to the whole central nervous system

Receptors HRH1 and HRH2 are located postsynaptically and mediate excitatory actions of histamine. Activation of H1 receptors elicits increase in intracellular calcium (Ca^{2+}) and decreases potassium (K^+) conductances. Activation of H2 receptors stimulates adenylyl cyclase and triggers cyclic adenosyl monophosphate (cAMP)-dependent activation of protein kinase A which inhibits Ca^{2+}-activated K^+ channels. The HRH3 receptors act as somatodendritic or presynaptic receptors that inhibit adenylyl cyclase and N- and P-type Ca^{2+} channels, and activate K^+ currents. Presynaptic HRH3 heteroreceptors inhibit release of other neurotransmitters, including glutamate, γ-aminobutyric acide (GABA), acetylcholine, norepinephrine and serotonin. Histamine also increases Ca^{2+} influx via N-methyl d-aspartic acid (NMDA) receptors both by binding to the polyamine modulatory site and by HRH1 receptor-triggered receptor phosphorylation [87].

The major site for HRH4 expression is on hematopoietic cells but recent publications suggest that it is also expressed in the CNS. Immunological and pharmacological studies have suggested that HRH4 is expressed on fibres emanating from brain regions, most likely the hippocampus or thalamus [88]. Other publications suggest it is expressed in subpopulations of neurons within the sensory dorsal root ganglia and dorsal horn of the spinal cord and in restricted cortical thalamic areas of the brain [89]. So far, HRH4 expression in the brain is still controversial. Several groups could not detect mRNA of this receptor in the CNS [5, 28], while other research groups reported the presence in various parts of the CNS, including amygdala, cerebellum, hippocampus, caudate nucleus, substantia nigra, thalamus and hypothalamus [7, 8]. The reason for this discrepancy is still unclear, which makes this topic an attractive target for further investigation.

The pattern of expression of HRH4 also shows some similarities to the HRH3, such as overlap in entorhinal cortex and deep laminae of the cortex, but contrasts particularly in the relatively high expression in the thalamus [88]. With respect to the phylogenetic tree, since HRH3 and HRH4 exhibit close relationships it indicates that overlapping structure-affinity-efficacy relationship and pharmacology of their ligands can exist [90]. HRH3 and HRH4 antagonists are currently in preclinical development and in early stages of clinical trials for a range of cognitive, psychotic, sleep and inflammatory disorders, obesity and cancers [90].

1.3.2 Polymorphisms on the HRH4 gene

Considering the biological significance of histamine, polymorphisms of histamine-related genes are potential genetic factors likely to influence a disease or a therapy. Single nucleotide polymorphisms can show an effect at a phenotypic level. In humans, these SNPs occur in at least 1% of the population with at a frequency of 1 every 100 to 300 bases along the 3 billion-base human genome [84].

An interesting polymorphism for this receptor is in the promoter region of the HRH4 which is located 300 base pairs upstream to the ATG start codon. Even though these polymorphisms do not lie on an exon it is known that non-coding regions can still influence biological regulation [64]. The region is shown in Figure 1.5 below. This polymorphism at the HRH4 promoter region is highly complex and is represented by a CAA microsatellite triplet repeat that varies in length and nucleotide.



Figure 1.5 NCBI snapshot image for the VNTR region

Where blue is INDEL (insert-deletion mutation); and red is SNP (single nucleotide polymorphism)

This region is shown in Figure 1.5, taken from NCBI GRCh37.p10 primary assembly. The region is marked and cross referenced in Table 1.5 for major characteristics and rs-numbers of how it is currently reported.

Table 1.5 VNTR polymorphism in the promoter region HRH4

The marker on the first column corresponds to polymorphisms in the figure above

Label	dbSNP	Position in sequence	Type of mutation	Alleles	Deletion length
A	rs143647150	22,040,304 - 22,040,315	INDEL	-/AACAACAACAAC	12
B	rs200585616	22,040,304 - 22,040,306	INDEL	-/AAC	3
C	rs147906468	22,040,306 - 22,040,306	INDEL	-/AAC	1
D	rs 71987380	22,040,314 - 22,040,314	INDEL	-/CAA CAA CAA	1
E	rs374842887	22,040,331 - 22,040,331	INDEL	-/A	1
F	rs 67648501	22,040,332,-,22,040,332	INDEL	ACAACAACAA/C	1
G	rs367737858	22,040,335,-,22,040,345	INDEL	-/ACAACAACAAC	11
H	rs 72065050	22,040,339 - 22,040,339	INDEL	-/CAA CAA CA	1
I	rs201123969	22,040,344 - 22,040,345	INDEL	-/AC	2
J	rs 665411	22,040,344 - 22,040,345	SNP	A/C	2
K	rs 56887672	22,040,342 - 22,040342	INDEL	-/C/CA	1
L	rs201228052	22,040,345 - 22,040346	INDEL	-/AAA	2
M	rs 71925612	22,040,346 - 22,040,347	INDEL	-/AAA	2
N	rs 7288702	22,040,344 - 22,040,344	SNP	A/C	1
O	rs 670327	22,040,345 - 22,040,345	SNP	-/A/C	1
P	rs 60525833	22,040,344 -22,040,345	INDEL	-/C	1

This type of polymorphism is known as a variable number tandem repeat polymorphism (VNTR). This is one of the most common forms of length polymorphism and can be found in coding and non-coding regions of the genes. Microsatellite polymorphisms, i.e. 2-4 base pair-long tandem repeats, spread throughout the whole genome and can be used as genetic markers in the whole genome search, as well as in mapping of disease gene loci [91]. Currently, the NCBI published this sequence with of 14 triplet (CAA) repeats (Assembly GRCh37.p10). Preliminary unpublished data suggested that the microsatellite triplet found in 10-14 triplet repeats is more common in schizophrenic patients than controls [92]. Genotyping this region in 138 samples showed a difference in allele frequency (Table 1.6).

Table 1.6 ***VNTR allele frequency in Schizophrenic patients and controls***

VNTR length Genotypes	Allele frequency	
10	17.0%	Controls: 55% Patients: 73%
12	0.4%	
13	27.5%	
14	20.6%	
15	25.7%	Controls: 45% Patients: 27%
16	2.9%	
17	2.9%	
18	2.2%	
19	0.7%	

This variation in allele frequency needs further investigation in order to establish a significant link. Linkage studies for this microsatellite repeat can be used to investigate a possible link to schizophrenia. The complexity of this polymorphism in the promoter region could also influence gene expression causing a functional impact on the organism.

Another difference that emerged from this study was that the polymorphism found in Exon 3, rs11662595 which attributes to the amino acid exchange from Histidine to Arginine is also more common in schizophrenia patients (Table 1.7)

Table 1.7 ***Genotype frequencies for rs11662595 in Schizophrenia patients***

Genotypes	Frequency	Controls	Patients
Homozygous A/A	75%	61%	84%
Heterozygous A/G	19%	27%	13%
Homozygous G/G	7%	13%	3%

This investigation analysed 133 samples. The genotype frequency for homozygous A/A is higher in Schizophrenia patients than in controls, which could suggest a possible linkage, however the control group did not reach a Hardy-Weinberg Equilibrium [$(p+q)^2 = 1$]. Further analyses n on a larger cohort would be needed to achieve a clearer picture.

1.3.3 Schizophrenia and polymorphisms

Schizophrenia is a complex disorder with a lifetime risk of about 1% characterised by positive (i.e. hallucinations, delusions and racing thoughts), negative symptoms (apathy, lack of emotion, poor or non-existent social functioning) and cognitive deficits (disorganised thoughts, poor memory, difficulty in concentrating and/or following instructions) [93]. It can be caused by both genetic and environmental factors and their interactions. Traditionally, research for this condition targets the neurotransmitter systems in the brain, especially those involving dopamine [94]. This complex disorder however has a clear absence of a single target gene or a biological marker, unlike other neurological disorders, such as Alzheimer's disease and Fragile X syndrome (FMR1gene). In fact, it was noted that microsatellite repeats $(CGG)_n$ and $(CAG)_n$ in the androgen receptor were identified as the cause for the genetic diseases in Fragile X syndrome and Kennedy's disease, respectively. Both loss-of-function and gain-of-function mechanisms are recognised from expanded trinucleotide repeats, and such microsatellites can be located in both coding and non-coding region of the gene [64].

Several linkage studies have been carried out to identify a possible genetic link between schizophrenia and polymorphisms. A very popular approach in the age of high-throughput screening is running a genome wide association study (GWAS). A GWAS is a test of association between genetic markers, such as SNPs, across the genome and disease. A large number of clear associations between specific chromosomal loci and a complex human disease, such as type 2 diabetes and rheumatoid arthritis have been discovered by GWAS [95]. Due to the main complexity of schizophrenia and the inadequate phenotype cases, GWAS in this type of psychiatric conditions have lagged behind those in other common diseases. Nevertheless, some results have emerged that gave a number of specific conclusions which show linkage to this disorder. Gene loci that have been identified to possess a strong association to schizophrenia are summarised in Table 1.8

Table 1.8 Loci from GWAS in Schizophrenia

Gene	Gene ID	Symbol	Location	PubMed ID
Zinc Finger Protein 804A	91752	ZNF804A	2q32.1	18677311 [96] and 20458322 [97]
Major Histocompaitibility Complex	3133	MHC	6p21.3-22.1	19571881 [98] and 19571809 [99]
Neuregulin 1	3084	NRGN	8p12	15581715 [100]
Transcription Factor 4	6925	TCF4	18q21.1	2171550 [101]
Calcium channel voltage-dependent, L-type, alpha 1c subunit	775	CACNA1C	12p13.3	19621016 [102]
Notch 4	4855	NOTCH4	6p21.3	10932176 [103]
Disrupted in Schizophrenia-1	27185	DISC1	1q42.1	19414483 [104]

Another study which emerged from genetic polymorphisms in schizophrenia patients was linked to the protein Synaptosomal-Associated Protein of 25kDa (SNAP25). This gene is located on chromosome 20p12.2 and has been ranked no.11 in the Genome Scan Met-Analysis of Schizophrenia [105, 106]. This association study was carried out in an Irish family sample and even though significant markers emerged, an attempt to replicate this finding in case-control samples resulted in no SNPs being significant. Despite this lack of replication, functional studies of associated variants observed in this study may be useful given the weight of previous expression studies and its role in schizophrenia.

To date, evidence linking HRH4 and schizophrenia is limited, however studies on histamine suggest it plays a role in the disease, where higher histamine levels were found in brains of schizophrenia patients, following post-mortem examination [22]. Pharmacogenetics is progressively becoming important aspect for patient's health, where the drug targets more efficiently and effectively the disease and provides personalised care. The search for genetic variations and their contribution to the disease for the HRH4 so far produced one finding for rs4483927, which as mentioned previously, could be a possible therapeutic marker in risperidone treatment. However, due to the similarity of HRH4 and HRH3 (which is principally expressed in the CNS), this new member is still under intense pre-clinical research for neurological/psychiatric conditions.

1.4 Aims and objectives

Preliminary unpublished results suggested that polymorphisms located on the HRH4 gene can be linked to schizophrenia. This needed to be confirmed and expanded by further analysis; hence the aims of this project consist of:

- Testing for association for polymorphisms on the HRH4 gene locus on a larger sample cohort using GWAS analysis conducted from the International Schizophrenia Consortium (ISC).

- Genotyping HapMap samples for the promoter region of HRH4 and establishing linkage disequilibrium (LD) with known SNPs at the gene locus.

- Cloning the promoter region of HRH4 for all different types of variable number of tandem repeat polymorphisms (VNTRs) and SNP and creation of a plasmid library

- Reporter gene assay for determination of functional impact of identified and cloned polymorphisms for HRH4.

Chapter 2

Materials and Equipment

2 Materials and Equipment

2.1 Reagents and kits

Reagents and kits used in the experiments are listed below:

Reagents and Kits	Catalogue number	Supplier
2-Log DNA Ladder	N3200L	New England Biolabs®
Agarose high resolution	A4718-100G	Sigma-Aldrich®
Agarose multi-purpose	BIO-41025	Bioline™
Ampicillin sodium salt	A0166-5G	Sigma-Aldrich
BspH1	R0517S	New England Biolabs®
Deoxynucleotide (dNTP) mix 10mM Solution	D7295	Sigma-Aldrich®
Deoxynucleotide Solution Set	N0446S	New England Biolabs®
DpnI enzyme	R0176S	New England Biolabs®
Dual Glo luciferase system	E2920	Promega
Dual-Glo® Luciferase assay kit	E2920	Promega
EcoR1 restriction enzyme	R0101S	New England Biolabs®
Endofree Plasmid Maxi kit	12362	Qiagen® Ltd.
Ethanol (Absolute Alcohol)	64175	Hazardous Material Facility
Ethylenediamine Tetraacetic Acid (EDTA)	E-4378	Sigma-Aldrich®
Gel Loading Solution	G2526-5ML	Sigma-Aldrich®
GeneScan™ ROX 350™ Size Standard	401735	Applied Biosystems™
Gentra Puregene Cell Kit	158745	Qiagen® Ltd.
Hi-Di™ Formamide	4440753	Applied Biosystems™

Reagents and Kits	Catalogue number	Supplier
Human Genomic DNA	BIO-35025	Bioline™
Isohelix® DNA swabs pack	SK-2	Cell projects Ltd™
Isopropanol	67630	Hazardous Material Facility
Kanamycin Sulphate	K1377-5G	Sigma-Aldrich®
LB Broth EZMix™ Powder	L7658-1KG	Sigma-Aldrich®
MinElute Gel extraction kit	28606	Qiagen® Ltd.
MinElute® Reaction Cleanup kit	28204	Qiagen® Ltd.
Msc1	R0534S	New England Biolabs®
Mwo1	R0573S	New England Biolabs®
Nco1	R0193S	New England Biolabs®
***PfuUltra*® high-fidelity enzyme**	600380	Stratagene - Agilent Technologies
pGL3 basic vector	E1751	Promega
pGL3-SV40 vector	E1741	Promega
Phusion® High-Fidelity DNA Polymerase	M0530S	New England Biolabs®
pRL-SV40 vector	E2231	Promega
QIAamp® DNA Investigator kit	56504	Qiagen® Ltd.
QIAprep® Spin Miniprep Kit	27106	Qiagen® Ltd.
QIAquick® Gel Extraction kit	28704	Qiagen® Ltd.
QIAquick® PCR Purification kit	28104	Qiagen® Ltd.
Quick Ligation kit	M2200S	New England Biolabs®
SYBR® Safe DNA Gel Stain	S33102	Invitrogen™
Taq DNA Polymerase	M0273X	New England Biolabs®
TE Buffer RNase free 20X	6787-1	Molecular Probes®
Tris Base	161-0791	Bio-Rad ®

Reagents and Kits	Catalogue number	Supplier
Xho1 restriction enzyme	R0146S	New England Biolabs®
Zero Blunt® PCR Cloning kit with One Shot® TOP10 Chemically Competent *E.Coli*	K2700-40	Invitrogen™

2.2 Equipment and Manufacturers

Instrumentation	Manufacturer
ABI 3700XL Genetic Analyzer®	Applied Biosystems™
Agarose gel electrophoresis apparatus	Fisher Scientific™
Biological safety cabinets Class II	Nuaire™
Centrifuge 5804R	Eppendorf
Centrifuge Heraeus Fresco 17	Thermo Scientific™
ChemDoc® XRS	BioRad®
CO_2 air-jacketed Incubator	Nuaire™
DNA Engine Thermalcycler	BioRad®
Electrophoresis power supply E431	Consort GmbH™
FLUOstar OPTIMA	BMG Labtech
NanoDrop™ ND-1000 Spectrophotometer	Thermo Scientific™
Olympus CKX41 microscope	Olympus
Orbital Shaker	Sony
PCR Primus 96	PeqLab®
Synergy HT plate reader	BioTek
Thermomixer Comfort	Eppendorf
Vortex Genius 3	Ika®
Water bath	Precision Scientific™

2.3 Bacterial Strains

One Shot® TOP10 Competent Cells Invitrogen™

Genotype: F⁻ mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL endA1 nupG

BL21 (DE3) competent cells Stratagene Agilent

Genotype: (BF−*dcm ompT hsdS* rB−mB−)

2.4 Primers

2.4.1 Primers for HRH4

The primers below were designed to anneal to the HRH4 promoter in various regions, in addition different restriction enzymes or extensions for site-directed mutagenesis were added.

Oligo Name	Sequence (5'→3')	Tm (°C)	%GC content
AS10_for	GAAATTCAGCCCCTGTTACTCC (22)	60.3	50
AS11_rev	ACTGTGATGCATCCATCTCC (20)	57.3	50
AS16_for	CCTGAACCTGGGAGACAGAGG (21)	63.7	62
Seq1	GGTGATGAGTGCCTGGAATC (20)	64.6	55
Seq1-6FAM	[6FAM]GGTGATGAGTGCCTGGAATC (20)	64.6	52
Seq2F	TTGCAGCCACTGTTTTTCTG (20)	64.0	45
Seq3R	TGACCAAAGCATTTCCTAGC (20)	61.8	45
Seq4F	GAGAG**CTCGAG**ACCCCTGCTGTTCTTAAGC (**Xho1** restriction digest site)	71.0	57
Seq5R	CTTA**CCATGG**CACATCAAATGATGAAGAAG (**Nco1** restriction digest site)	70.0	42

Oligo Name	Sequence (5'→3')	Tm (°C)	%GC content
Seq7F	CAGATTGTACTGAGAGTGCACCATATG	66.4	44
Seq8F	GAGAG**CTCGAG**CTGCATCTTTTTGGCAGTG (**Msc1** restriction digest site)	79.9	54
LL SDM_For	**GTGCTAGCCCGGG****CTCGAG**AATAAATTTATA AGCAAATGG (**Xho1** restriction digest site; **pGL3 vector**)	67	45
LL SDM_Rev	GAAAAGCTTAAGAACAGCAGGGGTTGAGGG GG	66	53
XL SDM_For	**CGCGTGCTAGCCCGGG****CTCGAG**CTGCATCT TTTTGGCAGTG (**Xho1** restriction digest site; **pGL3 vector**)	75	63
XL SDM_Rev	**GCCATTTGCTTATAAATTTATTAAAATGAAG ATAAAAAGC**AATT (aligns with 3' end **-2772bp 10RpGL3**)	21	58
Ex1 SDM_For	GACCTTCTTCATCATTTG**ATGTGATGCCAGA TACTAATAGCAC** (Start of **luciferase gene** on pGL3)	65	40
Ex1 SDM_Rev	**CTTTATGTTTTTGGCGTCTTCCATGG**ACCCA CAAAGAAGTCAGAGATGG (Start of **luciferase gene** on pGL3)	70	45

2.4.2 Standard primers

Standard primers used for sequencing or analysis of transformants.

Oligo Name	Sequence (5'→3')	Tm (°C)	%GC content
RV primer 3 (for pGL3)	CTAGCAAAATAGGCTGTCCC	60.3	50
M13 Forward (for pCR-Blunt)	GTAAAACGACGGCCAG	57.6	56
M13 Reverse (for pCR-Blunt)	CAGGAAACAGCTATGAC	50.7	47

2.4.3 Primers for Beta-Actin

Designed primers used to check the quality and integrity of human genomic DNA extracted.

Oligo Name	Sequence (5'→3')	Tm (°C)	%GC content
b-actin_for_pink	GCCGTACCACTGGCATCGTG (20)	63.5	65
b-actin_rev_gre	CTCCTTAATGTCACGCACGATTTC (24)	61.0	46

2.5 Specialised Software and Databases

Software	Version	Manufacturer
A Plasmid Editor (ApE)	1.17	M.Wayne Davis
GeneMapper	3.00	Applied Biosystems™
Haploview	4.20	Board Institute

Database	Website
1000genomes	http://www.1000genomes.org/
HapMap Project	http://www.hapmap.org/
Sequence manipulation suite: Primer stats	http://www.bioinformatics.org/sms2/pcr_primer_stats.html
NCBI	http://www.ncbi.nlm.nih.gov/
Promoter 2.0	http://www.cbs.dtu.dk/services/Promoter/

2.6 Cell culture

Cell lines:

SH-SY5Y – –is a thrice cloned (SK-N-SH -> SH-SY -> SH-SY5 -> SH-SY5Y) subline of the neuroblastoma cell line SK-N-SH derived from human bone narrow and obtained as an aliquot from Centre for the Study of Neurological Disorders (CSND), Department of Physiology and Medical Physics, Royal College of Surgeons, Ireland.

HEK293T – is derivative from human embryonic kidney 293 cells, containing the SV40 T antigen a and obtained as an aliquot from School of Biochemistry and Immunology, Trinity Biomedical Sciences Institute, Trinity College Dublin.

Product	Catalogue number	Supplier
Dulbecco's Modified Eagle's Medium/ Nutrient Mixture F-12 Ham	D8437-500ML	Sigma-Aldrich®
Fetal Bovine Serum	F7524-500ML	Sigma-Aldrich®
Dulbecco Phosphate Buffer saline	D8537-500ML	Sigma-Aldrich®
Penicillin- Streptomycin	P4333-100mL	Sigma-Aldrich®

2.7 Preparation of solutions, plates and agarose gel

TAE (Tris/Acetate/EDTA) buffer:

Stock solution 50x:

- 242 g Tris base
- 57.1 mL Glacial acetic acid
- 37.2 g EDTA
- Dissolve in deionised water up to 1 L

Working solution was 1x at a pH~ 8.5 with final concentrations of:

- 40 mM Tris-acetate
- 2 mM $Na_2EDTA\ 2H_2O$

Agarose Gel Electrophoresis

A 1% agarose gel was prepared by weighing 0.5 g agarose powder and dissolving it in 50 mL 1x TAE buffer. The mixture was heated in a microwave and left to cool down, upon which 5 µL of SYBR® safe was added and mixed. The agarose was poured onto the tray with a well-comb and left to set. Once set, 1x TAE buffer was poured over it inside the electrophoresis apparatus and the comb was gently removed. The DNA to be analysed was mixed with the gel loading dye (10 µL and 2 µL respectively) and loaded in the wells. The gel loading dye gives density and colour to the loaded product and since dyes are negatively charged in

neutral buffers they move in the same direction as the DNA during electrophoresis. The DNA was measured against a 2-log DNA ladder (NEB) that was loaded in the first well. The 2-log DNA ladder gives a size range of 100 bp to 10 kb. The gel was set to run at a voltage of 90 V for 45 min. The PCR product was visualised using ChemDoc® XRS Molecular Imager (Bio-Rad) and Quantity One® imaging software version 4.6.

Kanamycin stock solution

A stock solution of 30 mg/mL was prepared by dissolving 300 mg Kanamycin (Sigma-Aldrich) in 10 mL of deionised sterile water. The stock solution was sterile filtered through a 0.2 µm filter and aliquots transferred into 1.5 mL eppendorf tubes. These aliquots were stored at -20°C for long term storage. Stock solutions were diluted 1:1000 in the final LB broth medium.

Ampicillin stock solution

A stock solution of 50 mg/mL was prepared by dissolving 500 mg Ampicillin (Sigma-Aldrich) in 10 mL of deionised sterile water. The stock solution was sterile filtered through a 0.2 µm filter and aliquots transferred into 1.5 mL eppendorf tubes. These aliquots were stored at -20°C for long term storage. Stock solutions were diluted 1:1000 in the final LB broth medium.

LB broth solution

The medium was prepared by dissolving 10.3 g LB broth EZmix® powder (Sigma-Aldrich) in 500 mL deionised water. This was autoclaved at 121°C for 20 min in a liquid cycle and allowed to cool down to approximately 50°C, upon which 500 µL of Ampicillin or Kanamycin stock solution was added (to a final concentration of 30 µg/mL). (Storage was at 4°C when not in use).

LB agar plates

To 500 mL of LB broth, 7.5 g agar was added. They were autoclaved at 121°C for 20 min in a liquid cycle and allowed to cool down to approx. 50°C, upon which 500 µL of either Ampicillin or Kanamycin stock solution was added (to a final concentration of 50 or 30 µg/mL respectively). The LB agar mix was poured in petri dishes under a laminar flow hood. The LB agar was allowed to set and harden and were then stored at 4°C inverted and dated.

2.8 DNA sequences

2.8.1 18RCpEX synthesis

The sequence below is the 18RC which was confirmed by Eurofins by sequencing:

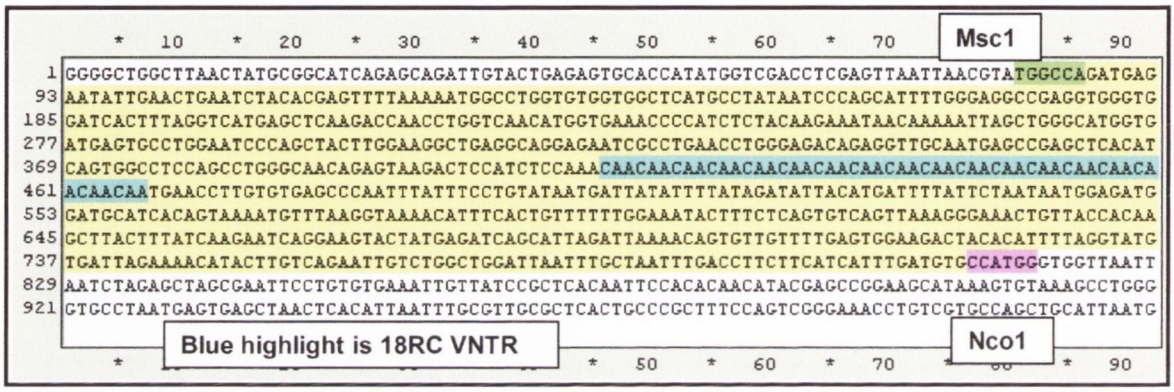


Figure 2.1 Synthesised sequence for 18RC

Restriction sites used for cloning are shown in this sequence, i.e. Msc1 (green) and Nco1 (prink). The sequence highlighted in blue represents the 18R VNTR region made of 18x CAA triplet repeat with the final nucleotide ending with '**C**AA'.

The vector specifications for this construct are shown below:

eurofins
mwg|operon

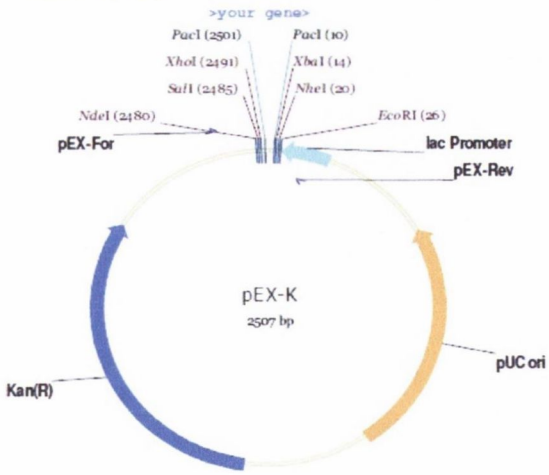
Gene Synthesis Quality Assurance Documentation

Order No.: **2780313/Sasse**

OI23757566

Plasmid Name:	pEX-K-18RC	**Internal Name:**	BE29-2
Gene Name:	18RC	**Gene Size:**	738bp
Vector Backbone:	pEX-K	**Antibiotic Selection:**	Kanamycin
Cloning:	via Type IIS restriction enzymes	**Quantity:**	1.1µg
Designation:	E. coli DH5alpha		
Genotype:	F⁻Φ80*lac* ZΔM15Δ(*lac* ZYA-*arg* F)U169*deo* R *rec* A1 *end* A1 *hsd* R17(r_k^-, m_k^+) *pho* A *sup* E44 *thi*-1 *gyr* A96 *rel* A1 λ⁻		

Plasmid Map



5' Restriction Site: Msc1
3' Restriction Site: Nco1
Cloning: via Type IIS restriction enzymes
(Type IIS sites not present in final plasmid)

MCS of pEX-K
Sequencing primers and *Pac* I
restriction sites underlined

GGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGC
TGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGgt
cgacCTCGAGttaattaaCGTA >your gene< GTGGttaattaaTCTAGA
gctagcGAATTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATA
CGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGA

Figure 2.2 *Vector specifications for 18RC*

2.8.2 19RCpEx synthesis

The sequence for 19RC confirmed by Eurofins is shown below:

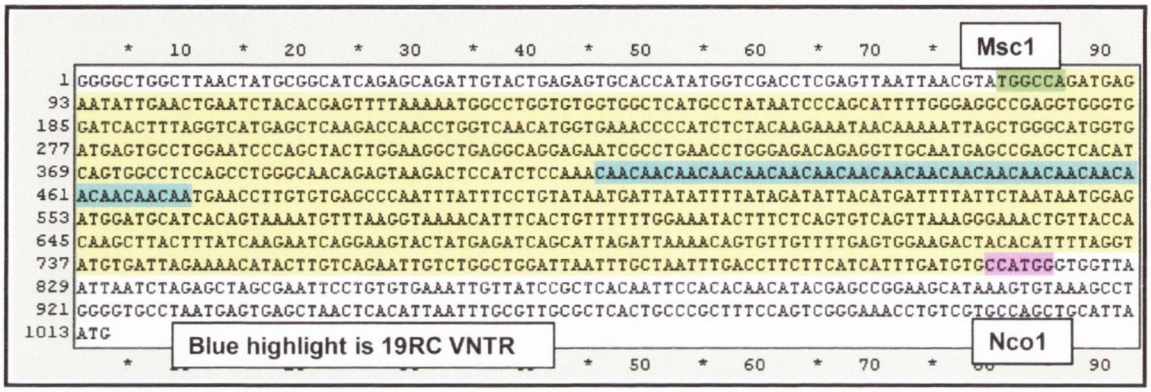


Figure 2.3 Synthesised sequence for 19RC

Restriction sites used for cloning are shown in this sequence, i.e. Msc1 (green) and Nco1 (prink). The sequence highlighted in blue represents the 19R VNTR region made of 19x CAA triplet repeat with the final nucleotide ending with '**C**AA'.

While the image below shows the vector characteristics:

eurofins
mwg|operon

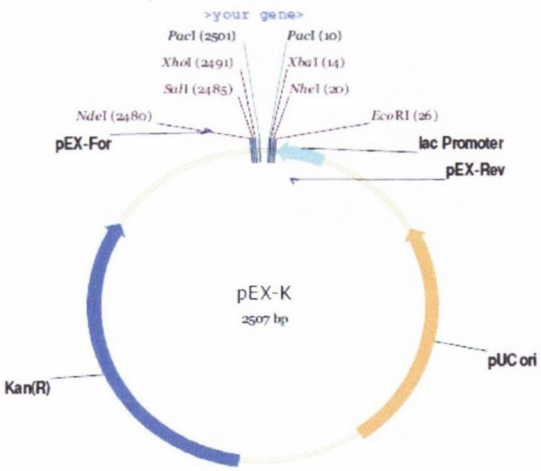
Gene Synthesis Quality Assurance Documentation

Order No.: **2777326/Sasse**

OI23725151

Plasmid Name:	pEX-K-19RC	**Internal Name:**	BD11-7
Gene Name:	19RC	**Gene Size:**	741bp
Vector Backbone:	pEX-K	**Antibiotic Selection:**	Kanamycin
Cloning:	via Type IIS restriction enzymes	**Quantity:**	2.7µg
Designation:	E. coli DH5alpha		
Genotype:	F⁻Φ80*lac* ZΔM15Δ(*lac* ZYA-*arg* F)U169*deo* R *rec* A1 *end* A1 *hsd* R17(r_k^-, m_k^+) *pho* A *sup* E44 *thi*-1 *gyr* A96 *rel* A1 λ⁻		

Plasmid Map



5' Restriction Site: Msc1
3' Restriction Site: Nco1
Cloning: via Type IIS restriction enzymes
(Type IIS sites not present in final plasmid)

MCS of pEX-K
Sequencing primers and *Pac* I
restriction sites underlined

GGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGC
TGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGgt
cgacCTCGAGttaattaaCGTA >your gene> GTGGttaattaaTCTAGA
gctagcGAATTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATA
CGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGA

Figure 2.4 Vector specifications for 19RC

2.8.3 XLpUC57 synthesis

Sequence sent for synthesis to Eurofins for the vector XLpUC57:

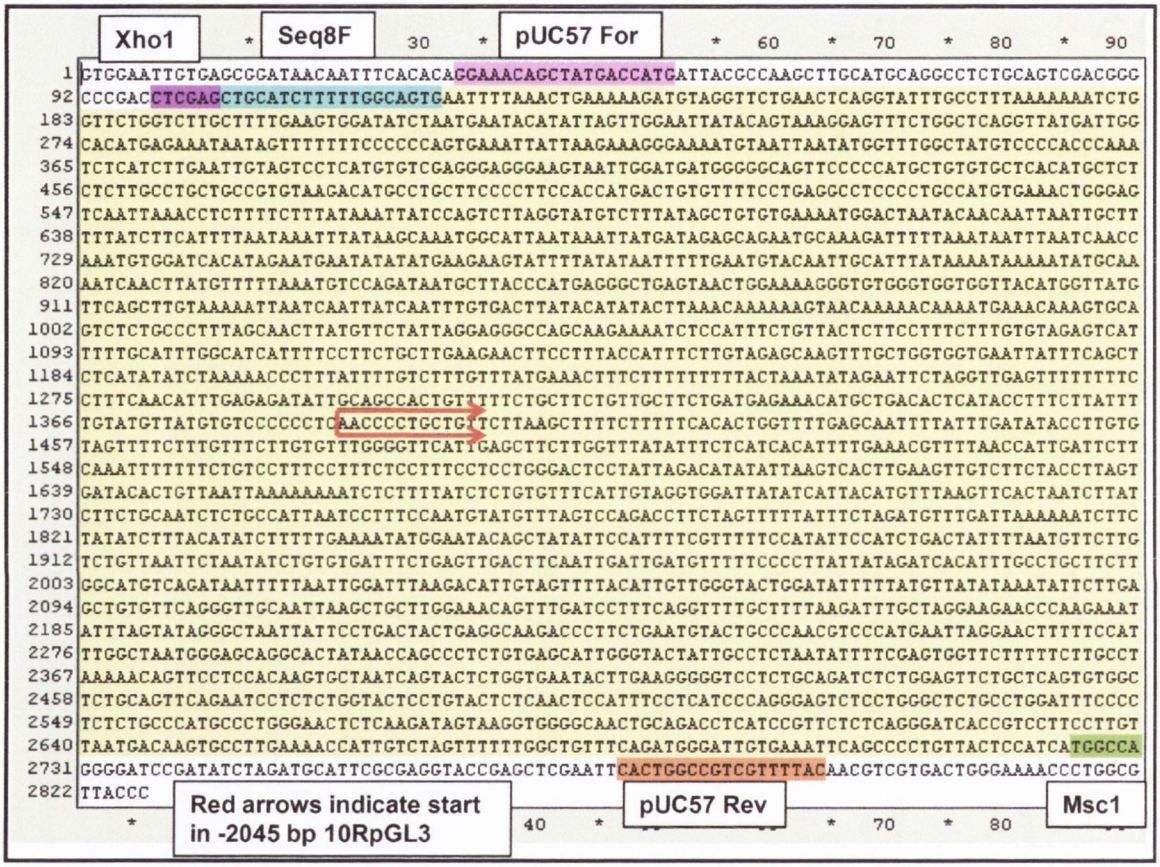


Figure 2.5 Synthesised sequence for XLpUC57

The primers Seq8F (blue), pUC57 For (pink) and pUC57 Rev (orange) were used to in order to amplify the region for cloning in -2045bp 10RpGL3 but the PCR was not successful. Hence, digestion from original plasmid was performed as described in Chapter 5.

Characteristics of vector:

eurofins
mwg|operon

Gene Synthesis Quality Assurance Documentation

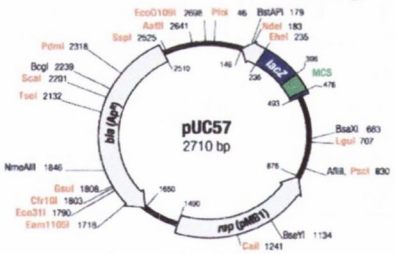
Order No.: **2878051/Sasse**

Plasmid Name:	pUC57-Extension LL promoter	**Internal Name:**	DA83
Gene Name:	Extension LL promoter	**Gene Size:**	2633bp
Vector Backbone:	pUC57	**Antibiotic Selection:**	Ampicillin
Cloning:	cloned blunt into SmaI	**Quantity:**	4.6µg
Designation:	E. coli DH5alpha		
Genotype:	F⁻ Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(r_k^-, m_k^+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ⁻		

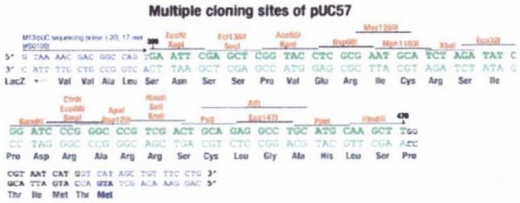
Plasmid Map

Insert: Extension LL promoter

5' Restriction Site: XhoI
3' Restriction Site: MscI
Cloning: cloned blunt into SmaI



Multiple cloning sites of pUC57



Please Note:

Verify sequence after each cloning step.
All material left at Eurofins MWG Operon will be discarded after 1 month.
Sequence analysis was done via doublestrand DNA sequencing. Sequence congruence was 100%.
Please find your sequence trace files in your personal account on our ECOM system (order ID 2878051).
The plasmid DNA has been lyophilised. We recommend to dissolve it in 10mM Tris buffer or TE.

Figure 2.6 Vector characteristics for pUC57

2.8.4 Mega-primer sequences

2.8.4.1 Mega-primer LL

This sequence was amplified by the forward primer LL_SDM_For (pink) having restriction site for Xho1 (blue) and the reverse primer LL_SDM_Rev (orange). The sequence highlighted in red over the reverse primer is the beginning of -2044 bp constructs.

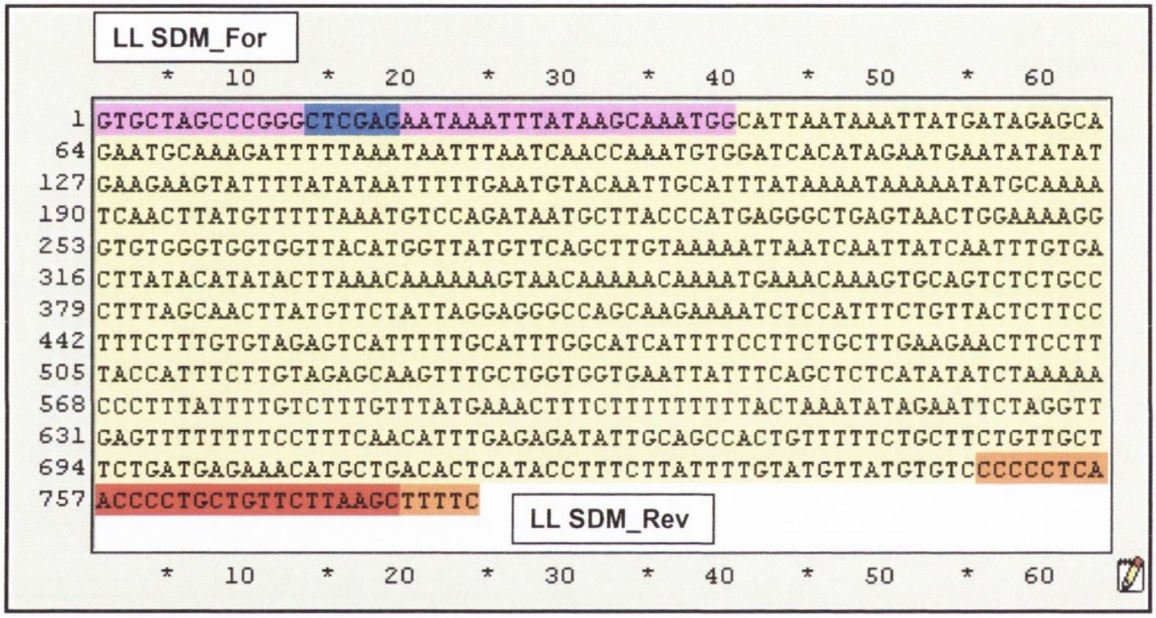


Figure 2.7 Sequence for amplified mega-primer LL

2.8.4.2 Mega-primer XL

This sequence was amplified by the forward primer XL_SDM_For (pink) having restriction site for Xho1 (blue) and the reverse primer XL_SDM_Rev (orange). The sequence highlighted in red over the reverse primer is the beginning of -2772 bp constructs.

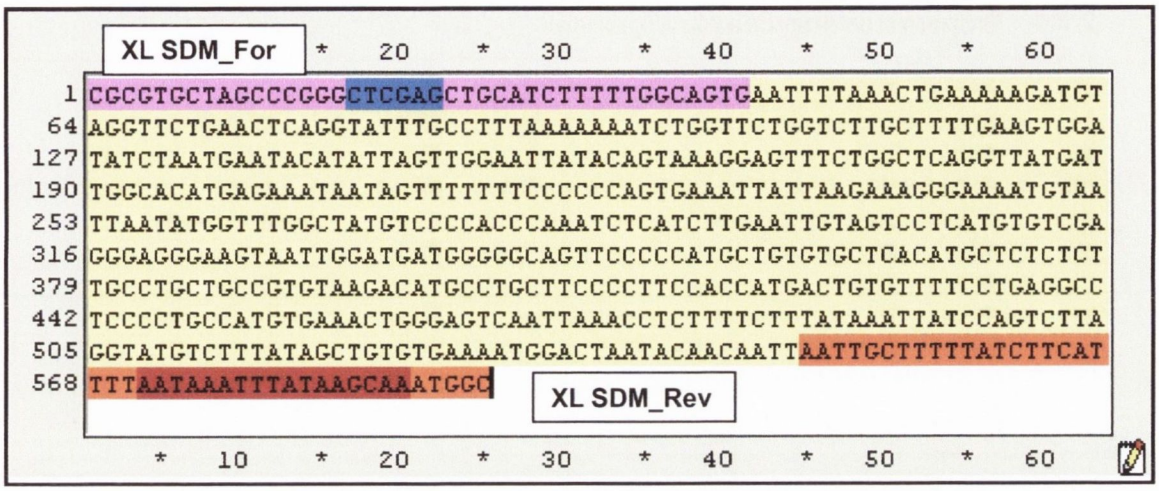


Figure 2.8 Sequence for amplified mega-primer XL

2.8.4.3 Mega-primer Ex1

This sequence was amplified by the forward primer Ex1_SDM_For (green) and the reverse primer Ex1_SDM_Rev (blue). The sequence highlighted in pink is the sequence for exon 1 HRH4, while the ATG in yellow represents the start codon of the luciferase gene..

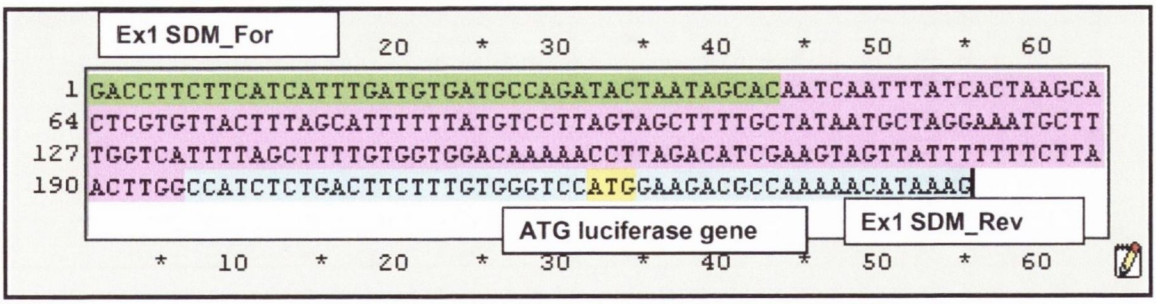


Figure 2.9 Sequence for amplified mega-primer Ex1

All clones constructed and their sequences can be found aligned to the proposed sequence in Appendix 3.

2.8.5 Proposed sequences for alignment

2.8.5.1 Proposed sequence for -2044 bp for the 10R VNTR in pGL3

Figure 2.10 shows the proposed sequence (used as a template for aligning sequencing results) for the -2044 bp 10RpGL3 clone.

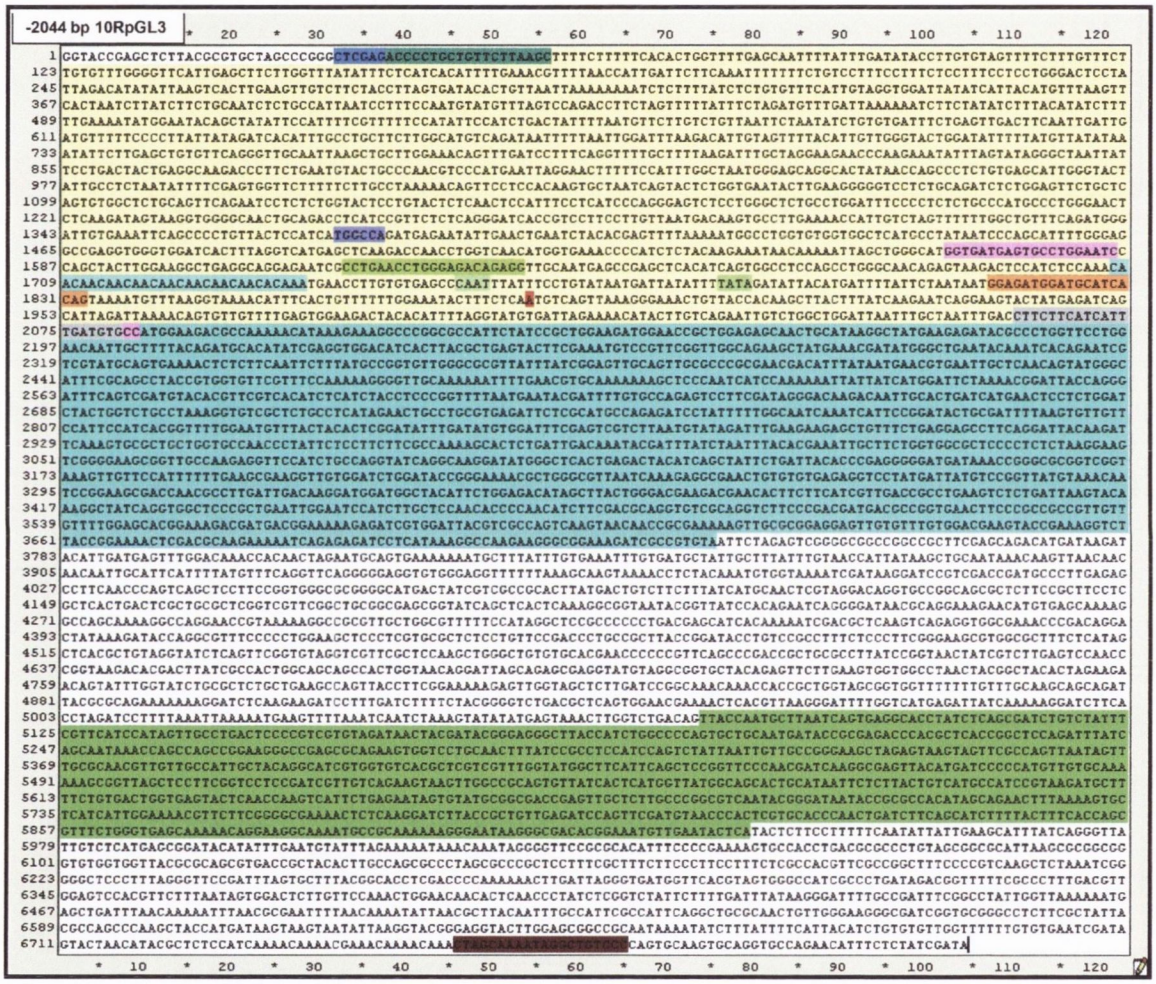


Figure 2.10 Proposed sequence for -2044 bp 10RpGL3

Starting chronologically the figure shows; Xho1 restriction site (in dark blue), Seq4F (in dark green), -2044 bp putative promoter sequence for HRH4 (in yellow), Msc1 restriction site (in purple), Seq1F (in pink), AS16For (in lime green), 10R VNTR region (in very light blue), CAAT box and TATA box (in green), AS11Rev (in orange), Seq5R (in grey with remaining nucleotide for Nco1 restriction site in pink), firefly luciferase sequence (in blue), ampicillin resistance gene (in green), and M13For (in brown).

The rest of the proposed sequences for the other genotypes are based on this and varying only in the VNTR region.

2.8.5.2 Proposed sequence for -3329 bp Ex1 for the 10R VNTR in pGL3

Figure 2.11 shows the proposed sequence (used as a template for aligning sequencing results) for the -3329 bp Ex1 10RpGL3 clone. Starting chronologically the figure shows; Xho1 restriction site (in pink), Longest length for the putative promoter region for HRH4 (in yellow), -2772 bp start (in red), -2044 bp (in red), Msc1 restriction site (in purple), Seq1F (in pink), AS16For (in lime green), 10R VNTR region (in very light blue), AS11Rev (in organe), Exon1 for HRH4 (in pink), firefly luciferase sequence (in blue), ampicillin resistance gene (in green), and M13For (in brown).

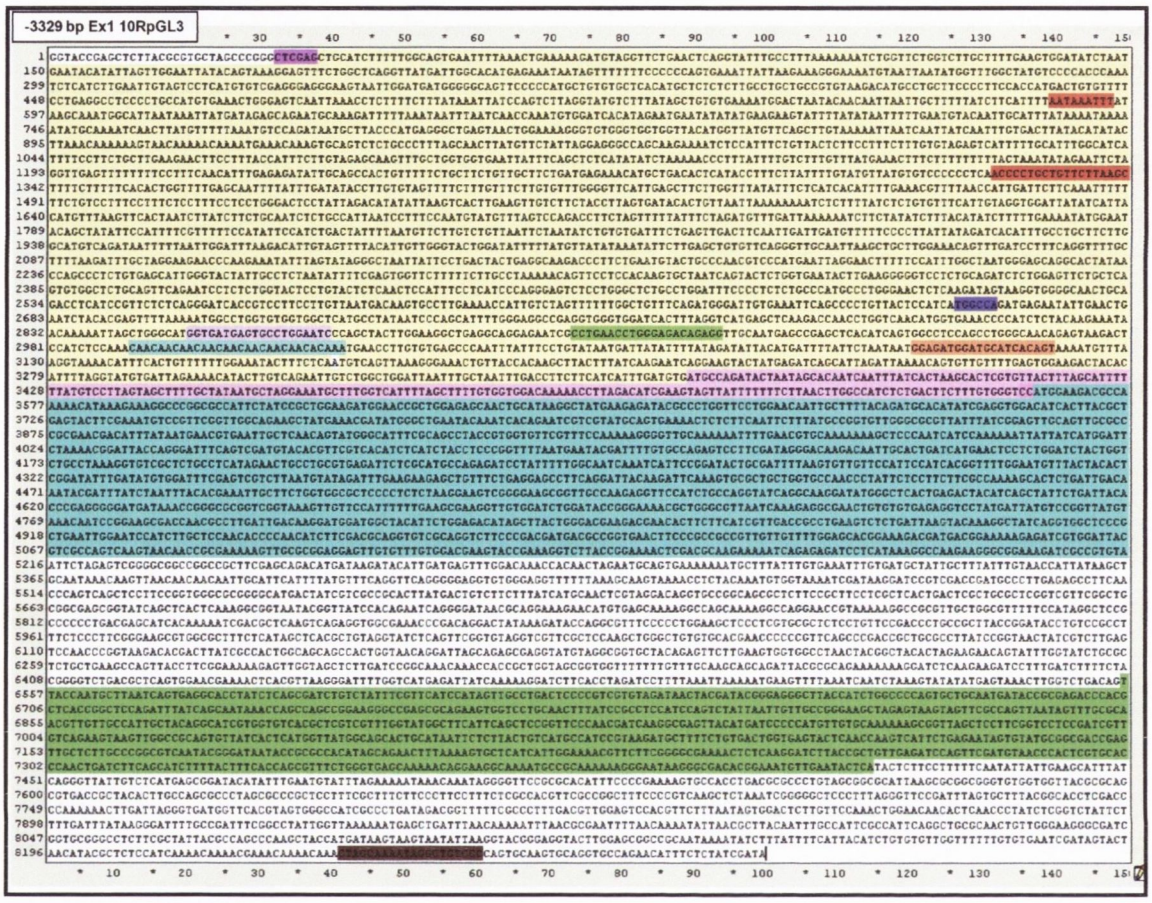


Figure 2.11 Proposed sequence for -3329 bp Ex1 10RpGL3

The rest of the proposed sequences for the other promoter lengths are based on this template and varied only in the VNTR region (for 13RA genotype).

All sequencing results are aligned to these proposed sequenced (Figure 2.10 and Figure 2.11) and shown in Appendix 3.

Chapter 3

Linkage Analysis for HRH4 polymorphisms

3. Linkage analysis for HRH4 polymorphism

Genome wide association studies (GWAS) have discovered a large number of clear associations between specific chromosomal loci and a complex human disease, such as type 2 diabetes and rheumatoid arthritis. A GWAS is a test of association between markers, SNPs, across the genome and disease. From the information collected on the International HapMap project, genetic variance at one locus can predict with high probability genetic variance at an adjacent locus in the human genome. By surveying the genome for common variability it can be associated with the risk of disease simply by genotyping approximately 500,000 carefully chosen markers in the genome of several thousand case subjects and control subjects [95]. In collaboration with the Neuropsychiatric Genetics Research Group in St. James' Hospital, Dublin, a study was performed to test for association of the SNPs located on the HRH4 in relation to schizophrenia.

This study also included identification of tag SNPs by performing a linkage disequilibrium study on the VNTR located in the promoter region. Linkage disequilibrium (LD) is the non-random association between two alleles at different loci, i.e., the co-inheritance of SNP alleles on haplotypes leads to associations between these alleles in the population [84]. In this manner, a non-functional allele can be indirectly associated to a disease phenotype if it is in LD with the biologically influential allele. This study of allelic association is a useful tool in identifying susceptibility alleles, particularly in complex diseases [107]. Tag SNPs for the highly polymorphic region of the promoter of the H4 receptor gene would also provide a simple tool to genotype samples and link to a disease, not only schizophrenia but many other heritable diseases.

3.1 Methods

3.1.1 Associations on the HRH4 Gene Locus

In a recent study, the International Schizophrenia Consortium (ISC) performed a genome-wide association analysis of 3,322 European individuals with schizophrenia and 3,587 controls. The ISC genotyped case-control samples for up to 1 million SNPs, augmented by imputed common HapMap SNPs which were analysed using Cochran-Mantel-Haenszel test and logistic regression [93].

Through collaboration with Dr. Derek Morris from the Neuropsychiatric Genetics Research Group in the Institute of Molecular Medicine in St. James' Hospital, the SNPs on the HRH4 were tested for association from the data published by ISC. The SNPs selected were 100 kb upstream and 100 kb downstream of the gene locus. Marker alleles from ISC were used and integrated with SNPs found on HRH4. The study included a chi-square analysis, *P*-value and odd ratio between the SNPs. Chi-square statistics and logistic regression are common analysis procedures for case-control designs. This prompted further analysis of the polymorphisms on this receptor gene, focusing on the variations in the promoter region to investigate whether this has an impact on other SNPs and on receptor expression levels.

3.1.2 Genotyping HapMap Samples

3.1.2.1 Polymerase Chain Reaction on HapMap Samples

HapMap samples were obtained from collaborative work with Dr. Derek Morris, Neuropsychiatric Genetics Research Group, Institute of Molecular Medicine, Trinity Centre for Health Sciences in St. James' Hospital.

An initial PCR was performed and optimised for the region of interest on 5 HapMap samples. The primers used were designed to be approximately 100 base pairs upstream and 100 base pairs downstream from the VNTR region. The forward primer Seq1_for and the reverse

primer AS11_Rev were used, which have a %GC content of 55% and 50% respectively. The PCR was performed in a 25 µL reaction volume and had the following master mix (Table 3.1):

Table 3.1 Master mix for genotyping HapMap samples

Component Master Mix	*25 µL reaction volume Reactions: x1*	*Final concentration per reaction*
10 µM Forward Primer: Seq1_For	0.5 µL	0.2 µM
10 µM Reverse Primer: AS11_Rev	0.5 µL	0.2 µM
10X *Taq* buffer	2.5 µL	1x
10 mM dNTP	0.5 µL	200 µM
Nuclease free H_2O	19 µL	NA

From the master mix, 19 µL was transferred into 0.2 µL PCR reaction tubes and 5 µL of template DNA was added. The PCR reaction tubes were placed in the thermocycler (PeqLab) and started with a *hot start* of 95°C for 8 min followed by the addition of 1.25U *Taq* Polymerase. Thermocycling conditions were as follows:

- Denaturation: 94°C for 30 s
- Annealing: 60°C for 30 s
- Extension: 68°C for 30 s

} 40 cycles

- Final elongation: 68°C for 5 min

The samples were retrieved from the thermocycler and ran on TAE 2% agarose gel stained with 5 µL of SYBR® safe for visualisation (Chapter 2, Section 2.7), to confirm that the PCR was successful. The gel was set to run at a voltage of 90 V for 45 min. The PCR product was visualised using ChemDoc® XRS Molecular Imager (Bio-Rad) and Quantity One® imaging software version 4.6. The band was then extracted from the gel using QIAgen gel extraction purification kit (Qiagen) as per manufacturer's instructions (Chapter 4, Section 4.1.1.4) and

sent to sequencing with AS16_for primer to Eurofins MWG (Germany) (Chapter 4, Section 4.1.1.9).

Once the sequence was confirmed, the PCR was performed on 90 HapMap samples (10 ng/μL). The forward primer was labelled with a fluorescent dye, 6FAM (Sigma-Aldrich), which has the same sequence as Seq1F but has attached to its 5'-end a 6FAM dye which is essential for determining the length of the PCR product in capillary electrophoresis. The PCR products were stored at -20 °C until ready for analysis.

3.1.2.2 Capillary Electrophoresis

Prior to electrophoresis, 2 μL of PCR product was diluted with 9.9 μL Hi-Di™ Formamide mixed with 0.1 μL GeneScan™ 350 ROX™ labeled size standard (ABI). The mix was denatured at 95°C for 3 min. Denatured samples were loaded on the capillary electrophoresis using Performance Optimised Polymer 6 (ABI). The allele image profiles were analysed using Applied Biosystems GeneMapper® version 3.0 software. Peak retention time was recorded for each sample to establish the PCR band size for the VNTR region. The peak at 280 bp was considered as the 14 triplet repeat as per NCBI database and which was confirmed by sequencing.

For each HapMap sample an allele image profile was generated which consisted of various length fragments of the PCR product. These length fragments were assigned according to the peak recorded from the capillary electrophoresis (CE), which gave the respective PCR band size. A homozygous sample would give one peak for one fragment length while a heterozygous sample would show two peaks with two different fragment lengths.

3.1.2.3 Linkage Disequilibrium Analysis

Linkage disequilibrium (LD) analysis between pairs of loci for the polymorphic region required the following steps in order to generate the required data to analyse:

1. Re-label alleles
2. Pedigree analysis
3. Establish allele frequency
4. Record genotypes for SNPs
5. Obtain the SNPs for the HapMap region
6. Run LD analysis

1. Re-label alleles

The peaks recorded in CE for each HapMap sample were relabelled in order to facilitate better grouping and recognition. The peak retention represented the length fragment of the allele. From previous optimisation results and sequencing it is known that a 280 bp fragment has a 14 triplet repeat, hence re-labelled 14R. The rest of the alleles were given the new label according to the distance travelled in the CE using the 280 bp length as a marking point, (i.e. 10R, 13R, 15R, 16R, 17R, 18R, 19R, respective of the VNTR length). Each HapMap sample allele was given the new label and recorded in a Microsoft Excel document.

2. Pedigree analysis

Pedigree analysis using these Mendelian principles is an essential aspect in genetic linkage studies, where it shows that a trait is passed on from the parents to the offspring and does not occur due to a mutation [108].

Linkage analysis attempts to locate a region of the genome that is co-inherited in families with the trait being studied. Mendel proposed that a hereditary factor for each hereditary trait is given by each parent to each of its offspring. The physical basis for this behaviour is the distribution of homologous chromosomes during meiosis, where one of each pair of homologous chromosomes (randomly chosen) is distributed to each haploid cell [108]. Hence it was important to check for Mendelian inheritance in HapMap samples and to analyse if the variation in the VNTR region is a family trait.

The HapMap samples have a family ID assigned to the parents and offspring where it links the sample name to the mother, the father and the offspring. This was used to analyse if the alleles Mendelise. The ones that did not follow a Mendelian trait of inheritance were omitted from the LD analysis.

3. Establish allele frequency

Once mendelisation was confirmed, allele frequency was calculated. The total number of the various VNTR length fragments was taken on a percentage of the total VNTR fragments. The high allele frequencies were taken into account and used for further analysis.

4. Record Genotypes for SNPs

The polymorphic VNTR length for each HapMap sample was grouped as a single polymorphism in order to be considered as a tag SNP. An allele was assigned to the length for the high allele frequency and given an rs number. The genotypes chosen for this analysis were 10R, 13R, 14R and 15R since from previous unpublished data, it was suggested that they are linked to schizophrenia [92]. The major allele frequencies that were considered for the purpose of this project were listed into four rs values and they were grouped as follows:

- rs unknown 1: When the 10R is the polymorphism it was assigned the allele A, and 13R, 14R, 15R, 16R, 17R, 18R, 19R were assigned the allele G
- rs unknown 2: When the 13R is the polymorphism it was assigned the allele A, and 10R, 14R, 15R, 16R, 17R, 18R, 19R were assigned the allele G.
- rs unknown 3: When the 14R is the polymorphism it was assigned the allele A, and 10R, 13R, 15R, 16R, 17R, 18R, 19R were assigned the allele G.
- rs unknown 4: When the 15R is the polymorphism it was assigned the allele A, and 10R, 13R, 14R, 16R, 17R, 18R, 19R were assigned the allele G.

Once these criteria were established, the genotype for each HapMap sample was recorded for that locus of the promoter region.

5. Obtain SNPs for HapMap samples

The SNPs for the CEU HapMap samples were downloaded from the International HapMap Project website. The SNPs ranged from 100 kb upstream to 100 kb downstream of the gene locus. This data was saved as a basic text format.

The SNPs found at the end of the VNTR were also downloaded from 1000genomes website where it gives the specific allele at the end of this region for each sample.

6. LD Analysis

The data saved in the text format mentioned in step 5 were integrated with the genotypes recorded in step 4 in Haploview software (version 4.1). Haploview calculates LD analysis and gives the r^2 and D' for each allele. The tandem repeats that were in LD with the SNPs found on the HRH4 were noted and tested for functional impact on the gene.

3.2 Results

3.2.1 Linkage analysis results

3.2.1.1 Association analysis for HRH4 gene locus

Using the data from ISC, in collaboration with Dr. Derek Morris from the Neuropsychiatric Genetics Research Group in St. James' Hospital an association test for the SNPs found on the HRH4 gene was performed (Table 3.2)

Table 3.2 Association analysis for HRH4 gene from ISC GWAS data

SNP	A1	A2	Base Position	Chi-Square	ISC p-value	Odds Ratio
rs593099	A	G	20246190	4.633	0.03137	0.8645
rs643829	A	C	20246258	4.275	0.03868	0.8697
rs12458456	C	A	20252498	2.002	0.1571	0.9275
rs599606	C	T	20259894	3.907	0.04809	0.8746
rs582234	A	T	20274465	0.0473	0.8278	0.9893
rs677688	C	G	20274541	4.465	0.0346	0.8668
rs518353	G	A	20280939	1.354	0.2445	0.922
rs10502444	A	G	20283377	0.08246	0.774	1.015
rs2739104	G	C	20283926	0.05732	0.8108	0.9651
rs16940754	G	A	20289412	0.04182	0.838	1.031
rs9963453	T	C	20297683	0.1911	0.662	1.026
rs4483927	G	T	20298398	2.429	0.1191	0.9108
rs16940762	A	G	20299831	0.1116	0.7383	1.02
rs17797945	T	C	20300654	7.061	0.00788	1.111
rs17203328	C	T	20305088	0.02211	0.8818	0.9891
rs11662595	G	A	20310968	2.547	0.1105	1.139
rs1421126	T	C	20311736	6.619	0.01009	1.097
rs9945837	T	C	20311818	0.06351	0.801	0.979
rs17797975	C	T	20313366	0.004478	0.9466	0.996
rs4800573	A	G	20313668	0.6451	0.4219	0.9443
rs527790	T	C	20314913	7.668	0.00562	1.103
rs487202	C	G	20316814	0.22	0.639	1.024
rs532622	T	G	20319272	0.001167	0.9728	1.002
rs673819	T	C	20319903	2.037	0.1535	0.919
rs12965720	C	T	20321889	1.736	0.1876	1.164
rs8094020	T	C	20322084	0.06229	0.8029	0.9868
rs7230451	G	T	20323391	0.643	0.4226	1.043

From this analysis, it was noted that very few SNPs reached genome-wide significance. The *P*-values highlighted in orange are SNPs have a low degree of association with the markers suggested by ISC having a p value smaller than the study's threshold of 0.05 and odds ratio close to 1.1. In GWAS analysis, typical *P*-values around 5 x 10^{-8} are considered "genome-wide significance" having odds ratio in the range of 1.1 to 1.5.

The SNPs mentioned in Section 1.3.2 did not show any significant value with the markers from ISC for schizophrenia. Further analysis was carried out for the VNTR in the promoter region since no SNP was picked up by this study as a significant marker.

3.2.1.2 PCR for HapMap samples

A PCR was run on five HapMap samples. The PCR product was loaded on a 2% agarose gel with SYBR safe DNA gel staining and visualised on BioRad® Molecular Imager having an exposure time to UV light of 0.70 sec.



Figure 3.1 Gel Electrophoresis for PCR for HapMap samples

Figure 3.1 shows the 280 bp product from Seq1-6FAM and AS11_rev from PCR. The samples used from the HapMap project were; NA12815, NA06993, NA12057, NA12264, NA12249 and are labelled samples 1, 2, 3, 4, 5 respectively. The positive control was genomic DNA (Bioline) and the negative control 5 µL sterile water.

Following gel extraction, the PCR product was sequenced. Figure 3.2 shows the sequencing result of NA12057, a homozygous triplet repeat for 14R but heterozygous for the A/C allele. After verifying successful PCR amplification, PCR products were subjected to CE.

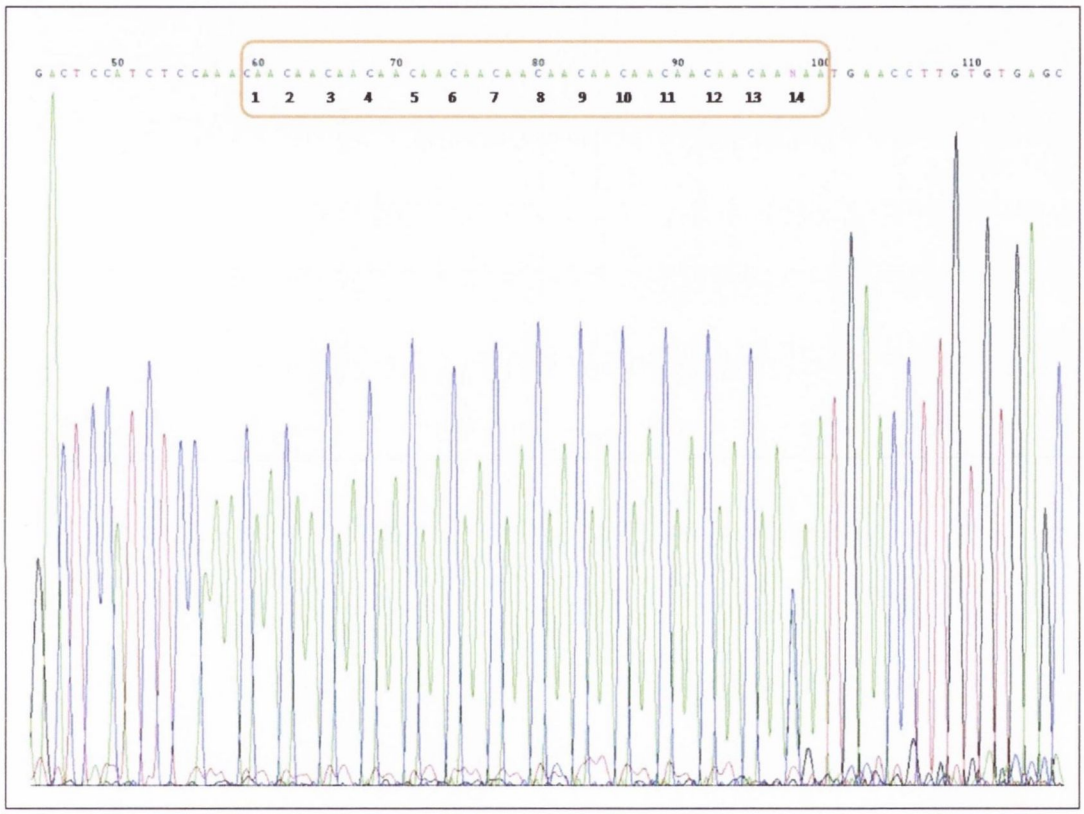


Figure 3.2 Sequencing results for sample 3, showing the 14 tandem repeat CAA/-AAA

3.2.1.3 Capillary Electrophoresis

The HapMap samples were run on capillary electrophoresis (CE) (ABI 3700XL Genomic Analyzer®) and analysed on GeneMapper software in a Microsatellite plot setting. The plot was based on fluorescence intensity against time and compared to a DNA fragment size standard for each electrophoresis run. GeneScan™ 350 ROX™ Size Standard (ABI) was the

size standard which sizes DNA fragments in the 35-350 bp range and provides 12 single-stranded labeled fragments of 35, 50, 75, 100, 139, 150, 160, 200, 250, 300, 340 and 350 bases. Each of the DNA fragments is labeled with the ROX™ fluorophore, which results in a single peak when run under denaturing or native conditions.

The peaks shown below represent the genotypes of four different VNTR alleles.

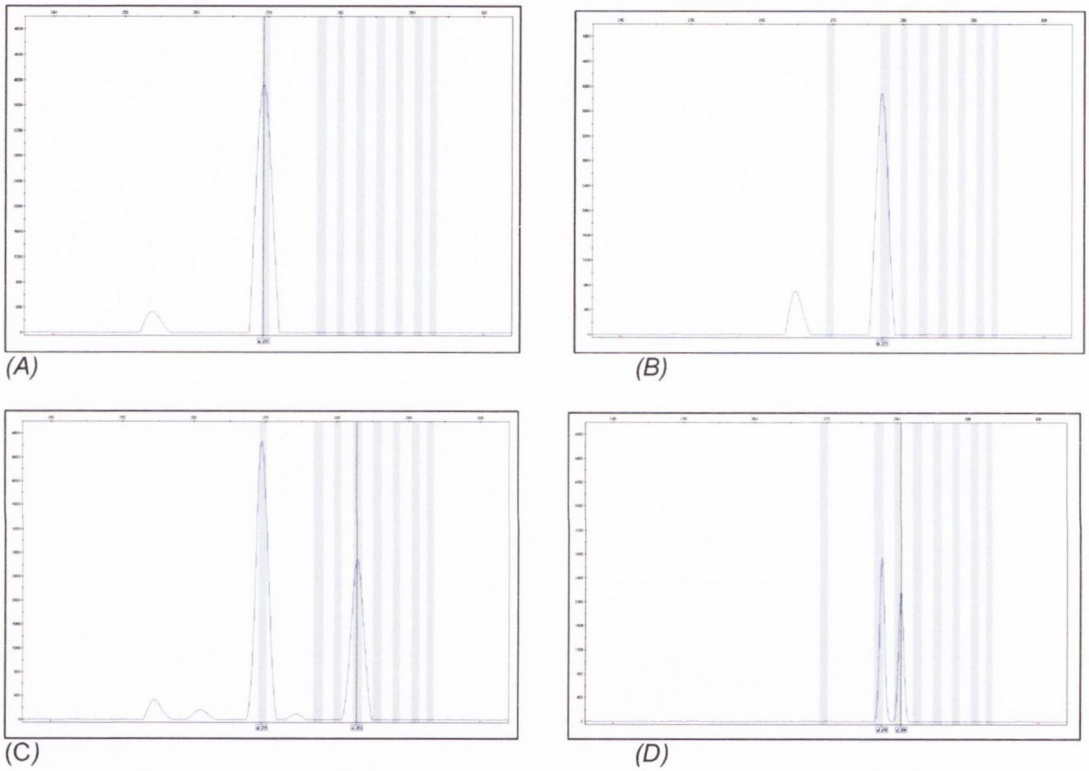


Figure 3.3 CE of HapMap samples showing 4 different genotypes.

(A) homozygous 10R genotype; (B) homozygous 13R genotype; (C) heterozygous 10R/15R genotype; (D) heterozygous 13R/14R genotype

Each major peak represents one allele. When only one peak is present on the microsatellite plot it signifies a homozygous genotype with the size standard marking the length of the VNTR. A homozygous genotype has the same number of tandem repeats on both alleles, these have the same retention time in the CE giving only one peak. When two peaks are present, it signifies a heterozygous genotype where the two alleles have different retention times, hence different length fragments. The retention time for each fragment in a CE run was

used to generate an allele image profile for each HapMap sample in order to establish the different fragment length polymorphisms and their frequencies in the samples.

3.2.1.4 Linkage Disequilibrium Analysis

Linkage analysis is based on non-random pattern between alleles. To establish such a pattern, alleles have to be based on a Mendelian trait (i.e., inherited with the family) and also on their frequency. Polymorphic variants can be associated with a condition not only because they are biologically causal, but because they are statistically correlated with a causal variant. This possibility arises because alleles at different loci are sometimes found together more or less often than expected based on their frequencies [109]. The image allele profile created from the CE was recorded on a Microsoft Excel document where eight length variants were detected for the VNTR region. The retention for each allele was relabelled according to the number of tandem repeats it represents. Table 3.3 shows the allele lengths recorded for the HapMap samples and the respective labelling.

Table 3.3 Relabeling alleles according to the VNTR length

Expected allele length	Allele length recorded in CE	VNTR length
268	270	10R
277	278	13R
280	280	14R
283	283	15R
286	286	16R
289	289	17R
292	291	18R
295	293	19R

Using this allele profile, pedigree analysis was carried out by linking the family ID to each genotype. From 90 HapMap samples only one family did not Mendelise, hence those 3

samples (mother, father and offspring) were omitted from LD analysis. Appendix 2 shows the pedigree analysis for each sample.

Allele frequency was established for the different length variations in the remaining 87 samples. The most common microsatellite triplet repeats were 10R, 13R, 14R and 15R which displayed an allelic frequency of 17%, 27%, 25% and 24%, respectively. Table 3.4 shows the allele frequencies for all the tandem repeats.

Table 3.4 Allele frequencies for the VNTR polymorphisms

VNTR length	Allele Frequency	Percent %
10R	**30**	**17.24**
13R	**47**	**27.01**
14R	**43**	**24.71**
15R	**41**	**23.56**
16R	5	2.87
17R	5	2.87
18R	2	1.15
19R	1	0.57
Total	**174**	**100**

These four fragment lengths with high allele frequencies were used for further analysis. In order to link these length polymorphisms to other SNPs on the HRH4, they was grouped and considered as a single nucleotide polymorphism at that locus.

Table 3.5 shows how these polymorphisms were sub-grouped as an individual polymorphism. The new sub group genotype was recorded for the 87 HapMap samples and is included in Appendix 2. The frequency of occurrence for this polymorphism was highest for 13R (45%) and lowest for 10R (28%) (Table 3.5).

Table 3.5 Sub-grouped genotypes and assignment of rs numbers.
Allele 'A' is the polymorphic allele and it is assigned to the VNTR length when considered as an SNP. The polymorphic frequency is calculated based on the 87 HapMap samples.

VNTR length	Allele assigned	rs number given	Polymorphic frequency (%)
Considering 10R as the SNP			
10R	A	rsunknown 1	28%
13R-19R	G		
Considering 13R as the SNP			
13R	A	rsunknown 2	45%
10R, 14R-19R	G		
Considering 14R as the SNP			
14R	A	rsunknown 3	41%
10R, 13R, 15R-19R	G		
Considering 15R as the SNP			
15R	A	rsunknown 4	37%
10R-14R, 16R-19R	G		

The SNPs genotyped for these samples were downloaded from the HapMap project website and ranges from 100 kb upstream and 100 kb downstream to the gene locus. Using these SNPs, the above data and Haploview software an LD plot was created. From these SNPs 10R and 13R had significant r^2 values which were greater than 8. The length variation for 14R and 15R had no significant r^2 value with any of the SNPs found on HRH4. Table 3.6 shows the polymorphisms which were in LD with 10R and 13R.

Table 3.6 R-squared values from LD plot for 10R and 13R

VNTR length	rs number	r^2
10R	rs687263	0.942
	rs677213	1.000
	rs571102	1.000
	rs628764*	1.000
	rs648171	1.000
	rs582234	1.000
	rs620931	1.000
	rs615283*	1.000
	rs616707	1.000
	rs661538	1.000
	rs687263	1.000
13R	rs17797945*	0.956

** Polymorphism lie on the HRH4 gene*

The SNPs rs687263, rs677213, rs571102, rs648171, rs582234, rs620931, rs616707, rs661538 and rs687263 that were found to be in LD with 10R are located on IMPACT gene (Gene ID: 55364). IMPACT also lies on chromosome 18 and is positioned 18q11.2-q12.1; upstream of the HRH4 (location 18q11.2). Figure 3.4 shows the ideogram of chromosome 18, where the red line marks the location of IMPACT and HRH4 and their proximity.

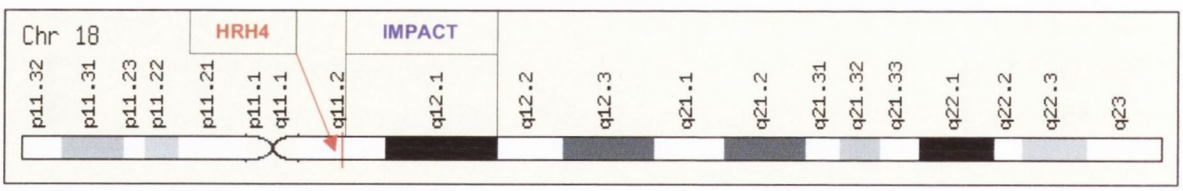


Figure 3.4 Location of IMPACT and HRH4 on Chromosome 18

The other two SNPs in LD with 10R, rs628764 and rs615283, are polymorphisms on the H4 receptor gene. The SNP rs628764 is located on intron 1 of the HRH4 gene which is a C/T

polymorphism whilst rs615283 is located on the promoter region of HRH4, 147 bp downstream the VNTR region and is polymorphic for A/G allele.

The LD plot for 13R in LD with rs17797945 is shown in Figure 4.9. Each diagonal box represents a SNP, with each box representing a pair-wise association of each SNP. Black squares indicate statistically significant allelic association between the pair of SNPs as measured by r^2 statistics; darker colours of black indicate higher values of r^2, up to a maximum of 1. White squares indicate pair-wise r^2 values of << 1 with no statistically significant evidence of LD.

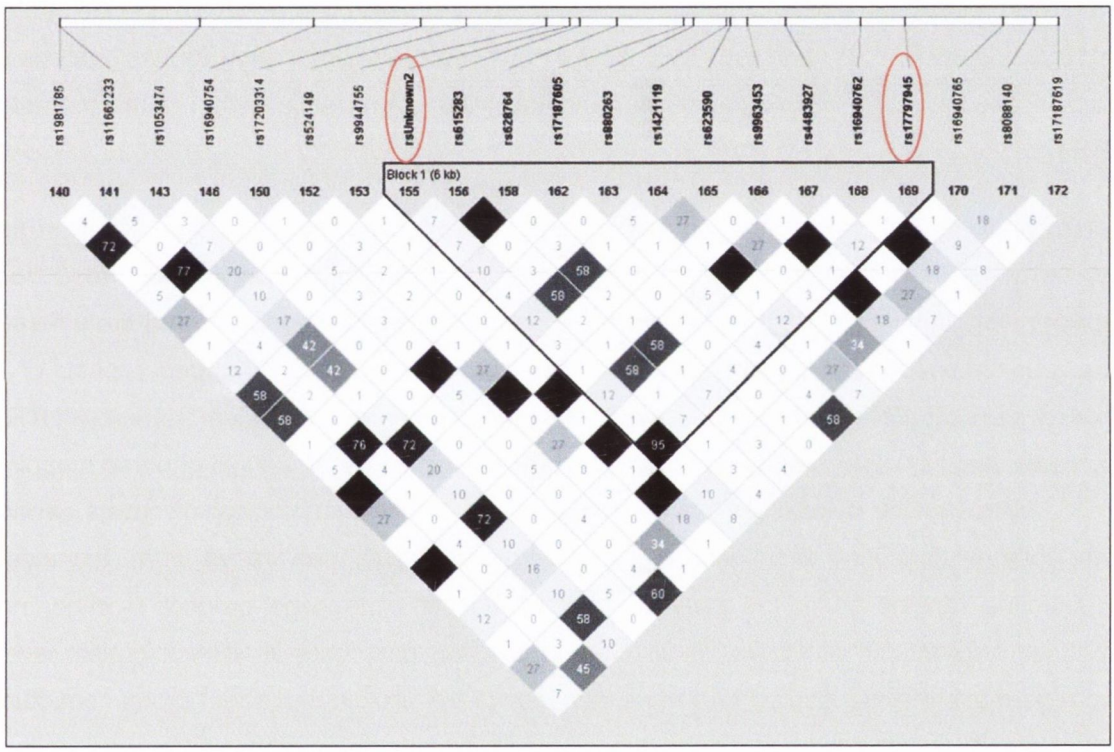


Figure 3. LD plot showing correlation between 13R and rs1779745 (r^2 0.956)

The SNP rs17797945 had a high LD value with 13R. This SNP exhibits a C/T polymorphism and is found on intron 1 of the HRH4. It was also found to be weakly associated with schizophrenia in the ISC GWAS (*P* = 0.007876).

3.3 Discussion

Schizophrenia is a complex disorder and it can be caused by both genetic and environmental factors [94]. Genome wide association studies (GWAS) in schizophrenia uncovered genetic loci which are thought to contribute to the disease (Table 1.8). In collaboration with the Neuropsychiatric Genetics Group in the Institute of Molecular Medicine in St. James' Hospital Dublin a possible association between SNPs and HRH4 gene was investigated. The linkage analysis data for the tag SNPs on the HRH4 tested for association with the ISC showed only a small association of low significance with schizophrenia (P <0.05). These SNPs were rs593099 (P 0.03), rs643829 (P 0.04), rs599606 (P 0.04), rs677688 (P 0.03), rs17797945 (P 0.01), rs142116 (P 0.01) and rs527790 (P 0.01). These results prompted further analysis of tag SNPs to provide a simple means of genotyping the polymorphic region of the promoter HRH4.

Genotyping this region showed the various lengths and variations as a final result of the various SNPs and insertion/deletions at this locus. Using the migration time recorded from capillary electrophoresis (CE), image allele profiles were generated and showed the different variations obtained from length repeats. Microsatellite repeats consist of tandem repeats of 2-6 base pairs of DNA, which are a type of variable number of tandem repeat (VNTR). Microsatellites are very common in eukaryotes and are typical landmarks in genome mapping [110, 111]. Microsatellite instabilities have been associated with neurological disorders, where both 'loss of function' and 'gain of function' mechanisms can result from expanded trinucleotide repeats [64]. In the case of 'loss of function', the gene product is either not produced or produced at a lower level. 'Gain of function' can occur at RNA or protein level, and repeat expansions located in a translated part of the protein can affect protein structure or function. The translated variation can lead to a change in amino acids which makes either the protein work in a different way or potentially affecting binding of therapeutic drugs. If the repeat expansion is in a non–translated region, it can affect the protein in an indirect method, where when the repeat is transcribed it can influence the function of the mRNA or other transcription factors [64]. These repeat instabilities are a product of DNA replication, where unpaired bases form loops resulting in expansions or contractions depending whether the loops are in the synthesised or in the template strand [64] (Figure 3.5). When the loop is

present of the template stand, contraction of occurs, while if the loop forms on the new synthesised strand, expansion occurs.

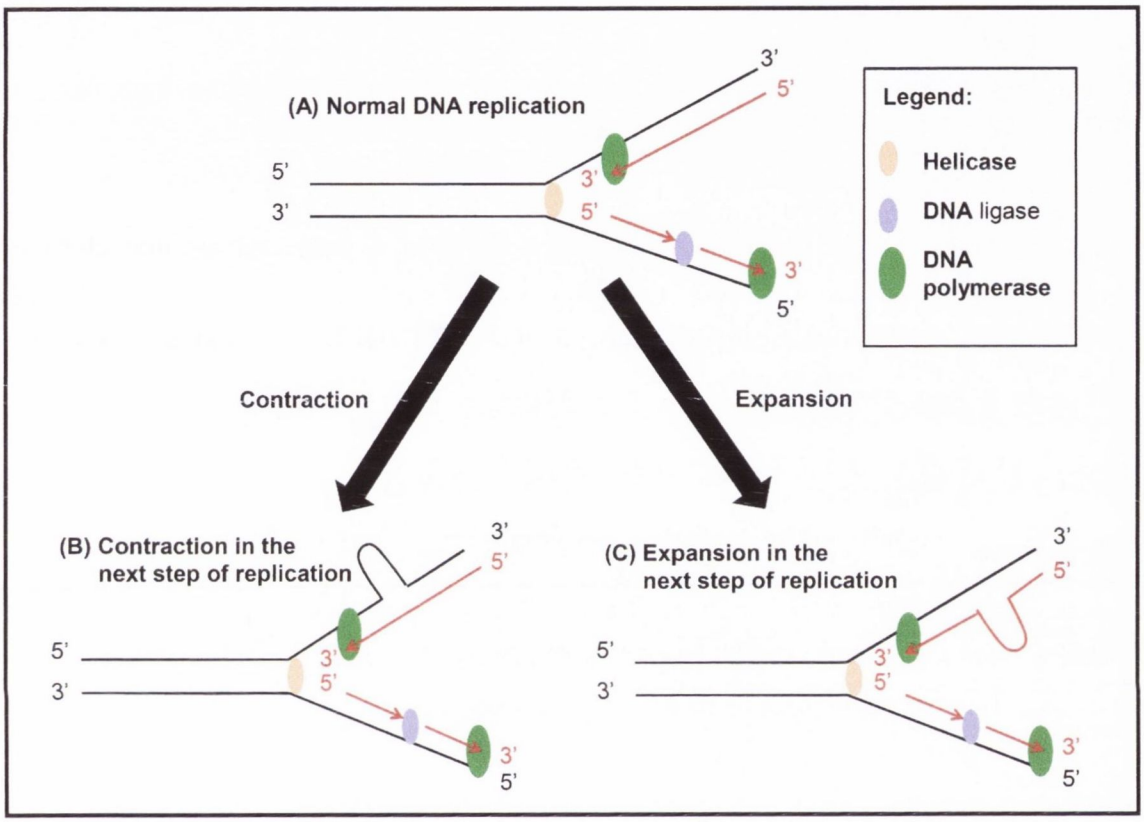


Figure 3.5 Formation of microsatellites

(A) During DNA replication helicase opens up the double strand and DNA polymerase synthesises the 5'-3' strand. (B) When loops are formed in the template, contraction occurs in the next step of replication, (C) When loops are formed in the synthesised strand, expansion occurs in the next step of replication (Figure adopted from [64]).

The DNA samples used in the linkage analysis study were checked for pedigree analysis. Pedigree analyses lies on the mendelian principle where the trait is passed on from the parents to the offspring. The genomic HapMap DNA used in this study was found to follow a mendelian trait of inheritance. The triplet micorsatellite repeats identified ranged from 10

repeats (10R) to 19 repeats (19R). The original DNA sequence found on the NCBI website is the 14R allele with the following ending to the triplet repeat:

1 2 3 4 5 6 7 8 9 10 11 12 13 14

...AAA **CAA CAA CAA CAA CAA CAA CAA CAA CAA CAA CAA CAA CAA CAA** TGA ...

Any other variation found from this sequence is listed as a polymorphism and given rs number in the dbSNP, as mentioned in Section 1.3.2. However, the first sequence for the HRH4 by Coge et al [112] was published as 10R for the VNTR variation and with a polymorphic end:

1 2 3 4 5 6 7 8 9 10

... AAA **CAA CAA CAA CAA CAA CAA CAA CAA CAC AAA** TGA...

This discrepancy in reported length of the VNTR can be explained by the fact that both variations have a high allele frequency of 17% (10R) and 24% (14R), however in this study it was concluded that the highest allele frequency is observed for the 13R variation, with 27%. This could suggest that the original sequence for this VNTR has a 13 microsatellite triplet repeat.

On comparing the allele frequencies obtained from the HapMap samples (n = 87) and the preliminary study from Dr. Sasse [92] on genotyping schizophrenia patients (n = 138), it can be observed that the allele frequencies between schizophrenia (SCZ) patients and HapMap are similar (Table 3.7), however this study should be repeated with a larger cohort and matching controls in order to confirm this finding. The only difference was observed for the 14R where it was found to be 24% in HapMap samples compared to SCZ patients with 20.6%. No triplet repeat length variation of 12R was noticed in the HapMap samples and 11R has never been observed

Table 3.7 Comparison between studies

VNTR length genotype	Study on SCZ patients %	Study on HapMap samples %
10R	17.00	17.24
12R	0.04	0
13R	27.50	27.01
14R	20.60	24.71
15R	25.70	23.56
16R	2.90	2.87
17R	2.90	2.87
18R	2.20	1.15
19R	0.70	0.57

Genotyping this polymorphic region also indicates that the SNPs reported in dbSNP (http://www.ncbi.nlm.nih.gov/snp/) for this region do not represent what is truly happening for this triplet microsatellite repeat. Hence it is proposed that this region is labelled in line with other microsatellites, e.g. in Fragile X gene (FMR1) and androgen receptor (AR). Figure 3.6 compares between NCBI reporting and the proposed reporting for this highly polymorphic region, where rs665411 is an **AC** to **CA** change which is only seen with the 10R length, and rs670327 is a **C** to **A** SNP which is seen at the end of the triplet repeat in other length variations.

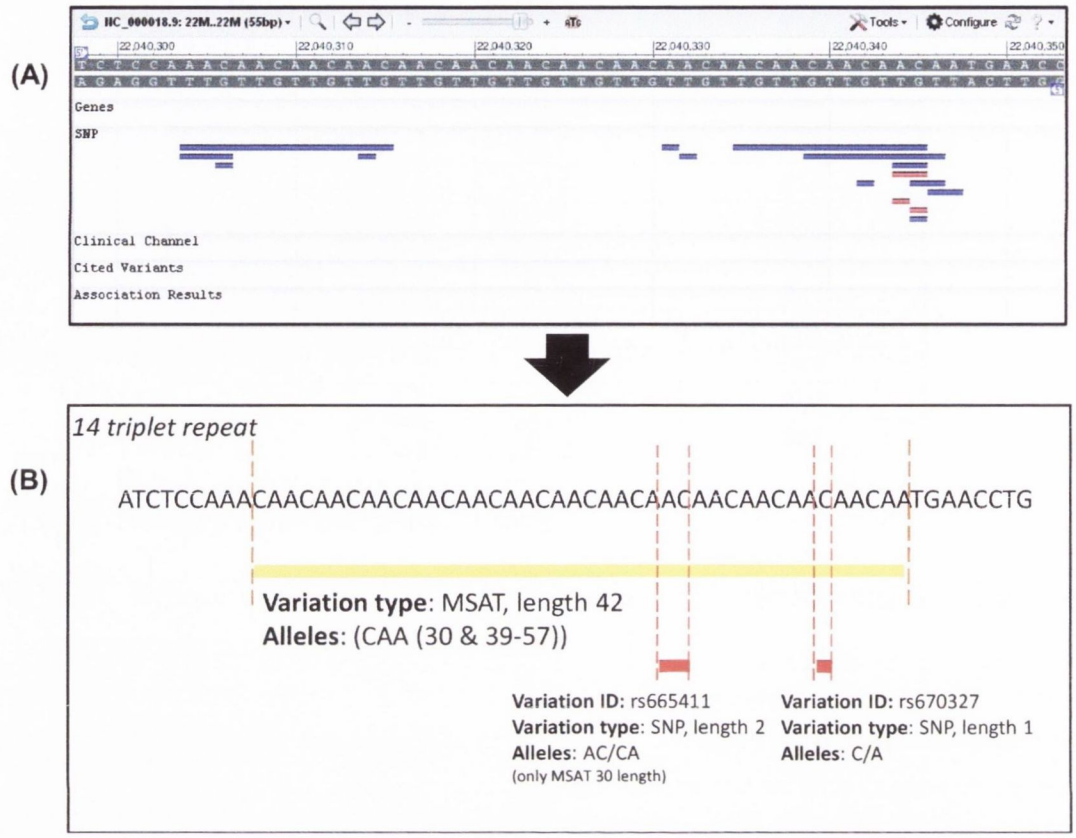


Figure 3.6 Proposed reporting for the VNTR region of HRH4 gene

(A) Shows how the VNTR region is reported in NCBI and (B) shows how it is proposed to be reported following genotyping this region

The specific nucleotide polymorphism for the HapMap samples was obtained by accessing the 1000genomes database. With this data, it was possible to identify the allele SNP found at the end of each length repeat. With this information, the nucleotide polymorphism for each variation was obtained as well as determination for homozygous or heterozygous alleles for the specific allele at each length. From this data, it was noted that the 10R variation length was always consistent with the ending established by Coge *et al* [112]. Three different SNPs were identified at the end of the length repeats:

...CAC AAA... * *only for 10R length*

...CAA CAA... } *For 13R – 19R lengths*

...CAA AAA... }

The repeats 10R, 13R, 14R and 15R were found to be high frequency alleles and hence were chosen for linkage disequilibrium analysis (LD). The LD study identified tag SNPs only for 10R and the 13R. The 13R had one SNP that exhibited LD of r^2 0.95 and is located in intron 1 of HRH4.

The GWAS analysis carried out in collaboration with St. James' Hospital on the HRH4 flagged this SNP, i.e. rs17797945, to be weakly associated with schizophrenia. Genome-wide studies are predicted on common-disease-common-variant hypothesis, mostly due to the idea that detection required a disease allele to be highly correlated with at least 1 SNP on the array. This hypothesis proposes that disease-related alleles are relatively common (frequency ≥ 0.01). An important aspect that emerged from GWAS is that the effect sizes conferred by such common alleles are typically small, with those identified having odds ratios (ORs) mainly in the range of 1.1 to 1.5 [113, 114]. The GWAS analysis on the SNPs located on HRH4 identified rs17797945 having a *P* value of 0.007876 and an OR of 1.111. From this GWAS and LD analysis, it was suggested that this SNP is weakly associated with schizophrenia and in high LD with the 13R variation length.

The 10R variation length was in LD with eleven SNPs, nine of which are found on IMPACT gene. This gene is located about 40 kb upstream of the HRH4 and closer to the centromere, so it is not surprising to see that there is a non-random association between the VNTR and SNPs on IMPACT. Human IMPACT encodes a protein with 320 amino acids and the main function of this protein is to act as a translational regulator that ensures constant high levels of translation under amino acid starvation. A study suggests that the human IMPACT gene is highly expressed in the brain and it was proposed that this gene can be directly involved in the pathogenesis of bipolar affective disorder [115]. It is interesting however to point out that while nine SNPs are in high LD with the 10R variation length for the VNTR, none of the other

microsatellite repeats are in LD with any SNPs within 100 kb of the previously mentioned eleven SNPs. The other two SNPs found in LD with 10R are polymorphisms found on the HRH4 gene. The SNP rs628764 is located in intron 1 and rs615283 lies on the promoter region, approximately 150 bp downstream to the VNTR triplet repeat. This polymorphism exhibits an exchange from G to A allele (and this was later on confirmed when the 10R clone was sequenced in Chapter 4, Section 4.2.1).

Linkage analysis on HRH4 from other research groups identified SNPs located on introns which are associated to diseases. For example, SNP rs4483927 found on intron 1 was associated with risperidone efficacy in schizophrenia and is potentially a biological marker for treatment [54]. A polymorphism located on intron 2, rs17187619, gave a strong *P* value (*P* = 0.002) in patients with or without infection-induced asthma [73]. Whilst another SNP in intron 1, rs623590, was described to be more common in breast cancer patients when found in a haplotype C-A-A with rs1162595 and rs1421125 [78]. Hence, polymorphisms that do not lie on a translation site (such as an exon) may still play a biological role and contribute to a disease or may influence treatment leading to personalised medicine.

3.4 Conclusion

The SNPs on the HRH4 locus analysed did not indicate a significant association to schizophrenia, however a weak association was found when analysed with tag SNPs by the international schizophrenia consortium. Since the VNTR region is designated as individual polymorphisms, given various rs-numbers, this type of linkage analysis would not provide the full picture for the VNTR region and its association to the disorder. A different way of reporting this region is also suggested which shows this region as a triplet microsatellite repeat and gives a clearer picture of this highly polymorphic area.

Hence, genotyping for this region was carried out through fragment analysis by capillary electrophoresis and a linkage disequilibrium study. Genotyping this region and looking at the allele frequency from the HapMap samples, showed that there does not seem to be an association between schizophrenia and a specific VNTR in the promoter region of the HRH4. On the contrary, looking into the linkages from the LD analysis for both 10R and 13R, it could possibly suggest a small degree of association but further investigation needs to be carried out in order to establish a clear profile this linkage might have with the disease or the gene expression itself.

The highly complex genotype profile obtained in this study generated interest to investigate if length polymorphisms and SNPs located on the VNTR region of the putative promoter region of HRH4 affect gene expression. Thus further analysis to study these genetic variations for possible functional impact on the gene was carried out. This was performed by constructing a library of the putative promoter region for all genotypes, including the different VNTRs (10R, 13R-19R) and the SNP variation at the end of the triplet repeat, in a reporter vector in order to analyse these clones on a cell-based assay for promoter activity.

Chapter 4

Construction of reporter plasmid library

4. Construction of reporter plasmid library

The previous Chapter 3, Section 3.1.1 included genotyping HapMap samples and identified a possible linkage to schizophrenia through GWAS. This resulted in obtaining a genotype profile for the highly polymorphic region in the potential promoter region of the HRH4, thus provoking further analysis as to identify if these genetic variations effect on gene expression. Assessing the functional significance for these genotypes required cloning all length variations of the promoter VNTR region, and their respective allele variant, into a standard vector as a first step. HapMap samples and human genomic DNA extracted from buccal swab samples were used to obtain the amplicons of different polymorphic forms of this potentially active promoter.

The predicted promoter region for the HRH4 ranged 2 kb upstream of the ATG start codon on exon 1 of the gene. This was considered as the main part of the promoter region according to a previous paper published by Coge *et al*[7]. Analysis for the region was carried out through bioinformatic analysis using TESS software and TRANSFEC [7]. After cloning the various VNTR genotypes for the promoter into a standard vector, point mutation was performed on each genotype by site-directed mutagenesis, where a single nucleotide is changed in order to obtain the other SNP allele variant. Once the library was constructed, the putative promoter regions were sub-cloned into a gene reporter vector that expresses the luciferase gene, pGL3-basic vector (Promega), which was later used in a luciferase assay to measure activity of the cloned promoter.

This chapter describes how the reporter plasmid library was constructed harbouring the human DNA encoding for this potential transcription site with the different polymorphisms. Various techniques were used, included cloning into a standard vector, sub-cloning into the reporter vector and point mutation with site-directed mutagenesis.

The description for constructing different promoter lengths is also explained, where site-directed mutagenesis using mega-primers is used to extend the 5'-end of the putative promoter. These promoter extensions are then tested on a cell-based assay described in Chapter 5.

4.1 Methods

4.1.1 Overview of methods

The four main genotypes with high allele frequency were constructed first. The template DNA used for the amplification step originated from HapMap samples or human genomic DNA extracted for a previous analysis of optimisation steps. Buccal cell samples are considered a reliable source of human DNA for PCR-based genotyping studies. However, even though it is a simple non-invasive procedure, ethical approval was required. Ethical approval was obtained for this anonymous procedure from the Faculty Research Ethics Committee (F.R.E.C.) of the School of Medicine (Faculty of Health Sciences) in Trinity College Dublin on 17-05-2010, Appendix 1.

The HapMap samples chosen for this cloning step were homozygous for length and SNP, and were NA12874 and NA11995 (10R and 15R respectively). The other two genotypes, 13R and 14R, were extracted from homozygous human genomic DNA samples. The length variations 16R and 17R used for cloning into a standard vector were a result obtained from student projects carried out under my supervision. Human genomic DNA was extracted from buccal swabs and using the same technique as for the other genotypes (10R, 13R, 14R, 15R). The longest genotype variations, 18R and 19R, were constructed by a different method where the synthesised genes were sub-cloned into the reporter vector where one of the sub-cloned vectors (10RpGL3) was used as the basic template.

Promoter length variations for 10R genotype were also performed in order to identify the promoter start site of the HRH4 gene. Hence, modification on the 10RpGL3 clone was achieved by conducting site-directed mutagenesis using mega-primers. This clone was used since this is the putative promoter described by Coge *et al* [7] had a 10R genotype (with 17% allele frequency in HapMap samples). A 3'-end modification on the 13RpGL3 clone was also performed, since 13R is the most common genotype (27% allele frequency in HapMap samples).

4.1.1.1 Genomic Human DNA Collection and Extraction

Buccal swabs were collected from students and postgraduates in Trinity College Dublin. To comply with the requirements from F.R.E.C., each participant was given a thorough explanation of both the research and the procedure of buccal swab collection. Sample collection was anonymous, so the data would not be traced back to an individual. The participants were given a minimum of seven day reflection period. On the day of swab collection, the participant signed an informed consent form and this was kept in a locked cabinet. Buccal swabs used were from Isohelix® SK-2 (Cell projects Ltd.). The participants were asked to rinse the mouth with water and rub the swab firmly against the inside of the cheek and gums for approximately 1 minute. The swab was transferred to a sterile sealed tube, the handle was detached from the swab, and stored at 4°C until ready for DNA extraction. Qiagen's QIAamp® DNA Investigator Kit was used for DNA extraction. Cell lysis was initiated by adding 20 µL Proteinase K and 400 µL Buffer ATL and mixing by pulse-vortexing for 10 s. The tube was placed in a thermomixer and incubated at 56°C with shaking at 900 rpm for 1 hr. The tubes were then briefly spun on a table-top centrifuge, 400 µL Buffer AL was added and mixed by pulse-vortexing for 15 s. To ensure efficient lysis, it was essential that the sample and Buffer AL were thoroughly mixed to yield a homogeneous solution. The 2 ml tube was placed again in the thermomixer, and incubated at 70°C with shaking at 900 rpm for 10 min. The tubes were again briefly spun on a table-top centrifuge and 200 µL ethanol (96–100%) was added and mixed by pulse vortexing for 15 s. The samples were centrifuged for 10 s. To bind the DNA to the column, 600 µL lysate was added to the QIAamp MinElute column (in a 2 ml collection tube) and centrifuged at 6000 x g (8000 rpm) for 1 min. The QIAamp MinElute column was placed in a clean 2 ml collection tube, and the collection tube containing the flow-through was discarded. The bound DNA was washed by adding 500 µL Buffer AW1 and centrifuged at 6000 x g (8000 rpm) for 1 min. The column was placed in a clean 2 mL collection tube, and the collection tube containing the flow-through was discarded. A second wash was performed by adding 650 µL Buffer AW2 and centrifuged at 6000 x g (8000 rpm) for 1 min. The column was placed in a fresh collection tube, 650 µl of ethanol (96–100%) was added and centrifuged at 6000 x g (8000 rpm) for 1 min. The column was again placed in a clean 2 mL collection tube and centrifuged at full speed (13,000 x g) for 6 min to dry the membrane completely. This step is necessary, since

ethanol carry-over into the eluate may interfere with some downstream applications. The column was then placed in a clean 1.5 mL microcentrifuge tube and incubated at room temperature (15–25°C) for 10 min or at 56°C for 3 min. The final elution of the DNA extract was made by adding 100 µL Buffer ATE to the centre of the membrane. The samples were incubated at room temperature (15–25°C) for 5 min and centrifuged at full speed (13,000 x g) for 3 min. This final incubation step before centrifugation generally increases DNA yield. The extracted genomic DNA was stored at -20°C until further analysis.

4.1.1.2 DNA quantification and qualification

The concentration and purity of the extracted genomic DNA was determined by using NanoDrop™ ND-1000 Spectrophotometer (Thermo Scientific). Before taking a measurement the instrument was calibrated by using 1 µL of RNase-free H_2O. The instrument was blanked with the buffer used to elute the DNA extract (ATE buffer). Then 1 µL of the DNA sample was measured and recorded. The concentration is a measure of absorbance at 260 nm and calculated using the Beer-Lambert equations, where the final reading is given in ng/µL. The NanoDrop™ also provides the ratio of sample absorbance at 260 and 280 nm. The DNA has a maximum absorbance of 260 nm whilst proteins have a maximum absorbance at 280 nm. Hence, the ratio of absorbance at 260 and 280 nm is used to assess the purity of DNA. A ratio of approximately 1.8 is generally accepted as "pure" for DNA. The readings were made in triplicate and an average was used to establish the concentration for each sample. DNA quantification was carried out at various steps of the procedure and unless specified, no deviation was made from this protocol.

A beta-actin amplification was also performed as a quality control on extracted human genomic DNA. The primers used were: forward primer b-actin-for-pink and reverse primer b-actin-rev-gre (sequence of all primers is found in Chapter 2, Section 2.4.3) that have a %GC content of 65 % and 46 % respectively. The same cycling conditions were used as mentioned in Section 3.1.2.1.

4.1.1.3 Amplification of the promoter region

This PCR was performed when amplifying the DNA sequence of the putative promoter region for cloning into pCR-Blunt standard vector and sub-cloning into pGL3 reporter vector. The DNA polymerase used was *Phusion* High-Fidelity (NEB) enzyme that produces a blunt-ended PCR product. All primers were verified on the bioinformatics website; Sequence Manipulation Suite: Primer stats, to check their characteristics (the sequence is found in Chapter 2, Section 2.4.1).

The PCRs were performed in a 50 µL reaction volume and the master mix used is shown in Table 4.1.

Table 4.1 Master mix for amplification of the putative promoter region

Component Master Mix	***50 µL reaction volume Reactions: x1***	***Final concentration per reaction***
10 µM Forward Primer:	2.5 µL	0.5 µM
10 µM Reverse Primer:	2.5 µL	0.5 µM
5X Phusion HF buffer	10 µL	1x
Phusion High-Fidelity DNA polymerase	0.5 µL	1.0 units
10 mM dNTP	1 µL	200 µM
50 mM $MgCl_2$	0.5 µL	0.05 mM (+ 1.5 mM from buffer = 2 mM total Mg^{2+})
Nuclease free H_2O	28 µL	NA

From the master mix, 45 µL was transferred into 200 µL PCR reaction tubes and 5 µL of template DNA was added and spun. The PCR reaction tubes were placed in the thermocycler (Bio-Rad®) and started with an initial denaturation step of 94°C for 1 min*. Thermocycling conditions were as follows:

• Denaturation:	94°C for 10 sec	35 cycles
• Annealing:	64°C for 10 sec	
• Extension:	72°C for 1min	
• Final elongation:	72°C for 7 min	
• Hold	4°C	

**The initial denaturation was changed to 94°C for 5 min when the amplification was performed using human genomic DNA.*

The samples were retrieved from the thermal cycler and ran on a 1% TAE agarose gel stained with SYBR® safe (Invitrogen) for visualisation with a 2-Log DNA ladder (NEB) as described in Chapter 2, Section 2.7. PCR products were stored at -20 °C until further analysis.

Method 1: Amplification for cloning into pCR-Blunt:

The forward primer Seq2F and the reverse primer Seq3R were used, which have a %GC content of 64% and 61% respectively.

Method 2: Amplification for sub-cloning from pCR-Blunt into pGL3 reporter vector

Primers were designed to amplify the region of the promoter from -2049 bp to -1 (where 'A' in the ATG start codon of Exon one of HRH4 is considered at +1). The forward primer Seq4F and reverse primer Seq5R were used which have 57% and 42% of %GC content and a T_m of 71°C and 70°C, respectively. The primers designed for the first amplification step included restriction sites for Xho1 in the forward primer and restriction site Nco1 in the reverse primer. These restriction sites were matched to the desired region needed for ligation into pGL3-basic vector (Promega) following restriction digest with the respective enzymes. This vector is a promoterless vector that has only the genetic code for luciferase, therefore the gene is not expressed unless a functional promoter is inserted upstream to the gene.

Method 3: Amplification for sub-cloning from pEX into pGL3 reporter vector

The forward primer Seq7F was used for amplifying the region for 18R and 19R and is located on the 5' end of pEX-K vector, and its oligo characteristics are mentioned in Chapter 2, Section 2.4. This primer is located approximately 80 bp upstream from the Msc1 restriction site. The reverse primer used was Seq5R which has the Nco1 restriction site as mentioned above (in Method 2).

4.1.1.4 DNA purification for further analysis

Two types of gel extraction kits were used according to procedure involved. The two methods were used depending on the final yield required. Gel extraction using method 2 produced a higher concentration of DNA, hence when a faint band was observed this method was adopted.

Method 1: Using QIAquick® gel extraction kit

The DNA samples were gel purified using QIAquick® gel extraction kit (Qiagen). After separation by agarose gel electrophoresis, the DNA fragment was excised from the agarose gel with a clean, sharp blade. The gel slice was weighed and 3 volumes of Buffer QG was added to 1 volume of gel (100 mg ~ 100 µL). It was incubated at 50°C for 10 min (until the gel slice was completely dissolved) and mixed by vortexing the tube every 2–3 min during the incubation to aid solvation of the gel. After the gel slice was dissolved completely, 1 gel volume of isopropanol was added to the sample and mixed. A QIAquick® spin column was placed in a 2 mL collection tube. To bind DNA, the sample was applied to the QIAquick® column, and centrifuged for 1 min. The flow-through was discarded and the QIAquick® column was placed back in the same collection tube. To remove all traces of agarose, 0.5 mL of Buffer QG was added to QIAquick® column and centrifuged for 1 min. To wash the PCR product, 0.75 mL of Buffer PE was added to QIAquick® column and centrifuged for 1 min. The flow-through was discarded and the column was centrifuged for an additional 1 min at 17,900 x g (13,000 rpm). The QIAquick® column was placed into a clean 1.5 mL microcentrifuge tube. Final elution of the DNA was made by adding 50 µL of Buffer EB (10 mM Tris·Cl, pH

8.5) to the centre of the QIAquick® membrane and centrifuging for 1 min. The DNA was quantified using NanoDrop™ ND-1000 Spectrophotometer (Thermo Scientific) as mentioned in Section 4.1.1.1. The DNA was stored at -20°C until further analysis.

Method 2 : Using MinElute gel extraction kit

The PCR product loaded on the gel was excised with a clean, sharp blade. The gel slice was weighed and a three-fold volume of Buffer QG was added to one-fold volume of gel (100 mg ~ 100 µL). It was incubated at 50°C for 10 min (until the gel slice has completely dissolved) and the tube vortexed every 2–3 min during the incubation to help dissolve the gel. After the gel slice has dissolved completely, one-fold gel volume of isopropanol was added to the sample and mixed. A MinElute spin column was placed in a 2 mL collection tube. To bind DNA, the sample was applied to the MinElute column, and centrifuged for 1 min. The flow-through was discarded and the MinElute column was placed back in the same collection tube. To remove all traces of agarose, 0.5 mL of Buffer QG was added to MinElute column and centrifuged for 1 min. To wash the PCR product, 0.75 mL of Buffer PE was added to MinElute column and centrifuged for 1 min. The flow-through was discarded and the column was centrifuged for an additional 1 min at 17,900 x g (13,000 rpm). The MinElute column was placed into a clean 1.5 ml microcentrifuge tube. Final elution of the DNA was made by adding 10 µL of 1x Buffer 3 (NEB). This buffer was the buffer used for ther restriction enzyme digest in the next step. Buffer 3 was applied to the centre of the MinElute membrane and centrifuged for 2 min. The DNA was quantified using NanoDrop™ (Thermo Scientific) as previously discussed in Section 4.1.1.1, using 1x buffer 3 as the band. When required for later analysis, the DNA was stored at -20°C.

Method 3: Using MinElute® Reaction Clean up

DNA purification following restriction enzyme digest was performed using MinElute® Reaction Cleanup Kit (Qiagen). The enzymatic reaction volume was adjusted to 20 µL volume with water which was followed by the addition of 300 µL buffer ERC. This volume was placed in

the MinElute® column in a 2 mL collection tube and centrifuged for 1 min at 13,000 rpm. The flow-through was discarded and 750 µL of buffer PE was added and centrifuged again for a further 1 min. The collection tube was emptied and centrifuged for another 1 min to remove any residual ethanol (in buffer PE). The MinElute® column was placed in a clean autoclaved 1.5 mL microcentrifuge tube where 10 µL autoclaved deionised water was added in the centre of the membrane and left to stand for 1 min. Final elution was performed by at 13,000 rpm for 2 min. The concentration of the digested DNA was measured with NanoDrop™ as described in Section 4.1.1.2 and stored at -20°C for further use.

4.1.1.5 Ligation

Once the desired product was obtained, it was ligated into the vector. The ligation enzyme used was T_4 DNA ligase (Invitrogen), unless the Quick ligation kit is used (Method 3). The recommended protocol from the manufacturer's instructions suggested a 10:1 molar ratio of insert:vector as a starting point for high ligation efficiency. Ligation ratio that proved to be successful for the longest length, 19R, was 6:1. Hence this was the only deviation made to procedure for this ligation. The formula below was used to estimate the amount of PCR product needed for a 10:1 molar ration of insert to 25 ng vector:

$$\text{X ng insert} = \frac{(\textit{ratio varied}^{\dagger})\ (\text{Y bp PCR product})\ (25\text{ ng vector})}{(*\text{ bp of vector})}$$

where X ng is the amount of PCR product of Y base pairs to be ligated for a 10:1 insert to vector molar ratio.

**This varied according to vector used. It was 3500 bp for pCR-Blunt and 4818 bp for pGL3.*

†This varied according to clone, ratio of 6:1 was used for the 19R and 10:1 was used for all others. A ratio of 3:1 was tried for the extension of the promoter, which is described in more detail in method 3 of this section.

Method 1: Ligation into pCR-Blunt

The vector used was pCR®-Blunt (Invitrogen) which is linear and contains blunt ends, compatible with PCR products amplified by *Phusion*® DNA polymerase. Ligation reaction was set up as follows:

Reagent	Volume
pCR®-Blunt (25 ng)	1 µL
Blunt PCR product	5 µL
10X Ligation Buffer (with ATP)	1 µL
Sterile water	2 µL
T4 DNA Ligase	1 µL
Total volume	10 µL

The ligation reaction was incubated at 16°C for 3 hours and subsequently stored at -20°C until ready for transformation.

Method 2: Ligation into pGL3

The double digested PCR (DD PCR) product and the pGL3-basic (DD pGL3) vector were ligated using their matching sticky ends created from the double digest. The volume of the final product in the reaction was different for each genotype depending on the concentration obtained after using the clean up kit. Ligation reaction was set up as follows:

Reagent	Amount
DD pGL3*	25 ng (volume adjust to concentration obtained per sample)
DD PCR product	Varied - see below (volume adjust to concentration obtained per sample)
10X Ligation Buffer (with ATP)	1 µL
Sterile water	Adjusted to total volume
T4 DNA Ligase	1 µL
Total volume	10 µL

The double digested PCR product varied in ng depending on clone:

- 108 ng in 10R, 13R, 14R, 15R
- 26ng for 18R
- 17ng for 19R

**The double digested pGL3 basic was used for 10R,13R,14R,15R and double digested 10RpGL3 was used for 18R and 19R*

The ligation reaction was incubated at 16°C for 3 h. This incubation time varied from 3 -16 h, as some genotypes (especially longer VNTR regions) proved to take longer to ligate. Ligation reaction was stored at -20°C until ready for transformation. The transformation into *E.Coli* was the same method as described in Section 4.1.1.6.

Method 3: Ligation using Quick Ligation kit (NEB)

The ligation kit used was Quick Ligation kit (NEB). The amount of DNA needed for ligation was calculated using the equation above. Hence for the ligation reaction 36 ng of digested DNA (from XLpUC57) was added to 50 ng digested 10RpGL3, where sterile water was added to reach a 10 µL volume. Following that, 10 µL of 2x ligation buffer was added together with 1 µL T_4 Ligase. The reaction mix was left for 5 minutes at 25°C, then chilled on ice and transformed into competent cells. Transformation was done using TOP10 *E.Coli* as described in Section 4.1.1.6, Method 1, and plated on LB Amp^+ agar plates. They were left to incubate at 37°C overnight.

4.1.1.6 Transformation into *E.Coli*

Method 1: Transformation into One Shot® TOP10 *E.Coli*

Transformation for each ligation reaction into chemically competent *E.Coli* was carried out. One Shot® TOP10 competent cells were thawed on ice and 25 µL was used for each transformation. Each ligation product was added by pippetting 1 µL of ligate to the competent

cells and mixed by gently tapping the tube. The vials were incubated for 30 min on ice and the rest of the ligation mixtures were stored at -20°C. The cells were then heat shocked for 45 s at 42°C without shaking and immediately placed on ice for 2 min. The medium S.O.C was brought to room temperature and 200 µL was added to each vial. The vials were placed in an orbital shaker at 37°C and shaken at 230 rpm for 90 min. Agar plates were placed in an incubator at 37°C for 30 min prior culture of cells. For each transformation 100 µL was spread on LB plates with the appropriate antibiotic (30 µg/mL kanamycin or 50 µg/mL ampicillin). Inverted plates were incubated overnight at 37°C.

Method 2: Transformation into BL21 (DE3)

Transformation for these bacteria was slightly different from TOP10 *E.Coli* where β-mercaptoethanol (β-ME) had to be used. These bacteria were used to obtain a *dcm*$^-$ plasmid. The plasmid was transformed by thawing the competent cells on ice. It was essential that 14-mL Falcon tubes were pre-chilled on ice. The cells were gently mixed and 100 µL of competent cells were added to the Falcon tube. The volume of 1.7 µL β-MEI (diluted from original kit 1:10) was added to the competent cells for a final concentration of 25nM β-ME. The tubes were swirled gently and incubated on ice for 10 minutes, and swirled gently every 2 minutes. Following that 50 ng of plasmid DNA was added to the transformation reaction and swirled gently. The reaction was incubated on ice for 30 min. The transformation reaction was heat-pulsed in a thermomixer at 42°C for 45 seconds. This was followed by incubation on ice for 2 minutes. Preheated SOC medium was added at a 0.9 mL volume to the reaction mix and incubated at 37°C for 1 hour with shaking at 225–250 rpm. Using a sterile spreader 200 µL of the transformed cells were spread onto pre-heated LB agar plates (Amp^+) and incubated overnight at 37°C. The following day the colonies were picked colonies, grown in 3 mL LB Amp^+ medium and incubated overnight at 37°C with shaking at 300 rpm. The plasmid was extracted by MiniPrep kit (Qiagen) as mentioned in Section 4.1.1.8.

4.1.1.7 Analysis of transformants - PCR on colonies

A PCR on colonies was performed in order to easily identify the bacterial cultures containing the desired clones. Once the first clone was identified, 10RpCR-Blunt, it was used as a positive control. Analysis for the 10R was performed by a different approach using restriction enzyme digest, which is described in Section 4.1.1.10. For PCR on colonies the reaction mixture shown in was Table 4.2 used.

Table 4.2 Master mix for PCR on colonies

Component Master Mix	*20 µL reaction volume Reactions: x1*	*Final concentration per reaction*
10 µM Forward Primer:	0.8 µL	0.2 µM
10 µM Reverse Primer:	0.8 µL	0.2 µM
10X Taq buffer	2 µL	1x
Taq DNA polymerase	0.2 µL	1.0 units
10 mM dNTP	0.4 µL	200 µM
Nuclease free H_2O	15.8 µL	NA

Each PCR tube had 20 µL Master Mix. A 20P pipette was set at 5 µL and used to pick a small part of the numbered colonies from the bacterial cultures. This was pipetted up and down in the master mix in each reaction volume and spin again. The following thermocycling conditions were used:

Initial Denaturation step:	94°C for 5 min	
Denaturation:	94°C for 30 s	35 cycles
Annealing:	58°C for 30 s	
Extension:	68°C for 3 min	

The samples were retrieved and run on a 1% agarose gel stained with 5 µL SYBR® safe (Invitrogen) for visualisation with a 2-Log DNA ladder (NEB). Colonies which showed a

positive band were picked and grown in 2 mL LB medium containing 30 µg/mL kanamycin or 50 µg/mL ampicillin. They were left overnight (12-16 h) in a shaking incubator at 300 rpm at 37°C. The following day the plasmid was extracted from the colonies using QIAprep® Spin Miniprep kit (Qiagen).

Primers used for PCR on colonies:

- pCR-Blunt clones: M13rev (forward) and AS11Rev (reverse)
- pGL3 basic clones: RV Primer 3 (forward) and AS11Rev (reverse)

4.1.1.8 Isolation of plasmids from *E.Coli*

Plasmid isolation was made by using QIAprep® Spin Miniprep kit (Qiagen). The overnight culture was pelletted by centrifugation at 8000 rpm for 3 min at RT. The pellet was re-suspended in 250 µL Buffer P1 and transferred to a microcentrifuge tube. The addition of 250 µL Buffer P2 rendered a clear mixture after a thorough mix by inverting the tube 4-6 times. This lysis step was not allowed to proceed for more than 5 min. Then 350 µL Buffer N3 was immediately added and mixed by inverting the tube 4-6 times. The tubes were then centrifuged for 10 min at 13,000 rpm in a table-top microcentrifuge. The supernatant was applied to QIAprep spin column by pipetting and centrifuged for 1 min and the flow-through was discarded. The QIAprep spin column was washed by adding 500 µL Buffer PB, centrifuged for 1 min where the flow-through was discarded. The spin column was then washed with another buffer by adding 750 µL Buffer PE. The tubes were centrifuged for 1 min and the flow-through was discarded. The tubes were centrifuged for a further 1 min to remove any residual wash buffer. The QIAprep column was placed in a clean 1.5 mL microcentrifuge tube, where 50 µL Buffer EB was placed carefully in the centre of the spin column for a final elution. It was allowed to stand for 1 min and eluted subsequently by centrifuging for 1 min. Quantification and qualification of the plasmid DNA was performed using the NanoDrop™ ND-1000 as mentioned in Section 4.1.1.2

4.1.1.9 Sequencing of clones

The plasmids were sent for sequencing to Eurofins MWG. Each sample was diluted to 50 ng/μL to comply with the sample submission guidelines for Euorfins MWG. Samples were sequenced using a forward primer depending on the required region to sequence.

Primers used in sequencing:

- M13rev (-29) used to confirm the correct orientation of the insert in pCR-Blunt
- AS16F to confirm the correct VNTR and end of promoter region.
- RV Primer3 used to confirm clones in pGL3 reporter vector.

4.1.1.10 Restriction enzyme digest

Method 1 : Using EcoR1 for analysis in pCR-blunt

The vector containing 10R was analysed using a restriction enzyme digest. Subsequent clones were analysed by PCR on colonies once a positive clone was obtained and used as a positive control in the PCR reaction. The plasmids were initially extracted by picking ten colonies, incubation in 3 mL LB Kan^+ medium and purified using a mini-prep kit as mentioned in Section 4.1.1.8. The plasmids were then subjected to a restriction enzyme digest to confirm the presence of the desired insert. The enzyme EcoR1 was used since it has a restriction site on the vector 3 bases upstream and 3 bases downstream the insert. The reaction was prepared on ice as follows:

Reactants	Amount per reaction (μL)
EcoR1 10x buffer	1
sterile water	7.5
EcoR1	0.5
Plasmid	1
Total volume	10

The enzyme digest was incubated at 37°C for 1 hour and then run on 1% agarose gel.

Method 2: Restriction enzyme double digest for pGL3

The amplified DNA was subjected to restriction enzyme digestion for the recognition sites Xho1 and Nco1, added by primers Seq4F and Seq5R respectively. The luciferase vector, pGL3 basic was also digested with these enzymes in order to create the sticky ends matching to the ones for the amplified products. The reactants were set on ice as described below:

Reactants	Amount per reaction (µL)
1 x buffer 3	5.5
Xho1	1
Nco1	1.5
pGL3-basic/ PCR product	2
Total volume	10

The reactants were left at 37 °C for 3 hours in a water bath. DNA purification was performed by using MinElute® Reaction Clean up kit as described in Section 4.1.1.4.

Method 3: Restriction enzyme digest for 18R and 19R

Both the 18R and 19R amplicons and the 10RpGL3 template were subjected to an enzyme digest to create the sticky ends. For amplicons purified and eluted with 1x buffer 4 (NEB), 2 µL Msc1 and 1.5 µL Nco1 was directly added to the PCR product. The digest for the template was done with the following reaction mix:

Reactants	Amount per reaction (µL)
10x buffer 4	2
Msc1	2
Nco1	1.5
Water	9.5
10RpGL3	5
Total volume	20

The enzyme digests were left incubated in a water bath at 37 °C for 3 h and subsequently purified by gel extraction with MiniElute gel extraction kit described in Section 4.1.1.4.

Method 4: Restriction enzyme digest for *dcm*⁻ plasmid and host plasmid

The *dcm*⁻ plasmid, XLpUC57, and the host vecotor 10RpGL3 were both digested with Xho1 and Msc1 to create matching complimentary ends. The digest for both vectors was performed as follows:

Reactants	Amount per reaction (µL)	
	XL pUC57	10RpGL3
1 x buffer 4	10.8	11.8
BSA	0.2	0.2
Xho1	1	1
Msc1	2	1
BspH1	1	
Total volume	20	20

The enzyme digest was incubated at 37°C for 3 hours and then run on 1% agarose gel. All the digestion reaction was run on the gel, which was followed by gel extraction for purification. This was done using the MiniElute gel extraction kit as mentioned in Section 4.1.1.4.

4.1.1.11 Site-directed mutagenesis – point mutation

Site-directed mutagenesis (SDM) was carried out by producing a point mutation in order to complete the library bank for the different genotypes and respective SNPs. This method included a nucleotide exchange for the CAA/-AAA triplet variation.

Primer design for point mutation

The primers were designed following the guidelines of the QuickChange II Site-Directed Mutagenesis Kit. The primers should be between 25 and 45 bases in length, with a melting temperature (T_m) of ≥78°C which is calculated as followed:

$$T_m = 81.5 + 0.41(\%GC) \quad (675/N) \quad \%\,mismatch$$

For calculating T_m:

- N is the primer length in bases
- Values for %GC and %mismatch are whole numbers

For pGL3 clones whose last triplet repeat was AAA, the nucleotide exchange was A→ C. Similarly for pGL3 clones whose triplet repeat was CAA, the nucleotide exchange was C→ A. The properties and sequence for these primers is shown in the figure below.

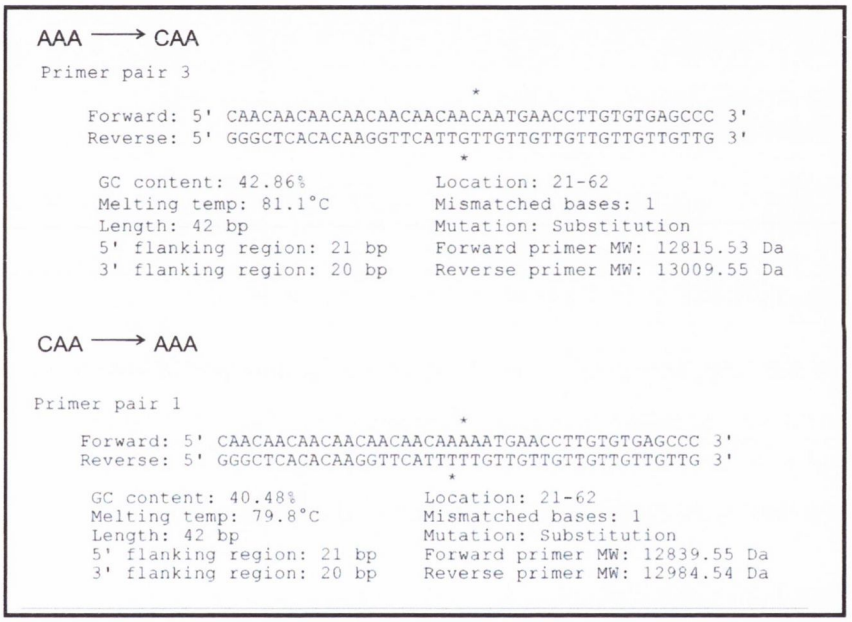


Figure 4.1 Primers for point mutation site-directed mutagenesis

PCR for point mutation site-directed mutagenesis

The PCR was performed on a vector containing the desired insert to be mutated. The first three vectors to be mutated were 13RApCR-Blunt, 14RCpCR-Blunt and 15RApCR-Blunt, and the primers designed above were used depending on the mutation desired on the respective

clone. The DNA polymerase used was *PfuUltra*® high-fidelity enzyme (Stratagene). It exhibits an average error rate 18-fold lower than *Taq* DNA polymerase which makes it ideal to amplify the whole plasmid. The reaction volume was set up to 50 µL and the master mix was prepared as shown in Table 4.3.

Table 4.3 Master mix for site-directed mutagenesis in point mutation

Component Master Mix	***50 µL reaction volume Reactions: x1***	***Final concentration per reaction***
10 µM Forward Primer: CAA_for or AAA _for	1 µL	0.2 µM
10 µM Reverse Primer: CAA_rev or AAA_rev	1 µL	0.2 µM
10X buffer	5 µL	1x
10 mM dNTP	1 µL	200 µM
Nuclease free H_2O	40 µL	

The above reaction was set on ice, and 48 µL of master-mix was added to each PCR tube. This was followed by the addition of 1 µL of plasmid DNA (25 ng/µL) and 1 µL (2.5 units/µL) of *PfuUltra*® high-fidelity polymerase. The tubes were spun and placed in the thermocycler (PeqLab) with a heated lid (110°C) under the following cycling conditions:

Segment 1: 95°C for 30 s

Segment 2:

Denaturation:	95°C for 30 s	12 cycles
Annealing:	55°C for 1 min	
Extension:	68°C for 6 min / 7 min*	
Final elongation:	68°C for 6 min / 7 min*	

The reaction was allowed to cool to 4°C for 2 minutes using ice and subsequently stored at -20°C until further analysis.

**Extension time and final elongation was different depending on the vector size. When a point mutation was performed on pCR-Blunt vectors, 6 min was used. A point mutation in pGL3 vectors had a final elongation and extension of 7 min.*

4.1.1.12 Parental plasmid digest after site-directed mutagenesis

The SDM PCR product also contained the parent plasmids which originated from *E.Coli*. One shot TOP10 *E.Coli* produce methylated and hemi-methylated DNA, unlike synthetic (PCR amplified) DNA which is not methylated. A Dpn1 digest is carried out on the SDM PCR product, where this enzyme digests the parent methylated DNA and leaves the synthesised mutated DNA intact. The PCR product was split into two microcentrifuge tubes, each with 25 µL. In one tube, 1 µL of Dpn1 (20 units/µL) enzyme was added. The other was be used to verify whether the PCR was successful. A positive control for parental digestion was set up as follows:

Reactants	Amount per reaction (µL)
Parent DNA template	1
Dpn1	1
10 x Buffer 4	1
Water	7
Total volume	10

The plasmids were left to digest in a 37 °C water-bath for 1 h.

4.1.1.13 Design and production of mega-primers for mutation

This type of mutation required inserting a long DNA sequence into the -2044 bp clones. The initial step for this mutation was to create mega-primers, where the desired DNA to insert in the clone was amplified and used as primers for the site-directed mutagenesis.

Primers for 5'-end extension to create mega-primers:

The primers for this initial PCR had the same basic idea, where the forward primer was designed based in order to have the first 20 bp annealing to the 10RpGL3 vector and the second 20 bp annealing to the insert to be amplified from XLpUC57. The primers were all analysed and checked for hair pin formations through Primer Stats (http://www.bioinformatics.org/sms2/pcr_primer_stats.html) and the key characteristics and sequence for these primers is listed in Chapter 2, Section 2.4.1. When creating mega-primers designed for -2772 bp 10RpGL3, the reverse primer annealed half way through the -2044 bp ending on both vectors and was 32 bp long. Whilst when creating mega-primers designed for -3329 bp 10RpGL3, the reverse primer annealed half way through the beginning of -2772 bp 10RpGL3.

Primers 3'-end extension to create mega-primers:

The forward primer was designed to anneal at the 5'-end of the promoter and extend (20 bp) into the 5'-end of exon 1 and the reverse primer anneal at the 3' end of the pGL3 vector where the ATG start codon of luciferase gene is located and extends into the 3' end of exon 1. The key characteristics of these primers are described in Chapter 2, Section 2.4.1

The master mix to generate the mega-primers is shown in Table 4.4.

Table 4.4 PCR for mega-primers

Component Master Mix	***50 µL reaction volume Reactions: x1***	***Final concentration per reaction***
10 µM Forward Primer:	2.5 µL	0.5 µM
10 µM Reverse Primer:	2.5 µL	0.5 µM
5X Phusion HF buffer	10 µL	1x
Phusion High-Fidelity DNA polymerase	0.5 µL	1.0 units
DMSO	0.5 µL	3%
10 mM dNTP	1 µL	200 µM
50 mM $MgCl_2$	0.5 µL	0.05 mM (+ 1.5 mM from buffer = 2 mM total Mg^{2+})
Nuclease free H_2O	28 µL	NA

The reaction was set on ice. In order to obtain a high concentration of the mega-primers three PCR reaction tubes were set up, and combined for visualisation and purification. All the above components were added to autoclaved centrifuge tubes, labelled and spun down. A volume of 45 µL of the master mix was added to each PCR reaction tube, 5 µL template DNA was added and spun again. The template DNA was XLpUC57 for thr 5'-end extension and human genomic DNA (Bioline) for the 3'-end extension. The DNA from Bioline was diluted 1:10 concentration. The negative control was autoclaved deionised water, where a PCR tube was set up and 5 µL was added instead of DNA. The PCR reaction tubes were placed in a thermocycler and started with an initial denaturation step of 98°C for 1 min (when XLpUC57 was used) or 5 min (when human genomic DNA was used). Thermocycling conditions were:

- Denaturation: 98°C for 10 s
- Annealing: 60°C for 30 s
- Extension: 72°C for 30 s

} 40 cycles

- Final elongation: 72°C for 7 min*
- Hold: 4°C forever

**A final elongation of 5 min was used when human genomic DNA was used.*

Following amplification, the three PCR reaction tubes were all loaded on the same well on a 1% agarose gel, stained with SYBR safe (Invitrogen). The product was extracted and purified by gel extraction using the Qiagen kit as mentioned in Section 4.1.1.4 (Method 2) and measured with the NanoDrop™ to determine concentration and purity.

4.1.1.14 Site-directed mutagenesis – Extension using mega-primers

The purified mega-primers were used as the primers for the next mutation step, where the 20 bp initially added on to the forward primer would anneal on to the 5'- or 3'-end of the 10RpGL3 vector and keep on extending into the desired mutating insert and finish amplifying the rest of the vector. This master mix for this SDM step was performed as shown in Table 4.5.

Table 4.5 Site-directed mutagenesis using mega-primers

Component Master Mix	***50 µL reaction volume Reactions: x1***	***Final concentration per reaction***
Mega-primers	*varied* µL (400ng)	400 ng
5X Phusion HF buffer	10 µL	1x
Phusion High-Fidelity DNA polymerase	1 µL	2.0 units
10 mM dNTP	1 µL	200 µM
Nuclease free H_2O	up to 49 µL	NA

The SDM was set on reaction on ice, where all the above ingredients were added in an autoclave centrifuge tube, label and spun in centrifuge. The volume of 49 µL master mix was added to the PCR reaction tubes and 1 µL (25 ng) of host vector was included. A negative control was also included where 1 µL of sterile deionised water was added instead of the DNA.

Host vector varied according to mutation desired:

- ✓ -2044 bp 10RpGL3 for -2772 bp 10RpGL3
- ✓ -2772 bp10RpGL3 for -3329 bp 10RpGL3
- ✓ -705 bp 10RpGL3 for -705 bp Ex1 10RpGL3
- ✓ -2044 bp 10RpGL3 for -2044 bp EX1 10RpGL3
- ✓ -2772 bp 10RpGL3 for -2772 bp Ex1 10RpGL3
- ✓ -3329 bp 10RpGL3 for -3329 bp Ex1 10RpGL3

After spinning in the centrifuge, the reaction tubes were placed in the thermocycler and set to the following cycling conditions:

Segment 1: 98°C for 1 min

Segment 2:

• Denaturation:	98°C for 1 min	5 cycles
• Annealing:	52°C for 1 min	
• Extension:	68°C for *varied**	

Segment 3:

• Denaturation:	98°C for 1 min	13 cycles
• Annealing:	55°C for 1 min	
• Extension:	68°C for *varied**	

**Both extensions time varied depending on total length of vector:*

- ✓ -2772 bp 10RpGL3: 7 min 30 s
- ✓ -3329 bp 10RpGL3, -2772 bp Ex1 10RpGL3 and -3329 bp Ex1 10RpGL3: 8 min 30 s
- ✓ -705 bp Ex1 10RpGL3 : 6 min
- ✓ -2044 bp Ex1 10RpGL3: 7 min

4.1.2 Construction of clones

The first clones to be constructed were the four main genotypes, i.e., 10R, 13R, 14R and 15R. Genotypes 16R and 17R were cloned from student projects. The method of cloning was the same for all genotypes. There were initially amplified from human genomic DNA then cloned into a standard vector, pCR-Blunt.

4.1.2.1 Cloning into pCR-Blunt

Cloning into pCR-blunt was performed for genotypes 10R,13R,14R,15R,16R,17R. The human DNA used for amplification of the desired region was either from HapMap samples or genomic DNA extracted from buccal swabs. The genotypes 10R and 15R were amplified from HapMap samples NA12874 and NA11995. Genotypes 13R, 15R, 16R and 17R were amplified from extracted DNA using the method mentioned in Section 4.1.1.1, then qualified (by a PCR for the Beta-Actin) and quantified (using the NanoDrop™ Spectrophotometer) using the method mentioned in Section 4.1.1.2.

The amplification for these genotypes was performed by running a PCR decribed in Section 4.1.1.3, Method 1. After the reaction was finished, the amplicon was purified by gel extraction by using the method described in Section 4.1.1.4, Method 1. Once purified and quantified by NanoDrop™ Spectrophotometer, the putative promoters were ligated into the blunt vector using the ligation Method 1 explained in section 4.1.1.5. The ligation ratio was 10:1

insert:vector and once the reaction was set, it was left to ligate at 16°C for 3 hours. Following that, the ligates were transformed into *E.Coli*, and this is explained in Section 4.1.1.6, Method 1.

Analysis of transformants was conducted by a PCR on colonies (Section 4.1.1.7). The colonies showing a positive band, after gel visualisation, were picked and grown in LB Kan^+. Following 16 hour incubation at 37°C with 300 rpm shaking, the plasmid was extracted from the bacteria by using a Miniprep kit, and the method is described in Section 4.1.1.8. The first clone identified, i.e., 10RpCR-blunt, was analysed by a restriction enzyme digest with EcoR1 (Section 4.1.1.10, Method 1). The samples sent for sequencing were prepared as described in Section 4.1.1.9 using M13rev (-29) and AS16F.

(Clones 16RApCR-Blunt and 17RApCR-Blunt were parts of an undergraduate research project under my supervision by senior freshman medicine students, Hazel O'Neil and Owen Thorpe.)

Summary of clones for section 4.1.2.1

Genotypes cloned into pCR-Blunt					
10RpCR-Blunt From HapMap NA12874	**13RApCR-Blunt** From extracted Human Genomic DNA	**14RCpCR-Blunt** From extracted Human Genomic DNA	**15RApCR-Blunt** From HapMap NA11995	**16RApCR-Blunt** From extracted Human Genomic DNA	**17RApCR-Blunt** From extracted Human Genomic DNA

4.1.2.2 Sub-cloning into pGL3 reporter vector

4.1.2.2.1 Sub-Cloning 10R,13R,14R, 15,16R,17R from pCR-Blunt

The promoter region for the four main genotypes cloned into pCR-Blunt; 10R, 13R, 14R, 15R, were sub-cloned into the reporter plasmid pGL3. Clones 16R and 17R proved to be difficult to sub-clone, thus the method outlined below was not used for these genotypes. The 17R in pGL3 was later obtained through a mutation which is outlined in Section 4.1.2.3.

The amplification from the pCR-blunt clones was conducted by using Method 2 in Section 4.1.1.3. The primers were designed to have restriction sites matching the desired region of insertion in pGL3-basic vector. The restriction sites used were Xho1 and Nco1, where after purification by gel extraction (Section 4.1.1.4, Method 2) the PCR product and pGL3-Basic were digested with these enzymes using the protocol mentioned in Section 4.1.1.10, Method 2. After enzyme digest, the product (i.e. the PCR and the pGL3-basic vector) was purified by a reaction clean up kit described in Section 4.1.1.10, Method 2.

The ligation of the double digested promoter region and the host vector, pGL3, was performed using a 10:1 molar ratio (similar to pCR-Blunt). The ligation reaction is discussed in Section 4.1.1.5, Method 2. Once ligated and transformated into *E.Coli* (Section 4.1.1.6), positive colonies were analysed by PCR on colonies (section 4.1.1.7) using RV Primer 3 and AS11Rev as forward and reverse primer respectively. The amplicon was visualised on an agarose gel (Chapter 2, Section 2.7). Colonies which gave a significant band for the desired length were picked and grown in 2 mL LB medium containing 50 µg/mL ampicillin. They were left overnight (12-16 hrs) in a shaking incubator at 300 rpm at 37°C. The plasmid was extracted from *E.Coli* after this incubation period using QIAprep® Spin Miniprep kit (Qiagen) as described in Section 4.1.1.8. The first pGL3 clone to be identified was 10RpGL3. This was used as a positive control for all other clones.

Following purification, potential positive plasmids were sent for sequencing to Eurofins MWG. The concentration for each sample was diluted to 50 ng/µL in order to comply with the sample submission guidelines for Euorfins MWG. They were sequenced using AS16For and RV Primer 3 to confirm the desired insert at the correct insertion into the luciferase pGL3 vector (Section 4.1.1.9).

4.1.2.2.2 Sub-cloning 18R and 19R from pEX

Identification for the rare and also longest triplet repeats, 18R and 19R, proved to more difficult. The sequence for both genotypes was sent to Eurofins MWG and was synthesised in a standard vector pEX-K. This is shown in Chapter 2, Section 2.8. The size synthesised was 738 bp and 741 bp for 18R and 19R respectively, where an Msc1 restriction site located 330 bp upstream from the start of the VNTR.

The sub-cloning for these genotypes hence adopted a different approach, therefore a new primer was designed to start amplification at this region primers were designed in such a way that after enzyme digestion, amplicons would have sticky ends matching the ones for the desired template. In this case, 10RpGL3 was used as the template instead of pGL3-basic, and the remaining part of the potential promoter region (1338 bp) located upstream of the VNTR was already present in the vector. The PCR to amplify the 738 and 741 bp region is mentioned in Section 4.1.1.3 Method 3. The amplicon was purified using the MinElute gel extraction kit (Section 4.1.1.4, Method 2) and digested with Msc1 and Nco1 as explained in Section 4.1.1.10 Method 3.

Two different insert to vector ratio were used for these genotypes, 10:1 and 6:1. Both ratios were tried on 18R and 19R, where 10:1 was successful for 18R and 6:1 was successful for 19R. The ligation reaction was performed under the same conditions mentioned in Section 4.1.1.5 method 2. What differed from previous ligations was the time and temperature, where the ligate was left overnight on a microcentrifuge rack on ice. The following day, the ligate was transformed into *E.Coli* (Section 4.1.1.6), plated on an Amp^+ plates and left to grow overnight in a 37 ^oC incubator.

The colonies grown on the Amp^+ plates were screened by PCR on colonies using the same protocol as for the other pGL3 vectors (Section 4.1.1.7). Positive colonies were picked, grown in 3 mL LB Amp^+ overnight in a shaking orbital at 37 ^oC and the plasmid was extracted from the bacteria (Section 4.1.1.8) the following day. The colonies were sent for sequencing to Eurofins MWG at a concentration of 50 ng/µL with AS16For.

Summary of clones for section 4.1.2.2

Genotypes cloned into pGL3 (reporter plasmid)					
10RpGL3 From 10RpCR-Blunt	**13RApGL3** From 13RApCR-Blunt	**14RCpGL3** From 14RCpCR-Blunt	**15RApGL3** From 15RApCR-Blunt	**18RCpGL3** From 18RCpEX-K	**19RCpGL3** From 19RCpEX-K

4.1.2.3 Point mutation for the VNTR genotypes

Once the reporter plasmids for the length variations were constructed, they were used as a template to derive the SNP for the allele towards the end of the promoter. Initially this method was used to create a point mutation only in pCR-Blunt vectors:

- ✓ 13RApCR-Blunt ⟶ 13RCpCR-Blunt
- ✓ 14RCpCR-Blunt ⟶ 14RApCR-Blunt
- ✓ 15RApCR-Blunt ⟶ 15RCpCR-Blunt
- ✓ 16RApCR-Blunt ⟶ 16RCpCR-Blunt
- ✓ 17RApCR-Blunt ⟶ 17RCpCR-Blunt

The design of primers and the PCR adopted is discussed in Section 4.1.1.11. Following amplification, the parent plasmid was digested in the reaction mixture using Dpn1 enzyme (Section 4.1.1.12).

The transformation protocol described in Section 4.1.1.5 was used for the following plasmids:

- SDM PCR product digested with Dpn1
- SDM PCR product not-digested
- Parent plasmid digested with Dpn1
- Negative control for PCR (H_2O)

} Controls for digest and PCR reaction

The transformants were plated on Kan^+ plates and incubated at 37°C overnight inverted for next day analysis.

For the following SDM point mutated clones (13R, 14R and 15R in pCR-Blunt) random colonies (max 10) were picked and grown in LB Kan^+ overnight and isolated next day using the Miniprep kit (Qiagen) (Section 4.1.1.8). These were then digested with EcoR1 to confirm the presence and size of the insert (Section 4.1.1.10). Colonies for 16R and 17R were screened using a PCR on colonies to check for potential mutated colonies (Section 4.1.1.7). The PCR was not used to identify the correct mutant, but mainly to omit plasmids that were elongated or shortened during mutation due to the nature of the triplet sequence. Positive colonies were picked, grown overnight in 2 mL LB with Kan^+ and purified by a miniprep (Section 4.1.1.8). The colonies were sent for sequencing to Eurofins MWG at a concentration of 50 ng/ µL with AS16For.

Successful point mutation vectors sub-cloned in the reporter plasmid

Following the successful SNP mutation in the clone that were in pCR-Blunt, the next step was to sub-clone them into the pGL3 vector that expresses the luciferase gene. This proved to be more problematic than in previous clones. The same method as mentioned in Section 4.1.2.2 was used. The only successful sub-clones were 14RA and 15RC into pGL3. The rest of the point mutations were performed again on the pGL3 vectors, since sub-cloning proved to be more difficult and time consuming (for 13R, 16R and 17R).

The 13RC, 18RA and 19RA in the reporter vector was constructed from the parent 13RApGL3, 18RCpGL3 and 19RCpGL3 respectively, whilst the 17RApGL3 resulted from a mutation while performing the SDM on the 18RCpGL3 (to get the 18RApGL3). The sub-cloning for 16RA and 16RC from pCR-Blunt into pGL3 was unsuccessful following various attempts and repetitions.

(Parts of this section were carried out through undergraduate research projects under my supervision: ERASMUS students, by Maximilian Schmitt and Jessica Krause, and undergraduate senior sophister pharmacy research project by Sarah Hosey).

Summary of clones 4.1.2.3

Point mutations in genotypes cloned in pCR-Blunt				
13RCpCR-Blunt From 13RApCR-Blunt	**14RApCR-Blunt** From 14RCpCR-Blunt	**15RCpCR-Blunt** From 15RApCR-Blunt	**16RCpCR-Blunt** From 16RApCR-Blunt	**17RCpCR-Blunt** From 17RApCR-Blunt
Followed by Sub-Cloning into pGL3				
14RApGL3 From 14RApCR-Blunt		**15RCpGL3** From 15RCpCR-Blunt		

Point mutations in genotypes sub-cloned in pGL3 (reporter plasmid)				
13RCpGL3 From 13RApGL3	**17RApGL3** From 18RCpGL3	**17RCpGL3** From 17RApGL3	**18RApGL3** From 18RCpGL3	**19RApGL3** From 19RCpGL3

4.1.2.4 Construction of promoter length variations

4.1.2.4.1 Construction of the short length promoter, -705 bp 10RpGL3

Initial experiments were carried out to generate the longest version of the promoter for HRH4, hence the method outlined below was performed having this in mind. The short version of the potential promoter was only achieved as a result of a non-conventional ligation from restriction sites.

The plasmid hosting the synthesised DNA (XLpUC57) was grown in a different bacterial strain (BL21(DE3), *dcm*$^-$). The nucleotides from the pUC57 at the Msc1 restriction site produced a methyl group that affected digestion by Msc1 digestion. The enzyme Msc1 is sensitive for bacterial methylation and has the recognition site TGG^CCA, where the CCA of the enzyme is followed by GG in pUC57. Hence the final sequence is TGG^**C**CAGG. This causes the '**C**' to be methylated when grown in TOP10 *E.Coli* since they are *dcm*$^+$ strain. The bacterial strain BL21(DE3) was used instead which are *dcm*$^-$ resulting in cleavage of the restriction site. Hence transformation was performed using Method 2 described in Section 4.1.1.6. Once the plasmid was extracted, it was ready for digestion with the two enzymes which produce the sticky ends to the extension insert, Xho1 and Msc1. This digestion was needed for both the XLpUC57 vector hosting the insert and the receiving plasmid, 10RpGL3. On digestion with these enzymes, the 10RpGL3 opens up and ligation for the 'extension promoter' is possible. The vector XLpUC57 also requires another enzyme for digestion, BspH1. This enzyme cuts in 3 points the vector however, it not does not affect the insert. This was needed since the vector cut by Xho1 and Msc1 is the same size as the insert cut out, and it would not have been possible to separate the pUC57 band from the desired insert. Hence this enzyme helps to break down the host vector pUC57 in smaller size and fragments and allows this easier differentiation from the desired insert. The digest for both vectors was performed using Method 4 described in Section 4.1.1.10.

After purification from agarose gels, ligation was the next step. Ligation was attempted initially using T_4 ligase from Invitrogen as described in Method 2 in Section 4.1.1.5. Various experiments were carried out and different vector:insert ratios were used with no success. However on changing the supplier to NEB, the Quick ligation kit recommended using a 3:1

ratio which was successfully tried. Ligation was performed as described in Section 4.1.1.5 Method 3. After transformation (Section 4.1.1.6, Method 1), the bacteria were plated on Amp^+ plates and incubated overnight. The colonies found the next day were analysed by PCR on colonies. The primers used were RV Primer 3 and AS11_Rev and the PCR was performed as in Section 4.1.1.7.

The band visualized on gel electrophoresis showed significant unexpected smaller bands. These were sent for sequencing following extraction by Miniprep Kit (Section 4.1.1.8) with the forward primer RV Primer 3. Sequencing results received from Eurofins MWG were analysed and showed the presence of a shorter version of the 10RpGL3. This is further discussed in the results section. The short promoter produced was kept and used for further analysis.

4.1.2.4.2 Extension of the 5'-end promoter

The method of cloning described above, which was originally designed to obtain the longest length for the promoter region of the HRH4, was never successful. Numerous attempts were made, either by the restriction enzyme method or by PCR directly from the pUC57 vector, using primers from the vector. Therefore, a new method was designed to extend the original promoter; -2044bp 10RpGL3. This method included creating mega-primers for the desired region and performing a site-directed mutagenesis by inserting this amplified extension.

The initial PCR to create the mega-primers is described in Section 4.1.1.13. The design of each primer is also explained in this section. The amplicon for each reaction was purified and once quantified with NanoDrop™ spectrophotometer, they were used in the second PCR for mutation. The amount of 400 ng of mega-primers was used in each site-directed mutagenesis and the protocol adopted is mentioned in Section 4.1.1.14. Following that, the parental DNA template was digested with Dpn1 as described in Section 4.1.1.12.

The transformation for these mutations was done on One Shot® TOP10 competent cells according to Section 4.1.1.5. The transformations for the samples and controls were conducted as follows:

- SDM PCR product digested with Dpn1
- SDM PCR product not-digested
- Parent plasmid (-2044 bp) digested with Dpn1
- Negative control for PCR (H_2O)

} Controls

Overnight incubation for the transformed competent cells showed a good number of potential positive colonies for all SDMs. They were screened by a PCR using RV Primer 3 and As11_rev (Section 4.1.1.7). The positive control for this PCR used was in each case the parent template of the SDM PCR, where the expected band would have an additional 762 bp or 554 bp (for -2772 bp 10RpGL3 and -3329 bp 10RpGL3 respectively) when compared to the original parent clone.

The colonies that looked promising were picked and grown overnight in LB Amp^+ medium, they were then purified Miniprep (Section 4.1.1.8) and sent for sequencing with RV Primer 3 to confirm the presence of the desired insert (Section 4.1.1.9).

(Parts of this section were carried out through undergraduate research projects under my supervision: ERASMUS students, by Aurelie Rogeau and Agathe Deniau).

4.1.2.4.3 Extension of the 3'-end of the promoter - Addition of exon 1

All promoter lengths of 10RpGL3, i.e. -705, -2044, -2772, -3329 bp, were modified to include exon 1 in each vector construct. This was also performed on -2044 bp 13RpGL3 (which has the highest allele frequency for all genotypes). The exon 1 was amplified from human genomic DNA (Bioline), and the sample used was from Bioline, where the original stock solution was diluted 1:10 (10 ng/µL). Mega-primers for this mutation were also generated by a PCR (Section 4.1.1.13), similar to the 5'-end extension. After purification they were used in a second amplification for the SDM. The method adopted for this plasmid modification is explained in Section 4.1.1.14.

Once the SDM PCR was finished, 25 µL of the reaction volume from the amplicon was removed and added to a fresh 1.5 mL centrifuge tube, where 1 µL of Dpn1 enzyme was added. The control used was set up as mentioned above (for the 5'-end extension), where the parent -2044 bp 10RpGL3 was used as an indication that the digestion was successful. Following digestion (Section 4.1.1.12) they were transformed into TOP10 *E.Coli* (Section 4.1.1.6, Method 1). The transformants, i.e., the digested SDM clone, digested parent clone, SDM reaction mix non-digested and the negative control from the SDM were grown on agar plates containing Amipicillin and incubated overnight at 37°C.

Colonies were screened with a PCR, where Seq1F and Ex1 SDM_Rev were used as primers. The same protocol for PCR on colonies was used as in section 4.1.1.7, but cycling conditions varied where the extension time was only 1 minute and 30 seconds since this PCR creates a smaller amplicon. Colonies showing a positive band were picked and grown overnight in LB Amp^+ medium. After incubation for 16 hours with orbital shaking at 37°C, the plasmid was extracted, purified (Section 4.1.1.8), and sent for sequencing at Eurofins MWG with Seq1F.

Summary of clones for section 4.1.2.4

Construction of promoter length variations from original clone						
-705 bp 10RpGL3 From -2044 bp 10RpGL3	**-2772 bp 10RpGL3** From -2044 bp 10RpGL3	**-3329 bp 10pGL3** From -2772 bp 10RpGL3	**-705 bp Ex1 10RpGL3** From -705 bp 10RpGL3	**-2044 bp Ex1 10RpGL3** From -2044 bp 10RpGL3	**-2772 bp Ex1 10RpGL3** From -2772 bp 10RpGL3	**-3329 bp Ex1 10RpGL3** From -3329bp 10RpGL3

4.2 Results

4.2.1 Cloning the HRH4 promoter in a standard vector

4.2.1.1 Human genomic DNA quantification and qualification

After achieving ethical approval from the ethics board, buccal swabs were collected from healthy volunteers. Genotypes identified from previous student projects were used as well. The DNA was extracted and purified using Qiagen's DNA investigator® kit. All measurements were done in triplicate and averaged. The genomic DNA isolated by the DNA investigator kit gave a consistent reading of approximate 30 ng/µL for each sample, Table 4.6 shows these readings.

Table 4.6 Human Genomic DNA quantification

Sample ID	ng/uL	A260	260/280	Average ng/µL
HG 13R.1	28.61	0.57	2.00	28.78
HG 13R.2	28.83	0.58	1.86	
HG 13R.3	28.91	0.58	1.90	
HG 14R.1	29.95	0.60	1.88	29.38
HG 14R.2	29.07	0.58	1.82	
HG 14R.3	29.11	0.58	1.94	

The quality of the human genomic DNA was tested by performing a PCR for the Beta-Actin gene and compared a commercially available human genomic DNA (Bioline). Figure 4.2 shows the amplified DNA having a band of 206 bp.

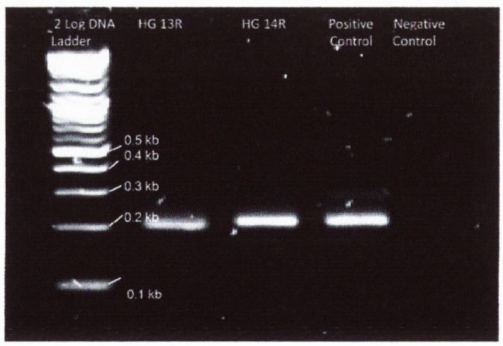


Figure 4.2 PCR qualification for Beta-Actin band

4.2.1.2 Amplification of the promoter region: 10R, 13R, 14R and 15R

The full promoter region was amplified with newly designed primers Seq2F and Seq3R and loaded on a 1% agarose gel stained with SYBR® safe (Invitrogen). The expected amplicons give a product size of 2247 bp for 10R, 2256 bp for 13R, 2259 bp for 14R and 2262 bp for 15R. The positive control used was genomic DNA (Bioline) and the negative control was 5 µL of sterile water. Figure 4.3 and Figure 4.4 show the visualised bands on Molecular Imager (Bio-Rad).

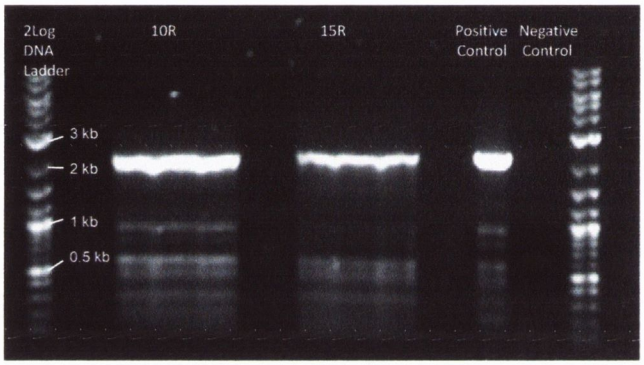


Figure 4.3 Agarose gel for PCR purification for 10R and 15R

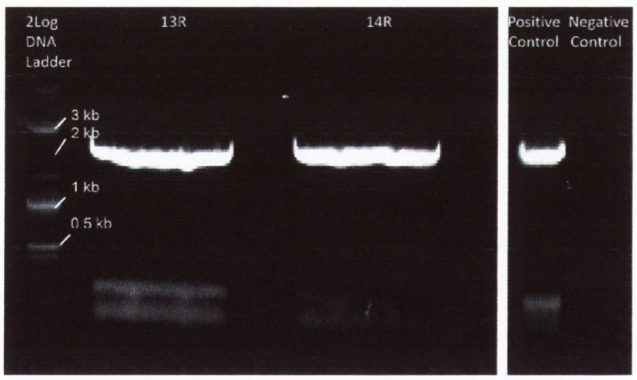


Figure 4.4 Agarose gel for PCR purification for 13R and 14R

The bands were excised and purified by gel extraction, and quantified using NanoDrop™ spectrophotometer to determine the concentration of the nucleic acids. The average DNA

concentration was ranging from 35 to 70 ng/μL,and 260/280 ratio is approximately 1.8 and this is shown in Table 4.7.

Table 4.7 DNA quantification after gel extraction

Sample ID	ng/uL	A260	260/280	Average ng/μL
10R.1	35.93	0.72	1.78	35.56
10R.2	38.45	0.77	1.79	
10R.3	32.32	0.65	1.79	
13R.1	52.28	1.05	1.92	41.07
13R.2	33.90	0.68	2.41	
13R.3	37.04	0.74	1.72	
14R.1	69.91	1.40	1.72	68.56
14R.2	68.31	1.37	1.78	
14R.3	67.46	1.35	1.76	
15R.1	43.00	0.86	1.65	43.26
15R.2	43.14	0.86	1.67	
15R.3	43.63	0.87	1.73	

4.2.1.3 Ligation into pCR-Blunt vector

The concentration of each PCR product was used to determine the amount needed to ligate it into the linearised vector. Using the equation mentioned in section 4.1.1.5, the amount of 161.36 ng of DNA was ligated with 25 ng vector (10:1 molar ratio). This ratio was used for all four genotypes. Figure 4.5 shows the end product of ligation for 10R insert into pCR-Blunt. The blunt PCR insert (2247 bp for 10R) ligates in between EcoR1 restriction sites of the linear vector (3,500 bp) making the product pCR-Blunt and the 10R insert a circular vector of 5759 bp as shown in Figure 4.5.

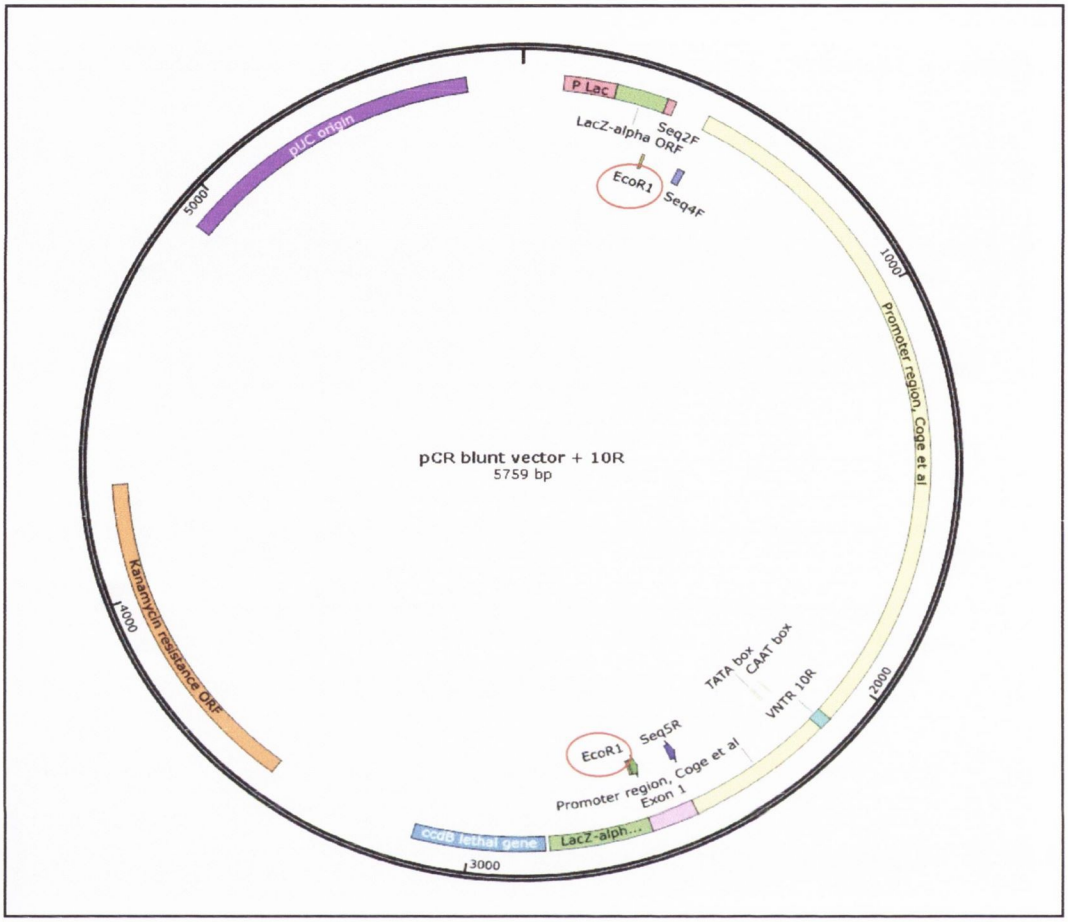


Figure 4.5 10R insert ligated into pCR-Blunt

4.2.1.4 Analyzing transformants for pCR-Blunt vectors

The first colonies obtained were directly picked and the plasmid was extracted. The plasmid clones were then subjected to restriction enzyme digestion with EcoR1 enzyme, where it cuts the vector at both ends of the insert (as shown in the Figure 4.5). Figure 4.6 shows the gel analysis from the enzyme digest. Since pCR-blunt vector is a linear vector, and the DNA polymerase used for the PCR produces blunt ended amplicons (i.e. no poly A tail is formed at the end of the amplified product), there was a possibility that the promoter could ligate in either orientation.



Figure 4.6 Restriction enzyme digest for plasmids with EcoR1 enzyme

The two colonies (#2 & #9) which showed a band at approximately 2259 bp, were sent for sequencing (circled in the figure above). The plasmids were sequenced with M13Rev (-29) which is a standard primer found in pCR-Blunt vector. It is located approximately 100 bp upstream to the PCR insert. The plasmids were also sequenced with AS16For primer, which is located 90 bp upstream to the VNTR, in order to confirm the correct genotype cloned. The sequencing results for the plasmids identified clone #9 for the 10R VNTR length. Plasmid #2 did not have the correct orientation even though they had the correct size product as seen in the restriction digest.

Figure 4.7 and Figure 4.8 show sequencing results, where Figure 4.7 (sequenced with M13Rev) represents the start of the 5' of the linearised vector leading into the start of the desired insert and Figure 4.8 (sequenced with AS16For) shows the genotype inserted in the clone. Sequencing result also confirmed the presence of rs615283 (A/G) polymorphism found in LD with the 10R in Chapter 3, and this is described more in detail in the discussion.

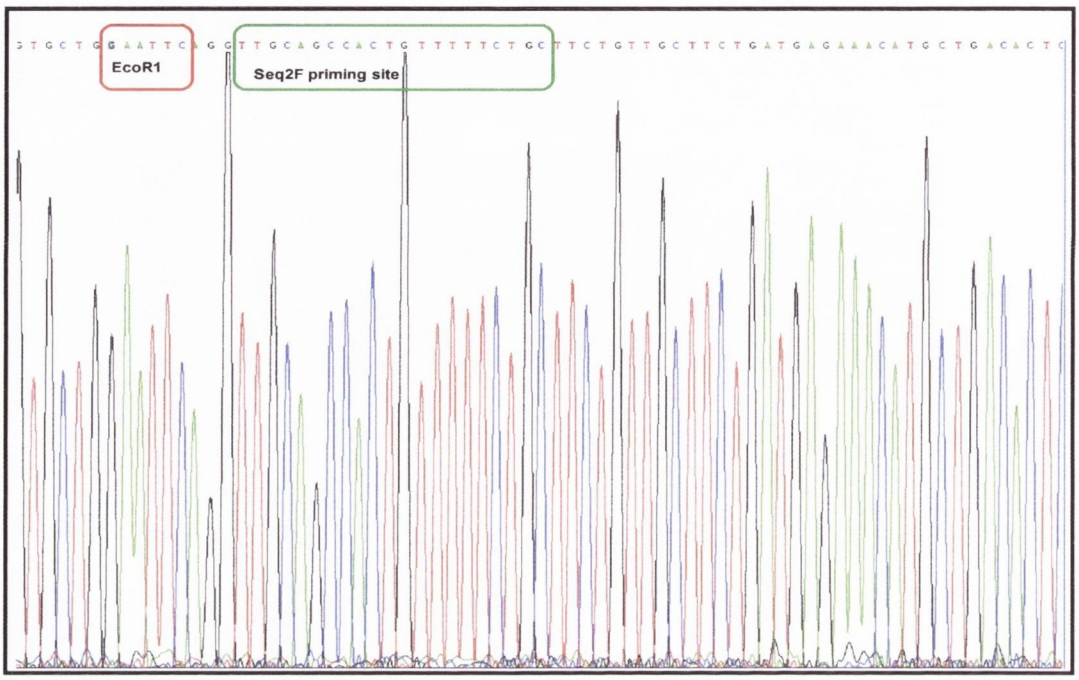


Figure 4.7 Sequencing results for 10RpCR-blunt clone with M13Rev (-29)

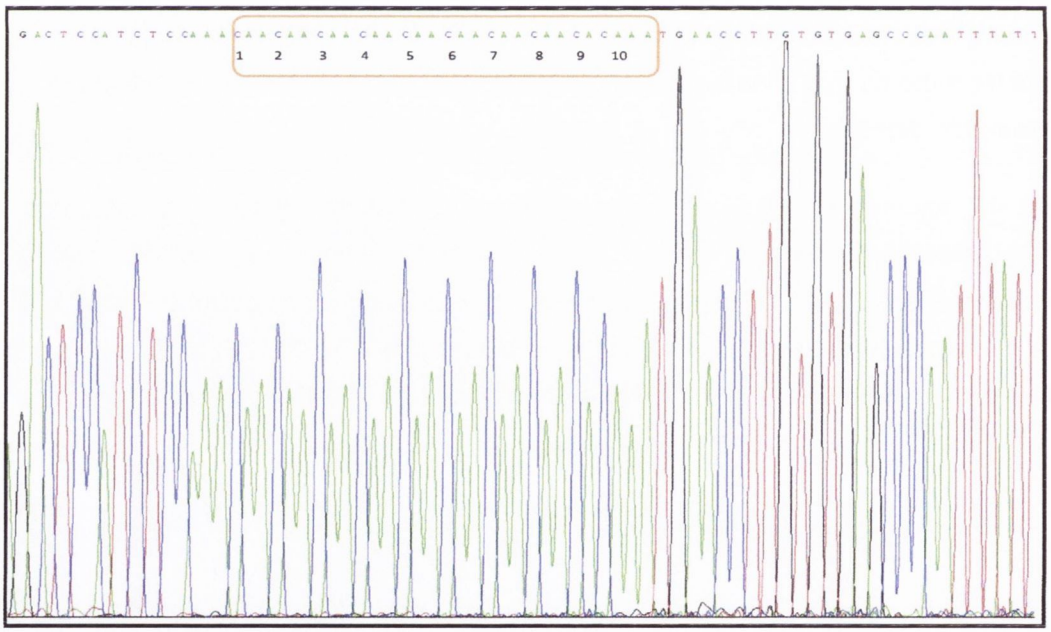


Figure 4.8 Sequencing result for 10RpCR-Blunt clone AS16For primer

The next clone to be identified was the 13R length repeat. Following the construction of 10RpCR-Blunt, which was used as a positive control in all subsequent experiments, a PCR on colonies was carried out to identify colonies with the desired clone before growing up individual colonies and extracting the plasmid. The primers used were M13Rev (having a forward priming site on the vector) and AS11_rev (having a reverse priming site on the insert).



Figure 4.9 PCR on colonies for 13R clone, using 10R as a positive control

The gel shown above in Figure 4.9 shows the colonies identified to harbour a potential promoter insert (2041 bp). These colonies were picked, grown in 2 mL Kan^+, plasmid purified and sent for sequencing. Table 4.8 shows the average concentrations for the plasmids.

Table 4.8 Plasmid concentrations for 13R colonies after mini-prep

Sample ID	ng/uL	A260	260/280	Average ng/µL
13R.1 #2	363.30	7.27	1.88	333.93
13R.2 #2	323.68	6.03	1.85	
13R.3 #2	314.80	6.30	1.85	
13R.1 #3	223.53	4.47	1.89	224.37
13R.2 #3	221.37	4.42	1.87	
13R.3 #3	228.22	4.56	1.90	
13R.1 #5	158.69	3.17	1.94	160.33
13R.2 #5	160.41	3.20	1.89	
13R.3 #5	161.88	3.23	1.88	
13R.1 #6	244.74	4.89	1.87	246.72
13R.2 #6	250.92	5.02	1.86	
13R.3 #6	244.50	4.89	1.87	

The colony circled in red (#2) in Figure 4.9 shows the colony that had the desired clone, which was confirmed by sequencing results as shown in Figure 4.10. Alignment for the sequencing results with the desired vector is found in Appendix 3.

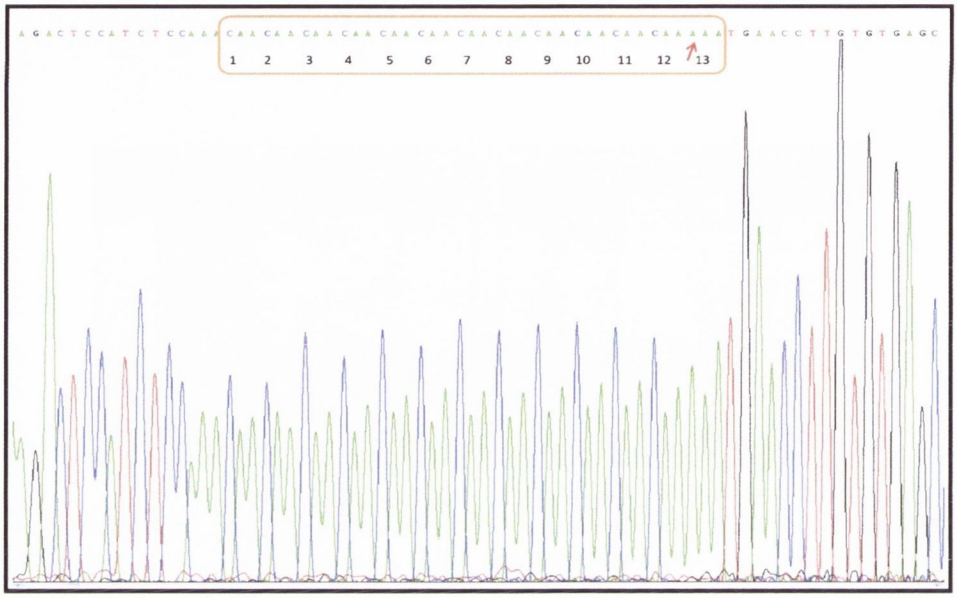


Figure 4.10 Sequencing result for 13RApCR-Blunt clone with AS16For

The next clone identified from the colonies was 15RA. The PCR on colonies showed a strong band for the amplicon with M13Rev and AS11_rev, similar to when compared with the 10RpCR-Blunt from the gel electrophoresis. And this is shown is the Figure 4.11.

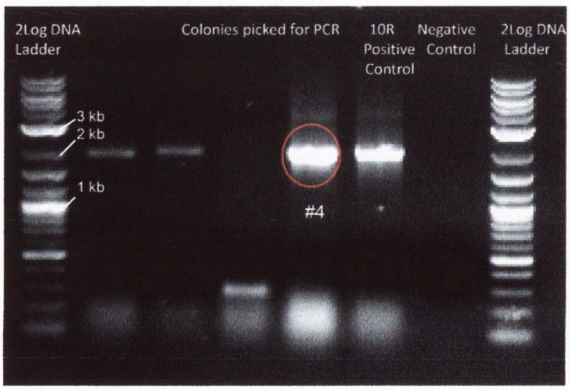


Figure 4.11 PCR on colonies for 15RA clone, using 10R as a positive control

Colony in well no.4 was the only colony that showed a strong significant band comparable to the 10RpCR-Blunt. It was picked grown in 2 mL LB Kan^+ and the plasmid was extracted and sent for sequencing. Table 4.9 shows the concentration obtained after purification.

Table 4.9 Plasmid concentrations for 15RA colonies after mini-prep

Sample ID	ng/uL	A260	260/280	Average ng/µL
15R.1 #4	336.70	6.73	1.85	338.57
15R.2 #4	340.98	6.82	1.85	
15R.3 #4	338.02	6.76	1.83	

Figure 4.12 shows sequencing results for 15RApCR-Blunt.

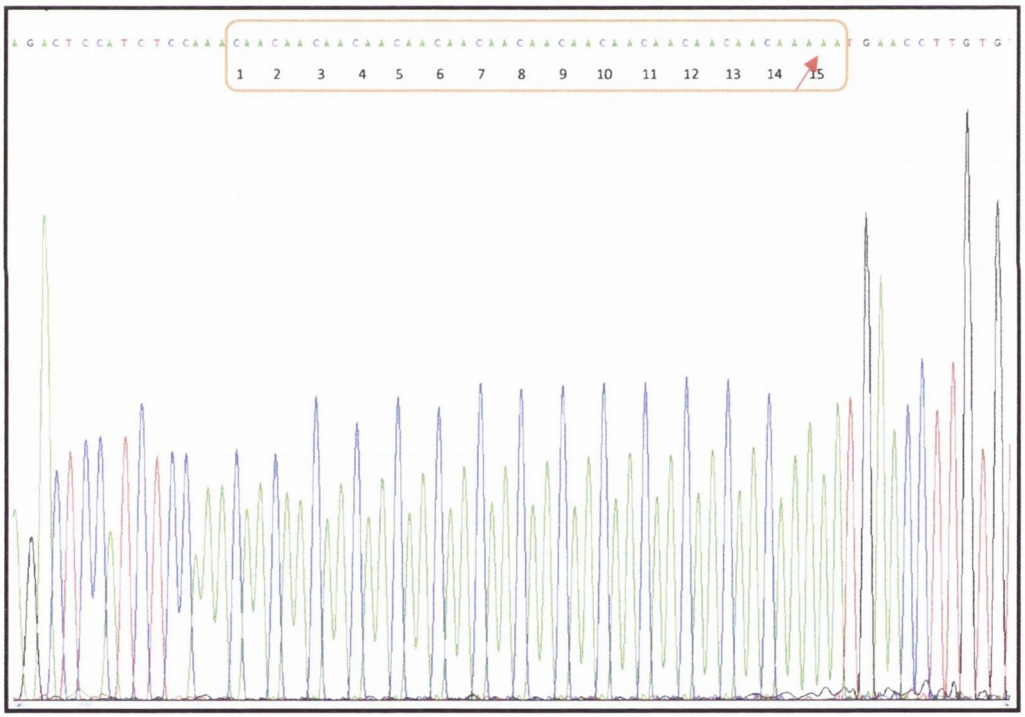


Figure 4.12 Sequencing result for 15RApCR-Blunt clone with AS16For

The last clone (having a high allele frequency genotype) to be constructed was 14RC, which was screened from more than 50 colonies, in which only 1 colony showed a faint band as

shown in Figure 4.13. This colony was extracted, purified for its plasmid, and sent for sequencing.

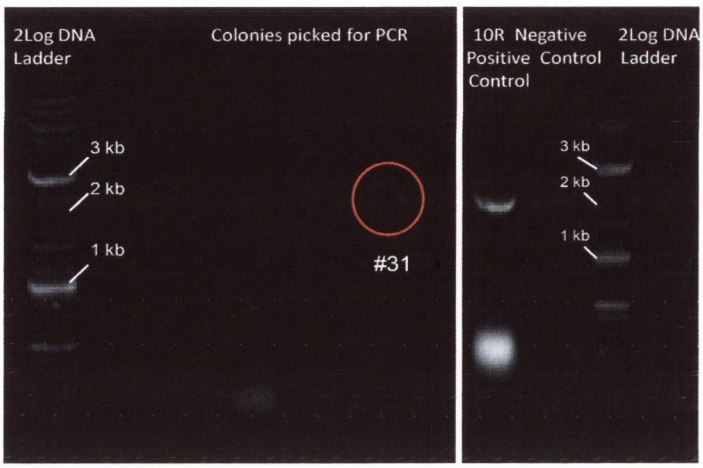


Figure 4.13 PCR on colonies for 14RC clone, using 10R as a positive control

The concentration of this plasmid was low when compared to previous purified clones, which is not surprising when compared to the faint band in the above gel. The colonies for this clone were also smaller in comparison to previous transformants. The plasmid concentration for 14R is shown in the Table 4.10.

Table 4.10 Plasmid concentrations for 14C colonies after mini-prep

Sample ID	ng/uL	A260	260/280	Average ng/µL
14R.1 #31	40.62	0.81	1.78	41.82
14R.2 #31	41.96	0.84	1.62	
14R.3 #31	42.88	0.86	1.60	

The plasmid was sequenced with AS16For primer and M13Rev (Figure 4.14) located on the vector.

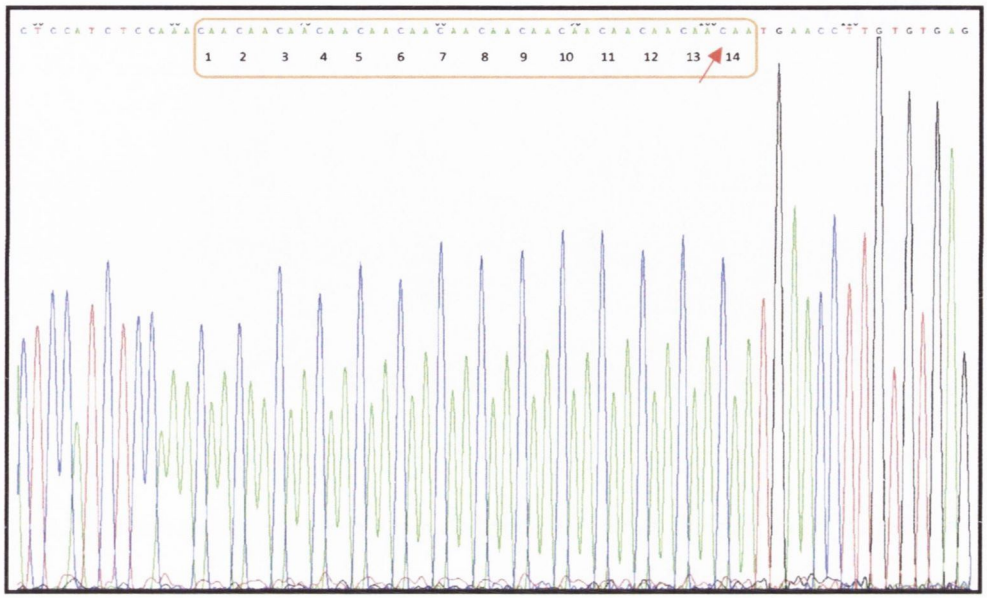


Figure 4.14 Sequencing result for 14RCpCR-Blunt clone with AS16For

4.2.1.5 Cloning results for 16R and 17R into pCR-Blunt

The human genomic (HG) DNA was extracted by undergraduate students as mentioned in the methods Section 4.1.2.2.1 and this DNA was used for cloning the 16R and 17R genotypes. The DNA was heterozygous for both human samples with a HG genotype of 13R/16R and 15R/17R. The region of interest was amplified as showin in Figure 4.15 and Figure 4.16.

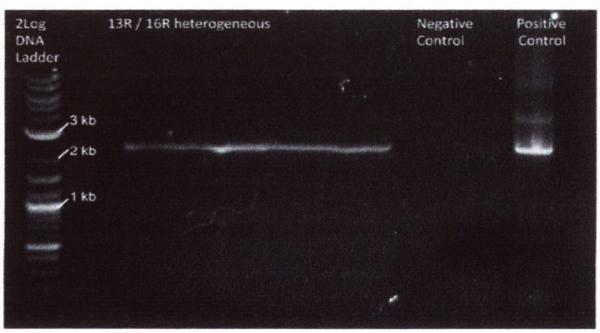


Figure 4.15 PCR band for 13R/ 16R amplicon

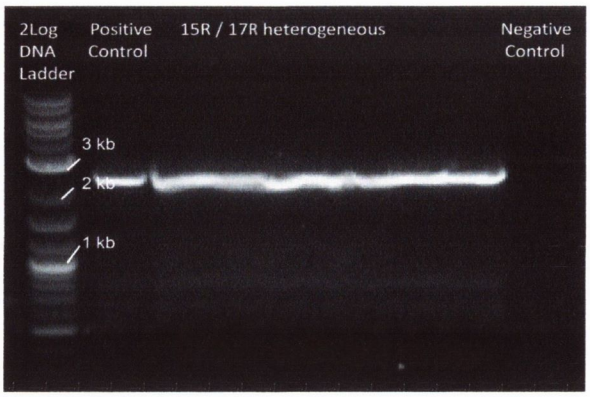


Figure 4.16 PCR band for 15R/ 17R amplicon

The DNA was excised and purified using the MiniElute gel extraction kit and Table 4.11 shows the concentration and quality of the DNA.

Table 4.11 Gel extraction for 13R/16R and 15R/17R

Sample ID	ng/uL	A260	260/280	Average ng/µL
13R/16R.1	44.64	0.89	1.82	42.83
13R/16R.1	48.53	0.97	1.69	
13R/16R.1	35.32	0.71	1.59	
15R/17R.1	23.81	0.48	1.64	22.96
15R/17R.1	22.38	0.45	1.47	
15R/17R.1	22.68	0.45	1.59	

Following ligation with pCR-Blunt and transformation into E.Coli, colonies were screened by PCR with RV Primer 3 and AS11Rev, where a band of around 1900 kb would identify a positive clone. The gel was prepared with 2% agarose, in order to have a higher resolution and separate (distinguish) between amplicons.

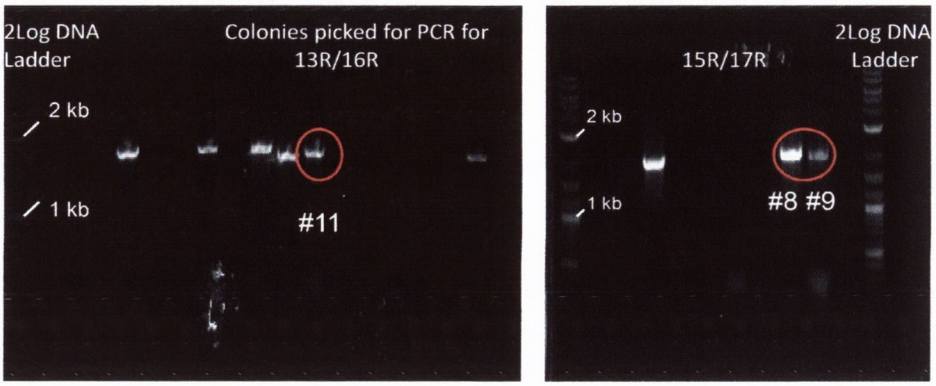


Figure 4.17 PCR on colonies for 16R and 17R clones

Figure 4.17 shows the band from the PCR on colonies, these were picked, purified by miniprep and sequenced with AS16For. Plasmid concentrations and quality are shown in Table 4.12.

Table 4.12 Plasmid concentration after miniprep for 16R and 17R in pCR-Blunt

Sample ID	ng/uL	A260	260/280	Average ng/µL
13R/16R.1	407.66	8.15	1.83	405.94
13R/16R.1	404.28	8.09	1.84	
13R/16R.1	405.87	8.12	1.83	
15R/17R.1	349.32	6.99	1.83	344.10
15R/17R.1	342.27	6.85	1.84	
15R/17R.1	340.70	6.81	1.86	

Sequencing results showed one positive colony for the 16R (#11), and two positive colonies for the 17R (#8 and #9). The rest of the colonies were clones from the other alleles (13R and 15R respectively). The sequence for these clones is shown in Figure 4.18 and Figure 4.19.

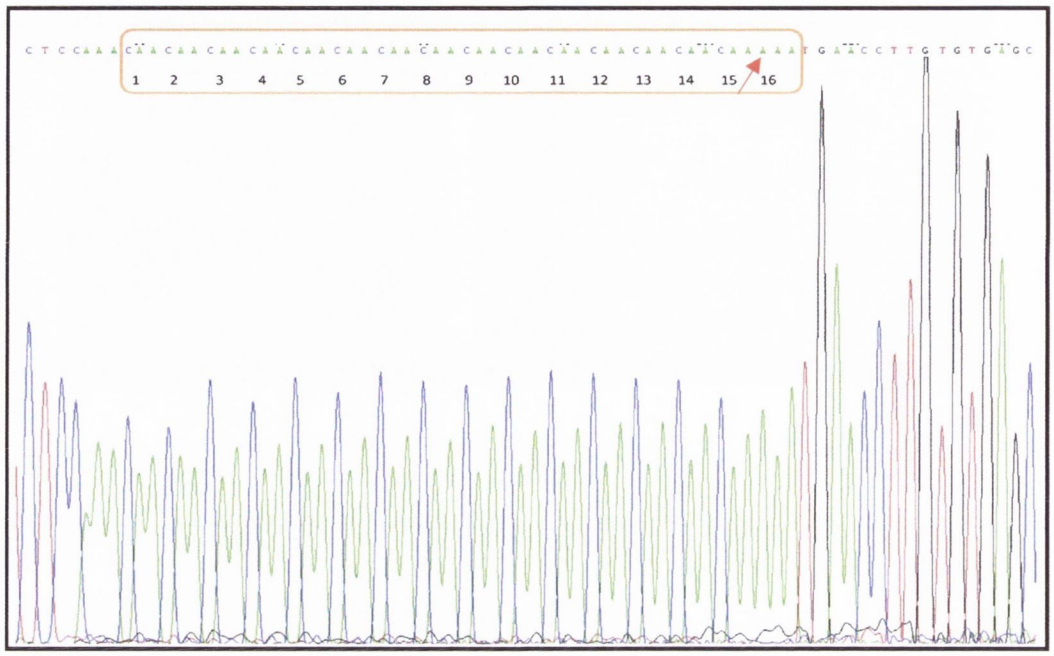


Figure 4.18 Sequencing result for 16RApCR-Blunt

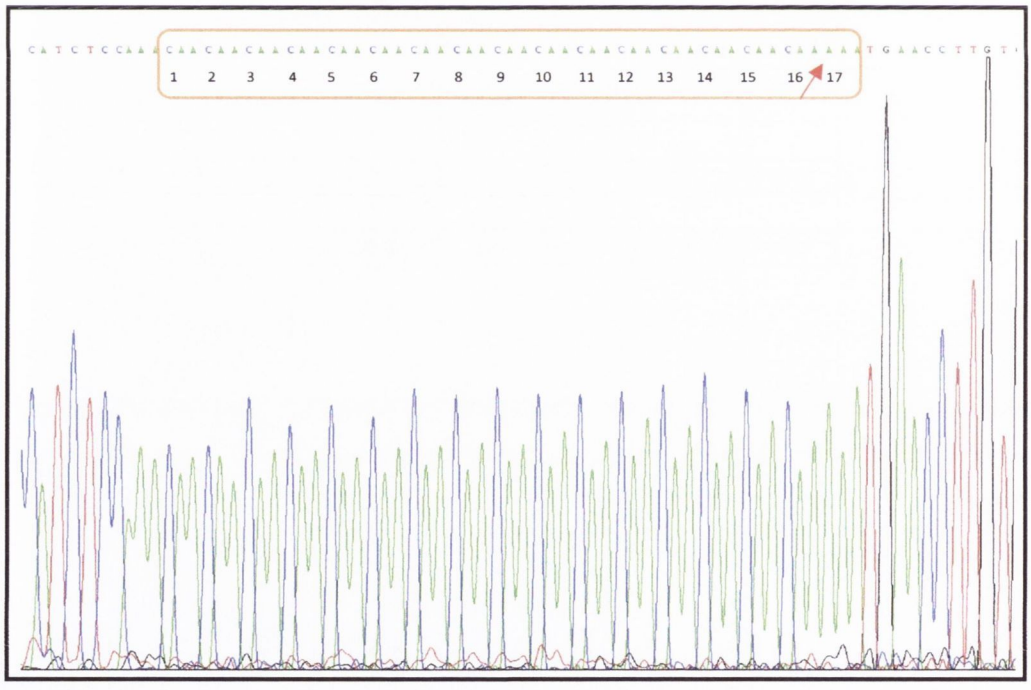


Figure 4.19 Sequencing result for 17RApCR-Blunt

4.2.2 Sub-Cloning from pCR-Blunt library into pGL3

4.2.2.1 Amplification of promoter region from clones in pCR-Blunt

A PCR for the desired region to be inserted into the reporter plasmid expressing luciferase was performed on the clones in the standard vectors. The clones in pCR-Blunt, mainly 10R, 13RA, 14RC, 15RA were amplified using Seq4F and Seq5R to produce a band size of around 2067– 2083 bp depending on the length repeat. PCR was very effective producing larger quantities of amplicon since the DNA template was a plasmid rather genomic DNA, hence the distinctive bands. And this is seen in the Figure 4.20 and Figure 4.21.

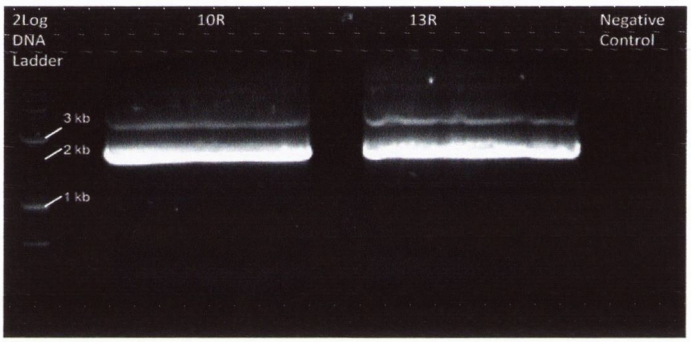


Figure 4.20 PCR product for 10R and 13R for pGL3

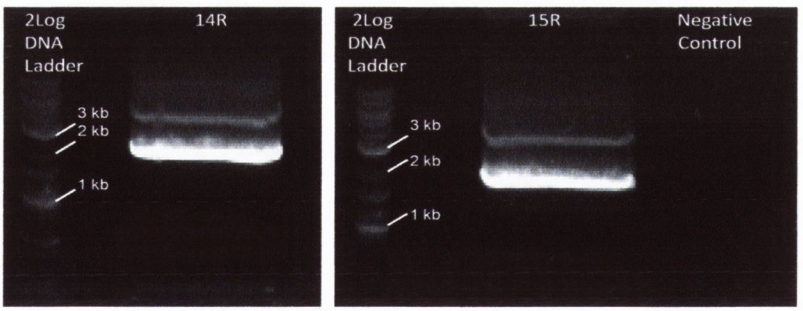


Figure 4.21 PCR product for 14R and 15R for pGL3

The amplified DNA was cut from the agarose gel, purified and quantified using NanoDrop™ and the concentrations and quality achieved are shown in Table 4.13.

Table 4.13 Gel extraction for PCR products into pGL3

Sample ID	ng/uL	A260	260/280	Average ng/µL
10R.1	173.27	3.45	1.88	175.66
10R.2	175.48	3.51	1.89	
10R.3	178.22	3.56	1.87	
13R.1	154.66	3.09	1.90	154.61
13R.2	153.89	3.08	1.91	
13R.3	155.29	3.11	1.90	
14R.1	60.09	1.20	1.99	60.75
14R.2	61.81	1.24	1.98	
14R.3	60.35	1.21	2.06	
15R.1	117.24	2.35	1.88	113.58
15R.2	110.80	2.22	1.88	
15R.3	112.70	2.30	1.87	

Following purification, the DNA bands were double digested (DD) with Xho1 and Nco1 to derive the sticky ends of Xho1 and Nco1, matching the pGL3 basic vector. Purification for the DNA was as seen in Table 4.14, where the concentration of the amplified DNA varied from 20 – 83 ng/µL.

Table 4.14 Double digest purification for PCR products and pGL3 basic

Sample ID	ng/uL	A260	260/280	Average ng/µL
DD 10R.1	21.82	0.44	1.82	20.97
DD 10R.2	20.74	0.42	1.93	
DD 10R.3	20.34	0.41	1.63	
DD 13R.1	89.91	1.80	1.75	83.17
DD 13R.2	81.21	1.62	1.66	
DD 13R.3	78.39	1.57	1.86	
DD 14R.1	46.74	0.94	1.69	42.72
DD 14R.2	51.03	1.02	1.57	
DD 14R.3	30.38	0.61	1.77	
DD 15R.1	50.25	1.00	1.81	48.42
DD 15R.2	47.11	0.91	1.63	
DD 15R.3	47.90	0.96	1.76	
DD pGL3.1	111.89	2.24	1.83	110.34
DD pGL3.2	109.51	2.19	1.84	
DD pGL3.3	109.61	2.19	1.84	

4.2.2.2 Ligation into pGL3 reporter plasmid

Once the sticky ends were created by enzyme digest for both purified PCR product and the host pGL3 vector, ligation was carried out with T4 ligase. The insert to vector ratio for the ligate used was 10:1, similar to pCR-Blunt ligation, where 107 ng of DNA was mixed with 25 ng pGL3. The resulting construct in the reporter plasmid is shown in Figure 4.22, with 10R as an example.

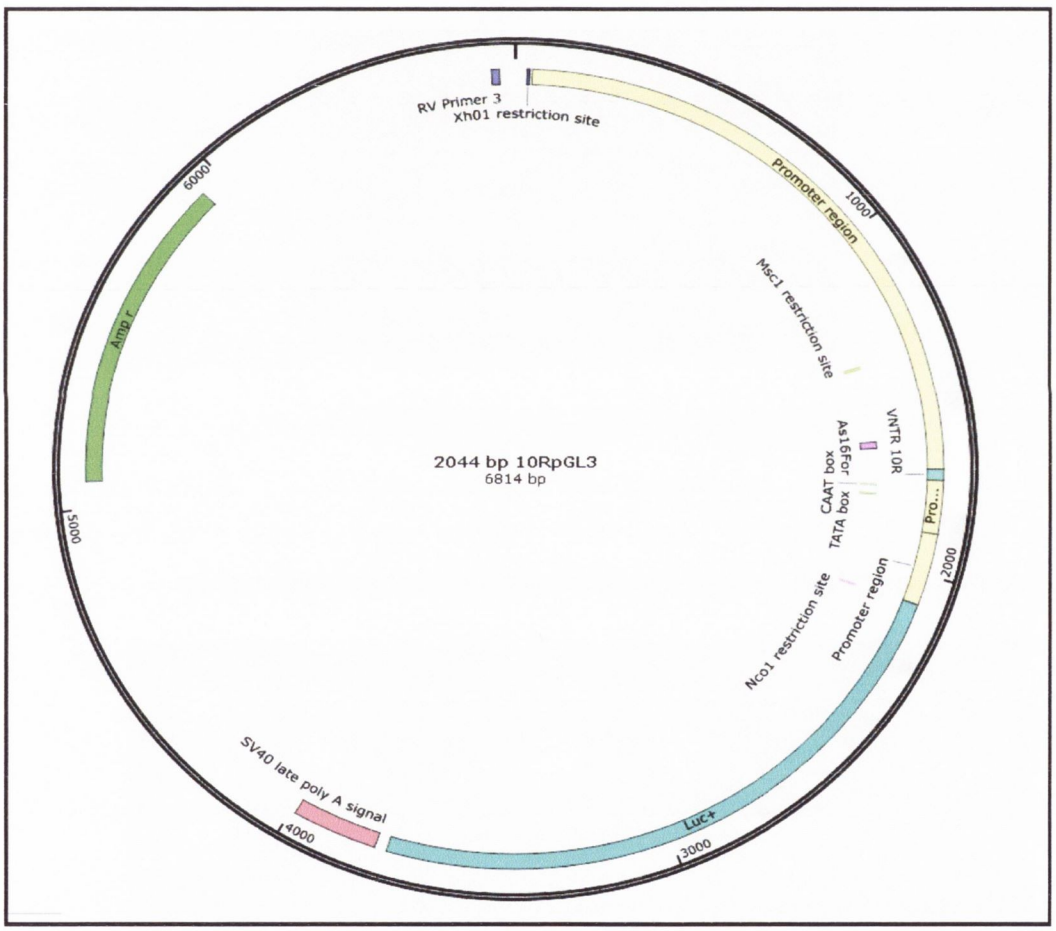


Figure 4.22 10RpGL3 construct in reporter plasmid

4.2.1.3 Analysing transformants in reporter plasmid

Ligates from the previous section were transformed into E.Coli, where the next day colonies were analysed by PCR for a successful amplicon. The primers used were located upstream of the desired insert on the pGL3 vector (RV primer 3) and on the ligated insert downstream to the VNTR. The band size expected was approximately 1.9 kb. The first colony identified was a single colony for 10RpGL3. A negative control (having water instead of DNA) was also loaded on the gel to confirm no contamination in the master mix.

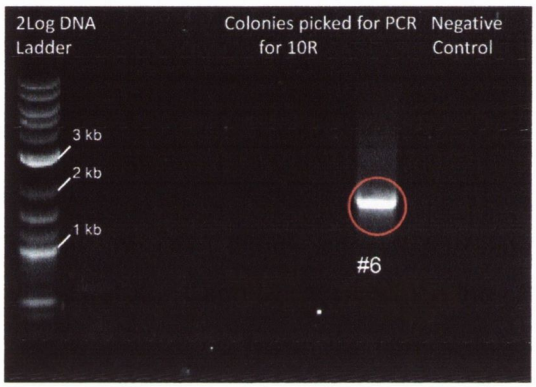


Figure 4.23 PCR on colonies for 10RpGL3

The next reporter vector to be identified was the 13RApGL3. Three colonies showed a noticeable band comparable to the control which is shown below in Figure 4.24. The positive control used was the 10RpGL3, since it was known to be a successful construct.



Figure 4.24 PCR on colonies for 13RApGL3

Following the identification of these two clones, the next ones to be singled out were 14RC and 15RA. The reporter vector having the 14RC genotype had two successful colonies, while 15RA produced seven colonies showing a convincing band (circled in Figure 4.25).

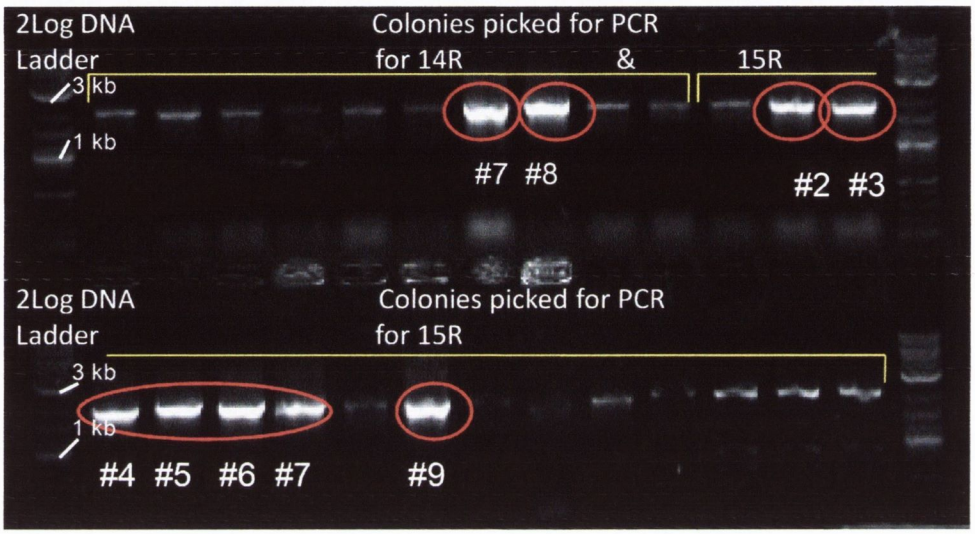


Figure 4.25 PCR on colonies for 14RCpGL3 and 15RApGL3

Sequencing results for the sub-clones in pGL3 reporter plasmid are shown in Figure 4.26 - Figure 4.29. The DNA sequenced with the forward primer AS16For was aligned with a template of the desired product to identify the presence of any unwanted mutations in the DNA, as a result of PCR and genetic manipulation. The alignment for 10R, 13RA, 14RC and 15RA is shown in Figure 4.26 - Figure 4.29. The proposed sequence is also shown in Chapter 2, Section 2.8.5 and full alignments for these results are found in Appendix 3.

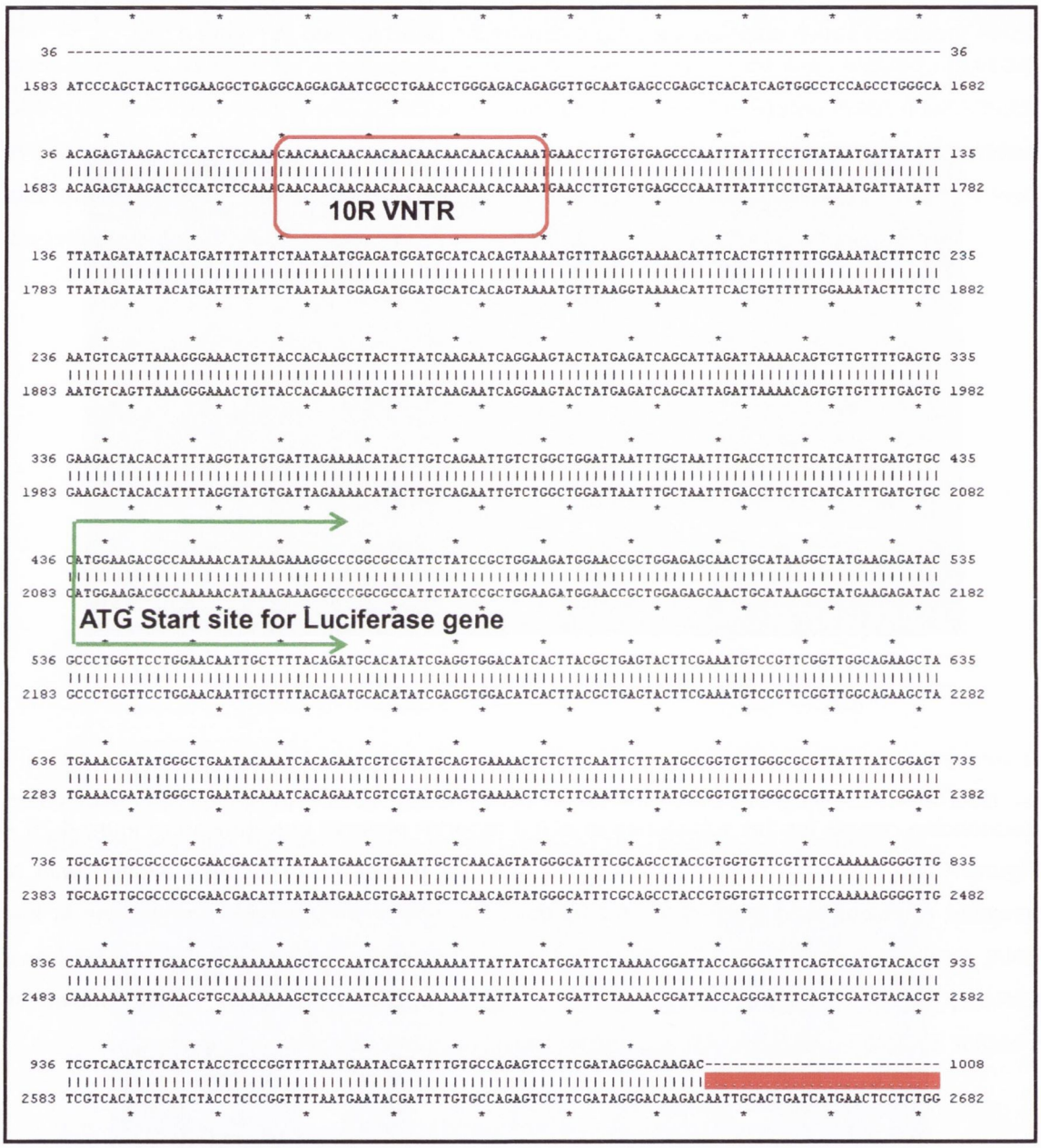


Figure 4.26 Alignment of 10RpGL3 with proposed sequence

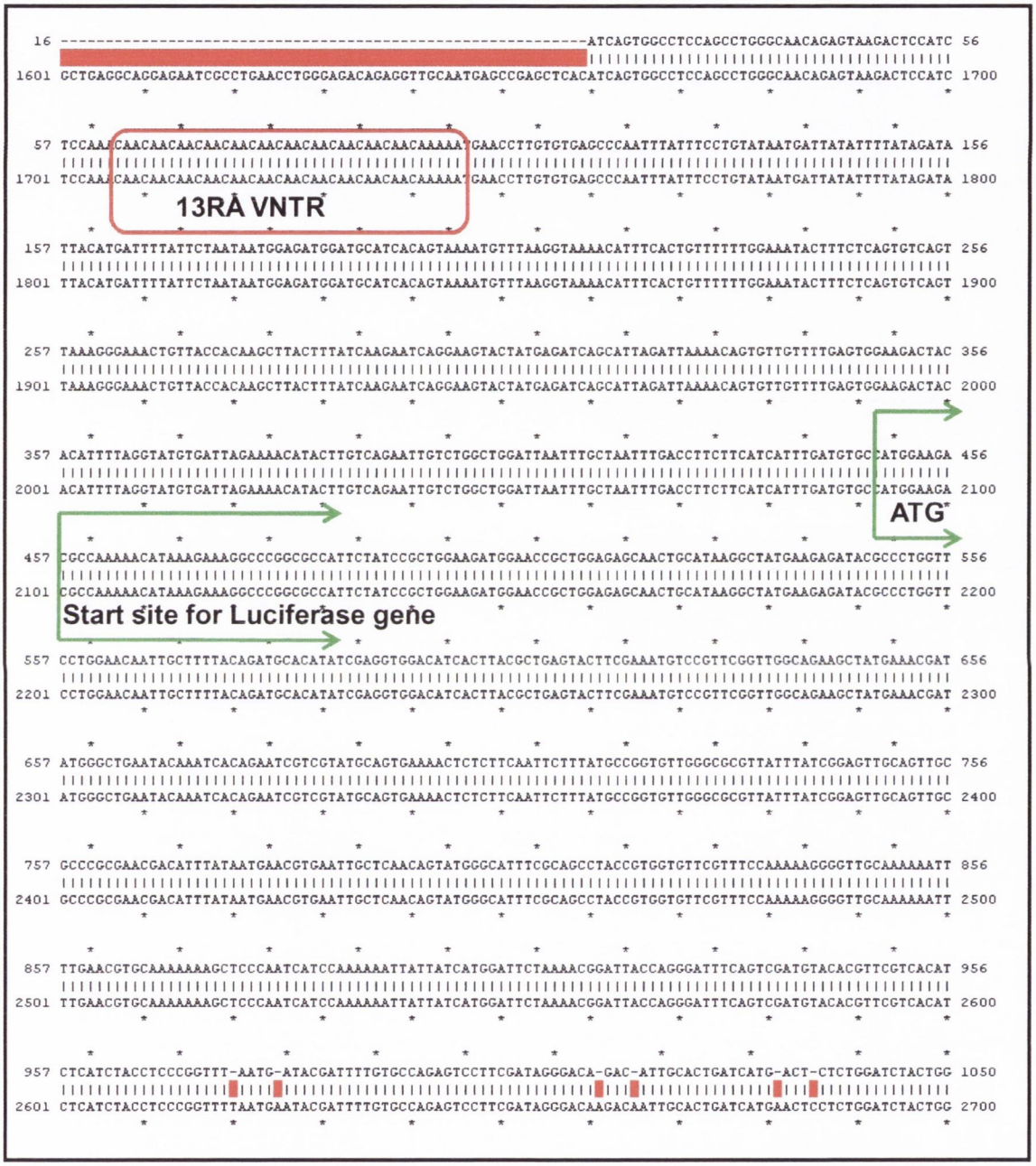


Figure 4.27 Alignment of 13RApGL3 with proposed sequence

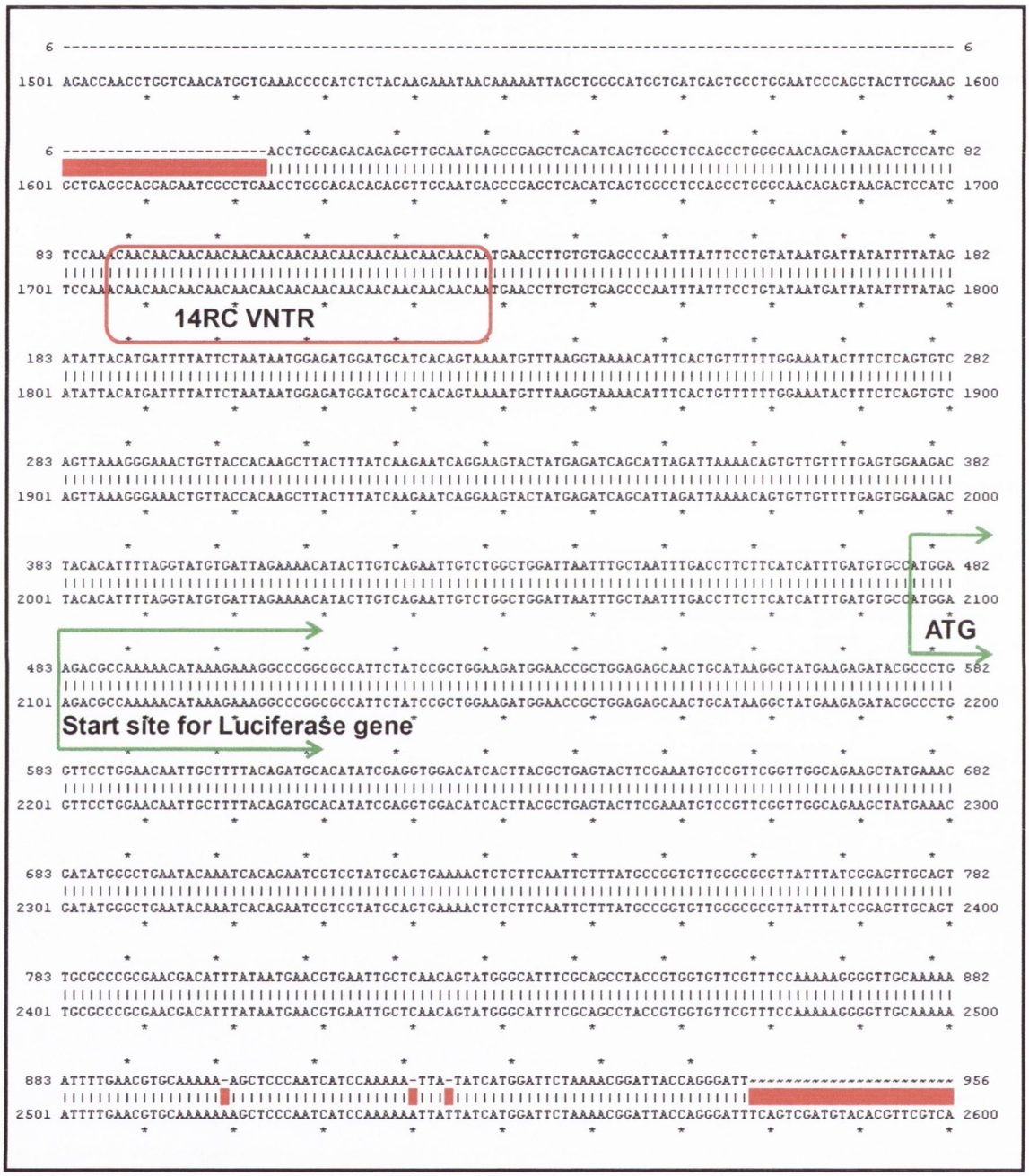


Figure 4.28 Alignment of 14RCpGL3 with proposed sequence

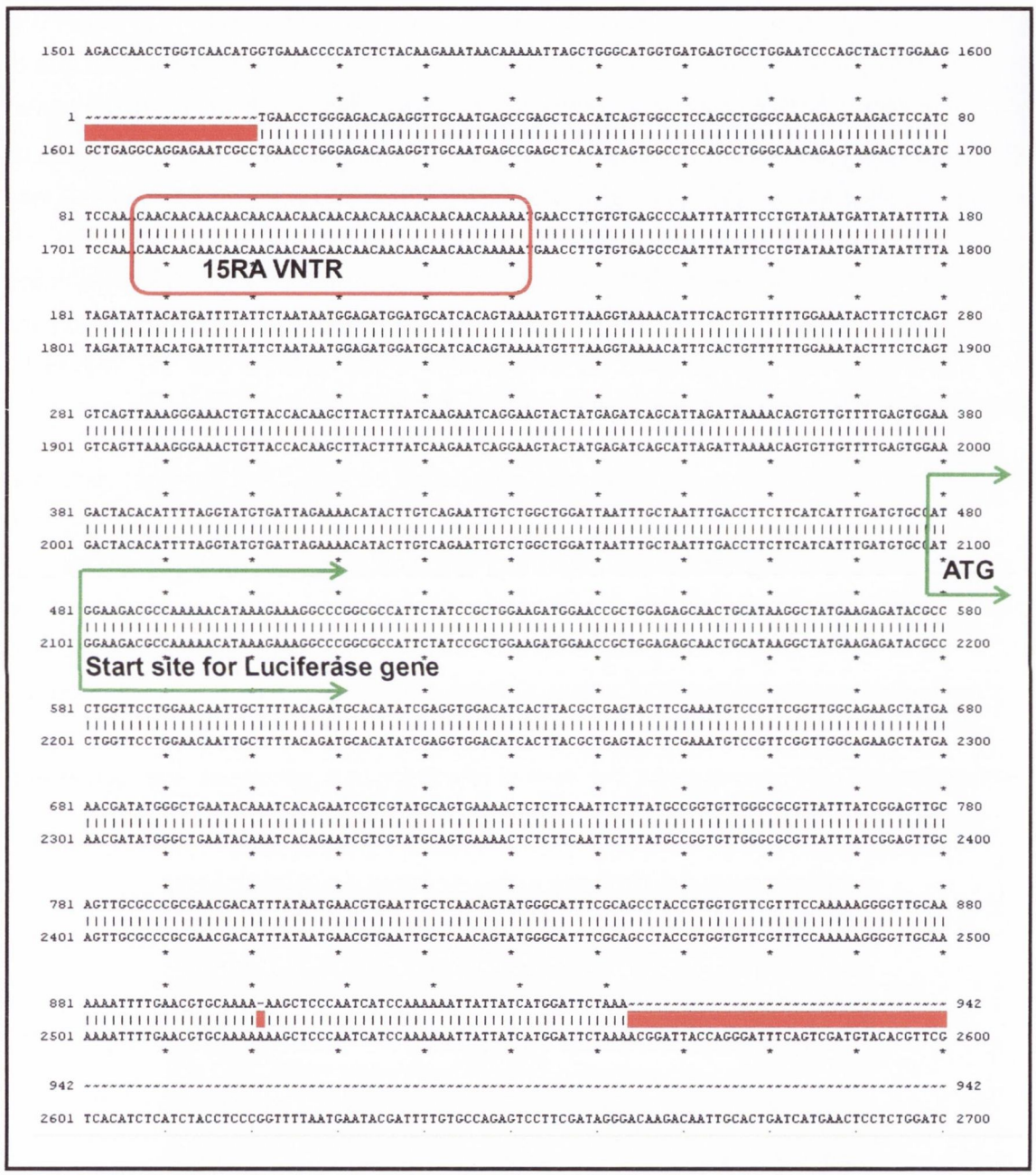


Figure 4.29 Alignment of 15RApGL3 with proposed sequence

4.2.1.4 Cloning results for 18RC and 19RC using 10RpGL3 as the host vector

After achieving positive clones (10R, 13R, 14R, 15R) in the reporter vector, the longer and rarer genotypes were attempted to be sub-cloned into pGL3. An issue with this step was to find homogenous DNA strands, especially since the ones identified from HapMap samples proved to be heterozygous. In order to overcome this problem, the DNA was synthesised by a gene synthesis service provided by Eurofins MWG. The DNA provided for synthesis was not the full promoter region as described by Coge *et al* [7], but 738 bp and 741 bp long for 18R and 19R respectively, where the 18R and 19R are 333 bp downstream from Msc1 restriction site. The sequence ordered was for the allele ending 'CAA' since analysis from the Eurofins labs confirmed it was a problematic site for synthesis, so to facilitate this, the CAA triplet would break the five A's formed with the AAA triplet variation.

The promoter region which was already previously cloned into pGL3, had an Msc1 restriction site 333 bp upstream from the VNTR. So this was indicated as a restriction site for ligation instead of the Xho1 used for the other pGL3 clones, and using the 10RpGL3 as the host vector rather than pGL3 basic, in order to retain the promoter sequence upstream of the Msc1 cloning site. The amplification of 18RC and 19RC from the synthesized vector required a different forward primer, Seq7F (Chapter 2, Section 2.4.1). This primer is located on the pEX-K vector. The band size to be excised and purified is respectively smaller (around 750 bp) than for the other clones since the desired region for both genotypes was cloned in a template already harbouring 5' upstream of the potential promoter (10RpGL3).

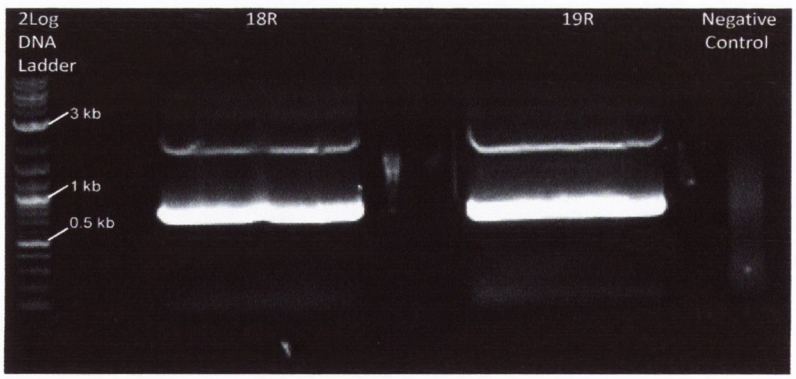


Figure 4.30 PCR product for 18RC and 19RC from pEX for insertion into pGL3 vector

The DNA band was cut, gel purified, and eluted with 1x buffer 4 in 10 µL volume. Due to the small final elution volume only two readings were performed with the NanoDrop™ and two molar ratios were used from this volume (10:1 and 6:1).

Table 4.15 PCR product for 18RC and 19RC after gel purification

Sample ID	ng/uL	A260	260/280	Average ng/µL
18R.1	188.53	3.77	1.85	191.08
18R.2	193.63	3.87	1.87	
19R.1	236.33	4.72	1.86	234.63
19R.2	232.93	4.65	1.87	

Following purification, the DNA and 10RpGL3 vector were subjected to restriction enzyme digest with Xho1 and Msc1 and subsequent agarose gel electrophoresis and purification. This time an ample concentration of DNA was obtained, when compared to previous enzyme purification where a substantial amount of DNA was lost.

Table 4.16 Double digest (Xho1 and Msc1) purification for 18RC, 19RC and 10RpGL3 host vector

Sample ID	ng/uL	A260	260/280	Average ng/µL
DD 18R.1	135.91	2.72	1.64	128.99
DD 18R.2	131.12	2.65	1.70	
DD 18R.3	119.93	2.40	2.08	
DD 19R.1	132.74	2.66	2.00	133.72
DD 19R.2	134.19	2.68	1.98	
DD 19R.3	134.23	2.65	1.95	
DD 10RpGL3.1	135.19	2.70	1.70	132.55
DD 10RpGL3.2	130.22	2.60	1.90	
DD 10RpGL3.3	132.23	2.62	1.92	

The consecutive step was a ligation of 18R and 19R digested products with the digested host template (from 10RpGL3), transform the ligate into competent cells and screening the colonies for a positive transformant.

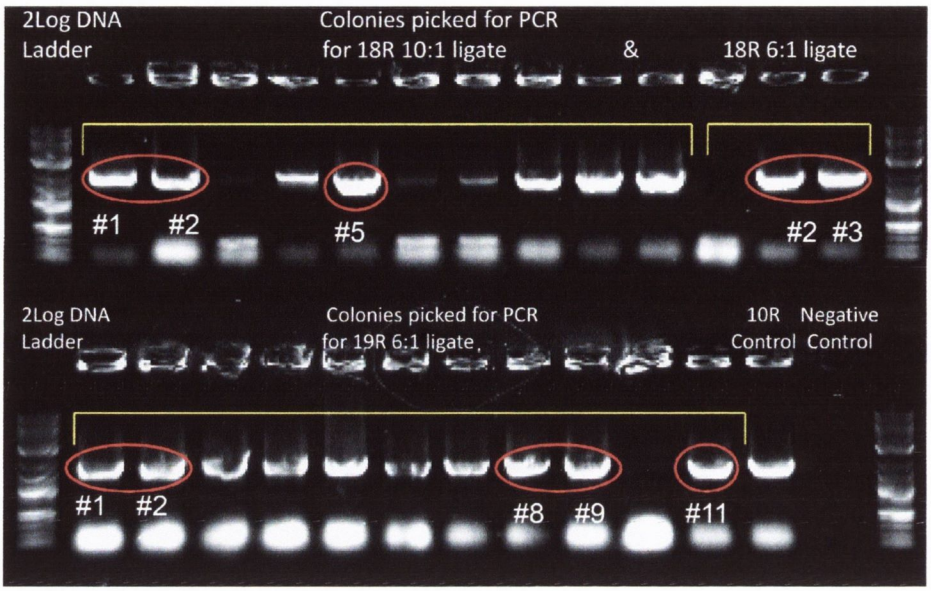


Figure 4.31 PCR on colonies for 18RC and 19RC in pGL3

The plates incubated with the transformed ligate were from a 10:1 and a 6:1 ratio for both genotypes. The 18RC had colonies in both plates, while the 19RC had colonies only in the 6:1 plate. Random colonies were selected for further analysis. The 6:1 plate for the 18RC had very few colonies; hence all there were picked for the PCR analysis.

The PCR on colonies using RV Primer 3 and AS11Rev gel electrophoresis is shown in Figure 4.31. The colonies (represented by the circled bands) were selected and were grown overnight in LB Amp^+ and purified by via Miniprep the following day. The concentration obtained for reach plasmid is shown in

Table 4.17. The isolated plasmids were then diluted to 50 ng/µL and sent for sequencing with AS16For primer to confirm positive clones.

Table 4.17 Miniprep colonies for 18RC and 19RC

Sample ID	ng/uL	A260	260/280	Average ng/µL
18R6-2.1	85.29	1.71	1.89	84.78
18R6-2.2	84.63	1.69	1.87	
18R6-2.3	84.41	1.69	1.91	
18R6-3.1	276.19	5.52	1.88	274.77
18R6-3.2	275.20	5.50	1.86	
18R6-3.3	272.92	5.46	1.88	
18R10-1.1	231.17	4.62	1.87	231.84
18R10-1.2	231.50	4.66	1.88	
18R10-1.3	232.84	4.66	1.88	
18R10-2.1	251.76	5.04	1.87	251.48
18R10-2.2	251.56	5.03	1.87	
18R10-2.3	251.13	5.02	1.87	
18R10-5.1	225.10	4.50	1.89	224.77
18R10-5.2	225.11	4.52	1.87	
18R10-5.3	224.11	4.48	1.89	
19R6-1.1	276.21	5.52	1.87	279.14
19R6-1.2	277.31	5.55	1.87	
19R6-1.3	283.91	5.68	1.87	
19R6-2.1	283.66	5.67	1.85	278.19
19R6-2.2	284.00	5.66	1.83	
19R6-2.3	266.92	5.34	1.88	
19R6-8.1	265.66	5.31	1.88	263.07
19R6-8.2	262.11	5.32	1.89	
19R6-8.3	261.44	5.23	1.88	
19R6-9.1	282.43	5.65	1.88	283.10
19R6-9.2	283.40	5.60	1.87	
19R6-9.3	283.47	5.67	1.87	
19R6-11.1	276.56	5.53	1.88	274.81
19R6-11.2	275.65	5.54	1.88	
19R6-11.3	272.23	5.45	1.88	

Sequencing results showed that some of the colonies were from the 10RpGL3, probably due to some undigested template DNA, either from the plasmid re-ligating on itself or an incomplete digestion. From the colonies picked, colony number 18R10-2 was the desired 18RCpGL3 clone, originating from the 10:1 ligation ratio. The correct colonies for the 19R genotype were colonies 1, 8 and 11 which gave a positive clone for the 19RCpGL3. The sequencing alignment for 18R10-2 and 19R6-1 are shown in Figure 4.32 and Figure 4.33. The full sequencing alignments for these results are found in Appendix 3.

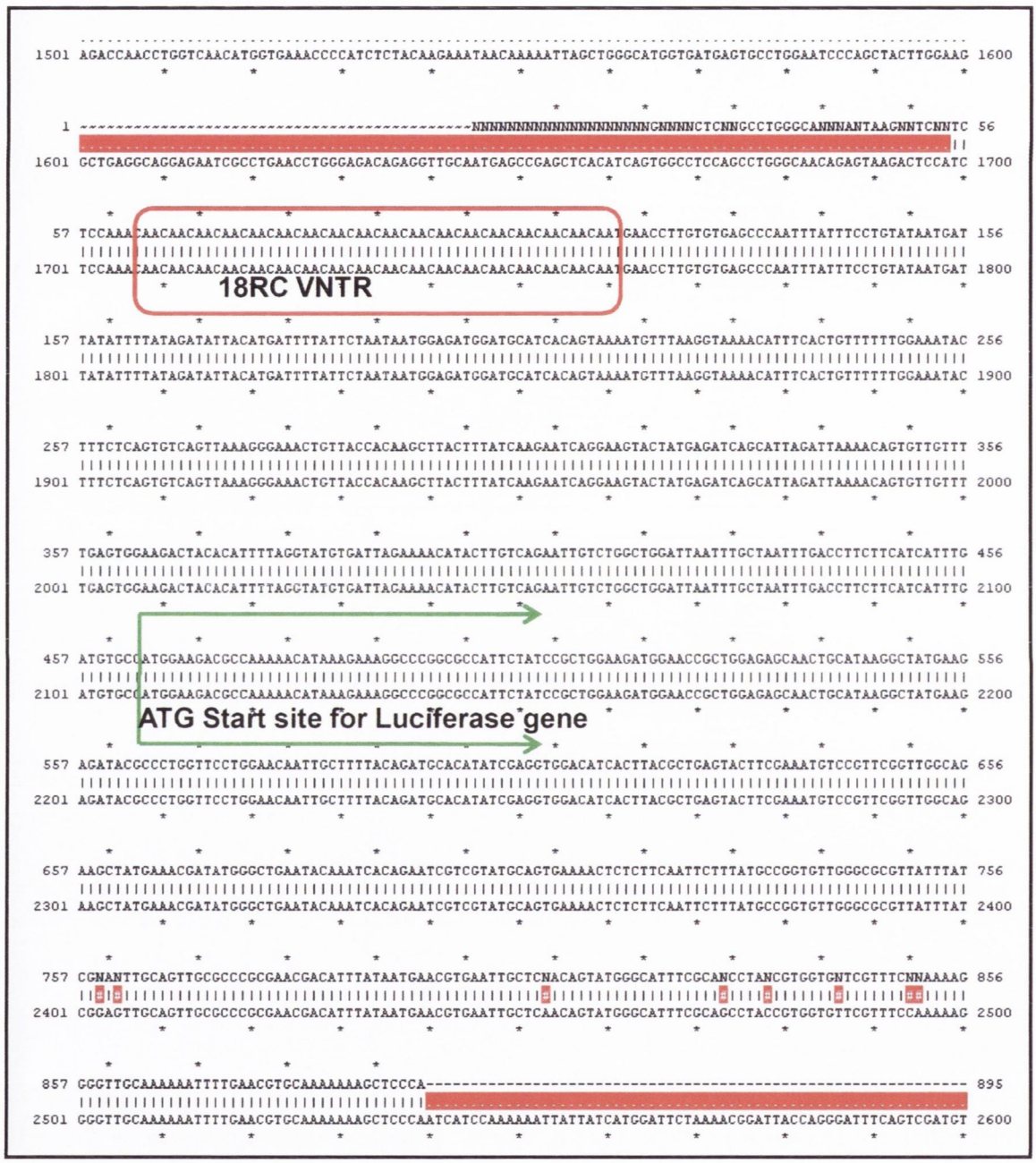


Figure 4.32 Alignment of 18RCpGL3 with proposed sequence

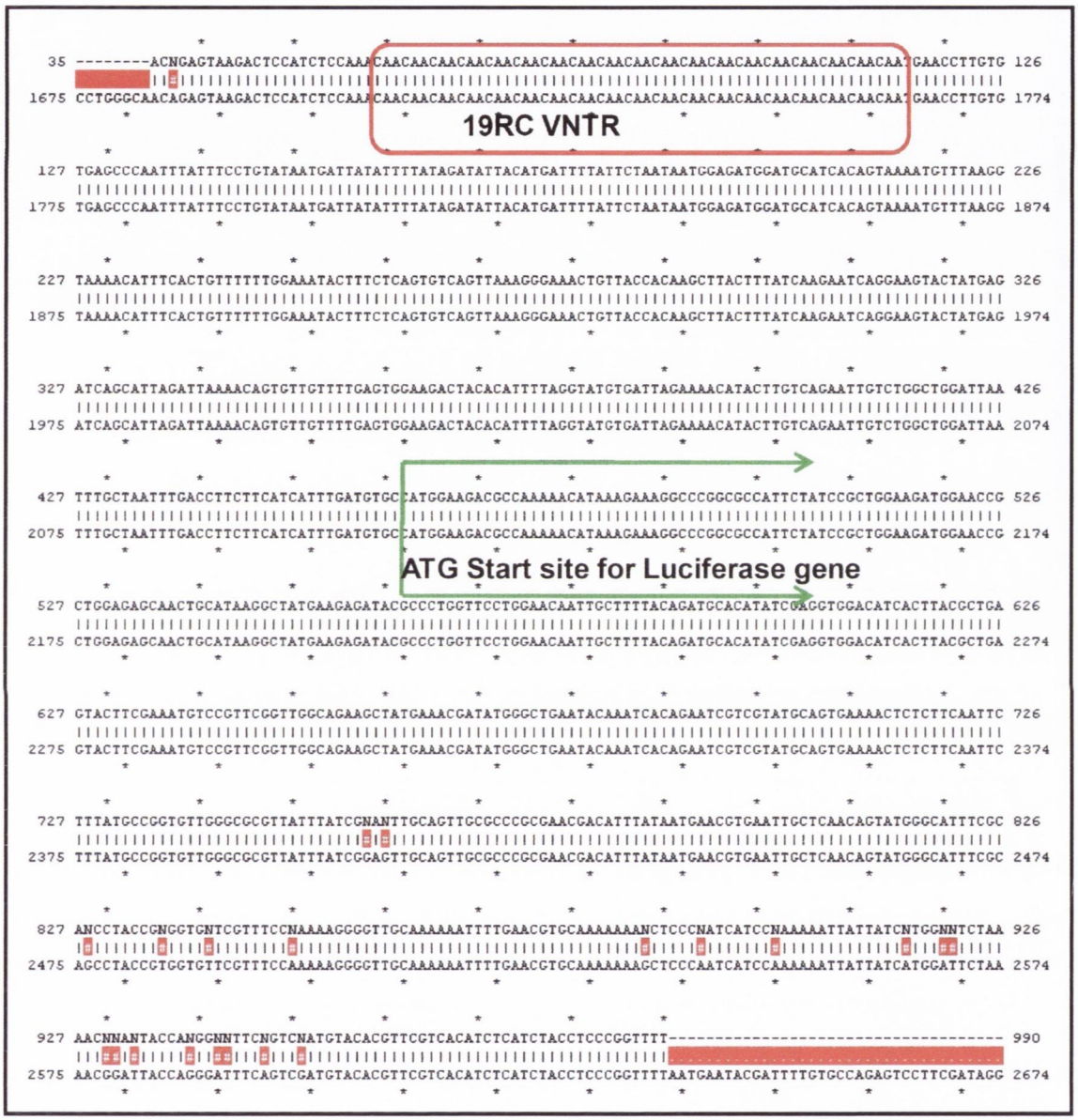


Figure 4.33 Alignment of 19RC with proposed sequence

4.2.3 Point mutation results on vector constructs

4.2.3.1 Point mutations in pCR-Blunt

Site-directed mutagenesis (SDM) was performed to create a point mutation in the clones, in order to obtain a complete collection of genotypes for the VNTR. Initial SDM was carried out on pCR-Blunt vectors, followed by sub-cloning into pGL3. This method was applied for:

pCR-blunt vectors:

- ✓ 13R**A** to 13R**C**
- ✓ 14R**C** to 14R**A**
- ✓ 15R**A** to 15R**C**
- ✓ 16R**A** to 16R**C**
- ✓ 17R**A** to 17R**C**

For each point mutation, once the PCR was ready, half of the PCR product was digested with Dpn1 to remove the methylated parent plasmid (i.e., the non-mutated clone). Following incubation in a water bath at 37 °C for 1 h, the product was transformed into *E.Coli*. Several controls were set up for this method: one of the controls consisted of the original parent plasmid digested with Dpn1, to confirm Dpn1 digestion was complete. Another control was transformation of the PCR product without Dpn1 digest, to confirm a successful amplification. Finally, the last control was transformation of the negative control for the PCR, which was essentially the master mix with water (no DNA). This confirms no contamination in PCR.

Once transformed, the colonies with the SDM plasmids were selected randomly for further analysis. The plates, following overnight incubation at 37°C are shown in Figure 4.34. This is an example of a 13R point mutation, which follows the same pattern for all SDM point mutations made for all clones.

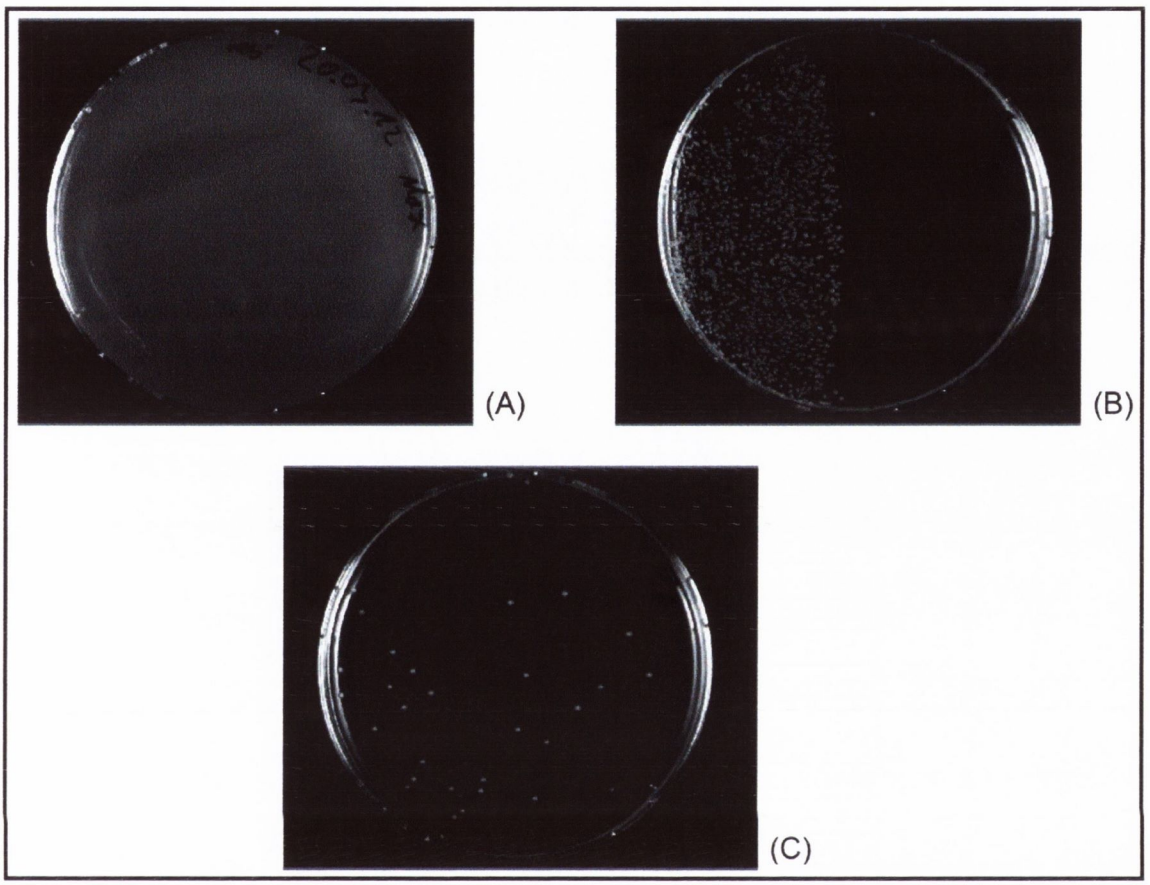


Figure 4.34 Bacterial plates for point mutations after site-directed mutagenesis

(A) Control for negative control from the PCR, (B) control split in two, where one side is PCR mutated vectors not digested by Dpn1 and the other side is the parent vector digested with Dpn1, (C) PCR mutated vectors digested with Dpn1 which were used for further analysis.

Random colonies were picked and grown overnight in LB Kan^+, then the plasmid was extracted and purified the next day. The quantity and quality (with NanoDrop™ measurements) for the selected colonies are shown in Table 4.18. The range of concentration obtained was between 300 - 400 ng/μL, where the quality (260/280) was approximately 1.8 for all plasmids.

Table 4.18 Miniprep for SDM point mutations for 13R, 14R and 15R

Sample ID	ng/uL	A260	260/280	Average ng/µL
13R4.1	391.48	7.83	1.86	374.95
13R4.2	367.18	7.34	1.88	
13R4.3	366.18	7.32	1.88	
13R5.1	388.02	7.76	1.89	401.47
13R5.2	392.31	7.85	1.90	
13R5.3	424.09	8.48	1.81	
13R6.1	431.53	8.63	1.79	432.35
13R6.2	425.53	8.51	1.80	
13R6.3	440.00	8.80	1.79	
14R1.1	281.94	5.64	1.87	286.66
14R1.2	282.77	5.66	1.87	
14R1.3	295.27	5.91	1.87	
14R2.1	276.53	5.53	1.87	285.25
14R2.2	288.28	5.77	1.87	
14R2.3	290.95	5.82	1.88	
14R3.1	209.48	4.19	1.88	210.92
14R3.2	210.46	4.21	1.87	
14R3.3	212.83	4.26	1.88	
14R4.1	237.77	4.76	1.89	253.56
14R4.2	259.56	5.19	1.96	
14R4.3	263.36	5.27	1.88	
15R6.1	335.16	6.7	1.87	342.68
15R6.2	357.95	7.16	1.88	
15R6.3	334.92	6.7	1.87	
15R7.1	358.06	7.16	1.87	361.56
15R7.2	379.46	7.59	1.87	
15R7.3	347.15	6.94	1.87	
15R8.1	345.84	6.92	1.86	330.46
15R8.2	316.77	6.34	1.88	
15R8.3	328.77	6.58	1.87	

The plasmid was then subjected to EcoR1 digest to confirm an insert in the vector Figure 4.37. The digested product would have a band size of approximately 2.5 kb. The agarose gel for 13R, 14R and 15R is shown Figure 4.37 - Figure 4.37.

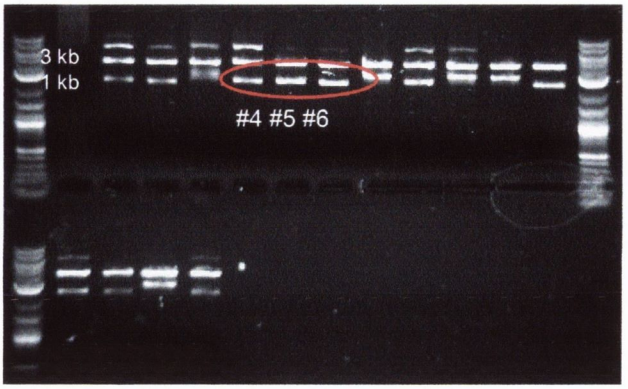


Figure 4.35 EcoR1 digest for potential point mutations for 13RpCR-Blunt

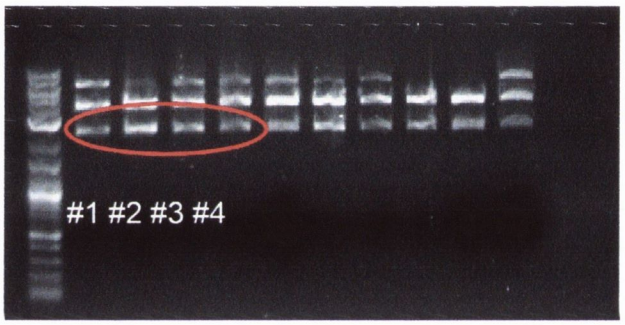


Figure 4.36 EcoR1 digest for potential point mutations for 14RpCR-Blunt



Figure 4.37 EcoR1 digest for potential point mutations 15RpCR-Blunt

Colonies that showed the correct band size were sent for sequencing with AS16For. The figures shown below reflect the changes made to the clone from the original template for 13R, 14R and 15R clones in pCR-Blunt.

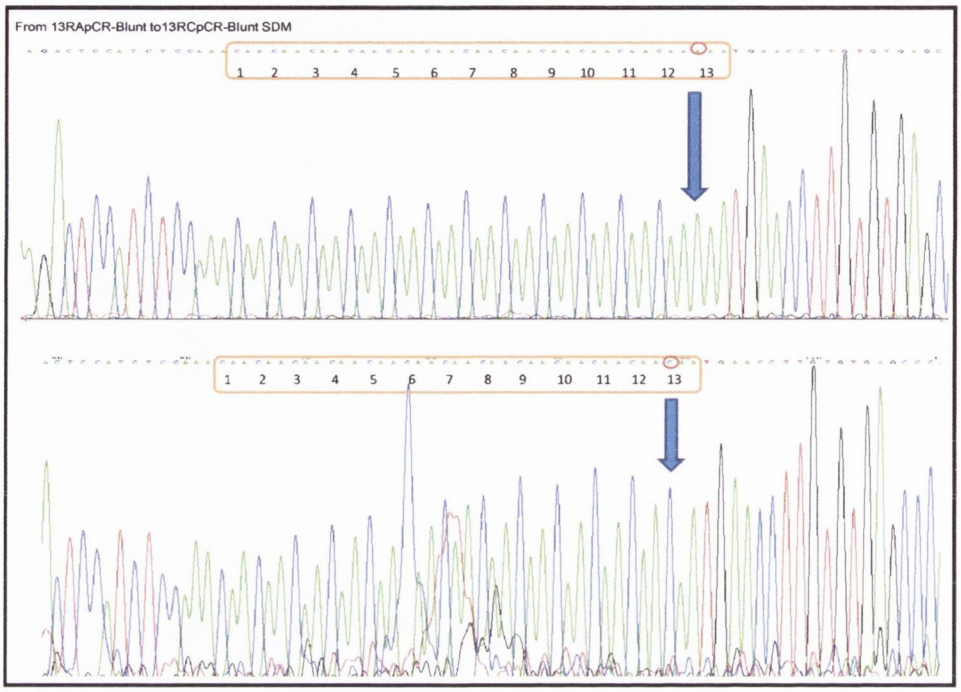


Figure 4.38 Sequencing result for 13RCpCR-Blunt

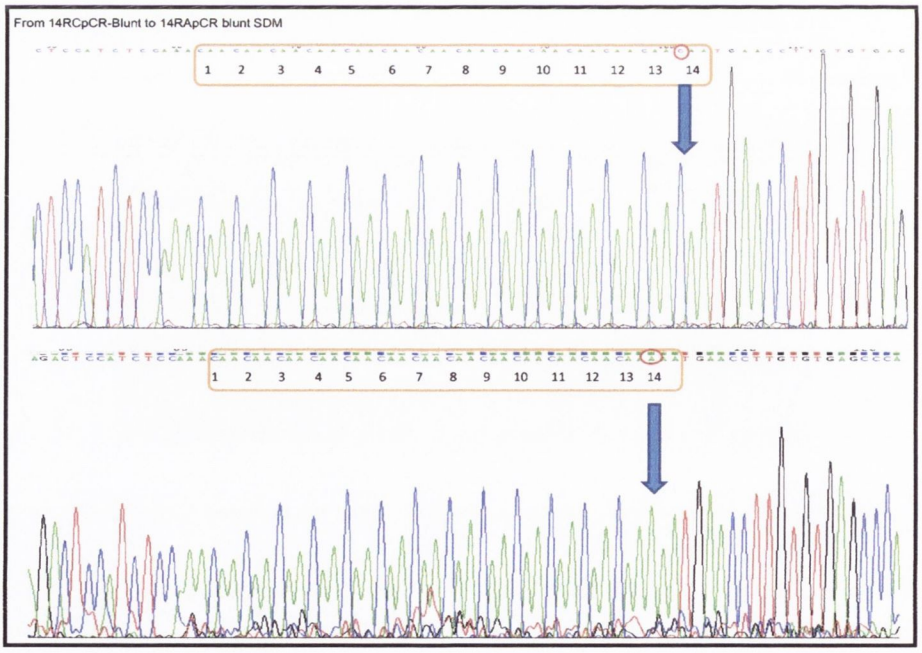


Figure 4.39 Sequencing comparison for 14RApCR-Blunt point mutation

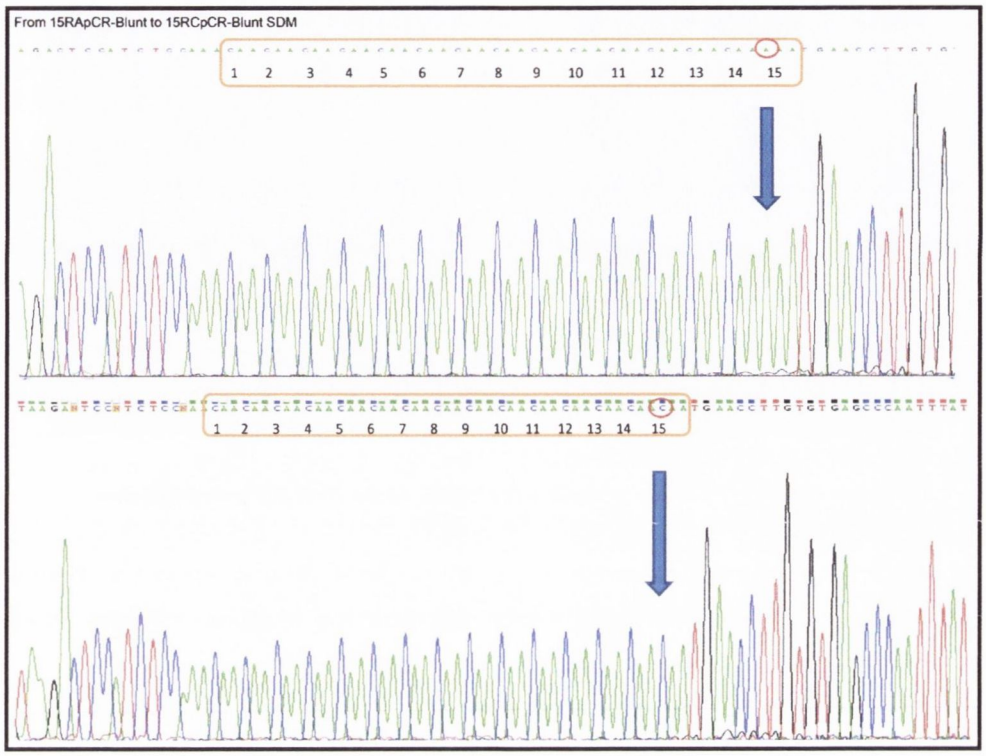


Figure 4.40 Sequencing comparison for 15RCpCR-Blunt point mutation

For 16RC and 17RC a different approach was taken. Once the PCR for the point mutation was completed, digested with Dpn1 to remove the methylated parent plasmid and transformed, a PCR on colonies was carried out using the M13rev and As11Rev primers. This PCR amplifies a region of around 2 kb. The gel electrophoresis is shown in Figure 4.41, where the first seven colonies were from the mutated 16R and the second seven colonies were from 17R. The second last band is from the positive control (10RpCR-Blunt) and the last is the negative control, showing no bands of contamination.

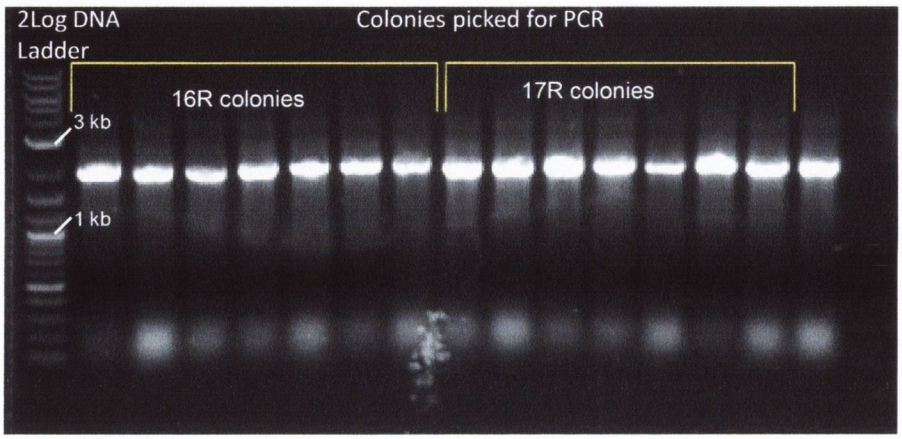


Figure 4.41 PCR on colonies for 16RC and 17RC SDM in pCR-Blunt

Since the bands showed a similar length, only two for each genotype was picked and grown overnight. This plasmid was purified the next day with the Miniprep kit from Qiagen and quantified. The results for the purification are shown in Table 4.19.

Table 4.19 Miniprep for SDM on 16RC and 17RC in pCR-Blunt

Sample ID	ng/uL	A260	260/280	Average ng/µL
16R1.1	254.29	5.09	1.83	255.18
16R1.2	255.67	5.11	1.82	
16R1.3	255.58	5.11	1.83	
16R2.1	296.27	5.93	1.83	291.92
16R2.2	271.00	5.42	1.76	
16R2.3	308.49	6.17	1.82	
17R1.1	213.87	4.28	1.82	211.30
17R1.2	208.09	4.16	1.85	
17R1.3	211.93	4.24	1.84	
17R2.1	396.32	7.93	1.83	390.18
17R2.2	387.48	7.75	1.83	
17R2.3	386.75	7.74	1.83	

Upon sequencing the plasmids, it was identified that the point mutation was successful for both genotypes, where the 'A' allele was changed to 'C' allele (Figure 4.42 and Figure 4.43).

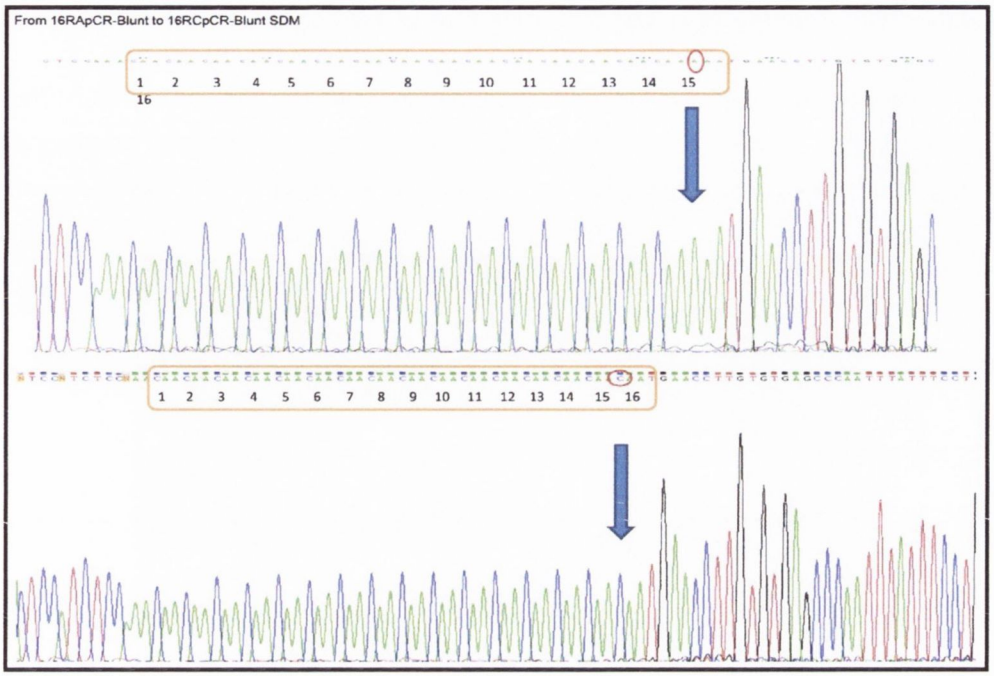


Figure 4.42 Sequencing comparison for the 16RCpCR-Blunt point mutation

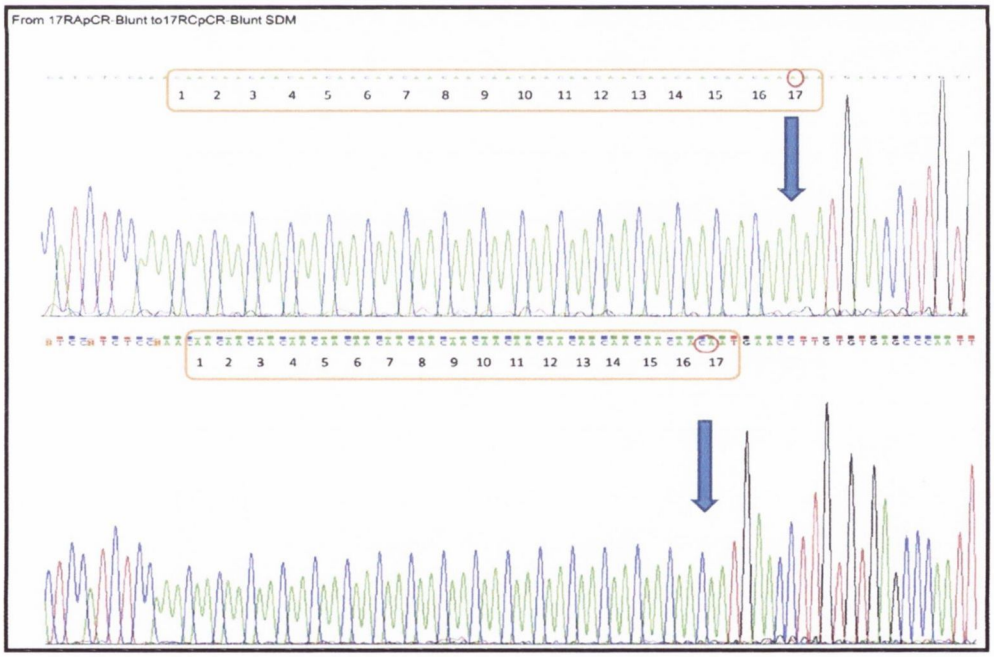


Figure 4.43 Sequencing comparison for the 17RCpCR-Blunt point mutation

4.2.3.2 Sub-cloning mutated plasmids into pGL3-basic

The next step was to sub-clone the mutated plasmids into the reporter vector. The same procedure was used as described in the previous section for sub-cloning (Section 4.1.2.2). The band size of approximately 2 kb was the desired product and is shown in Figure 4.44.

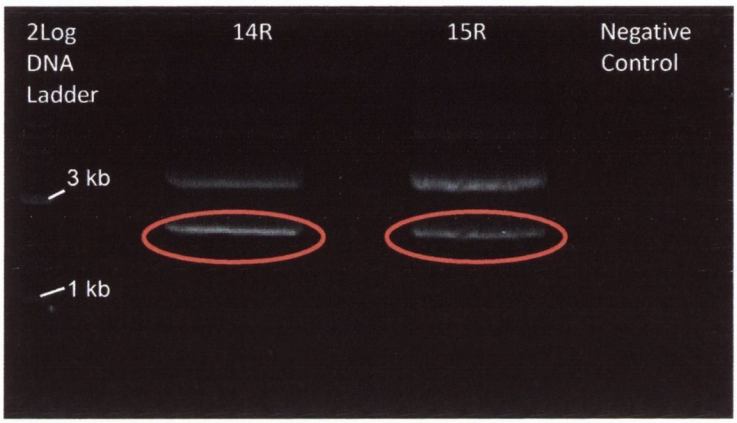


Figure 4.44 PCR for 14RA and 15RC for sub-cloning into reporter plasmid

The PCR amplicon from 14RA and 15RC was cut out of the gel and purified.

Table 4.20 PCR gel extraction for 14RA and 15RC

Sample ID	ng/uL	A260	260/280	Average ng/µL
14R.1	174.62	3.49	1.85	171.37
14R.2	172.87	3.46	1.61	
14R.3	166.63	3.33	1.87	
15R.1	60.09	1.20	1.99	60.75
15R.2	61.81	1.24	1.98	
15R.3	60.35	1.21	2.06	

Following PCR purification, 14RA, 15RC and pGL3-basic were subjected to double digestion with Xho1 and Nco1. This creates the sticky ends for ligation. The reaction mix was purified from enzymes and digested DNA, and results are shown below.

Table 4.21 Purification for double digested 14RA, 15RC and pGL3-basic

Sample ID	ng/uL	A260	260/280	Average ng/µL
DD 14R.1	41.66	0.83	1.66	40.39
DD 14R.2	38.82	0.78	1.72	
DD 14R.3	40.68	0.80	1.74	
DD 15R.1	43.48	0.87	1.68	39.22
DD 15R.2	38.57	0.77	1.78	
DD 15R.3	35.60	0.72	1.80	
DD pGL3.1	11.60	0.23	1.68	13.24
DD pGL3.2	13.52	0.27	1.46	
DD pGL3.3	14.60	0.25	1.70	

Ligation of these genotypes was similar to previously described clones, where a 10:1 insert to ligate ratio was used. The ligate was transformed in *E.Coli,* grown on Amp$^+$ plate and incubated over night and colonies were duly seen the next day. Ten colonies of each were screened by PCR, where only one colony showed a positive band for the 14RA and three colonies proved to have the desired insert for the 15RC (Figure 4.45).



Figure 4.45 PCR on colonies for 14RA and 15RC in pGL3

The circled colonies were grown overnight in 2mL LB Amp$^+$ and the plasmid was extracted the subsequent day. The purification results are seen in Table 4.22.

Table 4.22 Miniprep for 14RCpGL3 and 15RCpGL3 colonies

Sample ID	ng/uL	A260	260/280	Average ng/µL
14R.1	297.29	5.95	1.87	304.49
14R.2	309.55	6.19	1.91	
14R.3	306.64	6.13	1.86	
15R4.1	309.86	6.20	1.92	305.18
15R4.2	296.47	5.93	1.99	
15R4.3	309.21	6.18	1.94	
15R5.1	359.23	7.19	1.81	332.64
15R5.2	318.49	6.37	1.87	
15R5.3	320.20	6.40	1.87	
15R6.1	410.31	8.21	1.92	392.33
15R6.2	377.40	7.55	1.85	
15R6.3	389.29	7.79	1.85	

The plasmids were sequenced with RVprimer 3 (located on pGL3 upstream of the desired location of insert). Sequencing results proved the correct clone was obtained. The sequence alignments in Figure 4.46 shows the Xho1 restriction site upstream of the promoter region which continues to the potential promoter start site. The VNTR region is also seen with the desired point mutation.

In Figure 4.47, the alignement of 15RC is seen with the proposed sequence; however an SNP was noted upon sequencing. This SNP is rs10454097, which is a C/T polymorphism. Even though this clone had an SNP, it was still kept for further analysis since upon further investigation (through dbSNP database) it was noted that this SNP has a 50/50% occurrence in CEU HapMap samples.

The full sequencing alignments for these results are found in Appendix 3.

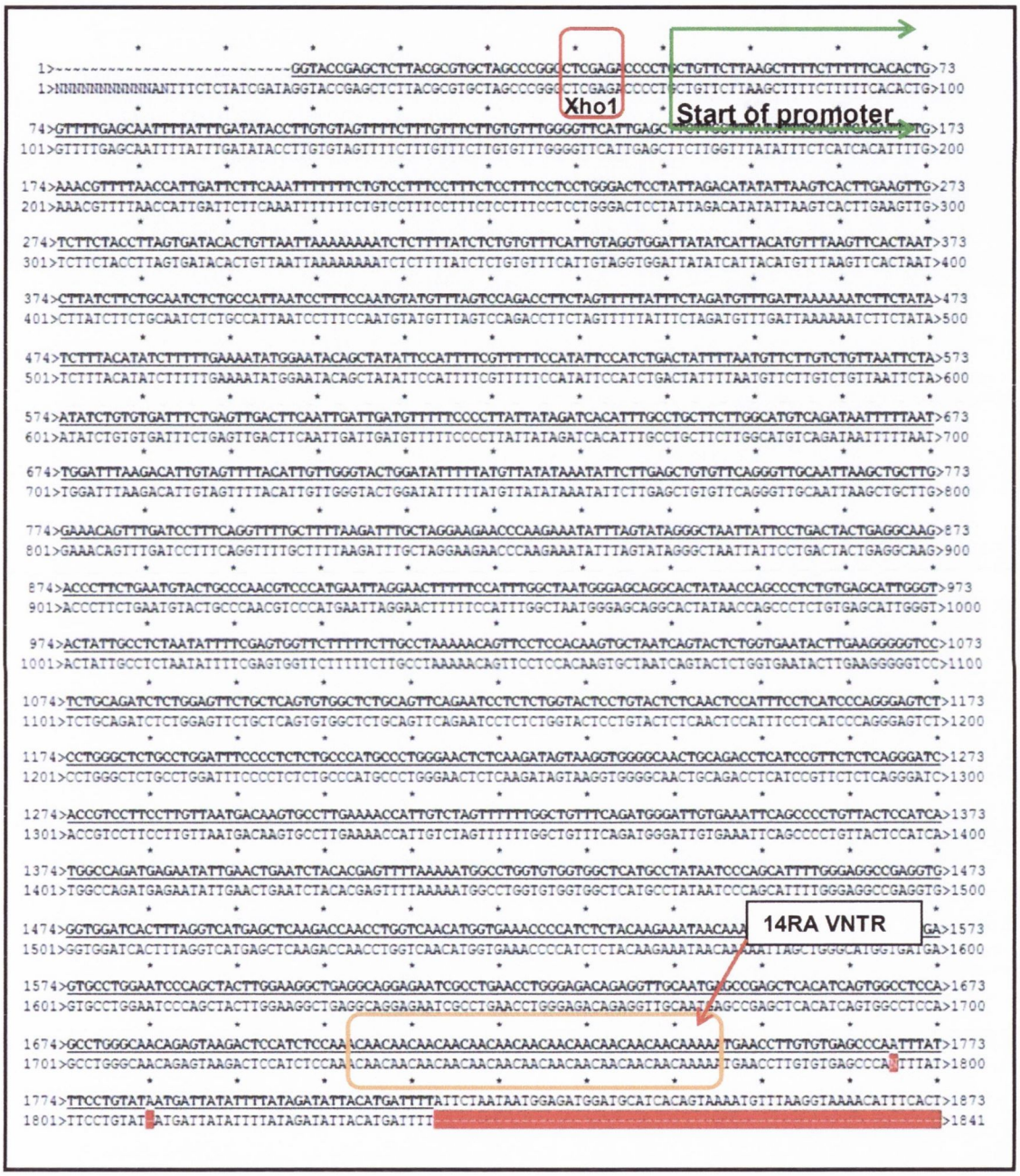


Figure 4.46 Alignment of 14RApGL3 with the proposed sequence

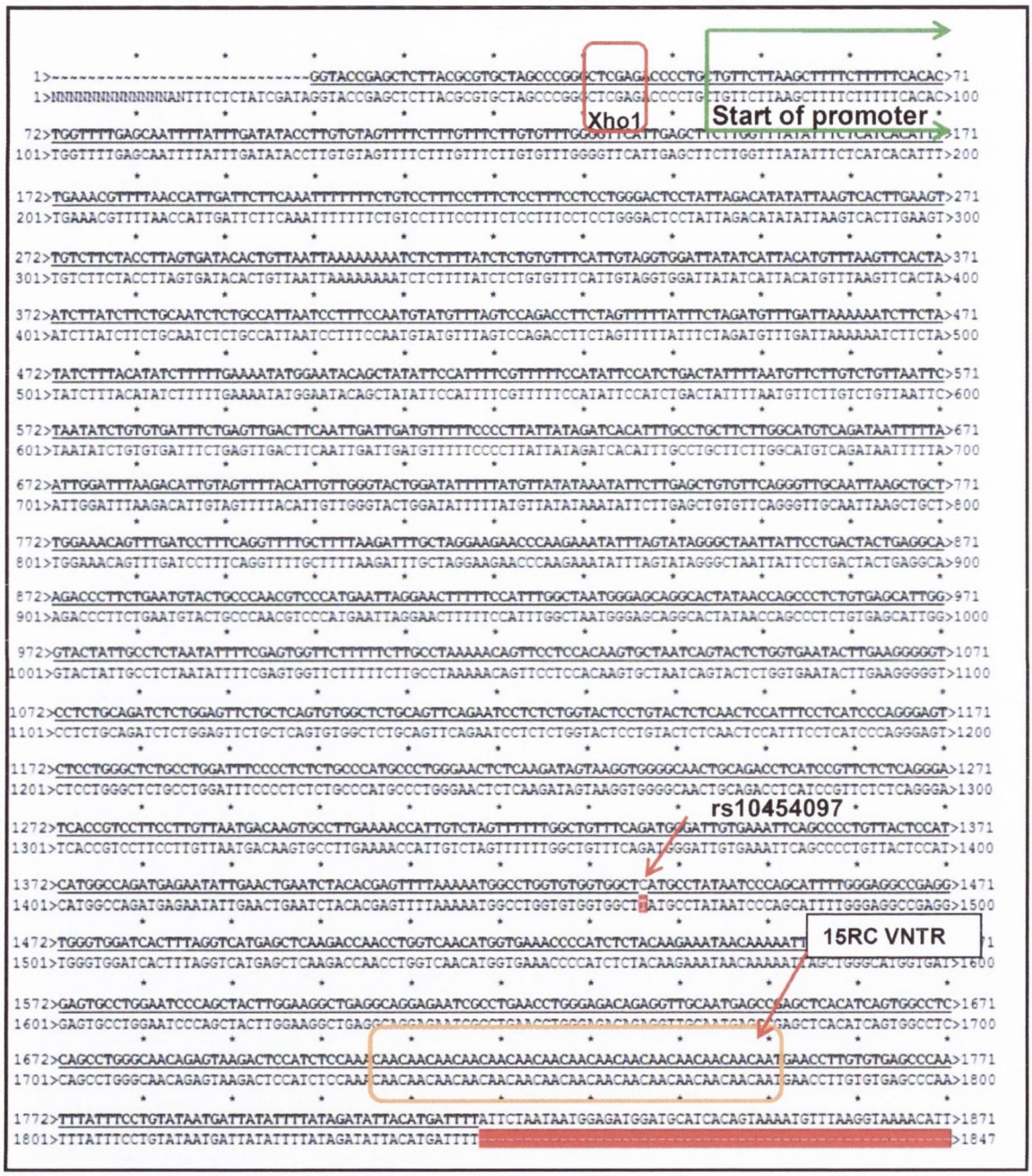


Figure 4.47 Alignment of 15RCpGL3 with a proposed sequence

Showing C/T SNP (rs1045097) having 50/50 % allele frequency

4.2.3.3 Point mutations in pGL3 reporter plasmid

The rest of the mutated clones, i.e. 13RC, 16RA/C and 17RA/C in pCR-Blunt were more difficult to sub-clone. After multiple attempts, it was decided to mutate the original 13RA clone in pGL3 instead. The clones for 16R and 17R will be described later in this chapter. The mutation for 18RA and 19RA was also conducted by this approach. The method adopted was the same to the ones from pCR-Blunt, with the exception that the cycling conditions had to be modified according to the length of pGL3. Screening the colonies from the PCR site-directed mutagenesis on 13R genotype, identified several positive colonies (Figure 4.48).

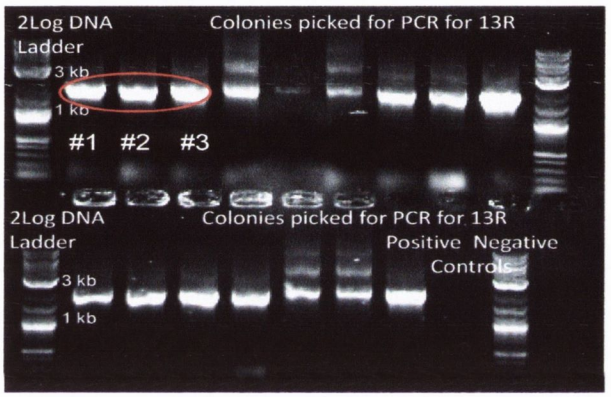


Figure 4.48 PCR on colonies for SDM on 13RCpGL3

The colonies circled in Figure 4.48 were chosen for sequencing. The purification concentration achived for each plasmid is shown in Table 4.23.

Table 4.23 Miniprep for 13RCpGL3 clones

Sample ID	ng/uL	A260	260/280	Average ng/μL
13R1.1	312.38	6.25	1.82	293.25
13R1.2	282.97	5.66	1.86	
13R1.3	284.39	5.69	1.85	
13R2.1	293.97	5.88	1.86	296.69
13R2.2	297.17	5.94	1.86	
13R2.3	298.92	5.98	1.85	
13R3.1	303.77	6.08	1.83	293.43
13R3.2	291.53	5.83	1.84	
13R3.3	284.98	5.70	1.85	

Two (#2 and #3) of the three sequenced clones proved to be correct mutants having the point mutation from an 'A' allele to a 'C' allele. The comparison between the proposed sequence and the new point mutated clone is shown in Figure 4.49 for clone #2.

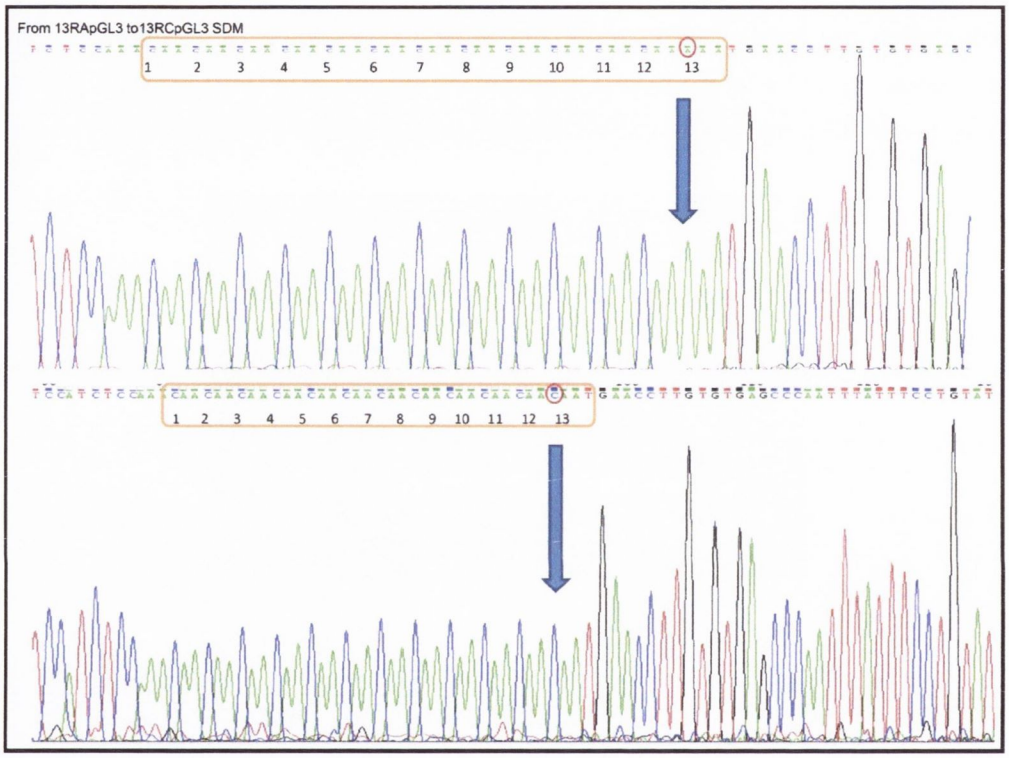


Figure 4.49 Sequence comparison for point mutation for 13RCpGL3

Mutation of 18RA and 19RA in pGL3 is described next. Upon transformation of the Dpn1 digested PCR product, colonies were obtained. The 18R genotype produced a substantial amount of colonies, whereas the 19R genotype only produced one single colony. Screening was done for 10 randomly selected colonies for 18R and the single colony for the 19R genotype. They all showed a promising insert size as seen in the Figure 4.50.



Figure 4.50 PCR on colonies for SDM for 18RApGL3 and 19RApGL3

The band size of the colonies was compared with the positive control in order to pick the correct length plasmids. The colonies selected (circled), were sent for sequencing with AS16For upon plasmid purification. The concentration achieved for each plasmid is seen in Table 4.24.

Table 4.24 Miniprep for 18RApGL3 and 19RApGL3 colonies

Sample ID	ng/uL	A260	260/280	Average ng/µL
18R1.1	151.02	3.02	1.86	223.35
18R1.2	260.56	5.21	1.87	
18R1.3	258.48	5.17	1.86	
18R4.1	311.78	6.24	1.86	309.86
18R4.2	310.85	6.22	1.86	
18R4.3	306.94	6.14	1.87	
18R5.1	313.67	6.27	1.86	316.42
18R5.2	318.66	6.37	1.86	
18R5.3	316.92	6.34	1.87	
19R.1	348.38	6.97	1.86	349.79
19R.2	349.87	7.00	1.86	
19R.3	351.13	7.02	1.86	

Sequencing resulted in a positive outcome, where both 18RA and 19RA mutations in pGL3 were obtained (Figure 4.51 and Figure 4.52).

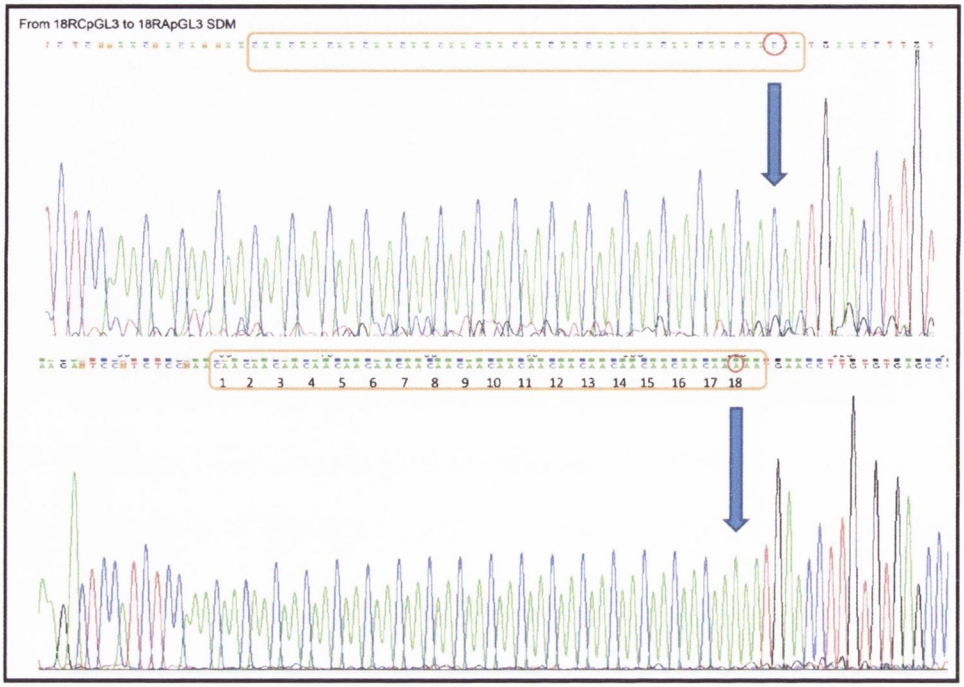


Figure 4.51 Sequence comparison for point mutation for 18RApGL3

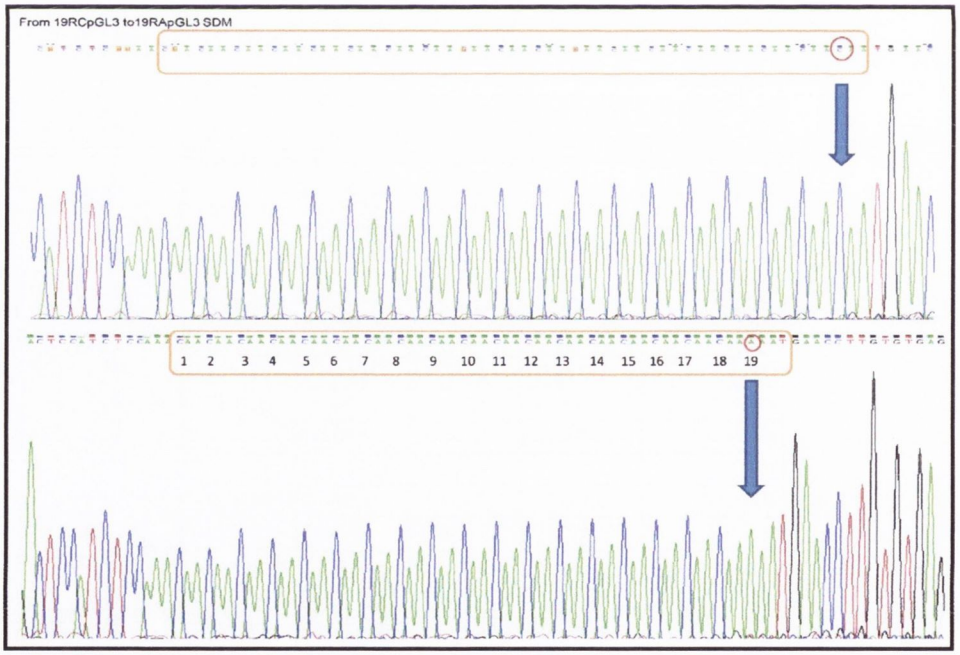


Figure 4.52 Sequence comparison for point mutation for 19RApGL3

From this result, it was also noted that one of the colonies from the mutated 18RA produced not only a nucleotide mutation but also a genotype mutation creating 17RApGL3. This was colony #5 in Figure 4.50.

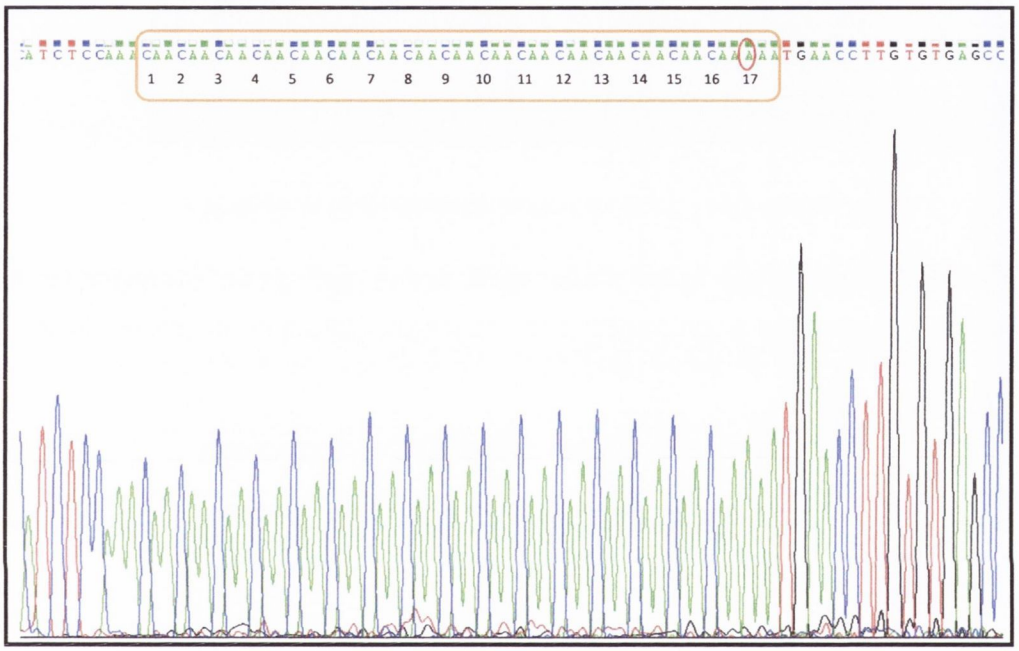


Figure 4.53 Sequencing result for 17RApGL3

Once this clone was identified, it was used to create the respective SDM for this genotype. Thus using this clone as a template, 17RApGL3, a point mutation using the CAA_For and CAA_Rev primers was performed. The PCR on colonies for this SDM showed a number of colonies that could be potentially carrying the desired mutant clone. The colonies were screened which is shown in Figure 4.54.

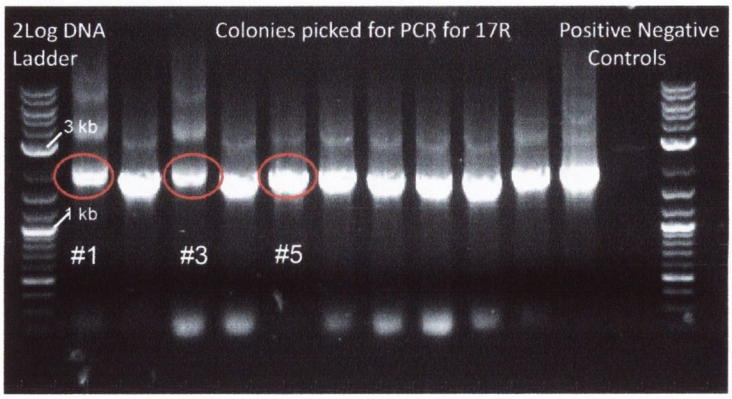


Figure 4.54 PCR on colonies for SDM for 17RCpGL3

The colonies shown in the figure above were picked and grown overnight for further sequencing analysis. The NanoDrop™ readings for quantification are shown in the table below:

Table 4.25 Miniprep for 17RCpGL3 colonies

Sample ID	ng/uL	A260	260/280	Average ng/µL
17R1.1	296.52	5.93	1.86	298.04
17R1.2	293.96	5.88	1.87	
17R1.3	303.64	6.07	1.84	
17R3.1	259.44	5.19	1.86	257.82
17R3.2	258.41	5.17	1.86	
17R3.3	255.60	5.11	1.88	
17R5.1	265.39	5.31	1.86	265.23
17R5.2	265.72	5.31	1.86	
17R5.3	264.57	5.29	1.86	

Following sequencing with AS16For, the point mutation for the 17RCpGL3 proved successful, as colony #5 had the desired mutation. A comparison of the parent plasmid to the newly formed nucleotide exchange clone is shown in Figure 4.55.

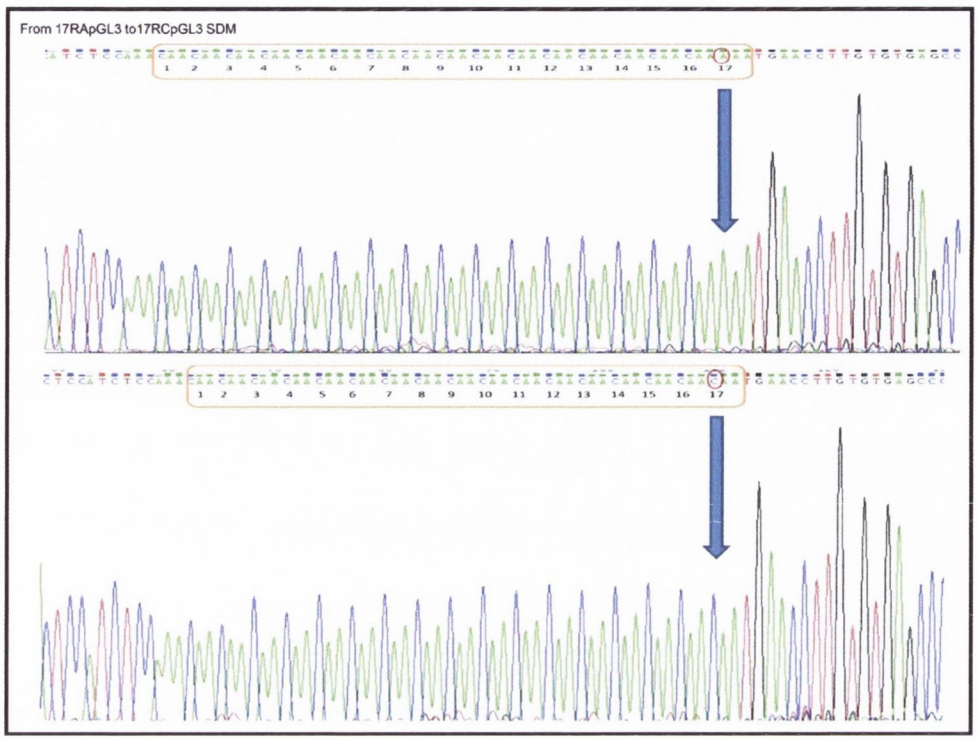


Figure 4.55 Sequence comparison for point mutation for 17RCpGL3

A legend of all clones with a description of method adopted to derive the clone is shown on page xxi.

4.2.4 Construction of different promoter lengths

In order to identify the main transcription start site of the promoter region, various promoter lengths had to be constructed in the reporter vector and then tested by on cells by running a luciferase assay. Depending on the strength of the promoter, the luciferase gene is expressed at different levels after trasnfection into mammalian cell line. The expression level of the luciferase directly translates into the bioluminescence emission which can be recorded and analysed. The extension sequence was synthesised by Eurofins MWG in a standard vector, named XLpUC57. This vector was used to clone in the sequence and this is described in this section. However, since cloning proved to be unsuccessful, a new method of mutagenesis was adopted.

4.2.4.1 Constructing the short length promoter: -705bp 10RpGL3

The promoter starting at -705 bp upstream of the ATG start codon for exon 1 of HRH4 was obtained a self-ligation of the digested -2044 bp 10RpGL3 (with Msc1 and Xho1). Hence, a shorter promoter resulted as described below.

4.2.4.1.1 Restriction enzymes digest for extension insert and 10RpGL3

The 10RpGL3 digested gives a 5472 bp band, while the digested XLpUC57 gives a 2627 bp for the insert and these are circled in the Figure 4.56 when visualised on an agarose gel.

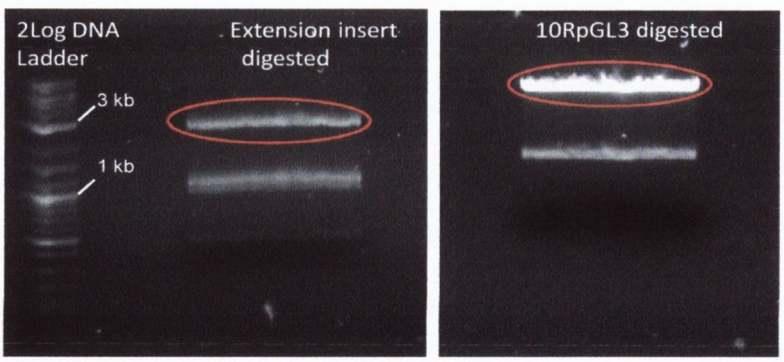


Figure 4.56 Restriction digest for extension insert and 10RpGL3

Purification by gel extraction gave a good concentration for the digested 10RpGL3 with Xho1 and Msc1 of about 90 ng/µL (Table 4.26). This was used to determine the volume need for ligation between the XL insert from digested XLpUC57.

Table 4.26 Purification for restriction enzyme digest for XLpUC57 and 10RpGL3

Sample ID	ng/uL	A260	260/280	Average ng/µL
XL insert.1	23.07	0.46	1.64	22.83
XL insert.2	21.85	0.44	1.74	
XL insert.3	23.56	0.47	1.76	
10RpGL3 digested.1	88.46	1.77	1.84	89.35
10RpGL3 digested.2	89.76	1.80	1.83	
10RpGL3 digested.3	89.84	1.80	1.85	

The ligation was performed using a different kit, purchased from NEB where quick ligation is done in 5 min compared to Invitrogen's kit which advised a ligation time between 1 hour and overnight.

4.2.4.1.2 Analysing transformants

After transformation, four colonies were seen the next day following incubation. A PCR on colonies was performed using RV Primer 3 and As11_Rev. It was expected to have a band size of 3178 (extension + -2044 bp 10RpGL3), however a band size of 500 bp was noticed. Figure 4.57 shows the resulting PCR for the colonies screened.



Figure 4.57 PCR on colonies for extension in 10RpGL3

The Msc1 restriction site to As11_Rev gives a band of 466 bp. This gave a suspicion that the digested 10RpGL3 ligated on itself. The colonies were grown in LB Amp^+ and extracted from the competent cells and purified. The purified plasmids were measure using NanoDrop™ and these readings are shown in Table 4.27. They were sent for sequencing with RV Primer 3 to check the identity of the band.

Table 4.27 Purified plasmids for short length colonies

Sample ID	ng/uL	A260	260/280	Average ng/µL
SL1.1	225.93	4.52	1.88	225.47
SL1.2	225.59	4.51	1.87	
SL1.3	224.90	4.50	1.86	
SL2.1	171.52	3.43	1.85	177.39
SL2.2	191.63	3.83	1.83	
SL2.3	169.02	3.38	1.87	
SL3.1	226.82	4.54	1.87	222.31
SL3.2	219.12	4.38	1.87	
SL3.3	221.00	4.42	1.88	
SL4.1	210.97	4.22	1.85	211.25
SL4.2	211.70	4.23	1.86	
SL4.3	211.09	4.22	1.84	

Analysing the sequence of these bands it showed that the sticky end produced by Xho1 and Msc1 ligated and formed a shorter version of the 10RpGL3. This resulted in a promoter region of -705 bp which was later used in the luciferase assay to test for promoter function.

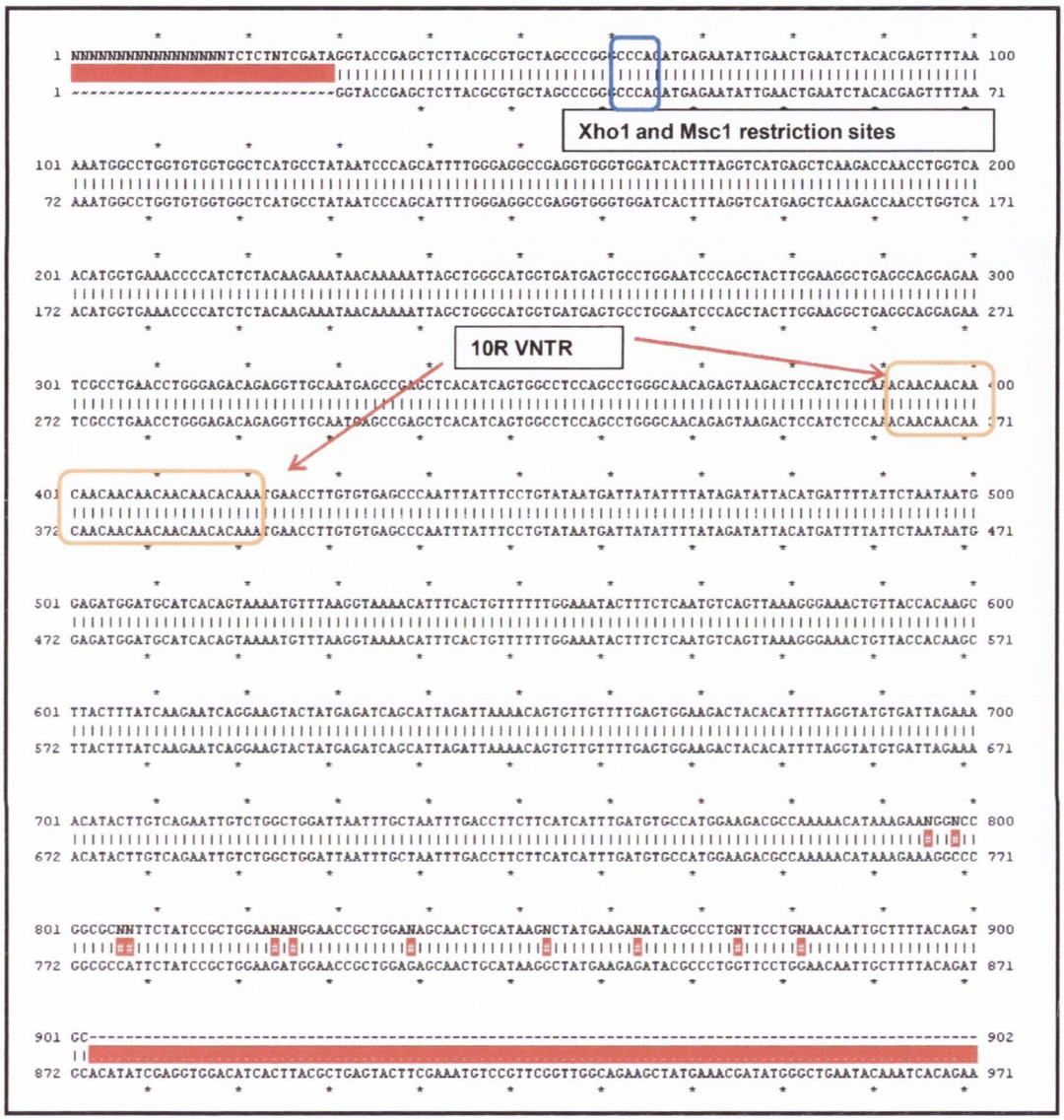


Figure 4.58 Alignment of -705 bp 10RpGL3 with proposed sequence

4.2.4.2 Constructing the long length promoter: -2772bp 10RpGL3

Cloning the other lengths (-2772 bp and -3329 bp) into -2044 bp 10RpGL3 from XLpUC57 proved to be difficult. Various conditions were changed and insert this longer part of the

potential transcription site into the original clone, but this turned out to be unsuccessful. Hence a different method was adopted. A mutation was done on the original clone, where a PCR-based site-direct mutagenesis step was used instead of ligation.

4.2.4.2.1 Amplification of the LL mega-primer

The first step was to amplify the desired region from XLpUC57. Amplification consisted of utilising primers that consist of two parts, one part anneals to the site of interest on XLpUC57 and the other part anneals to the original clone at the target locus. The mega-primer produced with LL SDM_for and LL SDM_rev primers shown below (Figure 4.59) gives a 780 bp band.

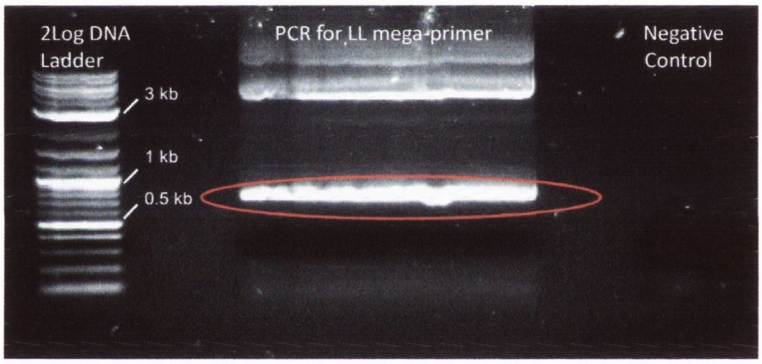


Figure 4.59 PCR for LL mega-primer from XLpUC57

The DNA was excised and purified by gel extraction giving a concentration of 86 ng/µL and good quality.

Table 4.28 Purified LL mega-primer from gel extraction

Sample ID	ng/uL	A260	260/280	Average ng/µL
LL Megaprimer.1	67.31	1.35	1.82	85.77
LL Megaprimer.2	90.40	1.81	1.75	
LL Megaprimer.3	99.61	1.99	1.70	

4.2.4.2.2 Analysis of mutated transformants with LL mega-primers: -2772 bp 10RpGL3

The newly amplified DNA was subjected to Dpn1 digestion to remove original parent template (-2044 bp 10RpGL3) from the PCR reaction mix. This was then transformed into TOP10 *E.Coli*, grown overnight and analysed with a PCR using RV Primer 3 (annealing to pGL3 vector) and AS11_rev (annealing on the insert) for potential positive clones. The expected band size of of 2772 bp.

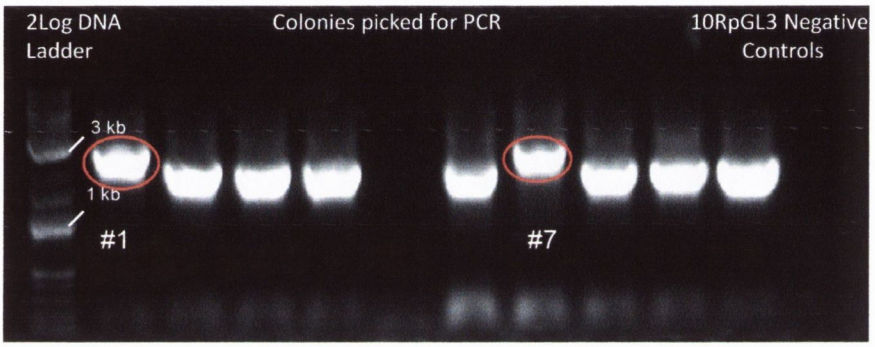


Figure 4.60 PCR on colonies for -2772 bp10RpGL3

Since the new addition to the 10R clone is 737 bp, two different band sizes were noticed when the gel was visualised under UV, where the original clone is 2044 bp and the mutated clone is 2772 bp. The PCR products showing a higher length are circled in red, colony #1 and colony #7. These two colonies were picked and grown overnight in LB Amp^+ and purified the next day using MiniPrep® kit from Qiagen.

Table 4.29 Miniprep for -2772 bp 10RpGL3 colonies

Sample ID	ng/uL	A260	260/280	Average ng/µL
#1 LL 10RpGL3.1	322.60	6.45	1.87	317.63
#1 LL 10RpGL3.2	315.73	6.32	1.87	
#1 LL 10RpGL3.3	314.57	6.29	1.88	
#7 LL 10RpGL3.1	333.80	6.68	1.87	325.77
#7 LL 10RpGL3.2	318.61	6.37	1.87	
#7 LL 10RpGL3.3	324.90	6.50	1.87	

The two colonies were sent for sequencing with RV Primer 3 and as seen from the alignment of template DNA with the sequencing result, it is shown that the mutation was successful for both and that the promoter was extended with 737 bp upstream on to the original clone, -2044 bp 10RpGL3. Figure 4.61 shows the sequencing result for colony #1.

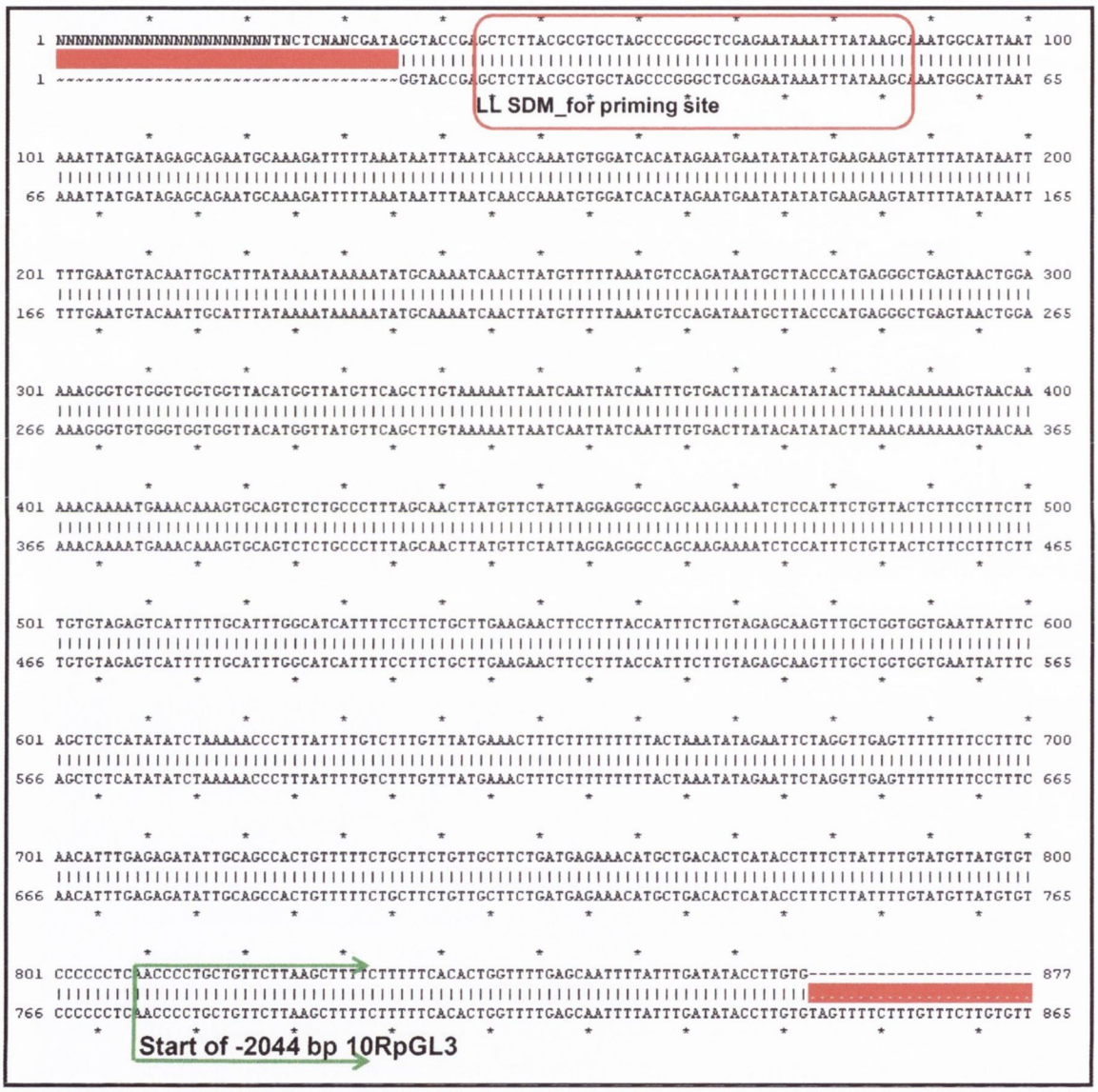


Figure 4.61 Sequencing result for -2772 bp 10RpGL3

4.2.4.3 Construction of the longest length promoter: -3329 bp 10RpGL3

Furhter extension to the previously cloned promoter was performed, -3329 bp 10RpGL3, where the 548 bp was added upstream of the -2772 bp of the promoter. This was calculated from Promoter 2.0 online database where this region gives a highly probable transcription start site.

4.2.4.3.1 Amplification of the XL mega-primer

The XL mega-primer was amplified using XL SDM_For and XL SMD_Rev, using XLpUC57 as the parent template. The amplicon is visualised on the molecular image produced a band approximately 600 bp (together with the full extension of the primers) as seen in Figure 4.62.

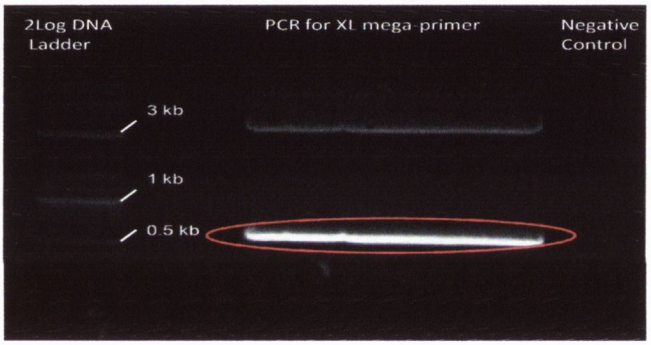


Figure 4.62 PCR for mega-primer for -3329 bp 10RpGL3

This band was excised out of the gel and purified to be used for the next PCR creating the mutation. The concentration for this amplicon was approximately 230 ng/µL as seen from Table 4.30, hence 1.7 µL (400 ng) was used for the site-directed mutagenesis PCR.

Table 4.30 Purified XL mega-primer after gel extraction

Sample ID	ng/uL	A260	260/280	Average ng/µL
XL mega-primer.1	226.69	4.534	1.83	229.55
XL mega-primer.2	230.56	4.611	1.83	
XL mega-primer.3	231.41	4.628	1.85	

4.2.4.3.2 Analysis of mutated transformants with XL mega-primers: -3329 bp 10RpGL3

The mega-primer used in the SDM added the extra part of 548 bp to the -2772 bp 10RpGL3 clone, where following that the parent plasmid was digested with Dpn1 and the PCR product transformed into the competent cells. The PCR on colonies, used to analyse transformants, gave a good indication that the mutation was successful. This is shown in the gel visualised below (Figure 4.63):



Figure 4.63 PCR on colonies for -3329 bp 10RpGL3

PCR was performed with RV Primer 3 and XL SDM_Rev which produces a band at approximately 650 bp. The first two colonies were picked and incubated for 16 hours in LB Amp$^+$ where the plasmid was then purified and sent for sequencing the next day with RV Primer 3. The concentration obtained after purification is shown in the table below (Table 4.31).

Table 4.31 MiniPrep for -3329 bp 10RpGL3

Sample ID	ng/uL	A260	260/280	Average ng/µL
XL1.1	354.04	7.08	1.87	335.25
XL1.2	324.26	6.49	1.86	
XL1.3	327.44	6.55	1.86	
XL 2.1	410.12	8.20	1.86	416.26
XL 2.2	410.31	8.21	1.86	
XL 2.3	428.34	8.57	1.86	

Sequencing for the new plasmids showed a successful mutation in the promoter region and Figure 4.64 shows the alignment with the desired template.

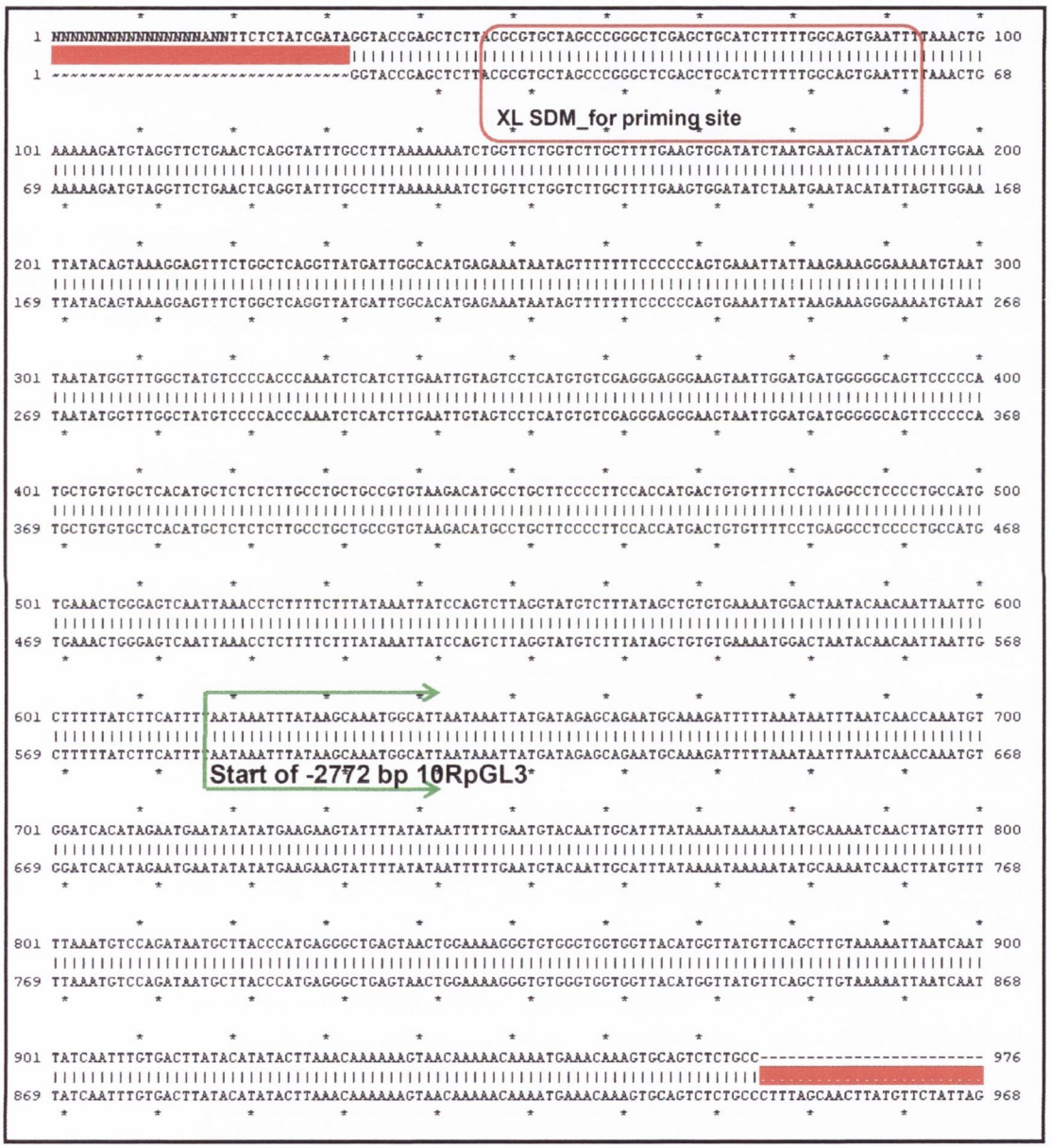


Figure 4.64 Alignment of -3772 bp 10RpGL3 with proposed sequence

4.2.4.4 Addition of Exon 1 to promoter length constructs

Modification to each promoter construct was conducted to identify a functional promoter. Exon 1 of HRH4 was added to them in order to see if this facilitates gene transcription. This modification was also conducted for the -2045 bp 13RpGL3 (the highest allele frequency genotype), in order to see if this created a difference in promoter activity.

4.2.4.4.1 Amplification of Exon 1 for a mega-primer

PCR included amplification from human genomic DNA and turned out to be highly efficient as minimal optimisation was required. As seen from the figure below, it gave a high concentration of PCR product with no side products. Exon 1 for HRH4 is small hence only an amplicon of approx 170 bp was produced.

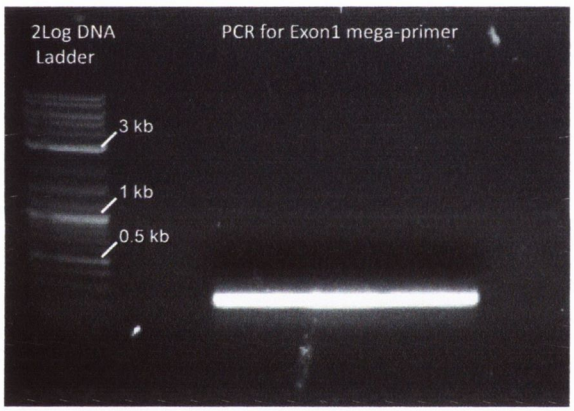


Figure 4.65 PCR for mega-primer for exon 1 extension

The band cut out from the gel gave a good concentration and purity of 243 ng/µL (Table 4.32).

Table 4.32 Purification of Exon 1 mega-primer after gel extraction

Sample ID	ng/uL	A260	260/280	Average ng/µL
Mega-primer Ex1.1	240.65	4.81	1.90	243.15
Mega-primer Ex1.2	242.28	4.85	1.88	
Mega-primer Ex1.3	246.52	4.93	1.87	

The amplification for exon 1 mega-primer had to be repeated a few times since it had to be added to all different promoter length constructs (and 13RpGL3). The resulting concentration for this mega-primer was consistent for all PCRs, where the DNA purification is shown in Table 4.32 which reflects of these results.

4.2.4.4.2 Analysing mutated transformats with Exon 1 mega-primers

The amplicons shown are the result of a PCR with Seq1F and Ex1 SDM_For, which produces a product of around 730 bp. Exon 1 extension was successful for the shortest length clone, i.e. -705 bp Ex1 10RpGL3. Figure 4.66 shows the positive colonies obtained after PCR analysis, where colonies #1 and #2 were picked, grown and minipreped.

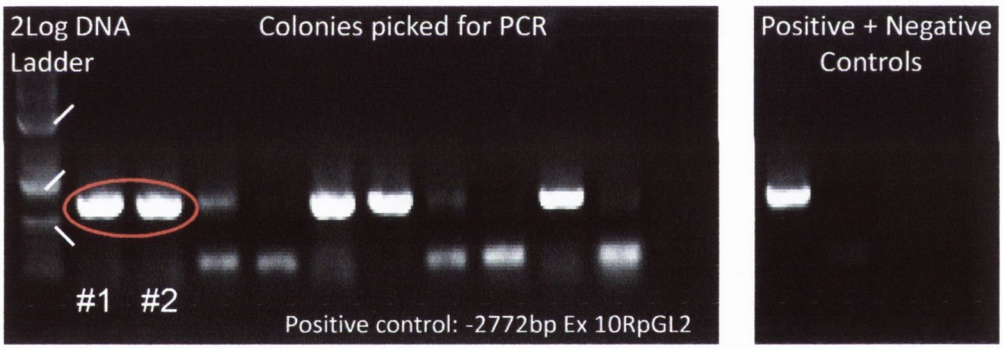


Figure 4.66 PCR on colonies for -705 bp Ex1 10RpGL3

Following that, the lengths for -2044 bp for both 10R and 13RA were performed and the colonies identified are shown in the next figure (Figure 4.67).

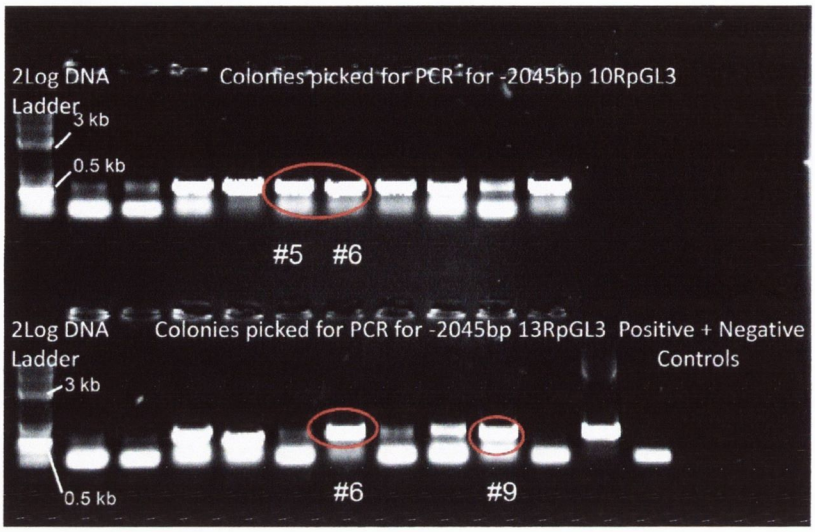


Figure 4.67 PCR on colonies for -2044 bp Ex1 10RpGL3 and -2044 bp Ex1 13RpGL3

Colonies #5 and #6 for 10RpGL3 were picked and colonies #6 and #9 were selected for 13RpGL3.

The clone identified for -2772 bp Ex1 10RpGL3 is seen below (Figure 4.68) show the resulting bands after screening the colonies.

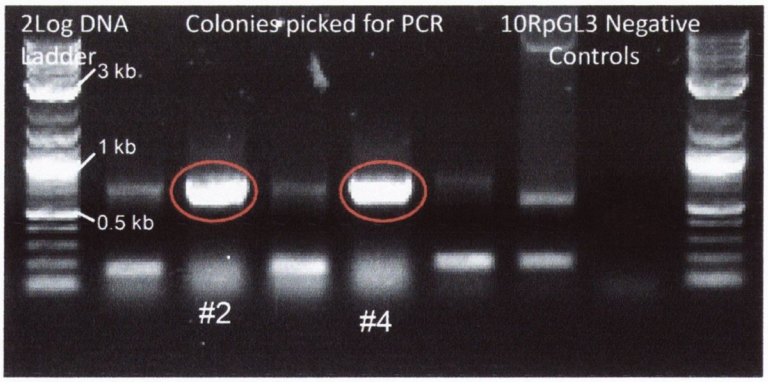


Figure 4.68 PCR on colonies for -2772 bp Ex1 10RpGL3

Colony #2 and #4 were picked, grown, purified and this is shown in Table 4.33.

The last extension was performed on -3329 bp 10RpGL3 and the PCR on colonies is seen in Figure 4.69.



Figure 4.69 PCR on colonies with 3329 bp Ex1 10RpGL3

The bands circled (Figure 4.66, Figure 4.68, Figure 4.67, Figure 4.69) were picked, grown and the plasmid was extracted with a MiniPrep. The resulting concentration is shown in the table below (Table 4.33).

Table 4.33 Miniprep for colonies with exon 1: -705, -2042, -2772, -3329 bp for 10RpGL3 & -2042 bp for 13RpGL3

Sample ID	ng/uL	A260	260/280	Average ng/μL
SL Ex1.1	334.56	6.69	1.85	333.08
SL Ex1.2	337.21	6.74	1.86	
SL Ex1.3	327.46	6.55	1.88	
SL Ex2.1	292.05	5.84	1.86	286.34
SL Ex2.2	282.99	5.66	1.87	
SL Ex2.3	283.99	5.68	1.88	
10R ML 5.1	339.56	6.79	1.89	339.85
10R ML 5.2	339.56	6.79	1.89	
10R ML 5.3	340.44	6.81	1.90	
10R ML 6.1	345.80	6.92	1.87	342.64
10R ML 6.2	341.02	6.82	1.87	
10R ML 6.3	341.10	6.82	1.88	
13R ML 6.1	325.13	6.50	1.87	322.56
13R ML 6.2	312.54	6.25	1.87	
13R ML 6.3	330.01	6.60	1.91	
13R ML 9.1	305.60	6.11	1.89	306.21
13R ML 9.2	303.78	6.08	1.90	
13R ML 9.3	309.26	6.19	1.88	
LL ex col2.1	343.94	6.88	1.87	343.26
LL ex col2.2	339.36	6.79	1.88	
LL ex col2.3	346.48	6.93	1.86	
LL ex col4.1	312.58	6.25	1.87	317.06
LL ex col4.2	319.60	6.39	1.86	
LL ex col4.3	319.00	6.38	1.86	
XL Ex9.1	447.86	8.96	1.85	442.09
XL Ex9.2	448.61	8.97	1.84	
XL Ex9.3	429.80	8.60	1.85	
XL Ex10.1	375.20	7.50	1.85	349.38
XL Ex10.2	336.09	6.72	1.86	
XL Ex10.3	336.86	6.74	1.86	

The sequencing results for all samples, i.e. -3329 bp, -2772 bp, -2044 bp, -705 bp Ex1 10RpGL3, and -2044 bp Ex1 13RApGL3 are shown in Figure 4.70, where the 3'-ending of the promoter region is shown extending into the ATG sequence of HRH4 exon 1 and further into the ATG codon for luciferase gene in pGL3 vector.

The last clone to be identified was the -3329 bp Ex 1 10RpGL3, hence the alignment for this clone is shown in Appendix 3. The proposed sequence used for alignment is shown in Chapter 2, Section 2.8.5



Figure 4.70 Sequence alignments for exon 1 on promoter length constructs

Where the proposed sequence (shown in Chapter 2, Section 2.8.5) is aligned for exon 1 of HRH4

4.3 Discussion

The human histamine 4 receptor gene has several single nucleotide polymorphisms (SNPs) and triple repeat variations (VNTR) on the regulatory site of the promoter region. Studying these repeats is beneficial in understanding the impact they have on transcription and possibly hint at a biological role.

Linkage analysis carried out using markers which were identified through genome-wide association studies by the international schizophrenia consortium (Chapter 3) identified SNPs which are weakly linked to schizophrenia and showed that there is a strong association between these markers and polymorphisms on HRH4. However, on further genotyping the VNTR region of the human histamine H4 receptor gene interesting insights into the potential promoter site indicating a highly complex polymorphic region were granted.

Initial focus for the VNTR variations on the promoter region for this protein focused on a potential relation to schizophrenia, but other diseases could be possibly linked to this interesting locus and warranted further investigations. Recent publication reported the association of SNPs on the HRH4 to Atopic Dermatitis [60]. One of the SNPs described (rs77485247) is located in intron 1 while others are located in exon 3 (rs74604924 and rs77041280) and *P* values of 0.002, 1.5×10^{-7} and 1.3×10^{-7}, respectively were reported. The same research group investigated SNPs for systemic lupus erythematosus. While no association was found between these 3 SNPs and this condition, they suggested that genetic variations in HRH4 copy number variations (CNV) contribute to the pathogenesis of systemic lupus erythematosus [60, 65]. With the increasing interest of genetic association to diseases, the HRH4 is joining the race for identifying polymorphism and their contribution to diseases in search for personalised medicines.

From the published work of Coge *et al* [7], it was suggested that the end of this VNTR (having CAC AAA end) lies on a recognition sequence for the transcription factor α-INF, which implies that HRH4 gene expression might be stimulated by factors such as interferon-α. A polymorphism in this region could affect important transcription processes in inflammatory responses and possibly other disorders. This reaffirms the importance of studying the response elicited by these variations in this important regulatory site for different genotypes.

Cloning the VNTR variations required identifying the start of the promoter region. The study of Coge *et al* [7], identified this potential regulatory site through bioinformatics analysis and hence was used as the template, i.e., -2044 bp. The initial step was to clone the insert into a standard vector in order to make the DNA easily available for further analysis and modifications. The vector chosen was pCR-Blunt (Invitrogen); a linearised vector possessing blunt ends. The vector also contains a lethal *ccd*B gene (control of cell death locus), which in case of self-ligation triggers this lethal gene which prompts DNA breakage and thus cell death. The standard vector chosen is a high copy number plasmid which allows for the host cell a high efficiency in replicating the plasmid. Figure 4.71 maps out the characteristics of the vector and identifies the key points of interest for the cloning method adopted.

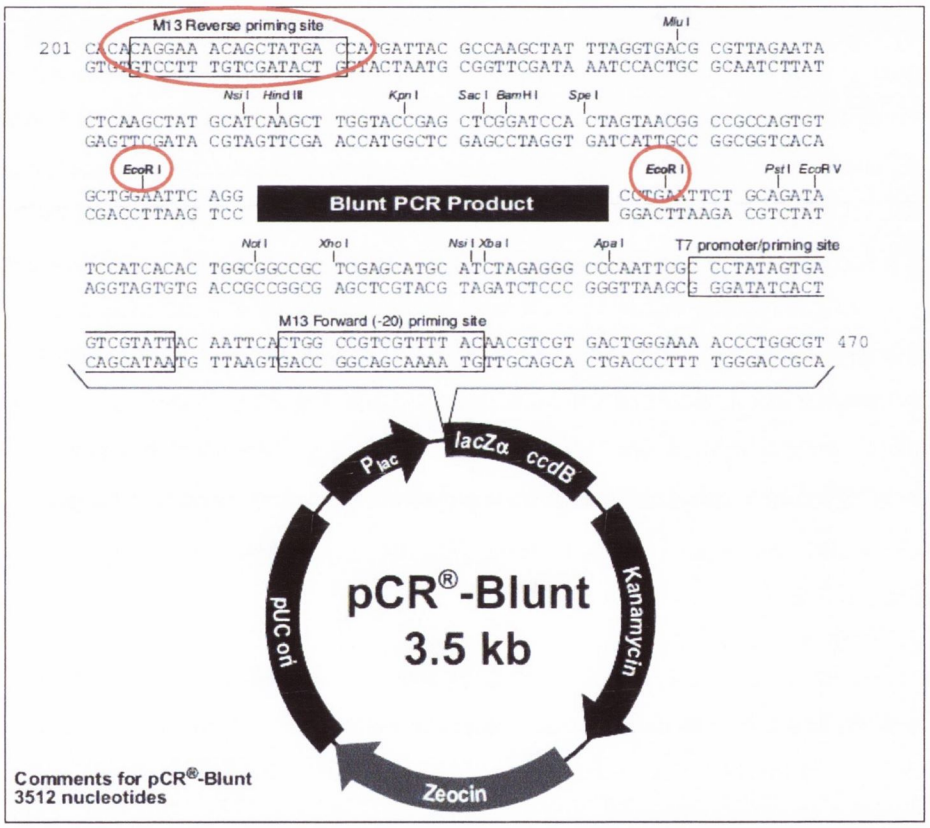


Figure 4.71 Structure and characteristics of pCR-Blunt vector

Showing the location for the inset; M13Rev forward primer in PCR on colonies and for sequencing; and EcoR1; restriction site digest

Initial amplification of the desired DNA region was performed on human genomic DNA. The DNA was either collected from volunteers in the School of Pharmacy and Pharmaceutical Sciences, following ethical approval or obtained from HapMap samples. The major advantage for using samples where the genotype was identified in previous studies was that it allowed for homozygous samples to be chosen. This facilitated the cloning process as the VNTR was controlled for and reduced the analysis time and screening process.

Primers designed for PCR on genomic DNA to target the region of interest gave a product of 2259 bp (based on 14R as per NCBI), which extended into exon 1 of HRH4. Since the vector (pCR-Blunt) chosen for cloning is a linear blunt-ended vector, the *Taq* polymerase used for the PCR was *Phusion*® DNA polymerase, an enzyme that does not add a polyA tail at the end of the amplicon, and hence it produces a blunt ended PCR product. Following purification of the amplicon, it was ligated into the host vector, pCR-Blunt. Various attempts were made to identify the ideal insert to ligate ratio and most successful proved to be the 10:1 insert:vector molar ratio. The nature of the triplet repeat could be a possible reason why ligation presented itself as a challenge. The repeated 'A' alleles along the VNTR and especially the A-tract located at the 3'-end of this repeat could be the basis for a supercoiling-dependent A-tract behaviour [116] which results in bent adenosine-tracts making it difficult to ligate. Hence, the ligation time and temperature proved to be somewhat difficult to optimise. Various ligation trials were conducted, including a variation temperature (ranging from room temperature to 4°C), duration ranging from 1 hour to overnight ligation, and also denaturating the DNA for 2-3 min at 95°C in order to uncoil the double stranded DNA from potential hinges (due to the CAA/ AAA triplet repeat). The optimised ligation conditions set were at 16 °C for approximately 3 hours.

Transformation into chemically competent cells was the step that followed, where potential colonies hosting the plasmid clone were analysed. As seen in Figure 4.71, EcoR1 is useful for screening as on incubation with EcoR1 in a suitable buffer the whole insert was cut out. In this way the first clone was identified. The cleaved product was visualised by agarose on a gel electrophoresis in order to verify that an insert of the expected band size is present in the plasmid. A restriction enzyme digest with EcoR1 will only confirm the product insert length but not the correct 5' to 3' end orientation. Since the ligation consisted of blunt ends between the insert and vector, the DNA could orientate itself in the opposite direction. Where a correct-

sized length band is observed, further sequencing analysis was carried out to confirm sequence and correct orientation of the insert. The correct clone was confirmed through sequencing with the standard primer M13Rev included in the pCR-Blunt construct (Figure 4.71). The sequencing results tend to give a good reading up to approximately 600-800 bp, and in order to confirm the correct VNTR was present, the clones were also sequenced with AS16F, a sequencing primer annealing approximately 90 bp upstream of the VNTR.

The first clone to be identified was the 10R length repeat. The sequencing result not only confirmed the correct 10R polymorphism but interestingly, it also showed the presence of rs615283 which exhibits A/G polymorphism. This polymorphism was one of the SNPs found to be in high correlation (r^2 = 1.00) with the 10R from the linkage disequilibrium study (Chapter 3).This SNP was also present in the promoter sequence published by Coge *et al*, confirming that the A allele of this SNP is in LD with 10R. This is shown in the alignment below (Figure 4.72):

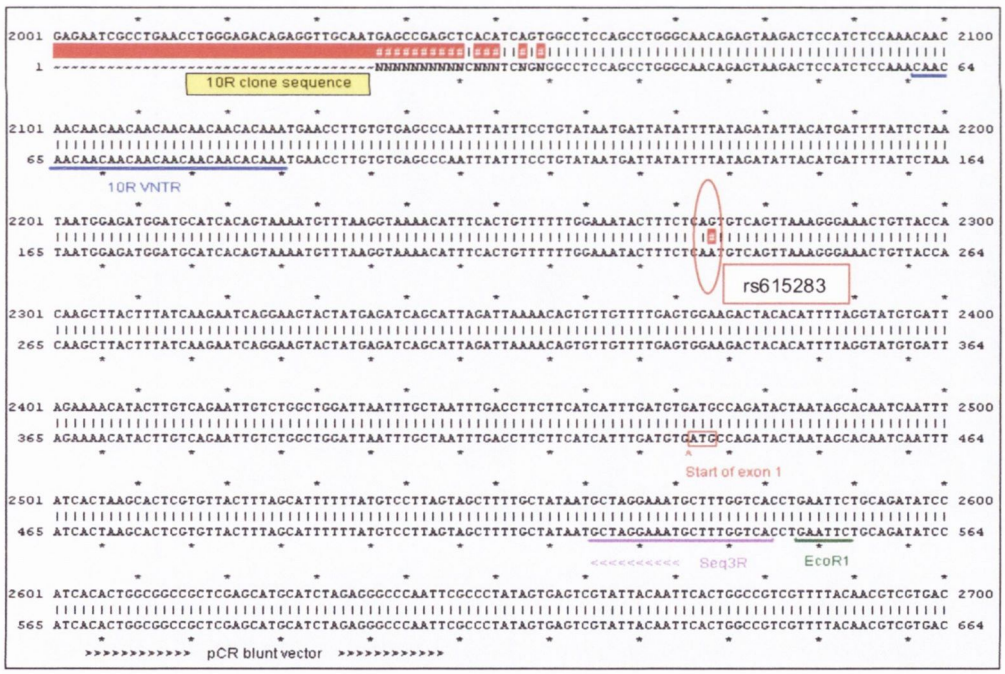


Figure 4.72 Tag SNP for the 10R genotype

(Aligned to the proposed sequence)

For none of the other genotypes (13R, 14R, 15R, 16R, 17R, 18R 19R) this polymorphism was observed, which further supports the hypothesis that this SNP is specific for the 10R repeat. Any variation (SNP) noticed on any of the cloned genotypes was analysed by using the dbSNP database from NCBI, and the allele frequency was taken into consideration. Where the SNP had a 50/50% frequency, the construct was used in further analysis.

The major steps of cloning into pCR-Blunt are shown in the figure below, where this potential promoter region is amplified and ligated into the standard vector.

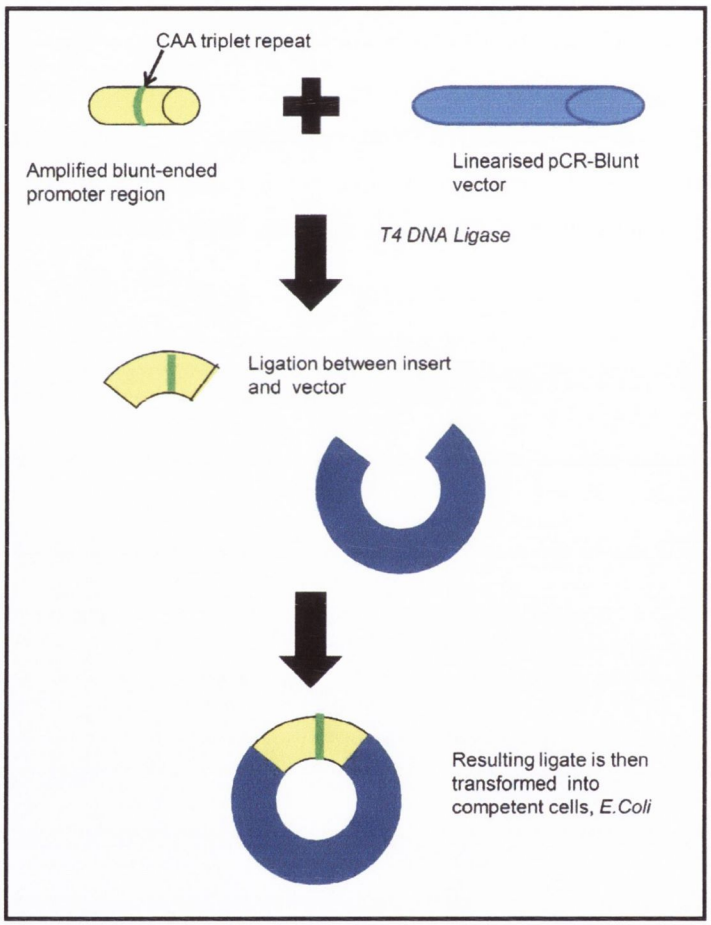


Figure 4.73 Cloning for the potential promoter region into pCR-Blunt

Genotypes having 10R, 13R, 14R, 15R, 16R and 17R length triplet repeat were successfully cloned into pCR-Blunt vector. This facilitated the PCR amplification for the sub-cloning into

the reporter plasmid, pGL3. This plasmid is a vector possessing the luciferase gene, which when transcribed displays bioluminescence which can be read, recorded and quantified. The plasmid chosen also had no promoter region, hence called pGL3-basic (Promega).

The clones in pCR-Blunt were designed to have a Xho1 restriction site on the 5'-end and a Nco1 site on the 3'-end. Hence these were used to further sub-clone into the reporter plasmid (pGL3-basic). The primers used for this amplification had to be primers that create a sticky ends matching the ones present on pGL3-basic. As shown in Figure 4.74, sub-cloning required well designed primers with restriction sites for Xho1 and Nco1 in order to have the regulatory region inserted near the ATG start codon of the luciferase gene.

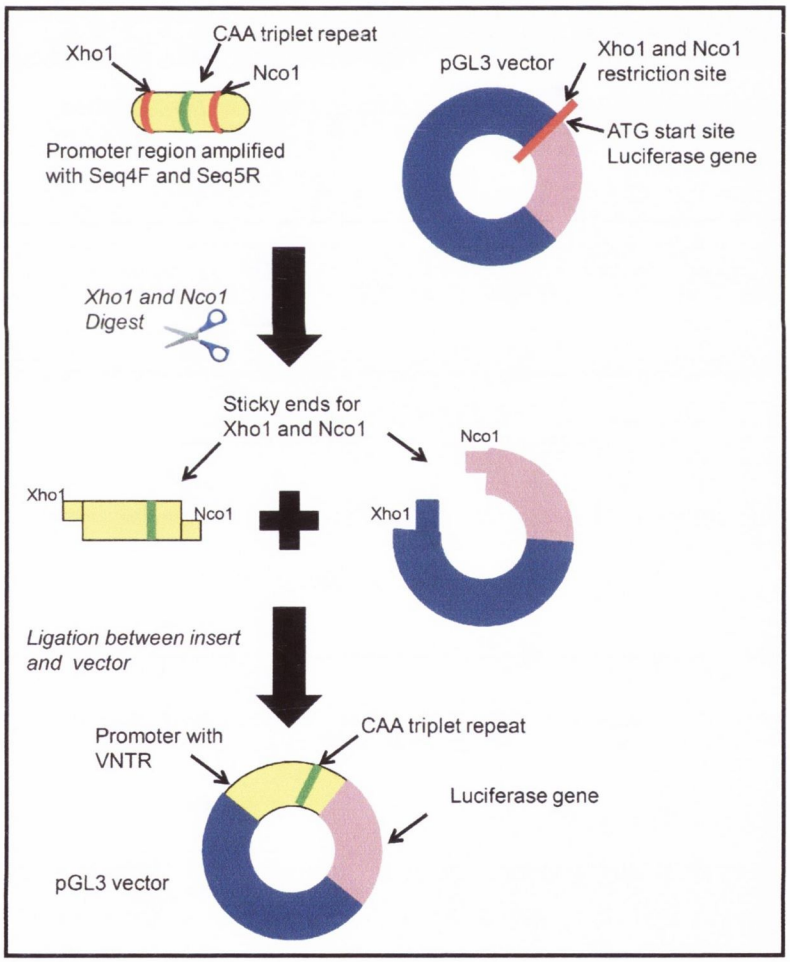


Figure 4.74 Sub-cloning the potential promoter region into a reporter plasmid

The method shown above was used for the majority of the genotypes. The two rare length repeats (18R and 19R) were more difficult to amplify and never actually found as homozygous alleles in any of the samples genotyped. Using the HapMap samples and data obtained in the analysis carried out in Chapter 3, these genotypes were found in a very low allele frequency. The 18R genotype was found only twice heterozygous with 14R (18R/14R) and the 19R genotype was found only once and heterozygous with 14R (19R/14R) (n=90). Hence, the sequence (shown in Chapter 2, Section 2.8) was given to Eurofins MWG and synthesised into a standard vector. A restriction site for Msc1 present 705 bp upstream to the ATG start site of luciferase in the 10RpGL3 containing the promoter region of HRH4 was used to insert the synthesised DNA and sub-clone these genotypes. This Msc1 restriction site, which is only found once in insert and plasmid is located 327 bp upstream to the VNTR region, hence made it a perfect ligation site. This is mapped in the figure below (Figure 4.75), where it shows the linear 10R promoter insert with the main characteristics.

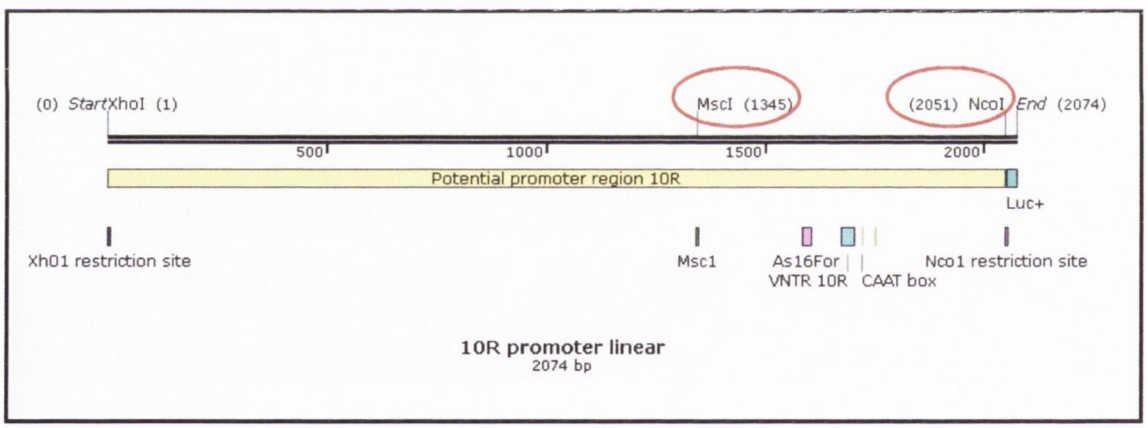


Figure 4.75 Schematic diagram for the 10R linearised promoter region

Sub-cloning of 18R and 19R was accomplished by using the 10RpGL3 as template. The vectors were digested with Msc1 and Nco1 to create matching sticky ends for ligation. The ligation for these two longer genotypes followed a slightly different protocol, where a ligation ratio of 6:1 (insert:ligate) was found to be optimal.

The construction of the human H4 receptor promoter genotype library progressed in creating this SNP on the first nucleotide of the last triplet repeat. This was accomplished by executing a point mutation that converts the last triplet repeats from **C**AA to **A**AA and vice versa. The primers designed had to comply with requirements described by Stratagene according to their guidelines in QuikChange II Site-Directed Mutagenesis Kit. The primers had to be between 25 and 45 bases in length, with a melting temperature (T_m) of ≥78 °C. Consequently, two primer sets were designed to produce these mutations, i.e. converting CAA to AAA and *vise versa*.

The primers used for this mutation are shown below with the respective mutation highlighted in red depending on the mutation needed:

CAA mutation

Forward: 5' **CAACAACAACAACAACAACAA<u>C</u>AA**TGAACCTTGTGTGAGCCC 3'

Reverse: 5' GGGCTCACACAAGGTTCA**TTGTTGTTGTTGTTGTTGTTG** 3'

AAA mutation:

Forward: 5' **CAACAACAACAACAACAACAAAAA**TGAACCTTGTGTGAGCCC 3'

Reverse: 5' GGGCTCACACAAGGTTCA**TTTTTGTTGTTGTTGTTGTTG** 3'

Another important aspect for this site-directed mutagenesis is that the parent DNA is grown in competent cells that produce methylation in plasmid DNA by expressing functional cystosine methyltransferase (*dcm*) and DNA adenine methyltransferase (*dam*). Methylation is a result of DNA replication in most prokaryotes and eukaryotes and it stabilises DNA [117]. These types of methylations are the most common post-replicative methylations, and differ from DNA synthesised in a PCR. As a result, in most DNA mutation methods the fact that the parent (plasmid) DNA is methylated but the PCR amplified DNA is not methylated is taken advantage of, where digestion with Dpn1 enzyme facilitates the disintegration of the parent

(prokaryotic) DNA template and leaves the mutated (synthesised) PCR product intact. Figure 4.76 shows the main steps required for this point mutation.

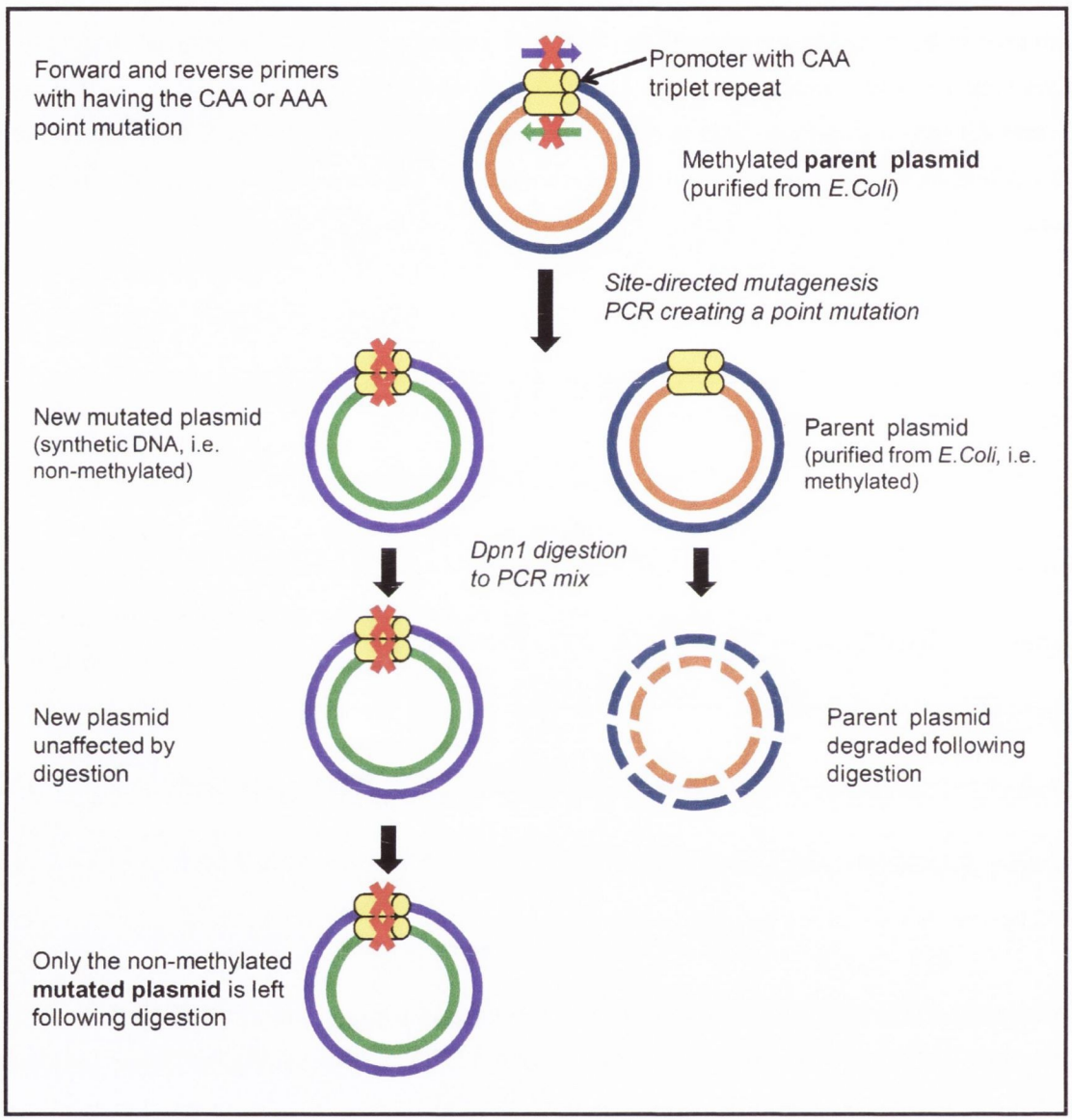


Figure 4.76 Site-directed mutagenesis creating a point mutation

(This figure does not show the nicked sites created after PCR; which are only restored after transformation into competent cells)

Surprisingly, this created various mutations with an extended triplet repeat, beyond 40 triplet repeats in some cases. Screening the colonies with PCR ensured that the band size was in the desired region. This is a possible explanation why the 17RA in pGL3 was achieved following a point mutation for 18RApGL3. This method, however, turned out to be more reliable in obtaining promoter constructs than sub-cloning from constructs in pCR-Blunt. Following various attempts to obtain the 13RCpGL3 from 13RCpCR-Blunt, it was decided to mutate the already sub-cloned 13RApGL3 with this technique. The 16RA and its respective mutated 16RC in pCR-Blunt failed to be sub-cloned into pGL3 vector after numerous attemps. The remaining clones in pGL3 were successfully mutated into the other (A or C) allele variant by site-directed mutagenesis.

A clones library for the different genotypes for the VNTR region of the putative promoter region was successfully achieved. However, the need to identify the functional promoter was required in order to confirm the proposed (10R) sequence (-2044 bp) as the functional promoter. This was performed by a cell-based assay which will be described in Chapter 5. This assay was being optimised while constructing the library, and the promoter length sequence of -2044 bp was not giving a strong indication that it is indeed the functional promoter. As a consequence, different promoter lengths were constructed in order to identify the active promoter of HRH4.

The promoter lengths were designed by using a bioinformatics database (Promoter 2.0). This bioinformatics site makes use of genetic alogarithms to discriminate between promoter and non-promoters [118]. Gene expression is controlled by a proximal promoter, where the pre-initiation complex (PIC) determines the location of transcription initiation for protein-coded genes [119]. The arrangement of transcription factor binding sites is specific for each gene, hence a promoter region is characterised differently for each gene [118]. Using bioinformatics tools such as Promoter 2.0, gives an indication of where this transcription site is located.

The region between IMPACT (the gene before HRH4) and HRH4 was obtained from the NCBI website as FASTA format and analysed in Promoter 2.0. The region between 22,033,495 and 22,040,692 bp was selected. The HRH4 gene starts at 22,040,593, hence in this analysis included the region for mRNA and DNA polymerase transcription. The ATG start codon for HRH4 is at 22,040,692 bp. Full promoter analysis showed that the area starting at

22,038,980 bp is highly likely to be an active promoter with a prediction score of 1.104. This region is -1681 bp upstream of exon 1. The sequence analysis was repeated with shorter lengths until the same score was achieved. The maximum length achieving the highest score (1.104) was at 22,037,352 bp, which is -3329 bp from the ATG start codon. Any upstream extension to this length was not improving the prediction score but rather decreasing it. Therefore, -3329 bp was considered as the longest length possible for the promoter region.

The first step was to create a longer version on to the -2044 bp promoter region, hence achieving -3329 bp. The PCR for this region proved to be poorly efficient to have a good concentration after purification, mostly due to the high AT rich area around that region. Hence the upstream region for the HRH4 was requested to be synthesised by Eurfoins MWG (shown in Chapter 2, Section 2.8). The co-ordinates for this region, according to NCBI (GRCh37.p10 primary assembly) are 22,037,352 – 22,040,203 bp, where the latter represents the area containing an Msc1 restriction site. This sequence was sent synthesised by Eurofins MWG into a standard vector, where Xho1 restriction site was introduced at the 5'-end and ended with the Msc1 restriction site at the 3'-end. This was named XLpUC57.

Initial thoughts for sub-cloning the synthesised DNA were to amplify the region using two standard primers found on the standard vector hosting the insert, pUC57 vector. Ligating this amplicon (after being digested with Xho1 and Msc1) was never successful, even though various conditions were tried, i.e., changing ligation time and temperature, and also trying different T4 ligases. Hence, this was changed to performing a restriction digest directly from the host vector (XLpUC57). Since the insert was synthesised beginning at the Msc1 restriction site on the 3'-end, and extending upstream to the Xho1 restriction site, it was possible to cut from both vectors and ligate them together. The first attempt failed to complete digestion on the XLpUC57. Following further investigation it was noticed that the reason why digestion was unsuccessful (the vector was only linearised) was because the first nucleotide from the pUC57 vector adjacent to the last nucleotide of the insert created a dcm^+ site when grown in TOP10 E.Coli. Most laboratory E.Coli strains have three common site-specific DNA methylases; Dam (Dam mehylase), SAM (S-adenosylmethionine) and Dcm (Dcm methylases). The Dcm methylates the internal cytosine residue in the sequence CCAGG and CCTGG (1,3) at the C5 position. The Msc1 restriction site present at the end of the insert was TGGCCA, and the nucleotide next to it (i.e. the start in pUC57) was GGG. Hence, the

combination of the sequence becomes TGG**CCAGG**G, where the second 'C' in CCAGG is methylated, and Msc1 is dcm sensitive hence cleavage is blocked by methylation. In order to overcome this non-cleavage of Msc1, the plasmid was grown in BL21(DE3). This bacterial strain is a Dcm^- strain which naturally lacks Dcm methylation. This proved successful in obtaining a complete cleavage at this restriction site and thus having the insert cut at both ends. However, once cut with Xho1 and Msc1, both the insert and vector produced a band approximately 2800 bp and 2700 bp respectively. It was important to separate bands by gel electrophoresis. To overcome this problem an extra restriction enzyme digest was added, BspH1. This enzyme cuts the vector (pUC57) in three sites but does not cut the insert. So the final product after digestion was easier to separate and hence to purify. Figure 4.77 shows an image of virtual triple digestion (Msc1, BspH1, Xho1) and the circled band is the insert desired for further work.

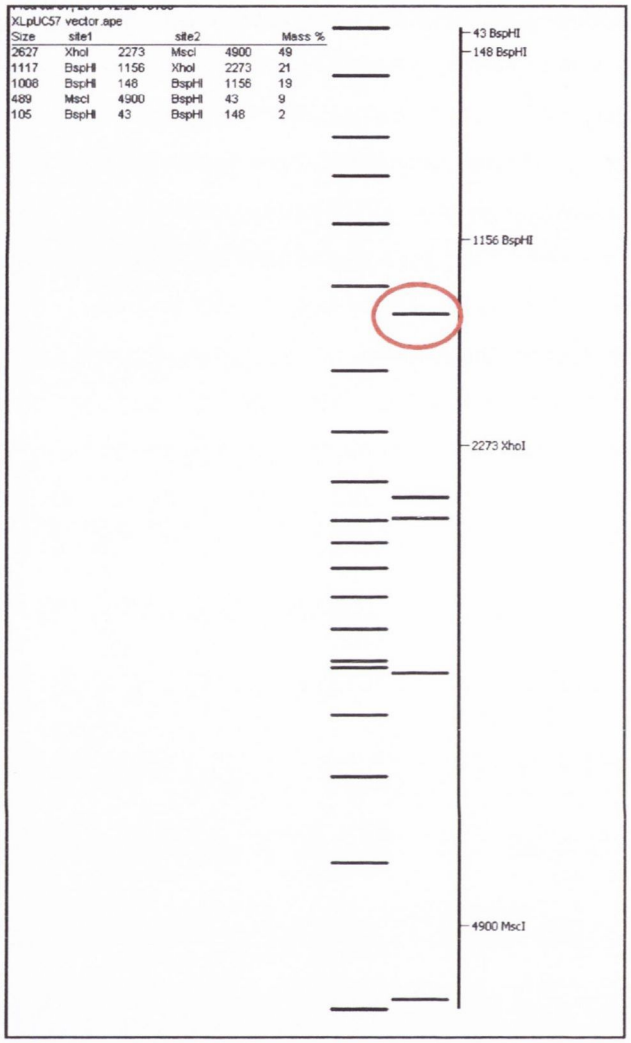


Figure 4.77 Virtual restriction enzyme digest for XLpUC57.

(The circled band represents the desired insert cut out from the host vector)

Using this insert, an attempt and ligate it into the 10RpGL3 (digested with Xho1 and Msc1) resulted in an unusual approach. Instead of having the 5'-end of the promoter attached to the digested clone, the 10RpGL3 re-ligated on itself using only 1 base pair for self-ligation. The Xho1 restriction site is C^TCGAC, while Msc1 restriction site is TGG^CCA. Hence, the sticky ends formed 'C' and 'CCA' which ligated creating 'CCCA'. This gave rise to -705 bp 10RpGL3, the shortest vector harbouring a potentially active promoter.

After achieving this short length of the promoter region, other attempts were made to try and insert the 2627 kb in the digested 10RpGL3. This was never successful after numerous attempts; hence a new approach was chosen using site-directed mutagenesis. This method included amplifying the region with specially designed primers; where the first 20 bp anneal to the final clone (10RpGL3) and the second 20 bp anneal to the desired insert to be added (in XLpUC57). The diagram show in Figure 4.78 represents the mechanistic approach for this mutation.

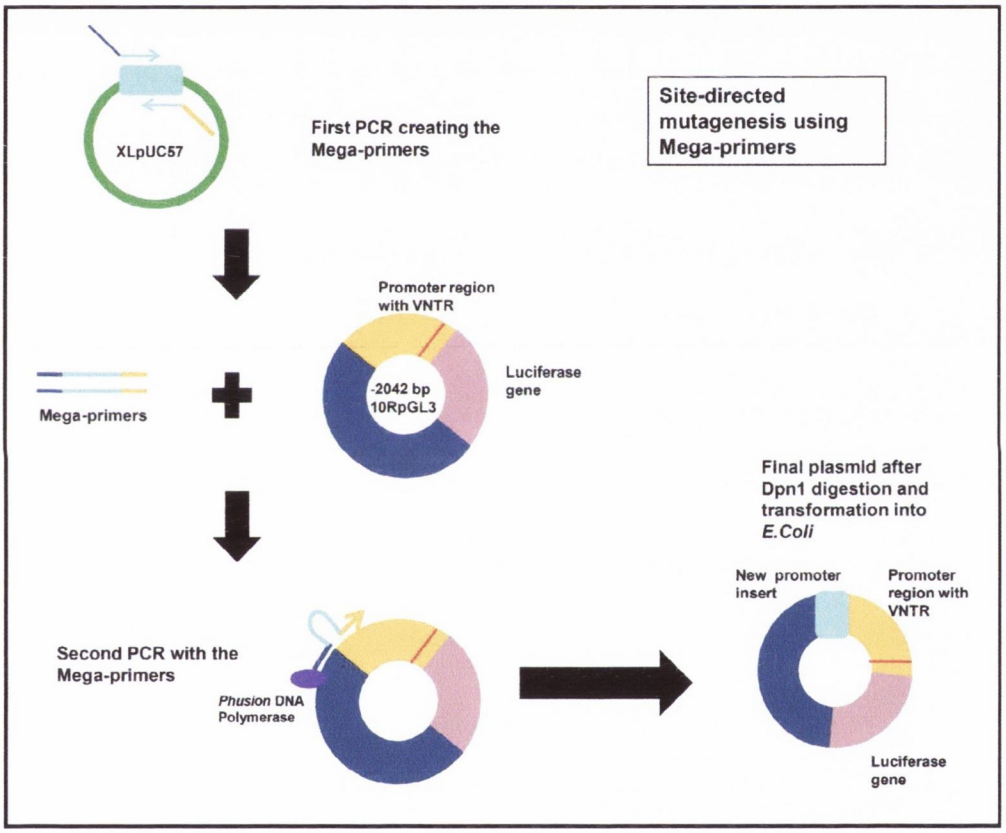


Figure 4.78 Site-directed mutagenesis mechanism using Mega-primers

All newly designed primers used to create mega-primers were analysed using bioinformatics database Primerstats (as mentioned in Chapter 2, Section 2.4), especially because since they are long primers, they can form hairpin formations or self-anneal. The primer pair which showed the more ambiguous statistics were the primer pair for XL SDM_For and

XL SDM_Rev, where the forward primer has more than 3 self-annealing base pairs and the reverse had more than 50% of the bases that self-anneal. However, since the PCR with these primers is performed on a plasmid DNA and the efficiency of the PCR should be sufficiently high the reaction was still successful with high concentration of mega-primer achieved.

This method of inserting a large sequence of DNA in the cloned vectors was employed to obtain all the lengths required to proceed in the luciferase assay for promoter identification. In parallel to this promoter extension, exon 1 was also added to these promoter lengths at the end of the transcription site. The sequence of exon 1 was also taken from the NCBI database starting at +1 to +194 bp. These constructs are used further in the next chapter and tested on the cell-based assay for promoter activity.

4.4 Conclusion

A plasmid DNA library encoding the putative regulatory of site the HRH4 was created through the use of molecular biology techniques. The polymorphic promoter constructs in the reporter plasmid will be further analysed in more detail in Chapter 5, where they are used in a reporter gene assay to identify a possible impact on gene expression and regulation. Through various studies, researchers have hypothesised that inappropriate repeat variation in coding regions could result in toxicity, incorrect folding or aggregation of a protein. In non-coding regions however, such repeats may be associated with increased instability or fragility of the DNA or in altering the binding properties of DNA binding proteins involved in DNA metabolism or transcription [64, 120]. Further analysis on this VNTR region in a regulatory site of the receptor can shed light on its characteristics and uncover possible functional implications on the protein expressed.

Chapter 5

Promoter analysis for HRH4

5 Promoter analysis for HRH4

Identification of the promoter region of a gene is of biological significance since transcriptional control can have direct implications on human health. Recent publications on the histamine receptors identified the decreased expression of HRH1 and HRH4 in colorectal tumours, where the mRNA for these two receptors was significantly reduced when compared to healthy controls [121]. One of the latest studies on HRH4 show that it is highly expressed in keratinocytes in patients suffering from atopic dermatitis and its stimulation might contribute to epidermal hyperplasia noticed in atopic dermatitis patients [122].Promoter prediction is based on transcription binding sites where algorithms are used in order to predict promoter length. However, no matter how extensive these predictions are, they can still vary from biological data, especially since not all identified binding sites are confirmed experimentally [123].

As mentioned in Chapter 4 (Section 4.3), the length of the promoter was postulated by using a bioinformatics database Promoter 2.0, where the highly predictable area was taken from the maximum value achieved in relation to the length of the promoter. The -2044 bp 10RpGL3 was used as the main promoter and changes were introduced to investigate different length variations. The putative promoter length described by Coge et al [7] shows major binding sites for transcription factors, including the CAAT and TATA box, where the CAAT box plays a role in promoter efficiency and the TATA box binds RNA polymerase II as well as a series of transcription factors to form an initiation complex for gene transcription [119]. The transcription binding sites predicted by this research group also describe an α-INF binding site at the end of the VNTR region 'CAC AAA TGA' for the 10R genotype, variations in this region could potentially affect the binding of this transcription factor. In Chapter 3, two other potential variations were described for the ending of the VNTR, i.e.,' CAA CAA TGA' and ' CAA AAA TGA' for genotypes with longer triplet repeats. Concentration of α-INF is affected by inflammatory conditions and its binding of this transcription factor to HRH4 is in turn affected by the genetic disposition of a patient, this HRH4 might contribute to severity of the disease and /or can serve as a genetic marker.

Based on this idea, the basic components for the various genotypes of the potentially functional active promoter were cloned into the reporter vector and activity of the promoter

was determined. This included the dual-luciferase assay on a mammalian cell line and measuring the activity of the gene transcribed. In order to identify the full promoter region of the HRH4, other lengths were cloned into the reporter vector. The putative promoter region was introduced before the ATG start codon on firefly luciferase and transiently transfected into a mammalian cell line, hence transcription of the protein is recorded as a light signal and represents activity of the promoter. The renilla luciferase, pRL-SV40, was used as an internal control which contains the SV40 enhancer elements and promoter and it is used as a normalising vector for transfection and luciferase activity. Optimisation for co-transfection between vector DNA and renilla ratio is required in order to ensure independent genetic expression between experimental and control reporter genes.

5.1 Methods

5.1.1 Promoter study for HRH4

Initially the promoter analysis was planned to be performed using a neuronal cell line, hence an aliquot for SH-SY5Y, a neuronal cell line originating from neuroblastoma was cultured. Optimisation for the luciferase assay required transfecting cells with pGL3-SV40 and pRL-SV40 (the control vectors), but bioluminescence readings on this cell line failed to be consistent for control transfection. Hence, the cell line was changed to cells which are easier to transfect, HEK293T. Both cell-based assays will be discussed.

5.1.1.1 Cell culture

SH-SY5Y cells were obtained from a research group from Dr Hans-Georg Koenig, Centre for the Study of Neurological Disorders (CSND), Department of Physiology and Medical Physics, Royal College of Surgeons in Ireland. Cells were cultured in a T75 flask at a seeding density of 75,000 cells/cm^2. Cells were grown in Dulbecco's modified Eagle's medium/Ham's F-12 (1:1 mix) (DMEM/F12) with 5 mM L-Glutamate, supplemented with 15% foetal bovine serum (FBS), 1% penicillin/streptomycin, and grown at 37°C in 5% CO_2 atmosphere.

HEK293T cells were obtained from the School of Biochemistry and Immunology, Trinity Biomedical Sciences Institute, Trinity College Dublin, Dublin 2, Ireland. Cells were cultured at a seeding density of 75,000 cells/cm^2 and were grown in Dulbecco's modified Eagle's medium/Ham's F-12 (1:1 mix) (DMEM/F12) supplemented with 10% foetal bovine serum (FBS) and 5 mM L-Glutamate. Cells were maintained at 37°C in 5% CO_2 atmosphere and the medium was exchanged every 48 hours until a confluent monolayer was reached after 5-7 days.

All solutions and equipment that came in contact with the cells were sterile using proper sterile technique and working in a laminar flow hood.

5.1.1.2 Preparation of DNA for transfection

The cloned reporter plasmids had to be re-grown in a larger volume and isolated in high purity from *E.Coli* using EndoFree Plasmid Maxi kit (Qiagen). A single colony was picked from a freshly streaked selective plate and inoculated a starter culture of 3 mL LB medium containing ampicillin. They were incubated for approximately 8 hours at 37°C with vigorous shaking (300 rpm). The starter culture was diluted by 1/1000 into LB Amp^+ medium, inoculated in 50 mL medium with 100 µl of starter culture (250 mL flasks), and grown at 37°C for 12–16 h with vigorous shaking (300 rpm).

The bacterial cells were harvested by centrifugation at 4000 x *g* for 15 min at 4°C. The bacterial pellet was resuspended in 10 mL buffer P1. The volume of 10 ml buffer P2 was added, mixed thoroughly, and incubated at room temperature (15–25°C) for 5 min. The reaction was not allowed to proceed more than 5 min. During the incubation the QIAfilter® Cartridge was prepared by screwing the cap onto the outlet nozzle of the QIAfilter® Maxi Cartridge and placing the QIAfilter® Cartridge in a 50 mL tube. Pre-chilled buffer P3 was added (10 mL), and mixed immediately. The lysate was poured into the barrel of the QIAfilter® Cartridge and incubated at room temperature (15–25°C) for 10 min. After incubation, the cap was removed from the QIAfilter® Cartridge outlet nozzle. Gently the plunger was inserted into the QIAfilter® Maxi Cartridge and the cell lysate was filtered into a 50 mL tube. A volume of 2.5 mL buffer ER was added to the filtered lysate, mixed by inverting the tube approximately 10 times, and incubated on ice for 30 min. After the addition of Buffer ER the lysate appears turbid, but becomes clear again during the incubation on ice.

QIAGEN-tip® 500 was equilibrated by applying 10 mL buffer QBT, and allowed the column to empty by gravity flow. The filtered lysate was applied to the QIAGEN-tip® and allowed to enter the resin by gravity flow.

The QIAGEN-tip® 500 was washed with 2 x 30 mL buffer QC. The DNA was eluted with 15 mL buffer QN in endotoxin-free/pyrogen-free plastic ware. The DNA was precipitated by adding 10.5 mL (0.7 volumes) room-temperature isopropanol to the eluted DNA. This was mixed and centrifuged immediately at 4,000 x *g* for 60 min at 4°C. Carefully the supernatant was decanted. The DNA pellet was washed with 5 mL of endotoxin-free room-temperature

70% ethanol and centrifuged at 4,000 x *g* for 60 min at 4°C. Carefully, the supernatant was decanted without disturbing the pellet.

The pellet was air-dried for 5-10 min and redissolved in 50 µL endotoxin-free buffer TE. The concentration and purity were determined by using NanoDrop™ spectrophotometer (section 4.1.1.2). The samples were stored at -20°C until further use.

5.1.1.3 Seeding and transfection

Seeding and transfection were performed following different protocols for the two cell lines. Cationic transfection was performed by using Lipofectamine2000, and the conventional approach recommended by manufacturers was adopted.

SH-SY5Y cell line

SH-SY5Y cells were seeded in a 96-well white plate. This was conducted by removing the cell medium and washing the cells twice with PBS and addition of 3 mL pre-warmed trypsin (1x). The cells were incubated at 37°C for approximately 5 min until the monolayer was detached. Trypsin was neutralised by the addition of 7 mL culture medium to the 75T flask. From the cell suspension, a small volume (approximately 100 µL) was removed to determine the total cell number by using a hemacytometer. The cells were centrifuged at 900 rpm for 5 min. The supernatant was removed, and fresh new cell medium without P/S was added to the cell pellet, where a cell suspension was obtained. The cell suspension was diluted to reach the required seeding density (3.5 x 10^4 cells/well) and 100 µL were pipetted into each well. The cells were incubated for 24 hours at 37°C in 5% CO_2 atmosphere.

An amount of 0.5 µL Lipofectamine 2000 was added into 25 µL Opti-MEM, and in a different centrifuge tube 0.15 µg of pGL3 constructs and 0.05 µg of pRL-SV40 were added to 25 µL Opti-MEM. These were left to incubate for 10 min at RT. The diluted lipofectamine was then added to the diluted plasmids and left for 20 min at RT to form complexes. Following that, 50 µL of the complexes was added to the cells and left to incubate at 37°C in 5% CO_2 atmosphere for 24 hours.

HEK293T cell line

Transfection for this cell line was slightly different, where the cells were seeded and transfected on the same day.

Centrifuge tubes were pre-labelled for diluted lipofectamine and diluted DNA. Lipofectamine was diluted by adding 0.5 µL into 25 µL Opti-MEM. Similarly the plasmids were diluted by using 0.15 µg of pGL3 constructs and 0.05 µg of pRL-SV40 and diluting them into 25 µL Opti-MEM. Both dilutions were left to incubate for 10 min at RT.

Cell culture medium from the culture vessel was removed and 3 mL pre-warmed trypsin (1x) was added. The flask was gently rocked until complete coverage of the cell monolayer. The cells were incubated at 37°C for approximately 8-10 min until the cells were detached. During this incubation period, the diluted lipofecatmine was added to the diluted plasmids and left to form cationic complexes for 20 min at RT.

Once the cells were detached, 7 mL pre-warmed medium was added to the 75T flask to neutralise the trypsin and create a cell suspension. A small volume (approximately 100 µL) was removed to determine the total number of cells using a hemacytometer. The cell suspension was transferred into a 15 mL centrifuge tube and spun at 900 rpm for 5 min. From the cell suspension (prior of spinning), 200 µL of cells were added to a new 75T flask, where fresh medium was added up to a volume of 10 mL and left to grow in the incubator for future experiments. After spinning the cell suspension, the medium was removed leaving a cell pellet at the bottom of the flask. This was resuspended into fresh medium, where an aliquot was taken and diluted to obtain the desired seeding density (2.5×10^4 cells/well) in a fresh tube. From this suspension, 100 µL was piptetted into each well of a 96 well plate. After 20 min incubation period, 50 µL of the lipofectmaine-DNA complex was added to each well. The plate was gently rocked to mix the seeded cells and incubated at 37°C in 5% CO_2 atmosphere and allowed to grow for 48 hours.

5.1.1.4 Dual-Glo luciferase assay

The transfected cells were seeded and transfected 24 hrs for SH-SY5Y or 48 hrs for HEK293T before performing the assay. 75 µL of the medium were removed in order to obtain a 1:1 ratio between medium and Luciferase Assay® substrate. The Stop and Glo® reagent was prepared fresh before the assay, where Stop and Glo® substrate was diluted 1:100 with its respective buffer, while the Luciferase Assay® substrate was allowed to equilibrate and reach room temperature. The cells were lysed with 75 µL Luciferase Assay® reagent per well and incubated for 10 minutes. The plate was read using FLUOstar OPTIMA (BMG Labtech) for emission*. After reading the plate, 75 µL of Stop and Glo® was added to each well and incubated for 10 minutes. Emission of renilla was read and recorded. The transcriptional activity of the promoter constructs were normalised with the renilla activities derived from pRL-SV40.

**Initial optimisation of the assay were carried out using Synergy HT (BioTek).as the bioluminescence plate reader. The rest of the experiments were then read on FLUOstar OPTIMA (BMG Labtech).*

5.1.1.5 Statistical analysis

The statistical significance of the difference in luciferase activity between each promoter constructs and control vector (pGL3-basic) was evaluated using a one-way ANOVA, followed by a Dunnett multiple comparison test, with $P < 0.05$ considered significant. Statistics were performed using GrapPad Prisim 5 version 5.01, where the *P* value is symbolised as follows:

P value	Wording	Summary
< 0.001	Extremely significant	***
0.001 to 0.01	Very significant	**
0.01 to 0.05	Significant	*
>0.05	Not significant	ns

These asterisks denoted above are used respectively in all graphs plotted.

5.2 Results

5.2.1 Dual-glo luciferase assay optimisation

Initial experiments and optimisation were carried out using pGL3-SV40 and pRL-SV40 on SH-SY5Y cell line. The clones for genotypes 10R and 13RA were also being tested to investigate if this was the functional promoter sequence as suggested by Coge *et al* [7]. Hence, the first samples to be purified by Maxi-prep were: -2044 bp 10RpGL3, -2044 bp 13RpGL3, pGL3-basic, pGL3-SV40 and pRL-SV40. Purification with this Qiagen Maxi-prep kit results in a higher plasmid concentration which is pyrogen and endotoxin free, suitable for transfection into mammalian cells. The resulting concentration is shown in Table 5.1.

Table 5.1 Endo-free maxi prep for optimisation assays

Sample ID	ng/uL	A260	260/280	Average ng/µL
-2044 bp 10RpGL3	4953.20	99.06	1.60	4966.00
	4956.38	99.13	1.58	
	4987.15	99.74	1.58	
-2044 bp 13RApGL3	5138.24	102.77	1.51	4950.79
	4759.65	95.17	1.68	
	4954.48	99.09	1.60	
pGL3-basic	3207.29	64.15	2.28	3287.84
	3339.07	66.78	2.27	
	3317.17	66.34	2.28	
pGL3-SV40	4836.55	96.73	1.96	4877.77
	4836.62	96.73	1.97	
	4960.14	99.20	1.95	
pRL-SV40	3493.07	69.86	2.26	3434.6
	3372.91	67.46	2.27	
	3437.83	68.76	2.25	

For initial experiments, two different ratios were tested for the luciferase:renilla vectors. The renilla is an asset to the experiment as it represents an internal control for transfection and normalising efficiency. Hence the reading for renilla should not be higher than the luciferase but high enough to give a significant reading to be used in the final calculations.

Controls for the experiment included:

- ✓ Untransfected cells
- ✓ Cells treated with lipofectamine without vectors
- ✓ Cells transfected with pGL3-SV40 only
- ✓ Cells transfected with pRL-SV40 only

These controls were used for all luciferase experiments.

The ratios tried were 10:1 and 3:1 for luciferase:renilla vectors. As example, one set of data is shown Table 5.2.

Table 5.2 Luciferase assay optimisation in SH-SY5Y (1)

Constructs	Luciferase	Renilla	L/R	SD	Average L/R	Normalised fold activity
pGL3-SV40 (10:1)	74	944	0.078	0.210	0.318	**750.94**
	156	388	0.402			
	335	709	0.472			
pGL3-SV40 (3:1)	491	4469	0.110	0.018	0.089	**209.96**
	291	3731	0.078			
	503	6401	0.079			
-2044 bp 10RpGL3 (3:1)	12	3678	0.003	0.001	0.003	**7.97**
	9	1923	0.005			
	6	2770	0.002			
-2044 bp 13RApGL3 (3:1)	9	1073	0.008	0.003	0.005	**12.51**
	3	1205	0.002			
	21	4199	0.005			
pGL3-basic (3:1)	6	9479	0.001	0.000	0.000	**1.00**
	6	20471	0.000			
	6	17457	0.000			
pGL3-SV40 (alone)	2834					
	1614					
	5375					
pRL-SV40 (alone)		43599				
		62025				
		39826				

Running these controls and samples on dual-glo luciferase assay gave controversial results. The controls for the experiment indicated that transcription worked for both luciferase and renilla, where the luciferase activity on its own gave an average reading of 3274 light units and renilla activity gave an average reading of 48483 light units.

Looking at the data obtained from the samples analysed gave a different picture. The luciferase activity appeared to be significantly lower than the renilla control and taking into account that the concentration of pGL3-SV40 is significantly higher than pRL-SV40 this prompted doubt in transfection efficiency. It is important to emphasise that both of these vectors are driven by the same viral promoter (SV40), thus making this difference in activity ambiguous. Figure 5.1 shows normalised data.

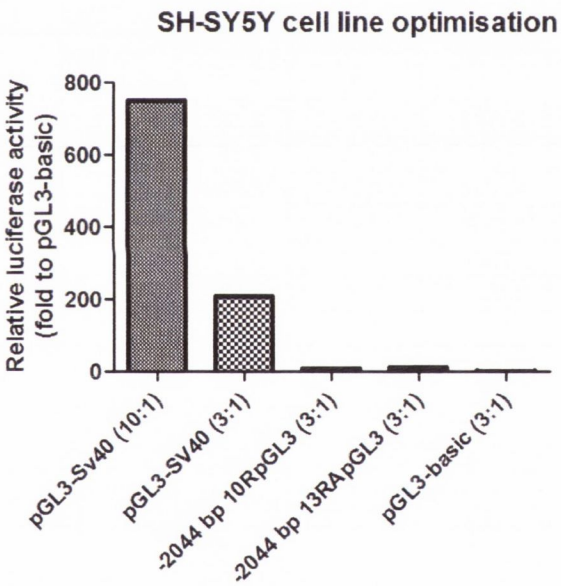


Figure 5.1 Luciferase activity for optimisation in SH-SY5Y cells

This was repeated a couple more times and the resulting reading varied, where in some cases even negative data for luciferase was obtained. It is already known that transfection in a neuronal cell line, especially post-mitotic neurons, is challenging and unreliable in some cases [124, 125], and since this assay requires transfection of two vectors it is even more

challenging for the cells to take up the two vectors. Hence, a different cell line was used as a comparison of reliability and transfection efficiency.

The two pGL3-SV40:pRL-SV40 ratios were again varied and the results were compared to previously obtained data from SH-SY5Y. Resulting data from this assay are shown in Table 5.3 and Table 5.4 for SH-SY5Y and HEK293T cells respectively.

Table 5.3 Luciferase activity for SH-SY5Y optimistation (2)

Constructs	Luciferase	Renilla	L/R	SD	Average L/R	Normalised fold activity
pGL3-SV40 (10:1)	4510	1114	4.048	2.589	6.790	**53.88**
	5075	712	7.128			
	10673	1161	9.193			
pGL3-SV40 (3:1)	8145	3464	2.351	0.322	2.531	**20.04**
	10091	4313	2.340			
	10438	3596	2.903			
pGL3-basic (3:1)	206	1644	0.125	0.007	0.126	**1.00**
	438	3658	0.120			
	603	4504	0.134			

Table 5.4 Luciferase activity for HEK293T optimisation

Constructs	Luciferase	Renilla	L/R	SD	Average L/R	Normalised fold activity
pGL3-SV40 (10:1)	20980	679	30.898	8.661	27.423	**3427.50**
	21270	1211	17.564			
	28128	832	33.808			
pGL3-SV40 (3:1)	38870	5598	6.944	1.079	7.363	**920.37**
	43740	6671	6.557			
	33660	3919	8.589			
pGL3-basic (3:1)	388	49963	0.008	0.001	0.009	**1.00**
	570	68201	0.008			
	597	60889	0.010			

This assay identified a difference in transfection activity and transcription. Table 5.3 shows that this time the luciferase activity was higher than the renilla activity in SH-SY5Y cells. However when the two cell lines are compared (Table 5.3 and Table 5.4), the luciferase activity in HEK293T is approximately 3 fold for 10:1 and 4 fold higher in 3:1 than SH-SY5Y and this is also represented in Figure 5.2.

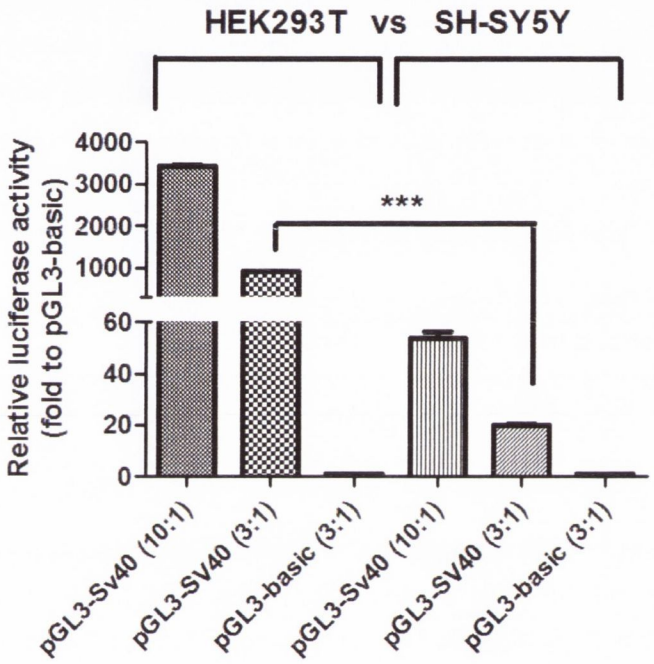


Figure 5.2 Optimisation of luciferase assay

Luciferase assay in HEK293T and SH-SY5Y cells with two ratios for control vectors where on comparing HEK293T vs SH-SY5Y cell lines in a 3:1 ratio pGL3-SV40:pRL-SV40 $P < 0.001$

Looking at the luciferase transcription for the two cells lines in 3:1 pGL3-SV40:pRL-SV40 ratio, a considerable difference of activity is noted (Figure 5.2). The same control vector in the 3:1 ratio gives a stronger signal when transfected in HEK293T cells when compared with SH-SY5Y cells. This control vector (pGL3) is driven by the same promoter (SV40) therefore activity should show a very strong significant difference. Indeed, using one-way ANOVA

followed by Dunnett's test, the difference between HEK293T and SH-SY5Y cells shows a strong significant difference of $P < 0.001$.

The objective for this study is to identify the functional sequence of the promoter and the measurement of luciferase activity is of crucial importance for analysis on the clones in pGL3 described in chapter 4, especially since it is not known how strong the HRH4 promoter is. Hence it was decided to conduct the Dual-Glo® luciferase assay on this cell line. The SV40 promoter is known to be a very strong promoter, and a signal of transcription was achieved even with a poor transfection efficiency cell line (SH-SY5Y). Since the promoter activity of the HRH4 is not known yet, it was essential that even a low level of bioluminescence detection was read and recorded.

5.2.2 Gene reporter assay for HRH4 promoter

5.2.2.1 Preparation of samples – plasmid purification

The plasmids, constructed in chapter 4 (Section 4.1.2), where re-grown in larger volumes in order to increase quality and concentration. The plasmids were also required to be endotoxin and pyrogen free (as mentioned in Section 5.2.1). Hence all plasmids (including controls) were purified using EndoFree Plasmid Maxi kit (Qiagen). Table 5.5 and Table 5.6 show the final concentrations obtained after purification from bacterial pellets.

Table 5.5 EndoFree plasmid purification for clones without exon 1 and controls

Sample ID	ng/uL	A260	260/280	Average ng/µL
-705 bp 10RpGL3	3008.24	60.17	1.83	3120.23
	3181.60	63.63	1.82	
	3170.84	63.42	1.82	
-2044 bp 10RpGL3	4953.20	99.06	1.60	4966.00
	4956.38	99.13	1.58	
	4987.15	99.74	1.58	
-2772 bp 10RpGL3	4595.08	91.90	1.68	3405.15
	3048.63	60.97	1.83	
	2571.73	51.44	1.84	
-3329 bp 10RpGL3	128.48	2.57	1.95	113.92
	114.51	2.29	1.97	
	98.78	1.98	1.94	
-2044 bp 13RApGL3	5138.24	102.77	1.51	4950.79
	4759.65	95.17	1.68	
	4954.48	99.09	1.60	
pGL3-basic	3207.29	64.15	1.82	3287.84
	3339.07	66.78	1.81	
	3317.17	66.34	1.82	
pGL3-SV40	3792.48	75.85	1.79	3831.00
	3794.39	75.89	1.79	
	3905.38	78.11	1.78	
pRL-SV40	1907.05	38.14	1.82	1744.03
	1680.96	33.62	1.84	
	1644.07	32.88	1.85	

The typical range of concentration for the maxi-prep was between 3 – 4 µg/µL (Table 5.5), where the renilla concentration was slightly different at a lower concentration (1.7 µg/µL). This did not affect adversely the transfection complexes since it is used in a lower concentration (0.05 µg) compared to the pGL3 vectors (0.15 µg). However, the other variable was the final concentration of the -3329 bp 10RpGL3 were the resulting concentration was 113.92 ng/µL. This resulted in insufficient DNA for a triplicate transfection for this promoter length. Hence the reading mentioned in section 5.2.2.2 was only performed once. The table below (Table 5.6) however shows a good concentration for all clones containing exon 1, where the average is around 2 µg/µL.

Table 5.6 EndoFree plasmid purification for clones with exon 1 and controls

Sample ID	ng/uL	A260	260/280	Average ng/µL
-705 bp Exon1 10RpGL3	2410.00	48.20	1.82	2428.73
	2470.29	49.41	1.82	
	2405.91	48.12	1.82	
-2044 bp Exon 1 10RpGL3	2629.22	52.58	1.80	2526.58
	2423.94	48.48	1.82	
	3570.71	71.41	1.79	
-2772 bp Exon 1 10RpGL3	2317.76	46.36	1.85	2358.88
	2304.97	46.10	1.85	
	24.53.92	49.08	1.85	
-3329 bp Exon 1 10RpGL3	2784.10	55.68	1.82	2645.76
	2593.51	51.87	1.81	
	2559.68	51.19	1.81	
-2044 bp Exon 1 13RpGL3	2792.45	84.18	1.73	2645.76
	2791.31	55.83	1.82	
	2768.35	55.37	1.81	
pGL3-basic	1769.49	35.39	1.87	1787.56
	1776.36	35.53	1.87	
	1816.84	36.34	1.87	
pGL3-SV40	3051.66	61.03	1.83	2862.00
	3748.55	54.97	1.83	
	2785.32	55.71	1.83	
pRL-SV40	4119.66	82.39	1.76	4045.40
	4373.27	87.47	1.72	
	3643.26	72.87	1.80	

Once the plasmids were purified, cells were seeded and transfected. As described in during optimisation (Section 5.2.1), the human embryonic kidney cell line, HEK293T, was used for this assay to perform all analyses.

5.2.2.2 Dual-Glo luciferase assay results

The first assay was run on the set of clones that had no exon 1 attached at the 3'-end of the promoter region. Cells were lysed with Luciferase® assay substrate and bioluminescence was read and recorded. Following that, Stop & Glo® (renilla) substrate was added and the reading was recorded. All values were normalised with the renilla values (L/R) and the fold activity was calculated by dividing the average of each promoter (L/R) with the lowest value (least active promoter based on average (L/R)), hence used as the baseline of activity and given the value of 1.00 as the fold activity. All other promoter constructs were compared with the least active construct; which was -2772 bp 10RpGL3. The controls for this assay were the pGL3-basic and pGL3-SV40, where the fold activity was compared to the promoterless vector (pGL3-basic) and the pGL3-SV40 was used as a positive control for the assay since it gives a high luciferase activity. The assay was performed in triplicate. The results for luciferase activity are shown in Table 5.7.

Table 5.7 Luciferase assay activity for constructs without exon 1

Constructs	Luciferase	Renilla	L/R	SD	Average L/R	Normalised fold activity
-3329 bp 10RpGL3	199	230	0.865	0.048	**0.868**	**11.27**
	188	205	0.917			
	194	236	0.822			
-2772 bp 10RpGL3	385	5528	0.070	0.008	**0.078**	**1.00**
	368	4318	0.085			
	400	5143	0.078			
-2044 bp 10RpGL3	643	1320	0.487	0.105	**0.472**	**6.09**
	630	1750	0.360			
	594	1044	0.569			
-705 bp 10RpGL3	480	1609	0.298	0.004	**0.294**	**3.81**
	465	1585	0.293			
	478	1647	0.290			
pGL3-basic	1517	3477	0.436	0.041	**0.431**	**5.56**
	1594	4108	0.388			
	1495	3190	0.469			
pGL3-SV40	5610	636	8.821	1.192	**9.047**	**116.73**
	4196	406	10.335			
	5030	630	7.984			

From these readings, it shows that the -3329 bp construct has the highest activity, however it is also noted that the renilla control vector was not giving a strong signal resulting in a high normalising ratio. In order to correct this, an average on all the renilla readings was taken and used to normalise the luciferase activity for -3329 bp 10RpGL3. The final fold activity is shown below, where the -3329 bp is the least active construct and this was considered as the baseline of activity where all the other promoter constructs are compared with.

Table 5.8 Final fold activity for promoter constructs without exon 1

Constructs	Luciferase	Renilla	L/R	SD	Average L/R	Normalised fold activity
-3329 bp 10RpGL3	199	2893	0.069	0.002	**0.067**	**1.00**
	188	2893	0.065			
	194	2893	0.067			
-2772 bp 10RpGL3	385	5528	0.070	0.008	**0.078**	**1.15**
	368	4318	0.085			
	400	5143	0.078			
-2044 bp 10RpGL3	643	1320	0.487	0.105	**0.472**	**7.05**
	630	1750	0.360			
	594	1044	0.569			
-705 bp 10RpGL3	480	1609	0.298	0.004	**0.294**	**4.39**
	465	1585	0.293			
	478	1647	0.290			
pGL3-basic	1517	3477	0.436	0.041	**0.431**	**6.44**
	1594	4108	0.388			
	1495	3190	0.469			
pGL3-SV40	5610	636	8.821	1.192	**9.047**	**135.03**
	4196	406	10.335			
	5030	630	7.984			

The data is shown as average normalised value with the respective standard deviation for the triplicate. After this normalisation operation, -3329 bp 10RpGL3 has a smiliar low L/R ratio to -2772 bp 10RpGL3. The highest activity of all constructs was obtained from the -2044 bp 10RpGL3, which is the promoter length described by Coge *et al* [7]. However, this clone failed to reach statistical significance (NS) when compared to the promoterless control (pGL3-basic). However, when compared to -705 bp 10RpGL3 (which is the second most active promoter for this set of clones), it reaches a strong significant difference of $P < 0.001$ to

0.01 when analysed with one-way ANOVA followed by Dunnett's test. This is shown in Figure 5.3. The controls for this set of experiments also show that the luciferase transcription was successful as the pGL3-SV40 reached a strong significant difference from the pGL3-basic ($P < 0.001$) (Figure 5.3). It is also interesting to note, that the renilla value for this highly active promoter is not as strong as for the other vectors.

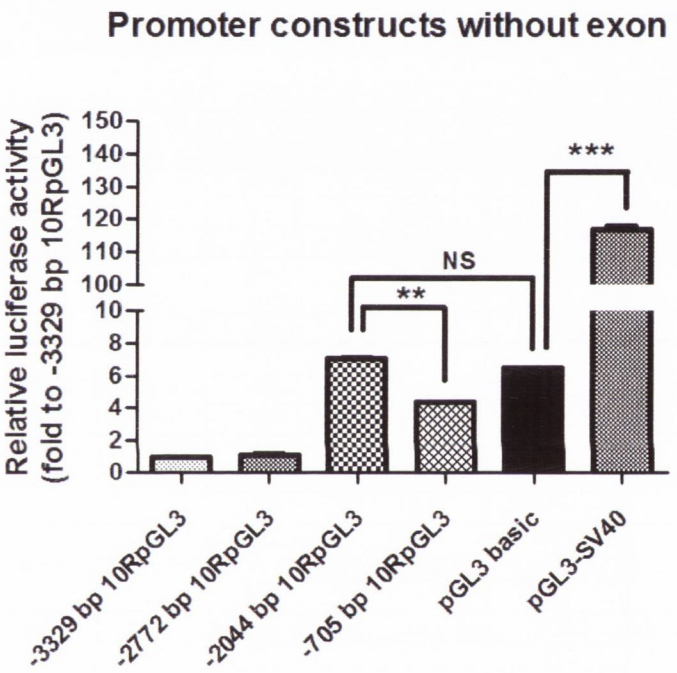


Figure 5.3 Promoter activity for 10RpGL3 construct without exon 1

Luciferase activity for promoter constructs without exon 1 where -2044 bp 10RpGL3 is not significantly (NS) different to pGL3-basic and Controls where pGL3-SV40 is strongly significant ($P < 0.001$) then pGL3-basic. Luciferase activity was normalised with the renilla level driven by SV40 promoter. Data is shown as the mean +/- SD where none of the constructs reached statistical significance when compared to the promoterless control, pGL3-basic.

The promoter constructs including exon 1 of HRH4 were run on the luciferase assay. The bioluminescence readings were recorded and the assay was performed in triplicate. The data obtained is shown in the table below (Table 5.9). This assay shows a very strong signal for one of the promoters (-705 bp) indicating a high activity. The bioluminescence shows that the shortest promoter construct, -705 bp Ex 1 10RpGL3, leads to a drastic increase in luciferase activity.

Table 5.9 Luciferase assay activity for constructs with exon 1

Constructs	Luciferase	Renilla	L/R	SD	Average L/R	Normalised fold activity
-3329 bp Ex1 10RpGL3	829	3841	0.216	0.037	**0.197**	**3.65**
	705	4563	0.155			
	622	2807	0.222			
-2772 bp Ex1 10RpGL3	225	4965	0.045	0.010	**0.054**	**1.00**
	242	4705	0.051			
	217	3353	0.065			
-2044 bp Ex1 10RpGL3	541	5255	0.103	0.010	**0.092**	**1.73**
	529	6194	0.085			
	503	5738	0.088			
-705 bp Ex1 10RpGL3	3781	276	13.699	0.265	**13.395**	**252.83**
	3186	240	13.275			
	3197	242	13.211			
pGL3-basic	1705	3073	0.555	0.023	**0.550**	**10.19**
	1761	3085	0.571			
	1560	2970	0.525			
pGL3-SV40	22735	3616	6.287	0.885	**6.041**	**112.26**
	10267	1515	6.777			
	11693	2311	5.060			

As seen in Figure 5.4, the shortest length sequence (-705 bp) shows a strong significant activity when compared to the strong viral promoter SV40, where $P < 0.001$. However, when taking a closer look, the renilla value is again lower than expected, especially when compared to the other renilla control readings (Table 5.9). This reading was adjusted to the average reading obtained from all the other promoter constructs and controls for renilla bioluminescence.

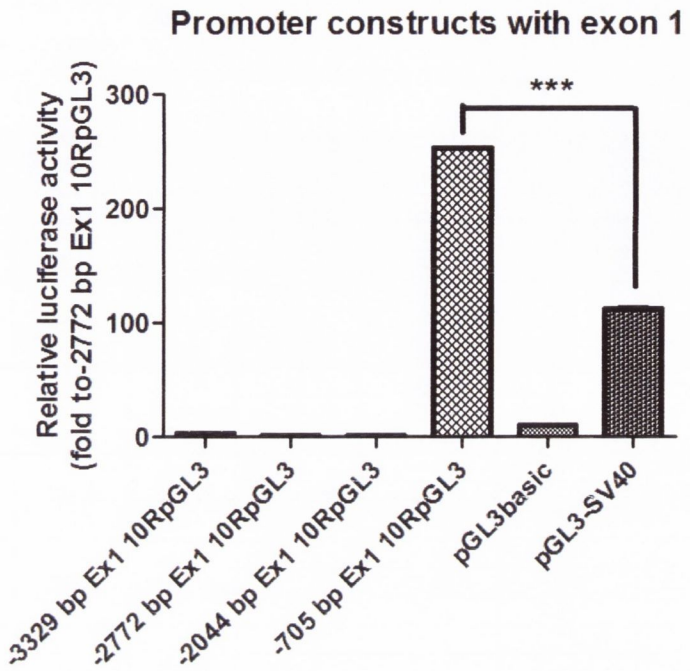


Figure 5.4 Promoter constructs with for 10RpGL3 with exon 1 with pGL3-SV40

*Where -705 bp Ex1 10RpGL3 shows a significant transcription activity (shown as mean +/- SD) when compared to pGL3-SV40 where ***P < 0.001*

When the average renilla value (3484 light units) was included for the -705 bp 10RpGL3, it still showed a high fold activity when compared to the other clones. The lowest value, i.e., transcriptional activity, for -2772 bp 10RpGL3 vector construct, was considered the baseline of activity. All fold activity values were compared with this baseline value, and this is shown in Table 5.10

Table 5.10 Final fold activity for promoter constructs with exon 1

Constructs	Luciferase	Renilla	L/R	SD	Average L/R	Normalised fold activity
-3329 bp Ex1 10RpGL3	829	3841	0.216	0.037	**0.197**	**3.65**
	705	4563	0.155			
	622	2807	0.222			
-2772 bp Ex1 10RpGL3	225	4965	0.045	0.010	**0.054**	**1.00**
	242	4705	0.051			
	217	3353	0.065			
-2044 bp Ex1 10RpGL3	541	5255	0.103	0.010	**0.092**	**1.73**
	529	6194	0.085			
	503	5738	0.088			
-705 bp Ex1 10RpGL3	3781	3484	1.085	0.098	**0.972**	**18.7**
	3186	3484	0.914			
	3197	3484	0.918			
pGL3-basic	1705	3073	0.555	0.023	**0.550**	**10.19**
	1761	3085	0.571			
	1560	2970	0.525			
pGL3-SV40	22735	3616	6.287	0.885	**6.041**	**112.26**
	10267	1515	6.777			
	11693	2311	5.060			

The data was analysed using one-way ANOVA followed by a Dunnett's test for comparison between the clones and the control vector (pGL3-basic). After this normalising procedure, it is shown that the short construct, i.e. -705 bp Ex1 10RpGL3 is not stronger active than pGL3-SV30. However when compared to the promoter vector, pGL3-basic, it still shows to exhibit a considerable level of luciferase activity where it was found to be strongly significant with a $P < 0.001$ when compared to the blank control (Figure 5.5).

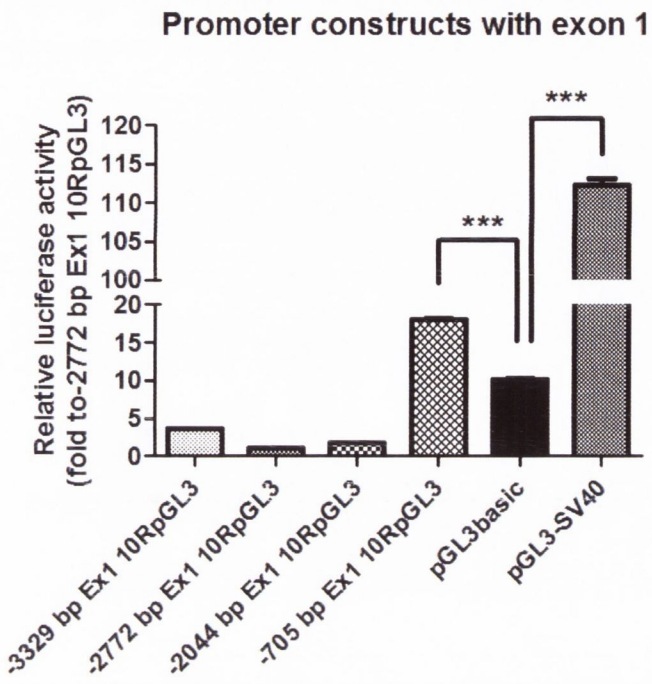


Figure 5.5 Promoter activity for 10RpGL3 construct with exon 1

Luciferase activity was normalised with renilla activity driven by SV40 promoter. Transcription activity is shown as mean +/- SD where -705 bp reached ***P < 0.001 when compared to the control, and Luciferase activity for controls where pGL3-SV40 reached ***P < 0.001 when compared to pGL3-basic, promoterless vector.

This set of experiments for -705 bp Ex1 10RpGL3 was repeated and even though a higher activity for this construct was observed when compared to pGL3-basic, statistical significance was not reached.

A comparison for clones having the -2044 bp promoter length (i.e. -2044 bp Ex1 10RpGL3, -2044 bp 10RpGL3, -2044 bp Ex1 13RApGL3, -2044 bp 13RApGL3) was performed in one experiment and the luciferase activity can be seen in Figure 5.6. The highest activity of the promoter is seen in the -2044 bp 10RpGL3 which has the 10R CAA repeat and does not contain exon 1 of the HRH4, however this was not significantly different from the blank vector (similar to Figure 5.3). This plot also shows a significant difference between 10R and 13R genotypes, and the construct having exon 1 of HRH4 in 10RpGL3 is less active than

13RApGL3 ($P < 0.01 – 0.001$). Comparing the two constructs without exon 1, i.e., 10RpGL3 and 13RApGL3, there is a trend that 10RpGL3 is more active; however this does not reach a significant difference when compared to the 13R genotype.

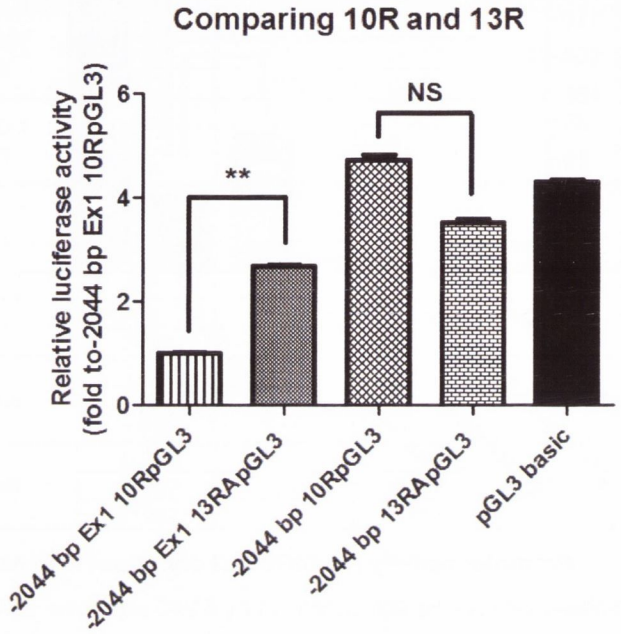


Figure 5.6 Luciferase activity for all -2044 bp constructs

*Luciferase activity was normalised with renilla control vector acitivity driven by SV40 promoter. Transcription activity is shown as mean +/- SD where -2044 bp Ex1 13RApGL3 reached **P < 0.01-0.001 when compared to -2044 bp Ex 1 10RpGL3 and -2044 bp 10RpGL3 is not significant (NS) compared to -2044 bp 13RpGL3.*

This is the first study to describe the promoter region of the HRH4 biologically in a cell-based assay, since previous studies were all bioinformatics analyses. Looking at the overall analysis of the data, it is suggested that the highest and most significant activity of the 10R promoter is achieved when it has the -705 bp sequence and this is only seen when exon 1 is present. On comparing two VNTR genotypes (10R and 13R) with the exon 1 attached a significant difference is observed, where the 13R shows a higher transcription level than 10R when the exon is included in the promoter construct. However this needs to be further looked into with the -705 bp sequence, since this was the most active promoter observed.

5.3 Discussion

Identifying the recognition start site of the promoter region requires analysis of various promoter lengths. The promoter lengths analysed were decided by using a bioinformatics database (Promoter 2.0). This bioinformatics software takes advantage of genetic algorithms to discriminate between promoters and non-promoters [118]. Gene expression is controlled by a proximal promoter, where the pre-initiation complex (PIC) determines the location of transcription initiation for protein-coded genes [119]. The PIC is a large complex of proteins which initiates RNA synthesis from a DNA template, and it includes RNA polymerase II and other general initiation factors such as TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH [119]. The arrangement of transcription factor binding sites is specific for each gene, hence a promoter region is characterised differently for each gene [118]. Using bioinformatics tools such as Promoter 2.0, gives an indication of where this promoter region is located.

The initial optimisation steps were performed on a neuronal cell (SH-SY5Y), which is a thrice-cloned sub-line of the bone marrow biopsy derived from SK-N-SH originating from human neuroblastoma (ECACC cat no 86012802). Due to a potential link of the triplet repeat and tag SNP to schizophrenia, it was initially thought to perform the assay on this neuronal cell line since some specific transcription factors could play a role in gene regulation. However, there was a difficulty in optimising transfection, especially since it is a double vector transfection and neuronal cell lines prove to be difficult to transfect. Transfection in neuronal cell lines is a known challenge for plasmid-based vectors in post-mitotic cells and generally results in lower transfection efficiencies [124]. Comparison between the SH-SY5Y cell line and HEK293T cell lines showed that transfection and luciferase activity in the latter cell line had a stronger (significant) read out. The signalling from luciferase is of crucial importance in this assay, hence it was decided to use the versatile HEK293T cell line and all further experiments were performed on transfected HEK293T.

The promoter region length constructs described in Chapter 4 were here tested on a cell-based assay to identify the active promoter of the gene coding for HRH4. The four length constructs for the 10R VNTR variation were run on a Dual-Glo® luciferase assay where two sets of constructs were tested, with and without exon 1 attached at the 3'-end of the putative promoter region, before the ATG start codon of luciferase. The resulting constructs are

illustrated in Figure 5.7, where all the promoter lengths tested are described in a simplified manner. The four different promoter lengths sequences (-705, -2044, -2772, -3329 bp) in the 10R CAA triplet repeat are represented in this figure and compared with the corresponding ones with exon 1 for the HRH4 gene. The 13R -2044 bp length with and without exon 1 is also shown, and all these promoter lengths and genotypes were the ones analysed for luciferase expression in this assay.

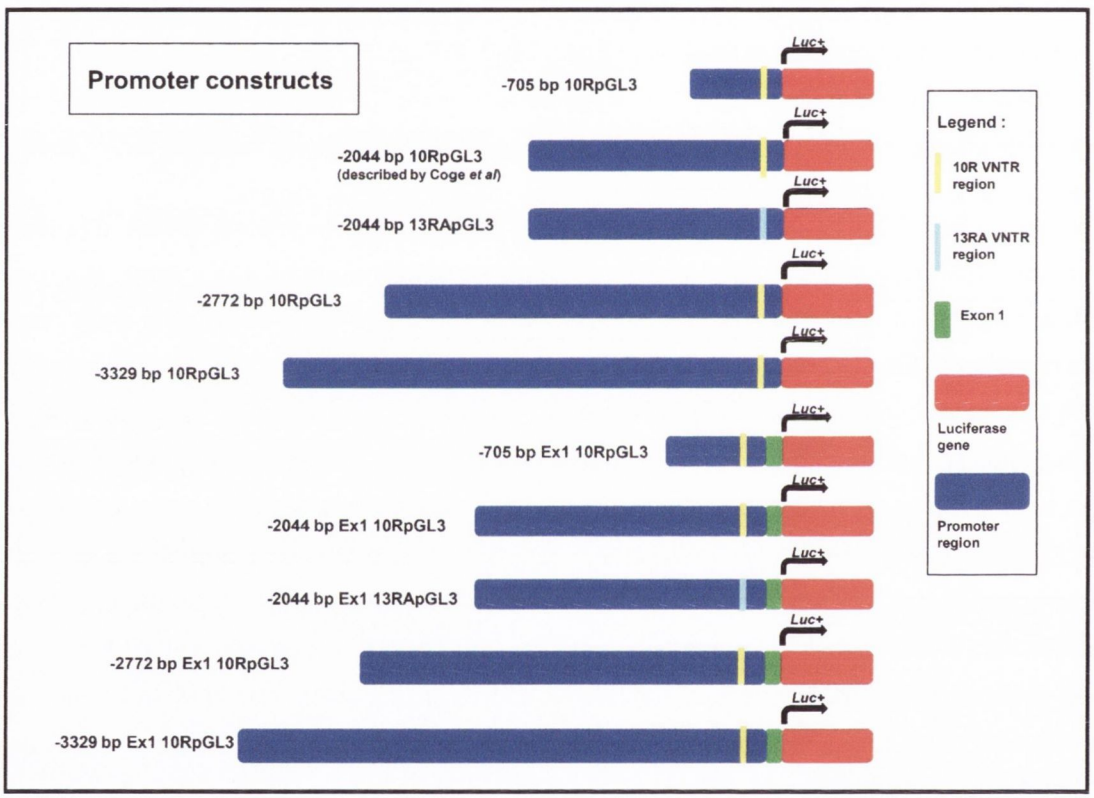


Figure 5.7 Promoter length constructs assayed for promoter activity

Firefly and renilla proteins are widely used as co-reporters, where in these experiments renilla was used as a normalising control vector. Normalising the expression of an experimental reporter to the expression of a control vector can help differentiate between specific and non specific cellular responses. The normalisation is also used to control transfection efficiency, and all values are expressed as the ratio between firefly and renilla bioluminescence.

Gene expression of firefly luciferase is here quantified as transcription activity for the promoter region. The firefly luciferase is commonly used as reporter gene for this type of analysis and tracks the physiological response, where the bioluminescence represents the immediate function upon translation. Thus, if the amount of luminescence from an experimental sample (promoter constructs) is greater than the luminescene from the control (pGL3-basic) an increase in transcription or translation has occurred. The mechanism of the luciferase assay is shown in Figure 5.8. The transcription or translation of the reporter molecules, such as firefly and renilla luciferase is immediately recorded because they do not need post-translational modification and thus are fully functional after translation. In order to have firefly luciferase, beetle luciferin, ATP, magnesium and oxygen is required, whilst renilla luciferase requires only coelenterazine and oxygen (Figure 5.8). Running the assay requires the addition of luciferase substrate which lyses the cell and drives the reaction shown in Figure 5.8. Bioluminescence produced by firefly luciferase is recorded. This assay provides a half-life of 2 hours for each signal i.e., firefly and renilla bioluminescence. Following the firefly luciferase measurement, the second reagent, Stop & Glo® substrate is added which simultaneously quenches the firefly luminescence and generates renilla luminescence.

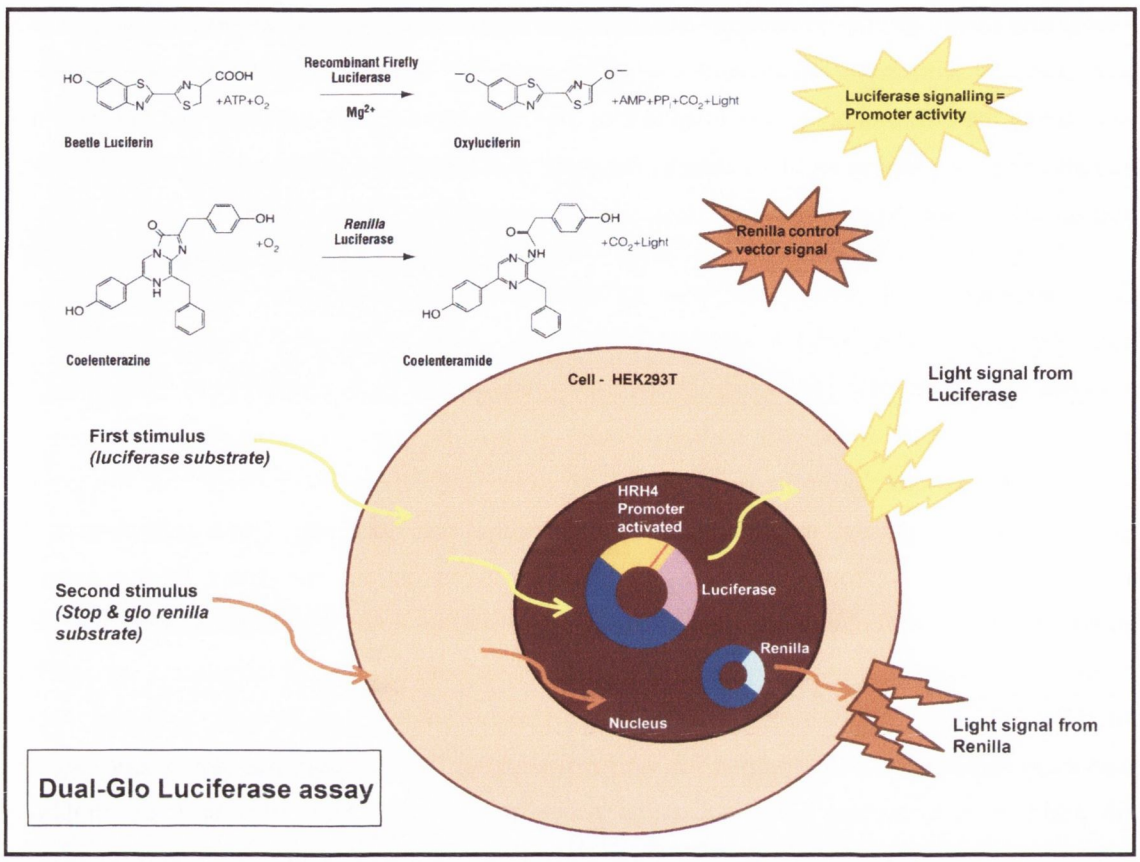


Figure 5.8 Luciferase assay for promoter identification

The promoter analysis carried out in this section shows that the sequence that exhibits highest transcriptional activity is observed for the shortest promoter, i.e. -705 bp to +194 bp, which it shows a significant difference ($P < 0.001$) when compared to the promoterless vector (pGL3-basic). This high promoter activity was only seen in this clone, -705 bp Ex1 10RpGL3. Interestingly, the promoter length -2044 bp 10RpGL3, i.e., the region described by Coge *et al* [7], showed a higher activity than the blank vector, it failed to produce a significant difference. The bioinformatics analysis carried out by previous research groups used databases such as TRANSFEC and TESS software, where the method employed uses known recognition binding sites of transcription factors. This *in silico* method is simulating how the cell works *in vivo*, however some of the binding sites used by these databases may

have very limited experimental evidence. This gives considerable doubts on the performance of algorithms used to predict promoter activity and recognition [123]. In addition, it is important to note that most *in silico* analysis focus on the upstream region of the gene, and it appears that at least some of the downstream sequence may also contribute to promoter function [123]. This could be a possible explanation for the discrepancies observed between *in silico* predictions and *in vitro* experimental analysis. None of the other promoter constructs analysed reached a statistically significant increase in activity when compared to the promoterless reporter pGL3-basic. From this promoter study, it can also be seen that, if the promoter lengths are compared with each other, activity (even though less than the blank vector) is still variable, where the clone -2772 bp Ex 1 10RpGL3 exhibits the least activity. This seemed to be reproducible in other experiments and significant.

The promoter predicted by Coge et al [7], ranging from -2040 bp to +3 (where ATG is the start codon in exon 1) is shown in Figure 5.9 and this sequence was used in the -2044 bp 10R pGL3 construct. It also shows that α-INF binding site described by this research group that binds on the '**A**AA TGA' ending of the VNTR sequence. Interferons (-α and –β) are cytokines primarily produced by plasmacytoid dendritic cells in response to viral and bacterial stimuli. They promote immunity through cytokine production, B-cell class switching and antibody production and have become recognised as important mediators in autoimmune disease [126]. This polymorphism, which lies in the promoter region of the gene, could possibly have an impact on the binding process of interferon when it is found as '**C**AA TGA'. But further analysis needs to be conducted to verify this. The sequence shown (Figure 5.9) also describes the TATA and GATA boxes, however it is also described that the sequence could be TATA-less since the putative transcription start site (i.e., 5'UTR) shown (-116 bp) is downstream to these regions.

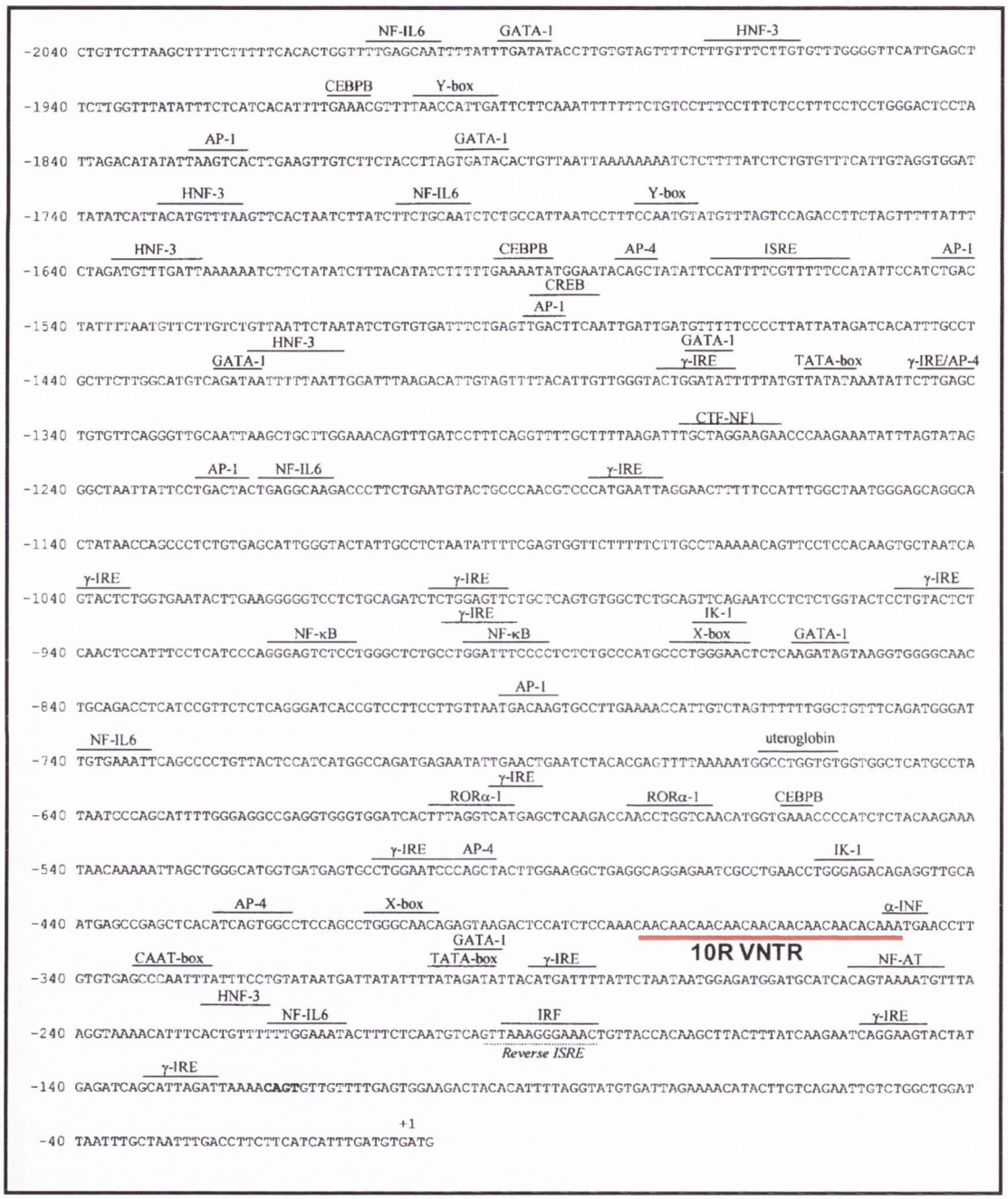


Figure 5.9 Proposed in silico promoter sequence by Coge et al [7]

The sequence which was found to be functional in this study ranged from -705 bp to +194 bp and this is shown in Figure 5.10 and is therefore proposed as the core promoter region for the HRH4 gene, were as per NCBI, -100 to -1 is the 5'-UTR of the gene.

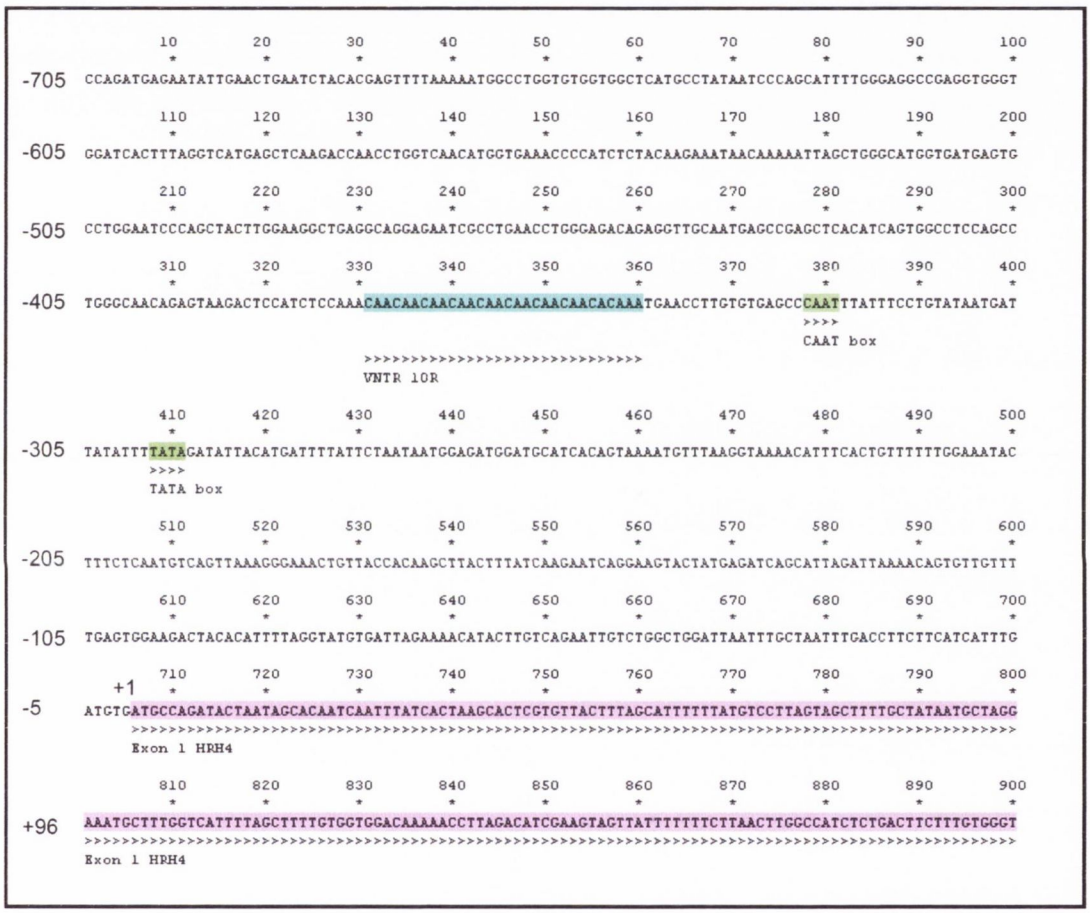


Figure 5.10 Proposed in vitro promoter sequence for HRH4

Where the sequence highlighted in blue shows the 10R VNTR region, the green shows the CAAT box and the TATA box and the pink shows the sequence of exon 1 for HRH4.

On comparing the transcriptional activity between the two genotypes, 10R and 13R, it was interesting to see a difference in activity. Initial data show that when the constructs -2044 bp 10RpGL3 and -2044 bp 13RApGL3 were compared, the 10R showed a higher level of activity, which however does not reach a significant value upon evaluation with one-way ANOVA statistics. On the other hand, when the -2044 bp Ex 1 10RpGL3 and -2044 bp Ex1 13RApGL3 are compared, the genotype with the 13R length variation seems more active and also gave a strong significant difference ($P < 0.01 – 0.001$) of activity between the two clones. Further analysis need to be conducted using the active promoter region (hence the -705 bp to +194 bp sequence) in order to confirm this.

5.4 Conclusion

In this study the promoter activity of various lengths of 5' extensions upstream of the ATG of HRH4 is described. The transcription factors involved in gene regulation are specific for each gene. Transcriptional initiation is the first step in gene expression and it generally represents the most important point of control. Through complex biological mechanisms genes can be turned on or off in a highly defined manner [123]. Hence, identification of this turning on and turning off of the gene is considered as an important biological process, especially when it can be linked to a disease or condition.

This study indicated that the difference in VNTR genotypes (10R and 13R) could possibly affect the promoter activity of the HRH4, where the 13R showed a significant difference in activity to the 10R. However this is only seen when the exon 1 sequence is attached. The activity of the short length promoter sequence (-705 bp) was also seen to be the most active promoter when exon 1 is attached. This data suggests that exon 1 has an important contribution to the activity of the promoter.

The results shown in this study also identify the recognition sequence for the 5'-end flanking region of the HRH4, hence the active promoter region for this protein. The sequence -705 bp to +194 bp was shown to be significantly active when compared to the promoterless control, suggesting the presence of the active promoter region. This is the first study that is performed to identify the active promoter of the HRH4.

Chapter 6

Conclusions

6 Conclusions

Gene expression and regulation play a crucial role in the function of the body and in human health. Gene transcription in eukaryotes is generally controlled by RNA polymerase II which binds upstream of the gene on the promoter region, and together with transcription factors forms the pre-initiation complex (PIC) which initiates transcription [119]. These general transcription factors (GTFs) are regulatory proteins which function is to activate (or inhibit) DNA transcription by binding to specific recognition sequences on genomic DNA [127]. Through complex mechanisms controlling these processes, genes can be turned on or off in a highly defined manner [123]. Hence transcriptional control can have direct impact on health since improper regulation could lead to up or downregulation of protein expression and diseases or symptoms of a disease.

Microsatellite repeats consist of tandem repeats of 2-6 base pairs of DNA and are a type of variable number of tandem repeat (VNTR) [110, 111]. Microsatellite repeat expansions can cause both 'loss of function' or 'gain in function' mechanisms, and these microsatellites can be found in coding and non-coding regions of the gene. Neurological disorders such as Fragile X syndrome and Kennedy's disease were found to possess such repeats consisting of $(CGG)_n$ and $(CAG)_n$, respectively, and these repeat expansions were identified to increase the risk of disorders when found in larger repeats [64]. Genotyping the VNTR region of the HRH4 identified a triplet repeat ranging from 10 to 19 CAA triplets. The VNTR region is located approximately 400 bp upstream to the ATG start codon of exon 1 of the HRH4 gene. Interestingly a transcription factor binding site for α-INF is located at the proximity of the last triplet repeat. A single nucleotide polymorphism was identified towards the end of the triplet repeat sequence and possible sequence combinations are: CAC AAA, CAA CAA or CAA AAA.

The first objective of this project was to test for association of polymorphisms found on the HRH4 gene locus to schizophrenia. The SNPs on the HRH4 gene were analysed for this possible linkage by using tag SNPs identified by the International Schizophrenia Consortium (ISC) through a genome-wide association study (GWAS). A few SNPs were flagged by this analysis; and a weak association to schizophrenia was observed. Linkage to the VNTR

region was not analysed directly, since this length variation is not represented as a single entity and is highly complex. Hence, genotyping this region through fragment analysis by capillary electrophoresis (CE) and linkage disequilibrium (LD) study was conducted. Fragment length analysis by CE indicated that the triplet repeat region consists of different lengths and allele frequencies including 10R (17%), 13R (27%), 14R (25%), 15R (24%), 16R (2.9%), 17R (2.9%), 18R (1%), 19R (0.6%). Triplet repeats for 11R and 12R were not observed. Furthermore, the LD study identified eleven SNPs that had a high LD with the 10R genotype, where two of them are located on the HRH4 gene. One SNP (rs628764) is located in intron 1 of HRH4 and the other SNP (rs615283) is found on the promoter region of HRH4. The high LD (r^2 = 1.00) of SNP rs615283 (G to A allele) to 10R observed in this study was also described by Coge *et al* [7]. A strong LD with the 13R genotype was observed for rs17797945, which was one of the SNPs flagged by the ISC study that had a weak association with schizophrenia (P = 0.007879). This suggests that the 13R genotype has a small degree of association with the disorder, but further analyses need to be performed in order to confirm this.

As previously mentioned, genotyping this region through fragment length analysis by CE identified the allele frequency of each VNTR length in HapMap samples. Comparing these with the unpublished study from Dr. Sasse [92], where schizophrenia patient samples were used, the allele frequencies seem to be similar. However, this needs to be repeated with a larger cohort and matching controls in order to confirm this. These results provoked further interest in studying the functional implication such a complex polymorphic region has on gene expression.

Another interesting point that emerged from this study was that this highly polymorphic region is not clearly represented in the NCBI website. The insert/deletions shown as polymorphisms in the dbSNP do not indicate what is actually seen once the DNA is genotyped. Hence it is suggested that this VNTR region is marked as a microsatellite triplet repeat, where rs665411 (seen only in the 10R) and rs670327 describe the SNP towards the end of the VNTR region.

The location of trinucleotide repeat expansions on the gene and the frequency of the expansion can affect the clinical outcome or contribute to a disease [64]. Repeats in the 5'-UTR or promoter region of the gene may play a role in gene regulation or translation, hence the

next objective of this study was to biologically analyse this VNTR region of the HRH4 gene as this could provide some interesting insights into the functional impact on the gene. Therefore, as a first step a plasmid library was constructed by constructing reporter vectors harbouring all the VNTR genotypes. The genotypes and polymorphisms identified were cloned into a standard vector (pCR-Blunt) and sub-cloned into a reporter vector (pGL3-basic). The cloned genotypes harboured the CAA triplet repeat which ranged from 10R and 13R to 19R (11R and 12R were not confirmed so far) (Table 6.1). The DNA used in these experiments was obtained from either HapMap samples or from genomic DNA extracted from buccal swabs of human controls. Techniques employed to create the library of plasmids, included cloning into a standard vector, sub-cloning into the reporter vector, point mutation by site-directed mutagenesis and promoter extension using mega-primers. Cloning the desired DNA sequences into vectors proved to be demanding hence various conditions and steps were changed in order to identify the correct clone. Ligation turned out to be one of the most challenging steps in the sequence of these experiments and ligation conditions were optimised with respect to length of ligation reaction, temperature and ratio of insert to vector. On the other hand, site-directed mutagenesis turned out to be a more reliable and very versatile method, for both point mutation and extension of the sequences using mega-primers. However, the drawback for this technique is that a parental template clone is needed in order to construct the variations.

Table 6.1 Plasmid library

Clones in the standard vector	Clones in the reporter vector	Promoter length variations
10RpCR-Blunt	-2044 bp 10RpGL3	-705 bp 10RpGL3
13RApCR-Blunt	-2044 bp 13RApGL3	-2772 bp 10RpGL3
14RCpCR-Blunt	-2044 bp 13RCpGL3	-3329 bp 10RpGL3
15RApCR-Blunt	-2044 bp 14RApGL3	-705 bp Ex1 10RpGL3
16RApCR-Blunt	-2044 bp 14RCpGL3	-2044 bp Ex1 10RpGL3
16RCpCR-Blunt	-2044 bp 15RApGL3	-2044 bp Ex1 13RApGL3
17RApCR-Blunt	-2044 bp 15RCpGL3	-2772 bp Ex1 10RpGL3
17RCpCR-Blunt	-2044 bp 17RApGL3	-3329 bp Ex1 10RpGL3
	-2044 bp 17RCpGL3	
	-2044 bp 18RApGL3	
	-2044 bp 18RCpGL3	
	-2044 bp 19RApGL3	
	-2044 bp 19RCpGL3	

The final constructs were cloned into the reporter vector, pGL3-basic, which is a vector that lacks a promoter and encodes for the luciferase gene, thus can provide a means to measure promoter activity. Gene expression of firefly luciferase, driven by a standard promoter is recorded and quantified as control for transfection and transcriptional activity. The construct library is shown in Table 6.1.

The putative promoter published by Coge *et al* [7] is the 2044 bp length sequence upstream of the ATG site, it contains 10 CAA repeats in the microsatellite, hence this was used to build the different promoter lengths in order to identify the functional promoter sequence. Modifications to this construct included 5'-end extensions for potentially longer promoter region and 3'-end extension which included exon 1 of the HRH4 gene. The length of the largest construct was established by using an *in silico* analysis obtained from a bioinformatics database (Promoter 2.0) which discriminates between promoter and non-promoters through genetic algorithms [118]. The -2044 bp 10RpGL3 construct was modified through site-directed

mutagenesis by using mega-primers. These were constructed to add the desired sequence upstream or downstream of the original putative promoter. The shortest putative promoter, i.e. -705 bp 10RpGL3 was obtained as a result of an unexpected self-ligation of the vector when it was digested with Xho1 and Msc1. In addition the construct -2044 bp 13RApGL3 (13 CAA microsatellite repeats) was modified by an extension of the 3'-end by the addition of exon 1. The 13R genotype has the highest allele frequency (27%) of all genotypes, hence this modification was analysed to investigate if this extension affects the functional activity of the promoter. Promoter length variations constructed are also shown in Table 6.1.

The last objective of the study was met by analysing selected constructs (i.e., all promoter length variations for 10RpGL3, -2044 bp 13RApGL3 and -2044 bp Ex1 13RApGL3) on a cell-based assay to assess promoter activity. The Dual-Glo® luciferase assay was used to analyse the constructs and to quantify the activity of the promoter. Previous linkage analysis highlighted that there is a weak association to schizophrenia, hence it was originally decided to use a neuronal cell line, since some transcription factors present in the cell could play a role in regulation. Hence, initial optimisation was conducted on SH-SY5Y cell line. Transfection in neuronal cell lines is known to be a challenging mechanism, and a double transfection proved to be not reliable and was providing inconsistent results. On comparison with HEK293T cell line it was decided to use the latter, since they proved to be transfected with a higher efficiency and produced a stronger read-out for luciferase activity.

Analysing these plasmids showed some interesting results. The promoter region described *in silico* by Coge *et al* [7], showed the strongest luciferase activity in constructs harbouring the upstream promoter sequence of HRH4 gene only. However, when compared to the promoterless vector, pGL3-basic, it did not reach statistical significance. On the other hand, looking at other vector constructs, the -705 bp Ex1 10RpGL3 vector, which is the shortest construct ranging from -705 bp to +194 bp, showed a strong significant activity ($P < 0.001$) when compared to the promoterless pGL3-basic.

A possible explanation for the differences between constructs in promoter activity could lie in the mechanism of action of a promoter. Eukaryotic promoters have a PIC (composed of RNA polymerase II and GTFs) which recognises the core promoter and initiates transcription [119], and this mechanism seems to be regulated by different combinations of transcription

elements, which are capable of directing assembly of this PIC and sustain basal transcription *in vitro* [123]. It is suggested that the downstream sequence may also contribute to the promoter function, and promoter prediction *in silico* methods tend to focus more on the upstream sequence [123]. The promoter sequence published [7] only investigated the upstream sequence of the HRH4, while in this study both upstream (i.e., -3329 bp to -1) and downstream (-3329 bp to +197 bp) sequences were considered, resulting in the -705 bp to +197 bp sequence having the strongest promoter activity.

Comparing the two genotypes, i.e., 10R and 13R, it was found that the construct -2044 bp 10RpGL3 had a higher level of activity than -2044 bp 13RApGL3; however this difference does not reach statistical significance on a one-way ANOVA statistics. On the contrary, examining the data from the other two constructs with exon 1 for these genotypes, i.e., -2044 bp Ex1 10RpGL3 and -2044 bp Ex1 13RApGL3, showed that the 13R promoter is more active than the 10R with a statistical significance of $P < 0.01 - 0.001$. Further experiments could clarify the role of exon 1 on expression level of the H_4 receptor.

Studying the polymorphisms on the promoter region of the HRH4 gene provided a good insight on a potential clinical implication. While SNPs ranging 100 kb upstream to 100 kb downstream did not show any significant linkage or association to schizophrenia, the 13R genotype showed a strong LD with an SNP which had a weak association to the disease ($P = 0.007879$). When analysed on a cell based assay, this genotype (13R; -2044 bp to +194 bp) also showed a stronger activity when compared to the 10R (-2044 bp to +194 bp) genotype. Looking into the promoter activity of the HRH4, the construct -705 bp Ex1 10RpGL3 (-705 bp to +194 bp), was the only one that produced a strong significant activity when compared to the promoterless control vector (pGL3-basic).

This project is the first study that analyses the promoter region of HRH4 for biological activity by a cell-based assay. The core transcription start site generally varies between different genes [119] and identifying the core transcription site of a gene is of biological importance since by this mechanism gene expression can be turned on or off [123]. This turning on and off of a gene is even more significant when it is linked to a disease since it could contribute to symptoms of the disease, especially if polymorphisms form part of this regulatory site. In this study the promoter sequence -705 bp to +194 bp was identified to be the functional promoter

of HRH4 gene, containing a region that harbours a microsatellite triplet repeat. Further in depth analysis of this VNTR region is suggested in order to produce further insights into the functional significance on HRH4 gene expression.

References

References

1. Leurs R, Vischer HF, Wijtmans M, de Esch IJ. En route to new blockbuster anti-histamines: surveying the offspring of the expanding histamine receptor family. *Trends Pharmacol Sci* 2011; **32**(4): 250-257.

2. Lim HD, Adami M, Guaita E, Werfel T, Smits RA, de Esch IJ, *et al.* Pharmacological characterization of the new histamine H4 receptor agonist VUF 8430. *Br J Pharmacol* 2009; **157**(1): 34-43.

3. Zhu Y, Michalovich D, Wu H, Tan KB, Dytko GM, Mannan IJ, *et al.* Cloning, expression, and pharmacological characterization of a novel human histamine receptor. *Mol Pharmacol* 2001; **59**(3): 434-441.

4. Nakamura T, Itadani H, Hidaka Y, Ohta M, Tanaka K. Molecular cloning and characterization of a new human histamine receptor, HH4R. *Biochem Biophys Res Commun* 2000; **279**(2): 615-620.

5. Oda T, Morikawa N, Saito Y, Masuho Y, Matsumoto S. Molecular cloning and characterization of a novel type of histamine receptor preferentially expressed in leukocytes. *J Biol Chem* 2000; **275**(47): 36781-36786.

6. Leurs R, Chazot PL, Shenton FC, Lim HD, de Esch IJ. Molecular and biochemical pharmacology of the histamine H4 receptor. *Br J Pharmacol* 2009; **157**(1): 14-23.

7. Coge F, Guenin SP, Rique H, Boutin JA, Galizzi JP. Structure and expression of the human histamine H4-receptor gene. *Biochem Biophys Res Commun* 2001; **284**(2): 301-309.

8. Liu C, Ma X, Jiang X, Wilson SJ, Hofstra CL, Blevitt J, *et al.* Cloning and pharmacological characterization of a fourth histamine receptor (H(4)) expressed in bone marrow. *Mol Pharmacol* 2001; **59**(3): 420-426.

9. Morse KL, Behan J, Laz TM, West RE, Jr., Greenfeder SA, Anthes JC, *et al.* Cloning and characterization of a novel human histamine receptor. *J Pharmacol Exp Ther* 2001; **296**(3): 1058-1066.

10. de Esch IJ, Thurmond RL, Jongejan A, Leurs R. The histamine H4 receptor as a new therapeutic target for inflammation. *Trends Pharmacol Sci* 2005; **26**(9): 462-469.

11. Kottke T, Sander K, Weizel L, Schneider EH, Seifert R, Stark H. Receptor-specific functional efficacies of alkyl imidazoles as dual histamine H3/H4 receptor ligands. *Eur J Pharmacol* 2011; **654**(3): 200-208.

12. Parsons ME, Ganellin CR. Histamine and its receptors. *Br J Pharmacol* 2006; **147 Suppl 1**: S127-135.

13. Black JW, Duncan WA, Durant CJ, Ganellin CR, Parsons EM. Definition and antagonism of histamine H 2 -receptors. *Nature* 1972; **236**(5347): 385-390.

14. Clinicaltrials.gov (2011). Proof of activity study of UR-63325 in allergic rhinitis induced by nasal challenge.

15. Salcedo C, Pontes C, Merlos M. Is the H4 receptor a new drug target for allergies and asthma? *Front Biosci (Elite Ed)* 2013; **5:** 178-187.

16. Garcia-Martin E, Ayuso P, Martinez C, Blanca M, Agundez JA. Histamine pharmacogenomics. *Pharmacogenomics* 2009; **10**(5): 867-883.

17. Agundez JA, Ayuso P, Cornejo-Garcia JA, Blanca M, Torres MJ, Dona I, *et al.* The diamine oxidase gene is associated with hypersensitivity response to non-steroidal anti-inflammatory drugs. *PLoS One* 2012; **7**(11): e47571.

18. Adams WJ, Lawson JA, Morris DL. Cimetidine inhibits in vivo growth of human colon cancer and reverses histamine stimulated in vitro and in vivo growth. *Gut* 1994; **35**(11): 1632-1636.

19. Kuefner MA, Schwelberger HG, Hahn EG, Raithel M. Decreased histamine catabolism in the colonic mucosa of patients with colonic adenoma. *Dig Dis Sci* 2008; **53**(2): 436-442.

20. Brown RE, Stevens DR, Haas HL. The physiology of brain histamine. *Prog Neurobiol* 2001; **63**(6): 637-672.

21. Passani MB, Blandina P. Histamine receptors in the CNS as targets for therapeutic intervention. *Trends Pharmacol Sci* 2011; **32**(4): 242-249.

22. Nuutinen S, Panula P. Histamine in Neurotransmission and Brain Diseases. *Adv Exp Med Biol* 2010; **709:** 95-107.

23. Fernandez TV, Sanders SJ, Yurkiewicz IR, Ercan-Sencicek AG, Kim YS, Fishman DO, *et al.* Rare copy number variants in tourette syndrome disrupt genes in histaminergic pathways and overlap with autism. *Biol Psychiatry* 2012; **71**(5): 392-402.

24. Ito C. The role of the central histaminergic system on schizophrenia. *Drug News Perspect*, vol. 17. 2004 Prous Science: Spain, 2004, pp 383-387.

25. Nakai T, Kitamura N, Hashimoto T, Kajimoto Y, Nishino N, Mita T, *et al.* Decreased histamine H1 receptors in the frontal cortex of brains from patients with chronic schizophrenia. *Biol Psychiatry*, vol. 30: United States, 1991, pp 349-356.

26. Prell GD, Green JP, Kaufmann CA, Khandelwal JK, Morrishow AM, Kirch DG, *et al.* Histamine metabolites in cerebrospinal fluid of patients with chronic schizophrenia:

their relationships to levels of other aminergic transmitters and ratings of symptoms. *Schizophr Res*, vol. 14: Netherlands, 1995, pp 93-104.

27. Basile VS, Masellis M, Potkin SG, Kennedy JL. Pharmacogenomics in schizophrenia: the quest for individualized therapy. *Hum Mol Genet* 2002; **11**(20): 2517-2530.

28. Nguyen T, Shapiro DA, George SR, Setola V, Lee DK, Cheng R, *et al.* Discovery of a novel member of the histamine receptor family. *Mol Pharmacol* 2001; **59**(3): 427-433.

29. Appl H, Holzammer T, Dove S, Haen E, Strasser A, Seifert R. Interactions of recombinant human histamine H_1R, H_2R, H_3R, and H_4R receptors with 34 antidepressants and antipsychotics. *Naunyn Schmiedebergs Arch Pharmacol* 2012; **385**(2): 145-170.

30. Orange PR, Heath PR, Wright SR, Pearson RC. Allelic variations of the human histamine H2 receptor gene. *Neuroreport* 1996; **7**(7): 1293-1296.

31. Orange PR, Heath PR, Wright SR, Ramchand CN, Kolkeiwicz L, Pearson RC. Individuals with schizophrenia have an increased incidence of the H2R649G allele for the histamine H2 receptor gene. *Mol Psychiatry* 1996; **1**(6): 466-469.

32. Ito C, Morisset S, Krebs MO, Olie JP, Loo H, Poirier MF, *et al.* Histamine H2 receptor gene variants: lack of association with schizophrenia. *Mol Psychiatry* 2000; **5**(2): 159-164.

33. Rosse RB, Kendrick K, Tsui LC, Fay-McCarthy M, Collins JP, Rosenberg P, *et al.* Famotidine adjunctive pharmacotherapy of schizophrenia: a case report. *Clin Neuropharmacol* 1995; **18**(4): 369-374.

34. Perlis RH, Adams DH, Fijal B, Sutton VK, Farmen M, Breier A, *et al.* Genetic association study of treatment response with olanzapine/fluoxetine combination or lamotrigine in bipolar I depression. *J Clin Psychiatry* 2010; **71**(5): 599-605.

35. Keeling BH, Vilarino-Guell C, Soto-Ortolaza AI, Ross OA, Uitti RJ, Rajput A, *et al.* Histamine N-methyltransferase Thr105Ile is not associated with Parkinson's disease or essential tremor. *Parkinsonism Relat Disord*, vol. 16. 2009 Elsevier Ltd: England, 2010, pp 112-114.

36. Palada V, Terzic J, Mazzulli J, Bwala G, Hagenah J, Peterlin B, *et al.* Histamine N-methyltransferase Thr105Ile polymorphism is associated with Parkinson's disease. *Neurobiol Aging*, vol. 33. A 2012 Elsevier Inc: United States, 2012, pp 836 e831-833.

37. Agundez JA, Luengo A, Herraez O, Martinez C, Alonso-Navarro H, Jimenez-Jimenez FJ, *et al.* Nonsynonymous polymorphisms of histamine-metabolising enzymes in patients with Parkinson's disease. *Neuromolecular Med* 2008; **10**(1): 10-16.

38. Marasovic-Susnjara I, Palada V, Marinovic-Terzic I, Mimica N, Marin J, Muck-Seler D, *et al.* No association between histamine N-methyltransferase functional polymorphism

Thr105Ile and Alzheimer's disease. *Neurosci Lett*, vol. 489. A 2010 Elsevier Ireland Ltd: Ireland, 2011, pp 119-121.

39. Nakamura S, Takemura M, Ohnishi K, Suenaga T, Nishimura M, Akiguchi I, *et al.* Loss of large neurons and occurrence of neurofibrillary tangles in the tuberomammillary nucleus of patients with Alzheimer's disease. *Neurosci Lett* 1993; **151**(2): 196-199.

40. Stevenson J, Sonuga-Barke E, McCann D, Grimshaw K, Parker KM, Rose-Zerilli MJ, *et al.* The role of histamine degradation gene polymorphisms in moderating the effects of food additives on children's ADHD symptoms. *Am J Psychiatry*, vol. 167: United States, 2010, pp 1108-1115.

41. Brioni JD, Esbenshade TA, Garrison TR, Bitner SR, Cowart MD. Discovery of histamine H3 antagonists for the treatment of cognitive disorders and Alzheimer's disease. *J Pharmacol Exp Ther*, vol. 336: United States, 2011, pp 38-46.

42. Millán-Guerrero RO, Baltazar-Rodríguez LM, Cárdenas-Rojas MI, Ramírez-Flores M, Isais-Millán S, Delgado-Enciso I, *et al.* A280V polymorphism in the histamine H3 receptor as a risk factor for migraine. *Arch Med Res* 2011; **42**(1): 44-47.

43. Gupta S, Nahas SJ, Peterlin BL. Chemical mediators of migraine: preclinical and clinical observations. *Headache* 2011; **51**(6): 1029-1045.

44. Wiedemann P, Bönisch H, Oerters F, Brüss M. Structure of the human histamine H3 receptor gene (HRH3) and identification of naturally occurring variations. *J Neural Transm* 2002; **109**(4): 443-453.

45. Flores-Clemente C, Osorio-Espinoza A, Escamilla-Sanchez J, Leurs R, Arias JM, Arias-Montano JA. A single-point mutation (Ala280Val) in the third intracellular loop alters the signalling properties of the human histamine H receptor stably expressed in CHO-K1 cells. *Br J Pharmacol* 2013.

46. Aamodt AH, Stovner LJ, Langhammer A, Hagen K, Zwart JA. Is headache related to asthma, hay fever, and chronic bronchitis? The Head-HUNT Study. *Headache* 2007; **47**(2): 204-212.

47. Mansfield LE. The role of antihistamine therapy in vascular headaches. *J Allergy Clin Immunol* 1990; **86**(4 Pt 2): 673-676.

48. Garcia-Martin E, Martinez C, Serrador M, Alonso-Navarro H, Navacerrada F, Agundez JA, *et al.* Histamine-N-methyl transferase polymorphism and risk for migraine. *Headache* 2008; **48**(9): 1343-1348.

49. Mancama D, Arranz MJ, Munro J, Osborne S, Makoff A, Collier D, *et al.* Investigation of promoter variants of the histamine 1 and 2 receptors in schizophrenia and clozapine response. *Neurosci Lett* 2002; **333**(3): 207-211.

50. Vehof J, Risselada AJ, Al Hadithy AF, Burger H, Snieder H, Wilffert B, *et al.* Association of genetic variants of the histamine H1 and muscarinic M3 receptors with BMI and HbA1c values in patients on antipsychotic medication. *Psychopharmacology (Berl)* 2011; **216**(2): 257-265.

51. Wu R, Zhao J, Shao P, Ou J, Chang M. Genetic predictors of antipsychotic-induced weight gain: a case-matched multi-gene study. *Zhong Nan Da Xue Xue Bao Yi Xue Ban* 2011; **36**(8): 720-723.

52. García-Martín E, Ayuso P, Luengo A, Martínez C, Agúndez JA. Genetic variability of histamine receptors in patients with Parkinson's disease. *BMC Med Genet* 2008; **9**: 15.

53. Wei Z, Wang L, Zhang M, Xuan J, Wang Y, Liu B, *et al.* A pharmacogenetic study of risperidone on histamine H3 receptor gene (HRH3) in Chinese Han schizophrenia patients. *J Psychopharmacol* 2012; **26**(6): 813-818.

54. Wei Z, Wang L, Yu T, Wang Y, Sun L, Wang T, *et al.* Histamine H4 receptor polymorphism: a potential predictor of risperidone efficacy. *J Clin Psychopharmacol* 2013; **33**(2): 221-225.

55. Banu Y, Watanabe T. Augmentation of antigen receptor-mediated responses by histamine H1 receptor signaling. *J Exp Med* 1999; **189**(4): 673-682.

56. Jutel M, Watanabe T, Klunker S, Akdis M, Thomet OA, Malolepszy J, *et al.* Histamine regulates T-cell and antibody responses by differential expression of H1 and H2 receptors. *Nature* 2001; **413**(6854): 420-425.

57. Idzko M, la Sala A, Ferrari D, Panther E, Herouy Y, Dichmann S, *et al.* Expression and function of histamine receptors in human monocyte-derived dendritic cells. *J Allergy Clin Immunol* 2002; **109**(5): 839-846.

58. Jutel M, Watanabe T, Akdis M, Blaser K, Akdis CA. Immune regulation by histamine. *Curr Opin Immunol*, vol. 14: England, 2002, pp 735-740.

59. Pino-Ángeles A, Reyes-Palomares A, Melgarejo E, Sánchez-Jiménez F. Histamine: an undercover agent in multiple rare diseases? *J Cell Mol Med* 2012; **16**(9): 1947-1960.

60. Yu B, Shao Y, Zhang J, Dong XL, Liu WL, Yang H, *et al.* Polymorphisms in human histamine receptor H4 gene are associated with atopic dermatitis. *Br J Dermatol* 2010; **162**(5): 1038-1043.

61. Akdis CA, Simons FE. Histamine receptors are hot in immunopharmacology. *Eur J Pharmacol*, vol. 533: Netherlands, 2006, pp 69-76.

62. Zampeli E, Tiligada E. The role of histamine H4 receptor in immune and inflammatory disorders. *Br J Pharmacol* 2009; **157**(1): 24-33.

63. Dijkstra D, Stark H, Chazot PL, Shenton FC, Leurs R, Werfel T, *et al.* Human inflammatory dendritic epidermal cells express a functional histamine H4 receptor. *J Invest Dermatol* 2008; **128**(7): 1696-1703.

64. Brouwer JR, Willemsen R, Oostra BA. Microsatellite repeat instability and neurological disease. *Bioessays* 2009; **31**(1): 71-83.

65. Yu B, Shao Y, Li P, Zhang J, Zhong Q, Yang H, *et al.* Copy number variations of the human histamine H4 receptor gene are associated with systemic lupus erythematosus. *Br J Dermatol* 2010; **163**(5): 935-940.

66. Baumer W, Stahl J, Sander K, Petersen LJ, Paps J, Stark H, *et al.* Lack of preventing effect of systemically and topically administered histamine H(1) or H(4) receptor antagonists in a dog model of acute atopic dermatitis. *Exp Dermatol* 2011; **20**(7): 577-581.

67. Jiang W, Lim HD, Zhang M, Desai P, Dai H, Colling PM, *et al.* Cloning and pharmacological characterization of the dog histamine H4 receptor. *Eur J Pharmacol* 2008; **592**(1-3): 26-32.

68. Yamada H, Tahara T, Shiroeda H, Hayashi R, Saito T, Nakamura M, *et al.* Effects of -1018G>A polymorphism of HRH2 (rs2607474) on the severity of gastric mucosal atrophy. *J Gastrointestin Liver Dis* 2012; **21**(2): 139-143.

69. Sande N, Nikulin M, Nilsson I, Wadstrom T, Laxen F, Harkonen M, *et al.* Increased risk of developing atrophic gastritis in patients infected with CagA+ Helicobacter pylori. *Scand J Gastroenterol* 2001; **36**(9): 928-933.

70. Ebule I, Longdoh A, Paloheimo I. Helicobacter pylori infection and atrophic gastritis. *Afr Health Sci* 2013; **13**(1): 112-117.

71. Choi JH, Kim SH, Suh CH, Nahm DH, Park HS. Polymorphisms of high-affinity IgE receptor and histamine-related genes in patients with ASA-induced urticaria/angioedema. *J Korean Med Sci* 2005; **20**(3): 367-372.

72. Ferreira MA, Zhao ZZ, Thomsen SF, James M, Evans DM, Postmus PE, *et al.* Association and interaction analyses of eight genes under asthma linkage peaks. *Allergy* 2009; **64**(11): 1623-1628.

73. Simon T, Semsei AF, Ungvári I, Hadadi E, Virág V, Nagy A, *et al.* Asthma endophenotypes and polymorphisms in the histamine receptor HRH4 gene. *Int Arch Allergy Immunol* 2012; **159**(2): 109-120.

74. Rivera ES, Cricco GP, Engel NI, Fitzsimons CP, Martin GA, Bergoc RM. Histamine as an autocrine growth factor: an unusual role for a widespread mediator. *Semin Cancer Biol* 2000; **10**(1): 15-23.

75. Medina VA, Brenzoni PG, Lamas DJ, Massari N, Mondillo C, Nunez MA, *et al.* Role of histamine H4 receptor in breast cancer cell proliferation. *Front Biosci (Elite Ed)* 2011; **3:** 1042-1060.

76. Medina V, Croci M, Crescenti E, Mohamad N, Sanchez-Jiménez F, Massari N, *et al.* The role of histamine in human mammary carcinogenesis: H3 and H4 receptors as potential therapeutic targets for breast cancer treatment. *Cancer Biol Ther* 2008; **7**(1): 28-35.

77. Martinel Lamas DJ, Croci M, Carabajal E, Crescenti EJ, Sambuco L, Massari NA, *et al.* Therapeutic potential of histamine H(4) receptor agonists in triple-negative human breast cancer experimental model. *Br J Pharmacol* 2013.

78. He GH, Lu J, Shi PP, Xia W, Yin SJ, Jin TB, *et al.* Polymorphisms of human histamine receptor H4 gene are associated with breast cancer in Chinese Han population. *Gene* 2013; **519**(2): 260-265.

79. Zhang C, Xiong Y, Li J, Yang Y, Liu L, Wang W, *et al.* Deletion and down-regulation of HRH4 gene in gastric carcinomas: a potential correlation with tumor progression. *PLoS One* 2012; **7**(2): e31207.

80. Arisawa T, Tahara T, Ozaki K, Matsue Y, Minato T, Yamada H, *et al.* Association between common genetic variant of HRH2 and gastric cancer risk. *Int J Oncol* 2012; **41**(2): 497-503.

81. Klepstad P, Fladvad T, Skorpen F, Bjordal K, Caraceni A, Dale O, *et al.* Influence from genetic variability on opioid use for cancer pain: a European genetic association study of 2294 cancer pain patients. *Pain* 2011; **152**(5): 1139-1145.

82. Laugsand EA, Fladvad T, Skorpen F, Maltoni M, Kaasa S, Fayers P, *et al.* Clinical and genetic factors associated with nausea and vomiting in cancer patients receiving opioids. *Eur J Cancer* 2011; **47**(11): 1682-1691.

83. Welsh MM, Karagas MR, Kuriger JK, Houseman A, Spencer SK, Perry AE, *et al.* Genetic determinants of UV-susceptibility in non-melanoma skin cancer. *PLoS One* 2011; **6**(7): e20019.

84. Consortium IH. The International HapMap Project. *Nature* 2003; **426**(6968): 789-796.

85. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, *et al.* The structure of haplotype blocks in the human genome. *Science* 2002; **296**(5576): 2225-2229.

86. Haas H, Panula P. The role of histamine and the tuberomamillary nucleus in the nervous system. *Nat Rev Neurosci* 2003; **4**(2): 121-130.

87. Benarroch EE. Histamine in the CNS: multiple functions and potential neurologic implications. *Neurology* 2010; **75**(16): 1472-1479.

88. Connelly WM, Shenton FC, Lethbridge N, Leurs R, Waldvogel HJ, Faull RL, *et al.* The histamine H4 receptor is functionally expressed on neurons in the mammalian CNS. *Br J Pharmacol* 2009; **157**(1): 55-63.

89. Strakhova MI, Nikkel AL, Manelli AM, Hsieh GC, Esbenshade TA, Brioni JD, *et al.* Localization of histamine H4 receptors in the central nervous system of human and rat. *Brain Res* 2009; **1250:** 41-48.

90. Chazot PL. Advances in histamine pharmacology reveal new drug targets. *Br J Pharmacol* 2009; **157**(1): 1-3.

91. Toth-Petroczy A, Szilagyi A, Ronai Z, Sasvari-Szekely M, Guttman A. Validation of a tentative microsatellite marker for the dopamine D4 receptor gene by capillary gel electrophoresis. *J Chromatogr A* 2006; **1130**(2): 201-205.

92. Sasse Astrid AJM Unpublished data.

93. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**(7256): 748-752.

94. Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, *et al.* Common variants conferring risk of schizophrenia. *Nature* 2009; **460**(7256): 744-747.

95. Hardy J, Singleton A. Genomewide association studies and human disease. *N Engl J Med* 2009; **360**(17): 1759-1768.

96. O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V, *et al.* Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet* 2008; **40**(9): 1053-1055.

97. Zhang R, Lu SM, Qiu C, Liu XG, Gao CG, Guo TW, *et al.* Population-based and family-based association studies of ZNF804A locus and schizophrenia. *Mol Psychiatry*, vol. 16: England, 2011, pp 360-361.

98. Hwang J, Pototschnig M, Lettow R, Zumofen G, Renn A, Gotzinger S, *et al.* A single-molecule optical transistor. *Nature* 2009; **460**(7251): 76-80.

99. Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I, *et al.* Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* 2009; **460**(7256): 753-757.

100. Arnold SE, Talbot K, Hahn CG. Neurodevelopment, neuroplasticity, and new genes for schizophrenia. *Prog Brain Res* 2005; **147:** 319-345.

101. Wei LN, Tsao JL, Chu YS, Jeannotte L, Nguyen-Huu MC. Molecular cloning and transcriptional mapping of the mouse cellular retinoic acid-binding protein gene. *DNA Cell Biol* 1990; **9**(7): 471-478.

102. Green EK, Grozeva D, Jones I, Jones L, Kirov G, Caesar S, *et al.* The bipolar disorder risk allele at CACNA1C also confers risk of recurrent major depression and of schizophrenia. *Mol Psychiatry* 2010; **15**(10): 1016-1022.

103. Wei J, Hemmings GP. The NOTCH4 locus is associated with susceptibility to schizophrenia. *Nat Genet* 2000; **25**(4): 376-377.

104. Schumacher J, Laje G, Abou Jamra R, Becker T, Mühleisen TW, Vasilescu C, *et al.* The DISC locus and schizophrenia: evidence from an association study in a central European sample and from a meta-analysis across different European populations. *Hum Mol Genet* 2009; **18**(14): 2719-2727.

105. Fanous AH, Zhao Z, van den Oord EJ, Maher BS, Thiselton DL, Bergen SE, *et al.* Association study of SNAP25 and schizophrenia in Irish family and case-control samples. *Am J Med Genet B Neuropsychiatr Genet* 2010; **153b**(2): 663-674.

106. Levinson DF, Levinson MD, Segurado R, Lewis CM. Genome scan meta-analysis of schizophrenia and bipolar disorder, part I: Methods and power analysis. *Am J Hum Genet* 2003; **73**(1): 17-33.

107. Bush WS, Haines J. Overview of linkage analysis in complex traits. *Curr Protoc Hum Genet* 2010; **Chapter 1:** Unit 1.9.1-18.

108. JD W, TA B, SP B, A G, M L, R L. *Molecular biology of the gene*. CSHL Press, 2008.

109. Goldstein DB, Weale ME. Population genomics: linkage disequilibrium holds the key. *Curr Biol* 2001; **11**(14): R576-579.

110. Queller DC, Strassmann JE, Hughes CR. Microsatellites and kinship. *Trends Ecol Evol* 1993; **8**(8): 285-288.

111. Varela MA, Amos W. Heterogeneous distribution of SNPs in the human genome: microsatellites as predictors of nucleotide diversity and divergence. *Genomics* 2010; **95**(3): 151-159.

112. Cogé F, Guénin SP, Rique H, Boutin JA, Galizzi JP. Structure and expression of the human histamine H4-receptor gene. *Biochem Biophys Res Commun* 2001; **284**(2): 301-309.

113. Corvin A, Craddock N, Sullivan PF. Genome-wide association studies: a primer. *Psychol Med* 2010; **40**(7): 1063-1077.

114. Owen MJ, Craddock N, O'Donovan MC. Suggestion of roles for both common and rare risk variants in genome-wide studies of schizophrenia. *Arch Gen Psychiatry* 2010; **67**(7): 667-673.

115. Kosaki K, Suzuki T, Kosaki R, Yoshihashi H, Itoh M, Goto Y, *et al.* Human homolog of the mouse imprinted gene Impact resides at the pericentric region of chromosome 18

within the critical region for bipolar affective disorder. *Mol Psychiatry* 2001; **6**(1): 87-91.

116. Tsen H, Levene SD. Supercoiling-dependent flexibility of adenosine-tract-containing DNA detected by a topological method. *Proc Natl Acad Sci U S A* 1997; **94**(7): 2817-2822.

117. Marinus MG, Casadesus J. Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more. *FEMS Microbiol Rev* 2009; **33**(3): 488-503.

118. Knudsen S. Promoter2.0: for the recognition of PolII promoter sequences. *Bioinformatics* 1999; **15**(5): 356-361.

119. Fickett JW, Hatzigeorgiou AG. Eukaryotic promoter recognition. *Genome Res* 1997; **7**(9): 861-878.

120. Usdin K. The biological effects of simple tandem repeats: lessons from the repeat expansion diseases. *Genome Res* 2008; **18**(7): 1011-1019.

121. Boer K, Helinger E, Helinger A, Pocza P, Pos Z, Demeter P, *et al.* Decreased expression of histamine H1 and H4 receptors suggests disturbance of local regulation in human colorectal tumours by histamine. *Eur J Cell Biol* 2008; **87**(4): 227-236.

122. Glatzer F, Gschwandtner M, Ehling S, Rossbach K, Janik K, Klos A, *et al.* Histamine induces proliferation in keratinocytes from patients with atopic dermatitis through the histamine 4 receptor. *J Allergy Clin Immunol* 2013.

123. Pedersen AG, Baldi P, Chauvin Y, Brunak S. The biology of eukaryotic promoter prediction--a review. *Comput Chem* 1999; **23**(3-4): 191-207.

124. Karra D, Dahm R. Transfection techniques for neuronal cells. *J Neurosci* 2010; **30**(18): 6171-6177.

125. Washbourne P, McAllister AK. Techniques for gene transfer into neurons. *Curr Opin Neurobiol* 2002; **12**(5): 566-573.

126. Becker-Merok A, Ostli-Eilersten G, Lester S, Nossent J. Circulating interferon-alpha2 levels are increased in the majority of patients with systemic lupus erythematosus and are associated with disease activity and multiple cytokine activation. *Lupus* 2013; **22**(2): 155-163.

127. Thomas MC, Chiang CM. The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol* 2006; **41**(3): 105-178.

APPENDIX 1

THE UNIVERSITY OF DUBLIN
TRINITY COLLEGE

SCHOOL OF MEDICINE
FACULTY OF HEALTH SCIENCES

Trinity College, Dublin 2, Ireland
Tel: +353 1 896 1476
Fax: +353 1 671 3956
email: medicine@tcd.ie

Professor Dermot Kelleher, MD, FRCPI, FRCP, F Med Sci
Head of School of Medicine
Vice Provost for Medical Affairs

Ms Fedelma McNamara
School Administrator
email: medschadmin@tcd.ie

Ms Stephany Micallef
School of Pharmacy and Pharmaceutical Sciences,
Trinity College,
Dublin 2

Monday, 17 May 2010

Study: Identification and Evaluation of Genetic Variations in Genes Related to G-Protein Coupled Receptors

Dear Applicant (s),

Further to a meeting of the Faculty of Health Sciences Ethics Committee held in December 2009, we are pleased to inform you that the above project has been approved without further audit.

Yours sincerely

Prof. Orla Sheils
Chairperson
Faculty of Health Sciences Ethics Committee

Cc
Dr Astrid Sasse,
School of Pharmacy and Pharmaceutical Sciences,
Trinity College,
Dublin 2

APPENDIX 2

Family ID	Sample Name	Fathers' ID	Mothers' ID	Marker	Allele 1	Allele 2	Mendelise?
1334	NA12239	0	0	HRH4	10R	10R	yes
1334	NA12146	0	0	HRH4	14R	15R	
1334	NA10847	NA12146	NA12239	HRH4	10R	15R	
1334	NA10846	NA12144	NA12145	HRH4	13R	14R	no
1334	NA12145	0	0	HRH4	10R	14R	
1334	NA12144	0	0	HRH4	15R	16R	
1340	NA07029	NA06994	NA07000	HRH4	14R	14R	yes
1340	NA06994	0	0	HRH4	14R	18R	
1340	NA07000	0	0	HRH4	14R	18R	
1340	NA07056	0	0	HRH4	13R	15R	yes
1340	NA07019	NA07022	NA07056	HRH4	10R	15R	
1340	NA07022	0	0	HRH4	10R	15R	
1341	NA06991	NA06993	NA06985	HRH4	14R	15R	yes
1341	NA06985	0	0	HRH4	13R	15R	
1341	NA06993	0	0	HRH4	10R	14R	
1341	NA07034	0	0	HRH4	14R	15R	yes
1341	NA07055	0	0	HRH4	15R	15R	
1341	NA07048	NA07034	NA07055	HRH4	15R	15R	
1344	NA10851	NA12056	NA12057	HRH4	13R	14R	yes
1344	NA12057	0	0	HRH4	14R	14R	
1344	NA12056	0	0	HRH4	13R	13R	
1345	NA07348	NA07357	NA07345	HRH4	13R	14R	yes
1345	NA07357	0	0	HRH4	13R	15R	
1345	NA07345	0	0	HRH4	13R	14R	
1346	NA12043	0	0	HRH4	10R	13R	yes
1346	NA10857	NA12043	NA12044	HRH4	13R	14R	
1346	NA12044	0	0	HRH4	14R	14R	
1347	NA11881	0	0	HRH4	15R	15R	yes
1347	NA11882	0	0	HRH4	10R	14R	
1347	NA10859	NA11881	NA11882	HRH4	14R	15R	
1349	NA11840	0	0	HRH4	14R	17R	yes
1349	NA11839	0	0	HRH4	10R	16R	
1349	NA10854	NA11839	NA11840	HRH4	10R	14R	
1350	NA11832	0	0	HRH4	10R	15R	yes
1350	NA10855	NA11831	NA11832	HRH4	13R	15R	
1350	NA11831	0	0	HRH4	13R	13R	
1350	NA10856	NA11829	NA11830	HRH4	13R	15R	yes
1350	NA11830	0	0	HRH4	10R	13R	
1350	NA11829	0	0	HRH4	13R	15R	
1358	NA12716	0	0	HRH4	10R	15R	yes
1358	NA12717	0	0	HRH4	13R	13R	
1358	NA12707	NA12716	NA12717	HRH4	10R	13R	

Appendix 2 – Pedigree Analysis

Family ID	Sample Name	Fathers' ID	Mothers' ID	Marker	Allele 1	Allele 2	Mendelise?
1362	NA10860	NA11992	NA11993	HRH4	14R	14R	yes
1362	NA11992	0	0	HRH4	13R	14R	
1362	NA11993	0	0	HRH4	14R	14R	
1362	NA11994	0	0	HRH4	13R	14R	yes
1362	NA10861	NA11994	NA11995	HRH4	13R	15R	
1362	NA11995	0	0	HRH4	15R	15R	
1375	NA10863	NA12264	NA12234	HRH4	10R	16R	yes
1375	NA12264	0	0	HRH4	10R	10R	
1375	NA12234	0	0	HRH4	14R	16R	
1408	NA12156	0	0	HRH4	13R	13R	yes
1408	NA12155	0	0	HRH4	13R	17R	
1408	NA10831	NA12155	NA12156	HRH4	13R	17R	
1408	NA10830	NA12154	NA12236	HRH4	10R	16R	yes
1408	NA12154	0	0	HRH4	15R	16R	
1408	NA12236	0	0	HRH4	10R	14R	
1416	NA10835	NA12248	NA12249	HRH4	13R	15R	yes
1416	NA12249	0	0	HRH4	15R	15R	
1416	NA12248	0	0	HRH4	13R	14R	
1420	NA10838	NA12003	NA12004	HRH4	13R	15R	yes
1420	NA12003	0	0	HRH4	13R	15R	
1420	NA12004	0	0	HRH4	13R	15R	
1420	NA12005	0	0	HRH4	10R	13R	yes
1420	NA12006	0	0	HRH4	14R	15R	
1420	NA10839	NA12005	NA12006	HRH4	13R	15R	
1444	NA12740	NA12750	NA12751	HRH4	14R	15R	yes
1444	NA12751	0	0	HRH4	15R	15R	
1444	NA12750	0	0	HRH4	14R	14R	
1447	NA12752	NA12760	NA12761	HRH4	10R	13R	yes
1447	NA12760	0	0	HRH4	13R	17R	
1447	NA12761	0	0	HRH4	10R	14R	
1447	NA12762	0	0	HRH4	13R	13R	yes
1447	NA12753	NA12762	NA12763	HRH4	13R	14R	
1447	NA12763	0	0	HRH4	14R	19R	
1454	NA12801	NA12812	NA12813	HRH4	13R	14R	yes
1454	NA12812	0	0	HRH4	10R	14R	
1454	NA12813	0	0	HRH4	13R	14R	
1454	NA12814	0	0	HRH4	14R	15R	yes
1454	NA12802	NA12814	NA12815	HRH4	10R	15R	
1454	NA12815	0	0	HRH4	10R	14R	
1459	NA12864	NA12872	NA12873	HRH4	13R	13R	yes
1459	NA12873	0	0	HRH4	13R	14R	
1459	NA12872	0	0	HRH4	13R	13R	

Appendix 2 – Pedigree Analysis

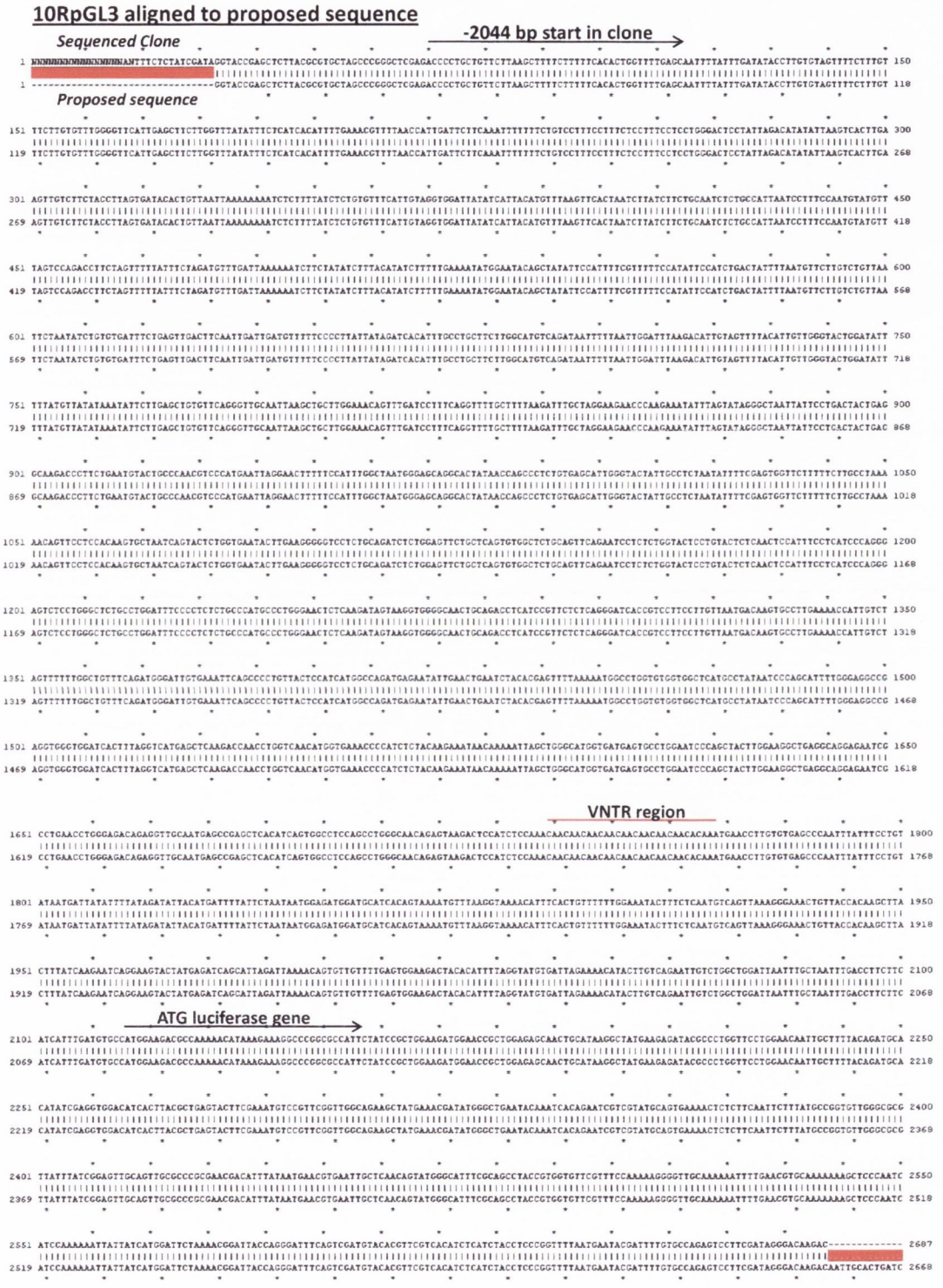
Family ID	Sample Name	Fathers' ID	Mothers' ID	Marker	Allele 1	Allele 2	Mendelise?
1459	NA12875	0	0	HRH4	15R	17R	
1459	NA12874	0	0	HRH4	10R	10R	yes
1459	NA12865	NA12874	NA12875	HRH4	10R	15R	
1463	NA12892	0	0	HRH4	13R	14R	
1463	NA12891	0	0	HRH4	10R	15R	yes
1463	NA12878	NA12891	NA12892	HRH4	10R	13R	

				Recode A"10R" v ALL		rsUnknown1	Recode A"12R" v ALL		rsUnknown2	Recode A"14R" v ALL		rsUnknown3	Recode A"15R "v ALL		rsUnknown4
				10R=A,13R-19R=G			13R=A,10R,14R-19R=G			14R=A,10R,13R,15R-19R=G			15R=A,10R,13R,14R, 16R-19R=G		
						A/G chr18			A/G chr18			A/G chr18			A/G chr18
Sample Name	**Marker**	**Allele1**	**Allele2**	**Allele1**	**Allele2**	**20294302**	**Allele1**	**Allele2**	**20294302**	**Allele1**	**Allele2**	**20294302**	**Allele1**	**Allele2**	**20294302**
NA11830	HRH4	10R	13R	A	G	AG	G	A	AG	G	G	GG	G	G	GG
NA10838	HRH4	13R	15R	G	G	GG	A	G	AG	G	G	GG	G	A	AG
NA12146	HRH4	14R	15R	G	G	GG	G	G	GG	A	G	AG	G	A	AG
NA10856	HRH4	13R	15R	G	G	GG	A	G	AG	G	G	GG	G	A	AG
NA12763	HRH4	14R	19R	G	G	GG	G	G	GG	A	G	AG	G	G	GG
NA12761	HRH4	10R	14R	A	G	AG	G	G	GG	G	A	AG	G	G	GG
NA10847	HRH4	10R	15R	A	G	AG	G	G	GG	G	G	GG	G	A	AG
NA12144	HRH4	15R	16R	G	G	GG	G	G	GG	G	G	GG	A	G	AG
NA12005	HRH4	10R	13R	A	G	AG	G	A	AG	G	G	GG	G	G	GG
NA07345	HRH4	13R	14R	G	G	GG	A	G	AG	G	A	AG	G	G	GG
NA12801	HRH4	13R	14R	G	G	GG	A	G	AG	G	A	AG	G	G	GG
NA12057	HRH4	14R	14R	G	G	GG	G	G	GG	A	A	AA	G	G	GG
NA10860	HRH4	14R	14R	G	G	GG	G	G	GG	A	A	AA	G	G	GG
NA12156	HRH4	13R	13R	G	G	GG	A	A	AA	G	G	GG	G	G	GG
NA07048	HRH4	15R	15R	G	G	GG	G	G	GG	G	G	GG	A	A	AA
NA11993	HRH4	14R	14R	G	G	GG	G	G	GG	A	A	AA	G	G	GG
NA06991	HRH4	14R	15R	G	G	GG	G	G	GG	A	G	AG	G	A	AG
NA12239	HRH4	10R	10R	A	A	AA	G	G	GG	G	G	GG	G	G	GG
NA12872	HRH4	13R	13R	G	G	GG	A	A	AA	G	G	GG	G	G	GG
NA07029	HRH4	14R	14R	G	G	GG	G	G	GG	A	A	AA	G	G	GG
NA10854	HRH4	10R	14R	A	G	AG	G	G	GG	G	A	AG	G	G	GG
NA12006	HRH4	14R	15R	G	G	GG	G	G	GG	A	G	AG	G	A	AG
NA12873	HRH4	13R	14R	G	G	GG	A	G	AG	G	A	AG	G	G	GG
NA12740	HRH4	14R	15R	G	G	GG	G	G	GG	A	G	AG	G	A	AG
NA12043	HRH4	10R	13R	A	G	AG	G	A	AG	G	G	GG	G	G	GG
NA10851	HRH4	13R	14R	G	G	GG	A	G	AG	G	A	AG	G	G	GG
NA06993	HRH4	10R	14R	A	G	AG	G	G	GG	G	A	AG	G	G	GG
NA12004	HRH4	13R	15R	G	G	GG	A	G	AG	G	G	GG	G	A	AG
NA11829	HRH4	13R	15R	G	G	GG	A	G	AG	G	G	GG	G	A	AG
NA12236	HRH4	10R	14R	A	G	AG	G	G	GG	G	A	AG	G	G	GG
NA12248	HRH4	13R	14R	G	G	GG	A	G	AG	G	A	AG	G	G	GG
NA12892	HRH4	13R	14R	G	G	GG	A	G	AG	G	A	AG	G	G	GG
NA12762	HRH4	13R	13R	G	G	GG	A	A	AA	G	G	GG	G	G	GG
NA12891	HRH4	10R	15R	A	G	AG	G	G	GG	G	G	GG	G	A	AG
NA07019	HRH4	10R	15R	A	G	AG	G	G	GG	G	G	GG	G	A	AG
NA10839	HRH4	13R	15R	G	G	GG	A	G	AG	G	G	GG	G	A	AG
NA10861	HRH4	13R	15R	G	G	GG	A	G	AG	G	G	GG	G	A	AG
NA12865	HRH4	10R	15R	A	G	AG	G	G	GG	G	G	GG	G	A	AG
NA12155	HRH4	13R	17R	G	G	GG	A	G	AG	G	G	GG	G	G	GG
NA10846	HRH4	13R	14R	G	G	GG	A	G	AG	G	A	AG	G	G	GG
NA12249	HRH4	15R	15R	G	G	GG	G	G	GG	G	G	GG	A	A	AA
NA12154	HRH4	15R	16R	G	G	GG	G	G	GG	G	G	GG	A	G	AG
NA11994	HRH4	13R	14R	G	G	GG	A	G	AG	G	A	AG	G	G	GG
NA11832	HRH4	10R	15R	A	G	AG	G	G	GG	G	G	GG	G	A	AG
NA12752	HRH4	10R	13R	A	G	AG	G	A	AG	G	G	GG	G	G	GG
NA12264	HRH4	10R	10R	A	A	AA	G	G	GG	G	G	GG	G	G	GG

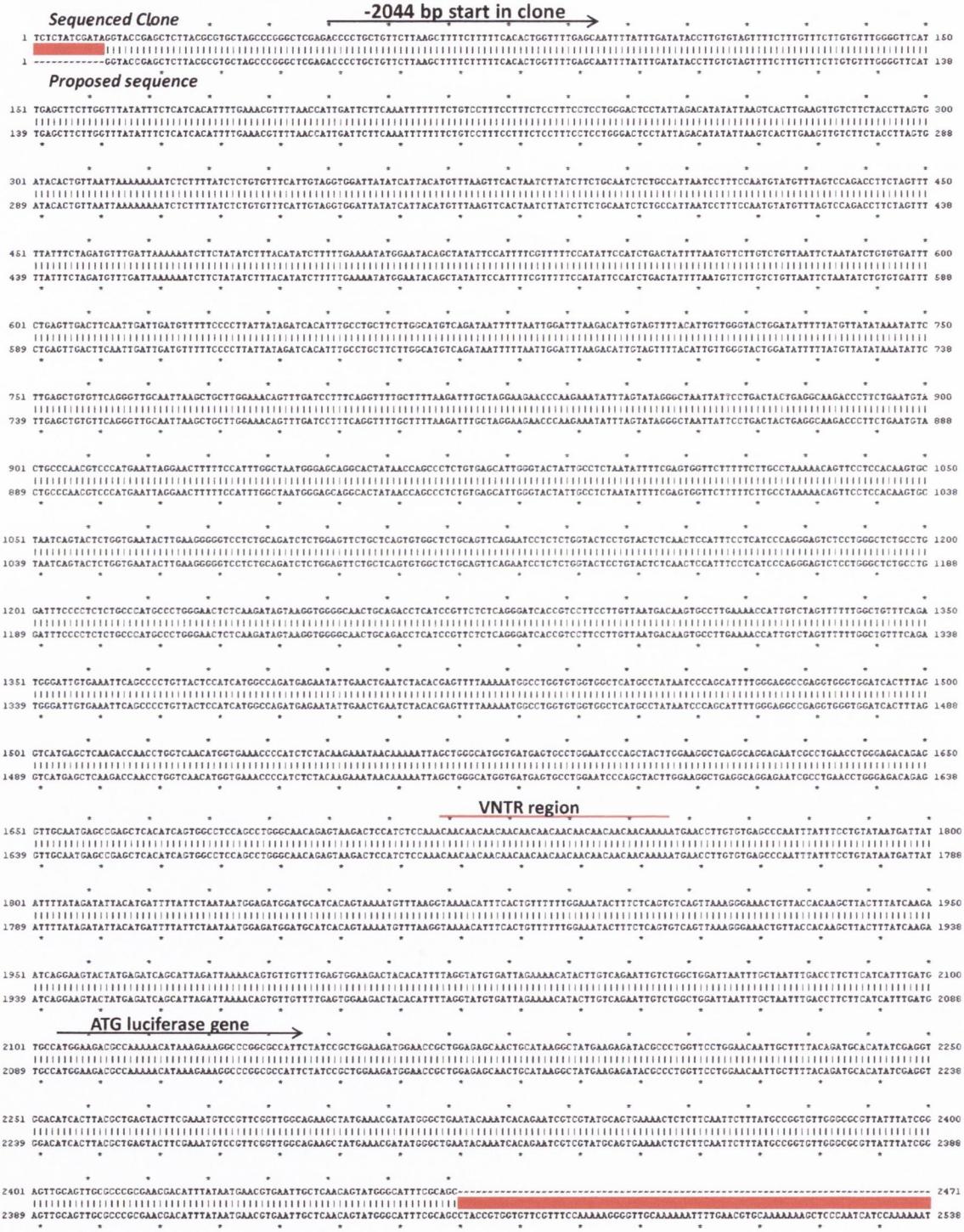
NA12815	HRH4	10R	14R	A	G	AG	G	G	GG	G	A	AG	G	G	GG
NA10859	HRH4	14R	15R	G	G	GG	G	G	GG	A	G	AG	G	A	AG
NA12145	HRH4	10R	14R	A	G	AG	G	G	GG	G	A	AG	G	G	GG
NA11840	HRH4	14R	17R	G	G	GG	G	G	GG	A	G	AG	G	G	GG
NA06994	HRH4	14R	18R	G	G	GG	G	G	GG	A	G	AG	G	G	GG
NA12864	HRH4	13R	13R	G	G	GG	A	A	AA	G	G	GG	G	G	GG
NA12753	HRH4	13R	14R	G	G	GG	A	G	AG	G	A	AG	G	G	GG
NA12875	HRH4	15R	17R	G	G	GG	G	G	GG	G	G	GG	A	G	AG
NA11992	HRH4	13R	14R	G	G	GG	A	G	AG	G	A	AG	G	G	GG
NA07348	HRH4	13R	14R	G	G	GG	A	G	AG	G	A	AG	G	G	GG
NA10835	HRH4	13R	15R	G	G	GG	A	G	AG	G	G	GG	G	A	AG
NA10830	HRH4	10R	16R	A	G	AG	G	G	GG	G	G	GG	G	G	GG
NA12707	HRH4	10R	13R	A	G	AG	G	A	AG	G	G	GG	G	G	GG
NA11839	HRH4	10R	16R	A	G	AG	G	G	GG	G	G	GG	G	G	GG
NA10863	HRH4	10R	16R	A	G	AG	G	G	GG	G	G	GG	G	G	GG
NA07000	HRH4	14R	18R	G	G	GG	G	G	GG	A	G	AG	G	G	GG
NA12234	HRH4	14R	16R	G	G	GG	G	G	GG	A	G	AG	G	G	GG
NA12717	HRH4	13R	13R	G	G	GG	A	A	AA	G	G	GG	G	G	GG
NA11881	HRH4	15R	15R	G	G	GG	G	G	GG	G	G	GG	A	A	AA
NA11831	HRH4	13R	13R	G	G	GG	A	A	AA	G	G	GG	G	G	GG
NA12751	HRH4	15R	15R	G	G	GG	G	G	GG	G	G	GG	A	A	AA
NA06985	HRH4	13R	15R	G	G	GG	A	G	AG	G	G	GG	G	A	AG
NA10855	HRH4	13R	15R	G	G	GG	A	G	AG	G	G	GG	G	A	AG
NA12716	HRH4	10R	15R	A	G	AG	G	G	GG	G	G	GG	G	A	AG
NA12750	HRH4	14R	14R	G	G	GG	G	G	GG	A	A	AA	G	G	GG
NA11995	HRH4	15R	15R	G	G	GG	G	G	GG	G	G	GG	A	A	AA
NA10857	HRH4	13R	14R	G	G	GG	A	G	AG	G	A	AG	G	G	GG
NA12813	HRH4	13R	14R	G	G	GG	A	G	AG	G	A	AG	G	G	GG
NA07055	HRH4	15R	15R	G	G	GG	G	G	GG	G	G	GG	A	A	AA
NA11882	HRH4	10R	14R	A	G	AG	G	G	GG	G	A	AG	G	G	GG
NA07357	HRH4	13R	15R	G	G	GG	A	G	AG	G	G	GG	G	A	AG
NA07056	HRH4	13R	15R	G	G	GG	A	G	AG	G	G	GG	G	A	AG
NA07034	HRH4	14R	15R	G	G	GG	G	G	GG	A	G	AG	G	A	AG
NA12056	HRH4	13R	13R	G	G	GG	A	A	AA	G	G	GG	G	G	GG
NA12814	HRH4	14R	15R	G	G	GG	G	G	GG	A	G	AG	G	A	AG
NA12878	HRH4	10R	13R	A	G	AG	G	A	AG	G	G	GG	G	G	GG
NA12760	HRH4	13R	17R	G	G	GG	A	G	AG	G	G	GG	G	G	GG
NA07022	HRH4	10R	15R	A	G	AG	G	G	GG	G	G	GG	G	A	AG
NA12003	HRH4	13R	15R	G	G	GG	A	G	AG	G	G	GG	G	A	AG
NA10831	HRH4	13R	17R	G	G	GG	A	G	AG	G	G	GG	G	G	GG
NA12802	HRH4	10R	15R	A	G	AG	G	G	GG	G	G	GG	G	A	AG
NA12812	HRH4	10R	14R	A	G	AG	G	G	GG	G	A	AG	G	G	GG
NA12044	HRH4	14R	14R	G	G	GG	G	G	GG	A	A	AA	G	G	GG
NA12874	HRH4	10R	10R	A	A	AA	G	G	GG	G	G	GG	G	G	GG

OMIT-DIDN'T MENDELISE

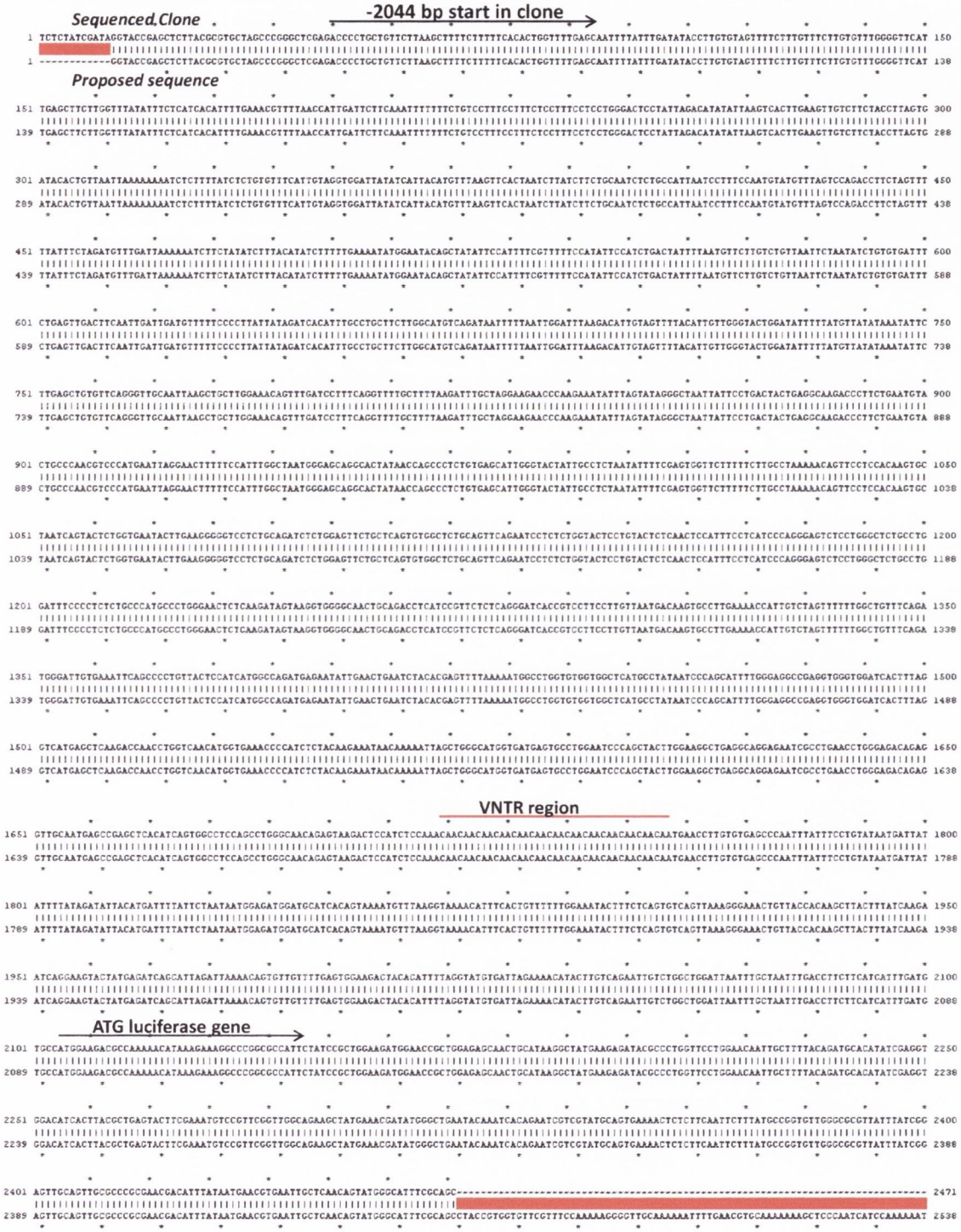
APPENDIX 3


10RpGL3 aligned to proposed sequence
Sequenced Clone
-2044 bp start in clone
Proposed sequence
VNTR region
ATG luciferase gene

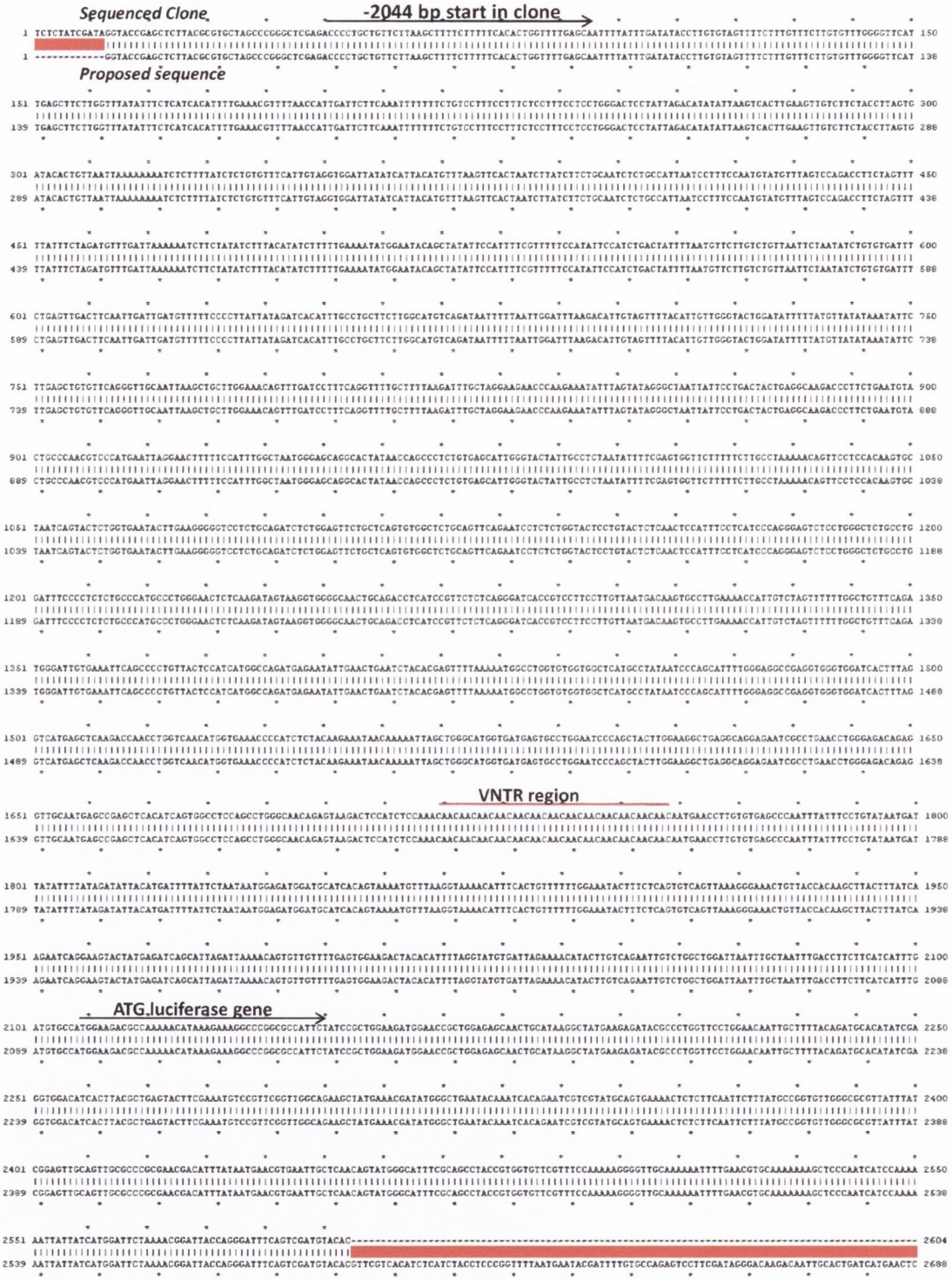
13RApGL3 aligned to proposed sequence



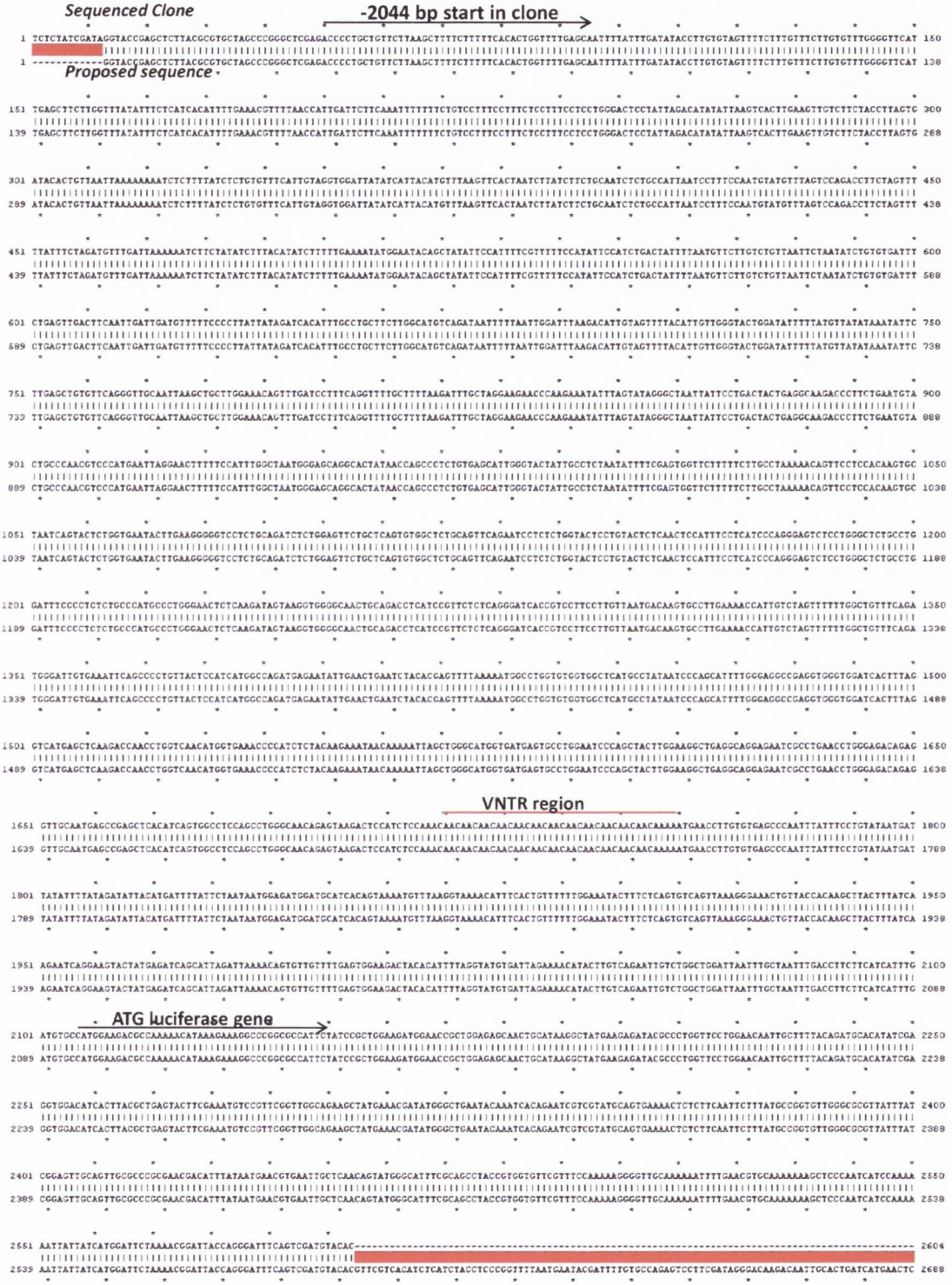
13RCpGL3 aligned to proposed sequence



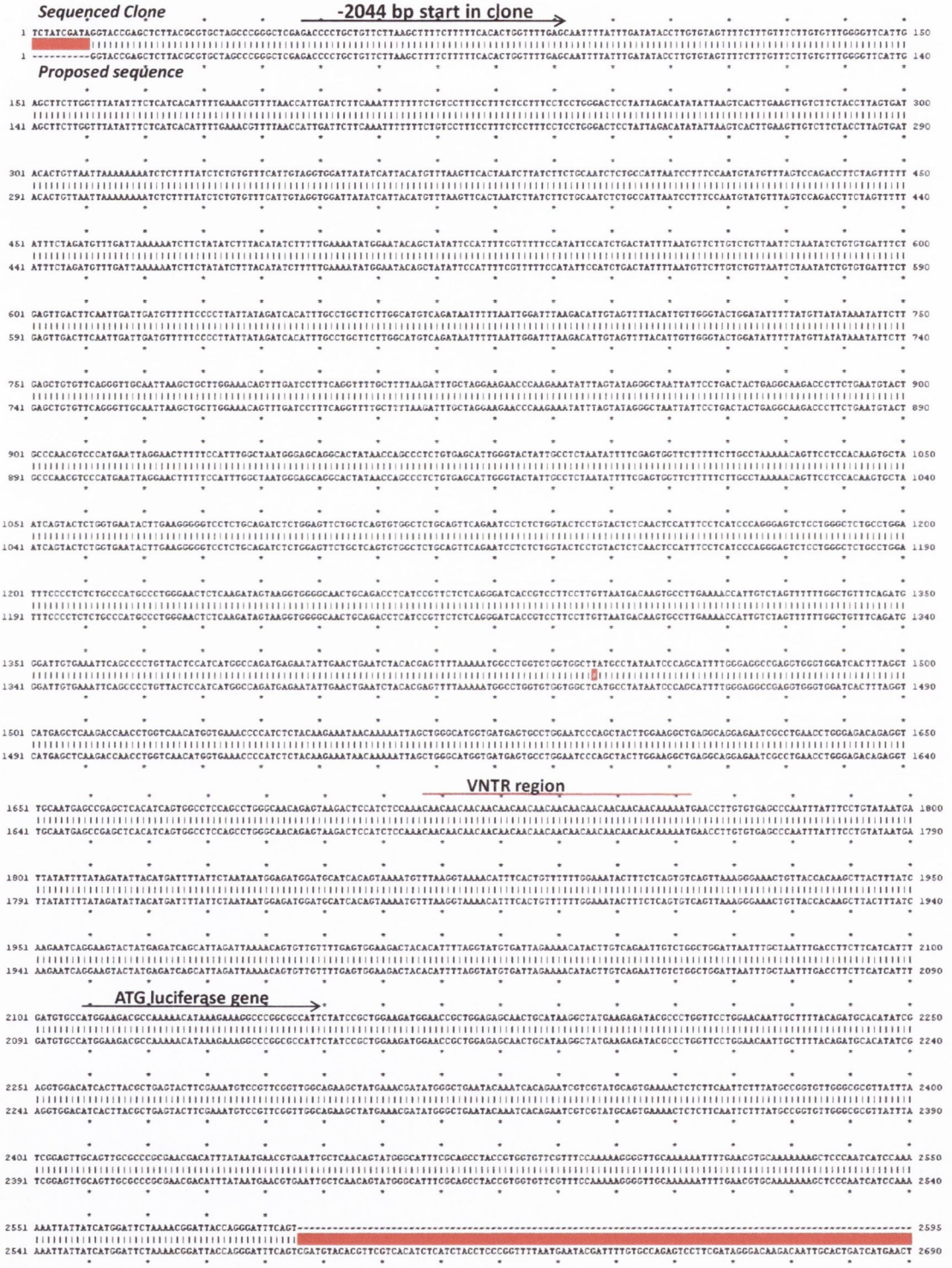
14RCpGL3 aligned to proposed sequence



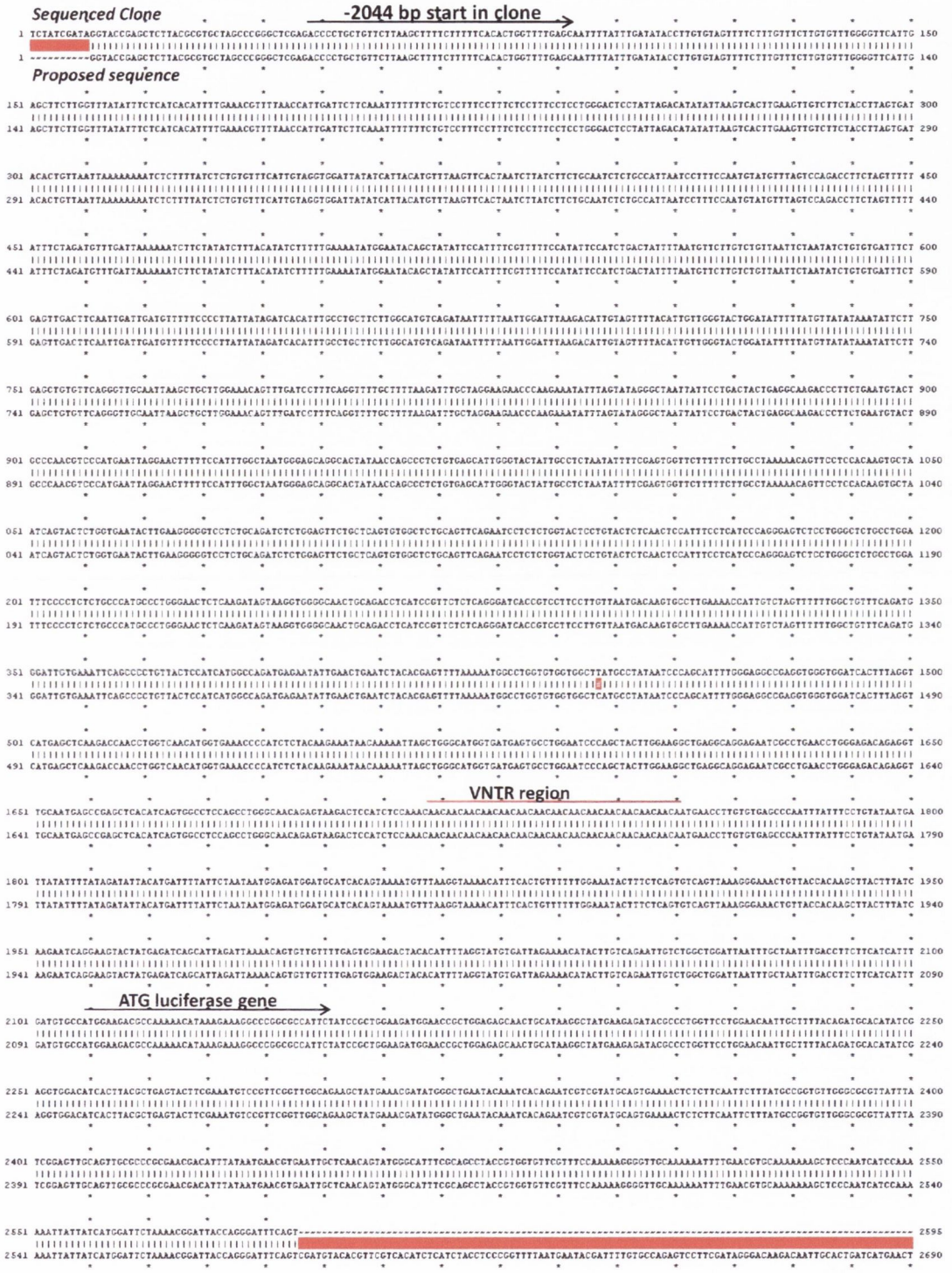
14RApGL3 aligned to proposed sequence



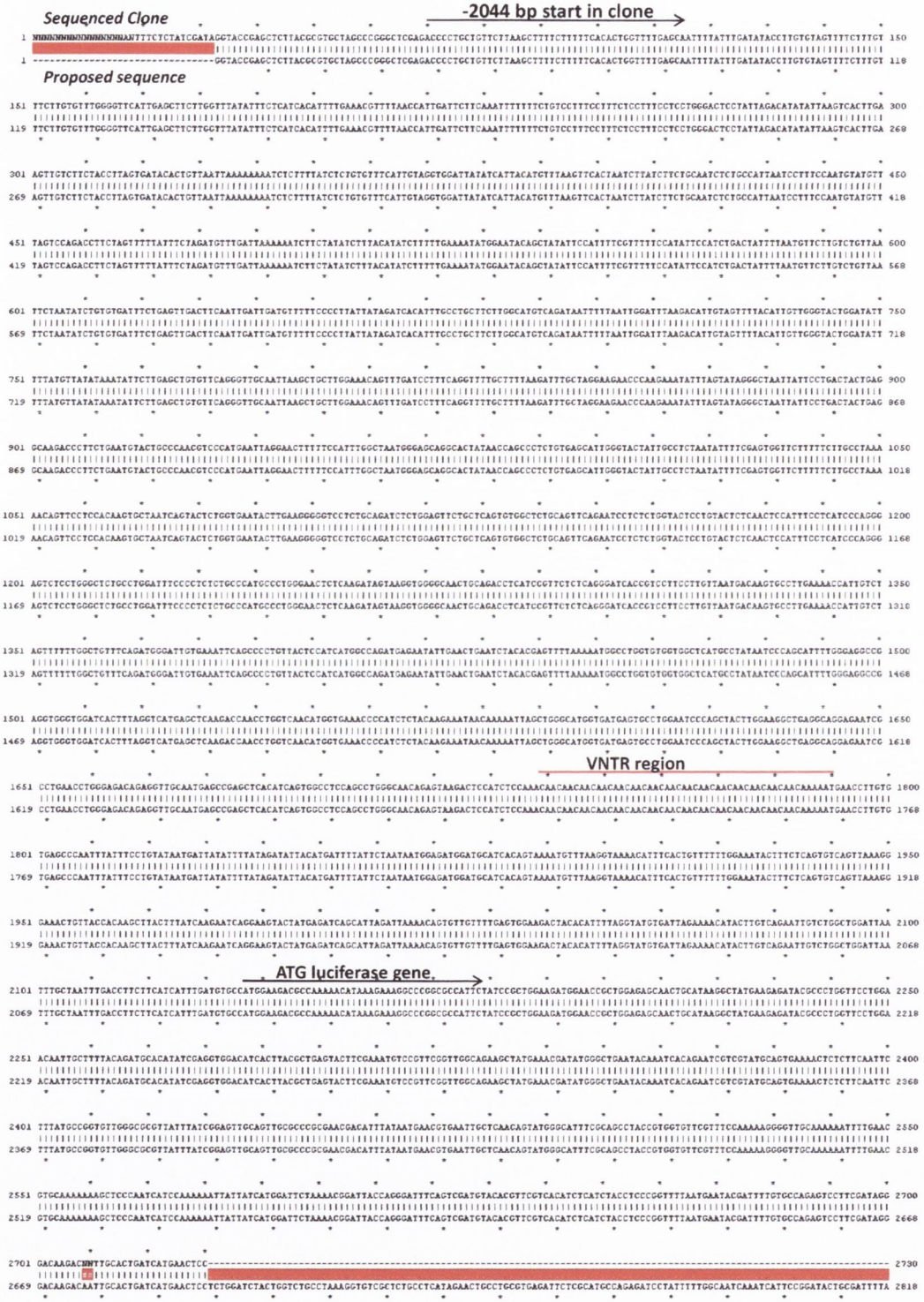
15RApGL3 aligned to proposed sequence



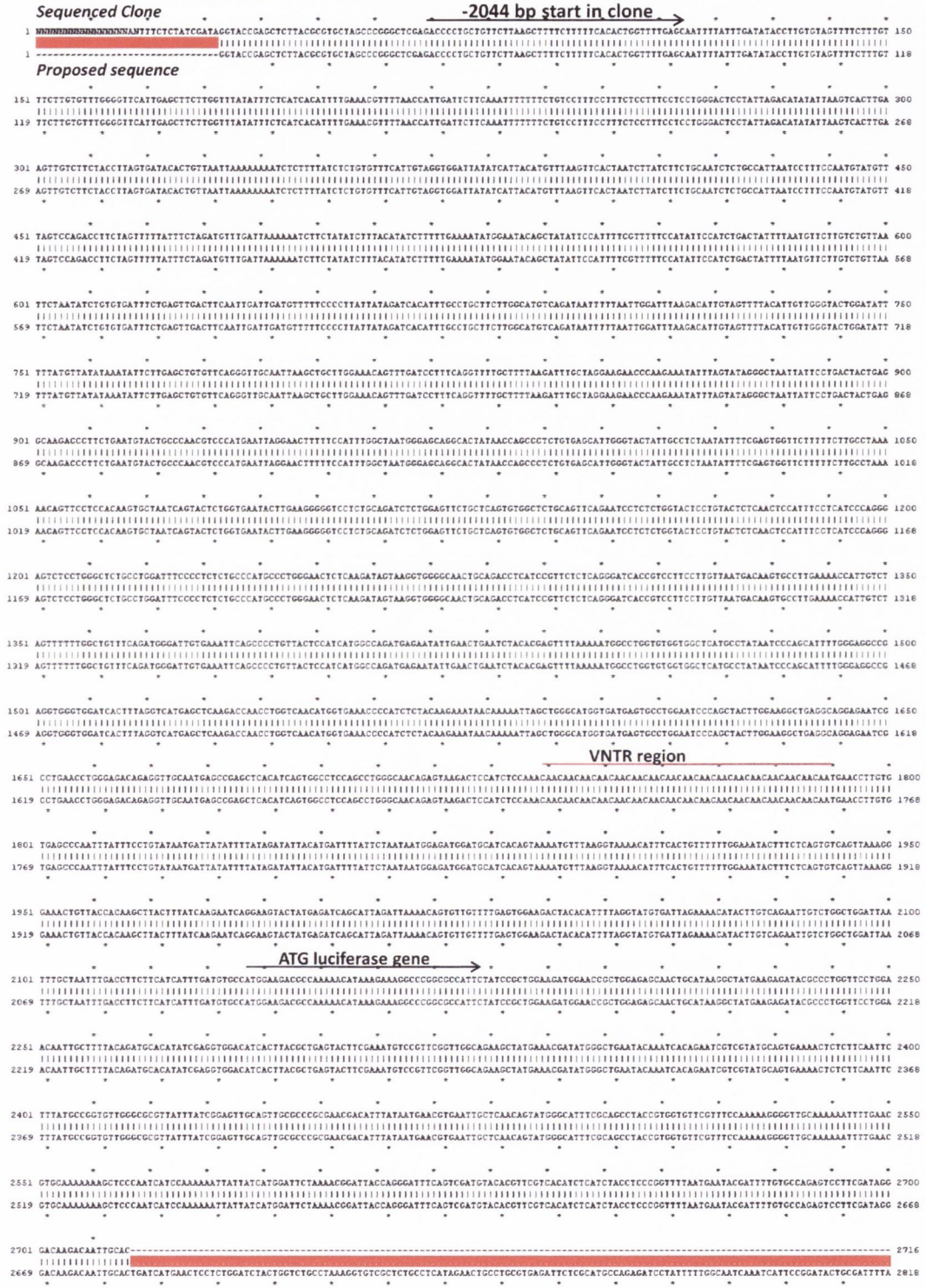
15RCpGL3 aligned to proposed sequence



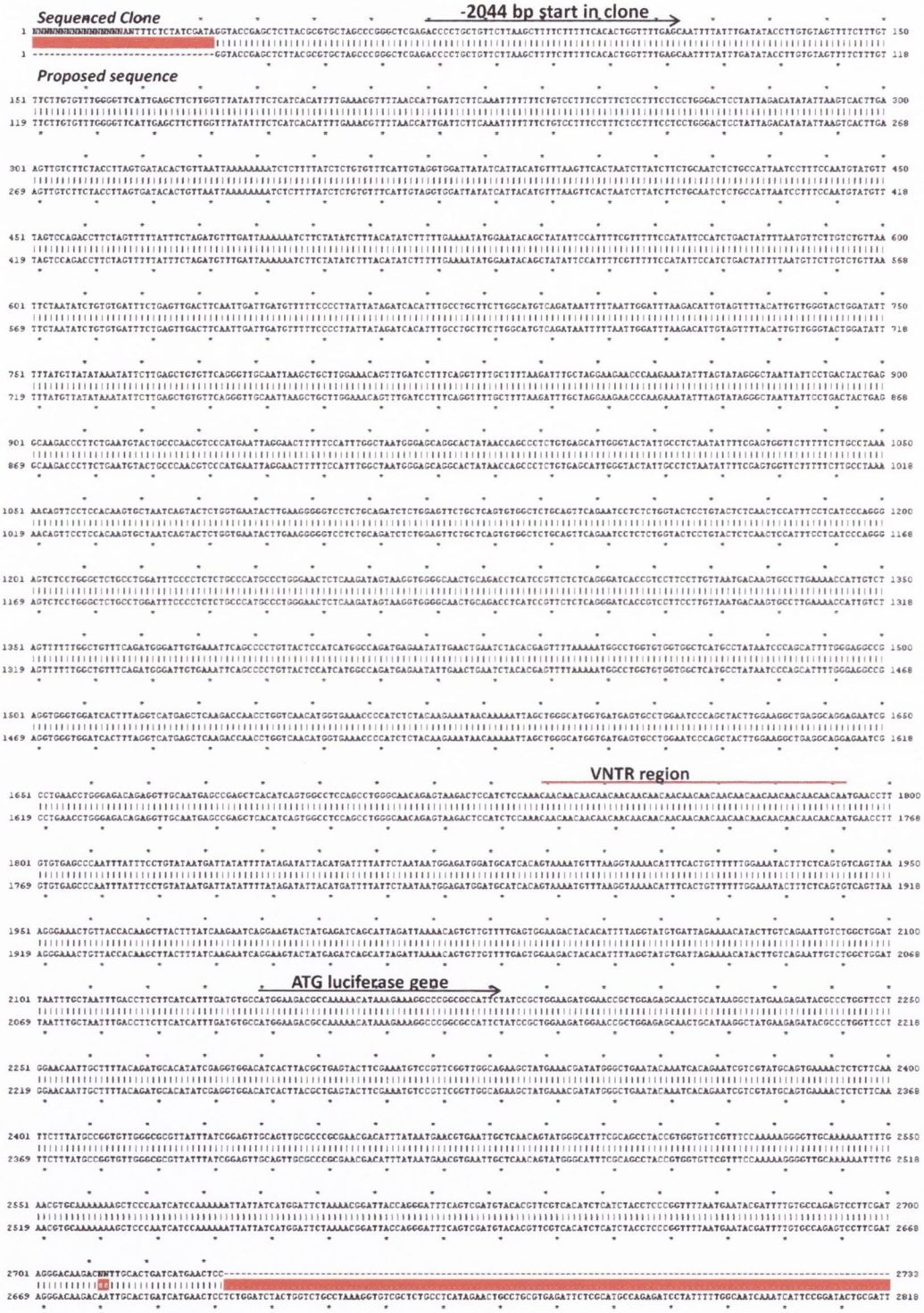
17RApGL3 aligned to proposed sequence



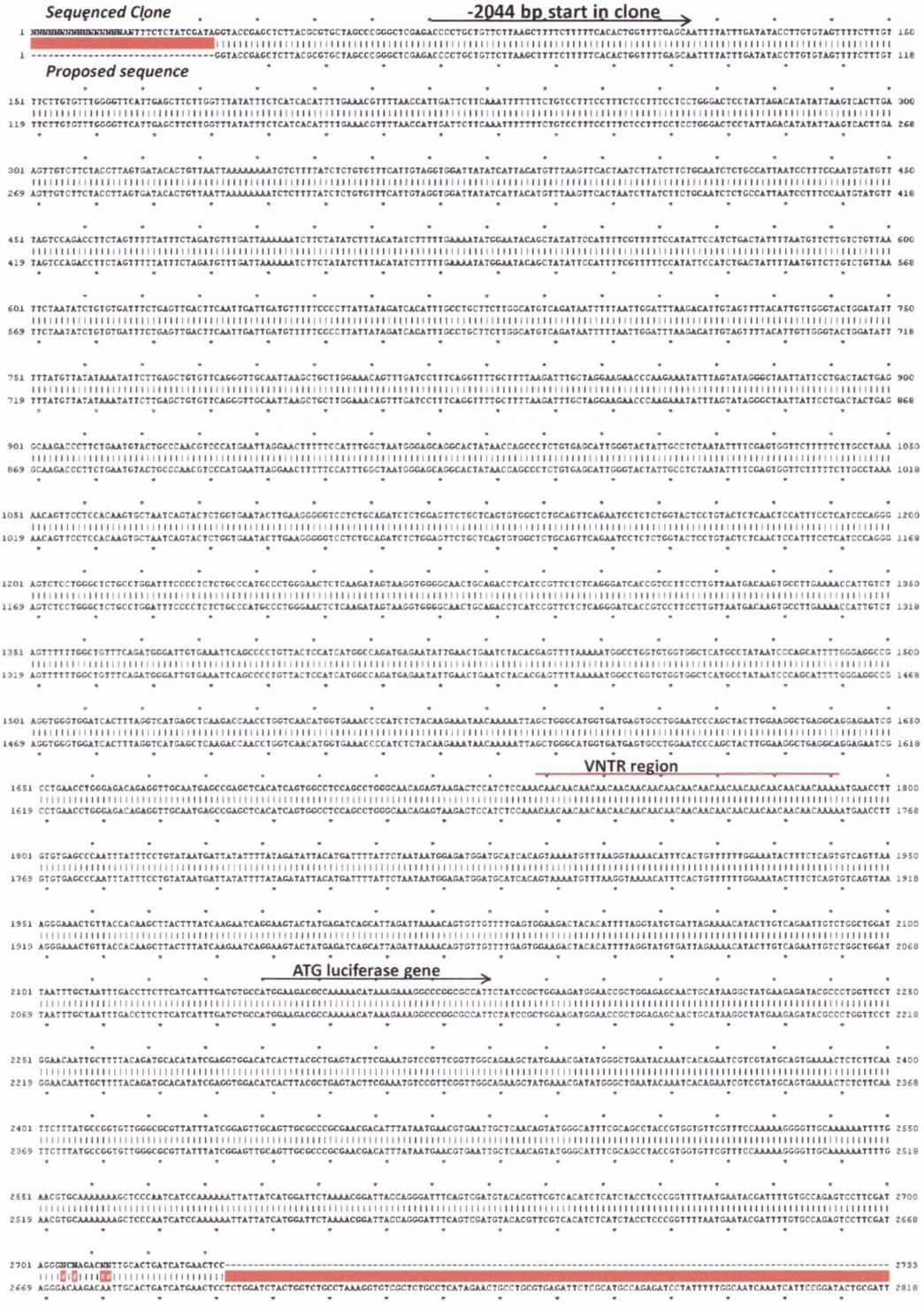
17RCpGL3 aligned to proposed sequence



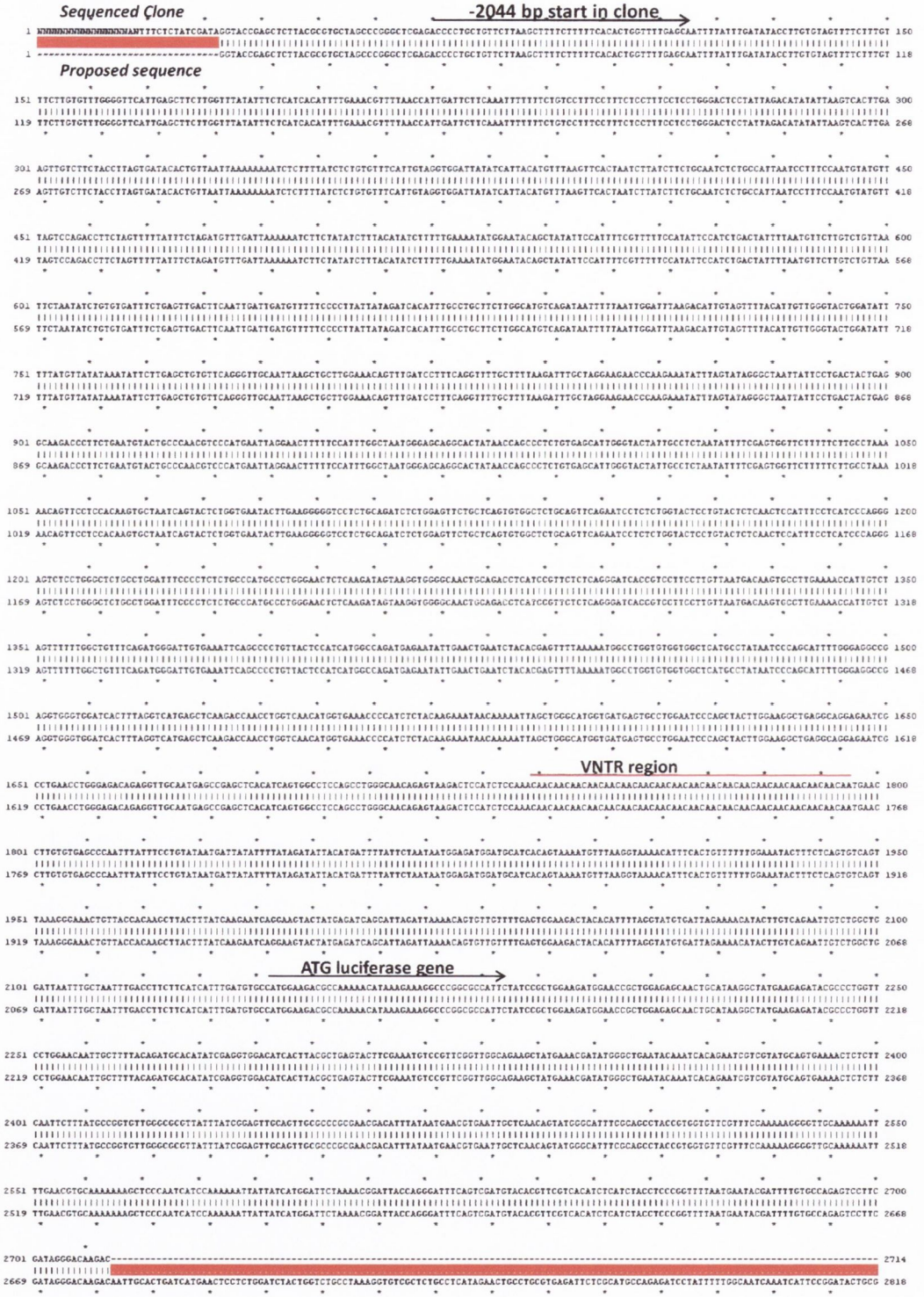
18RCpGL3 aligned to proposed sequence



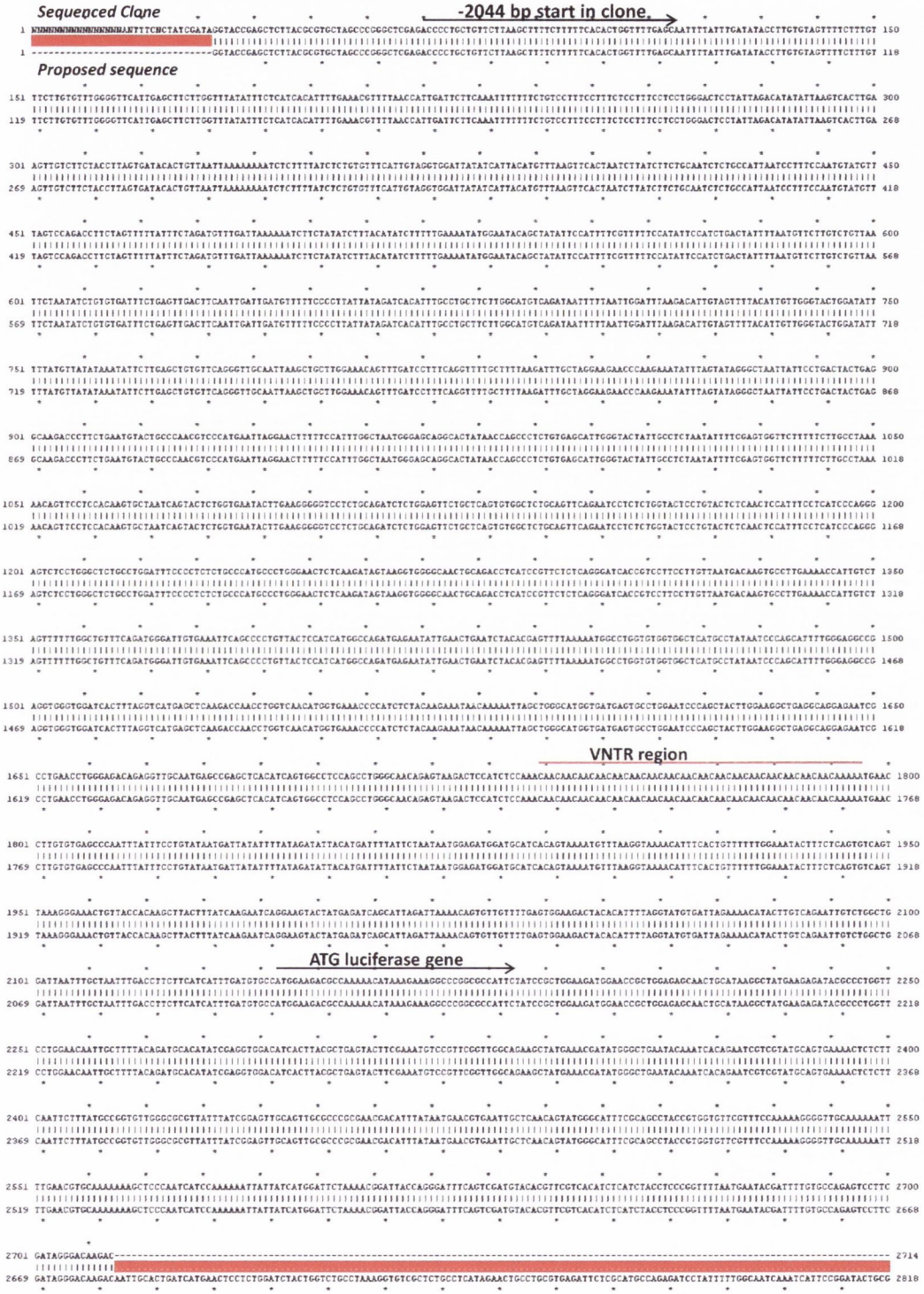
18RApGL3 aligned to proposed sequence



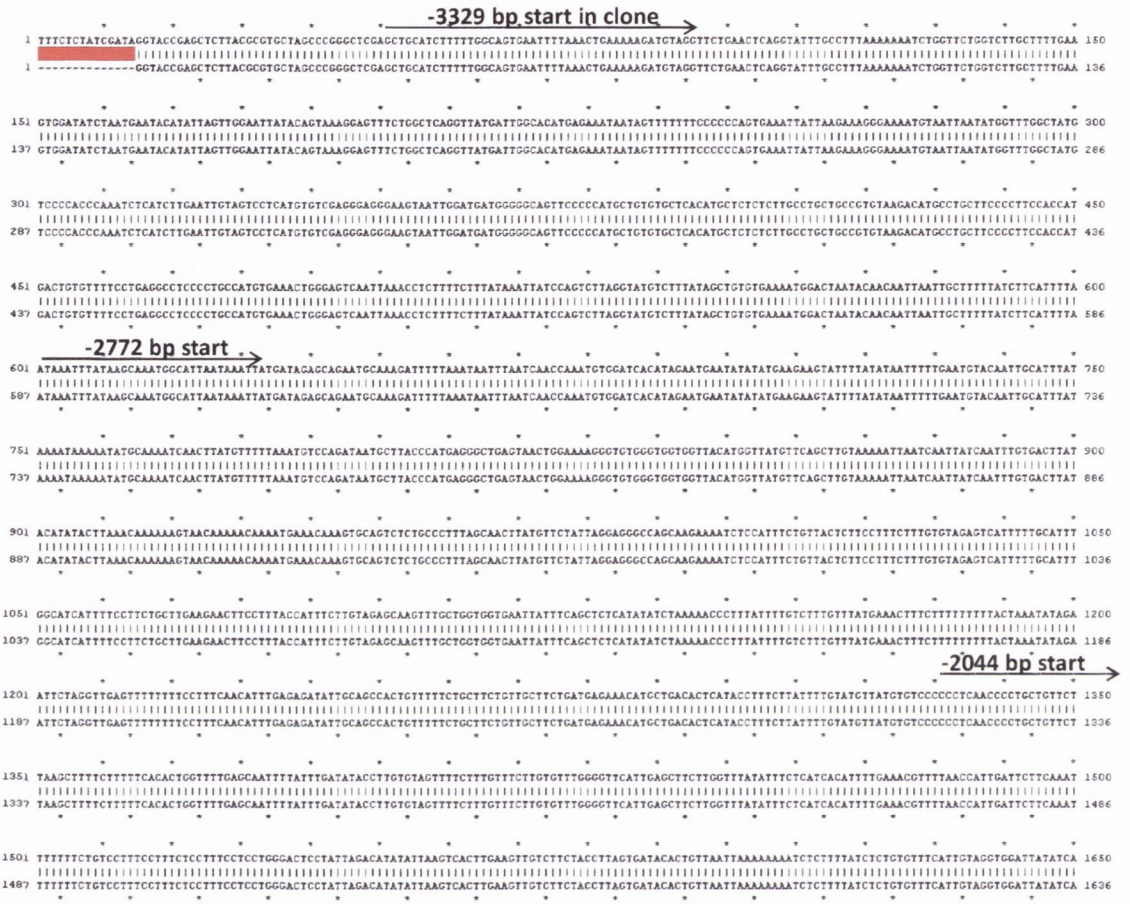
19RCpGL3 aligned to proposed sequence



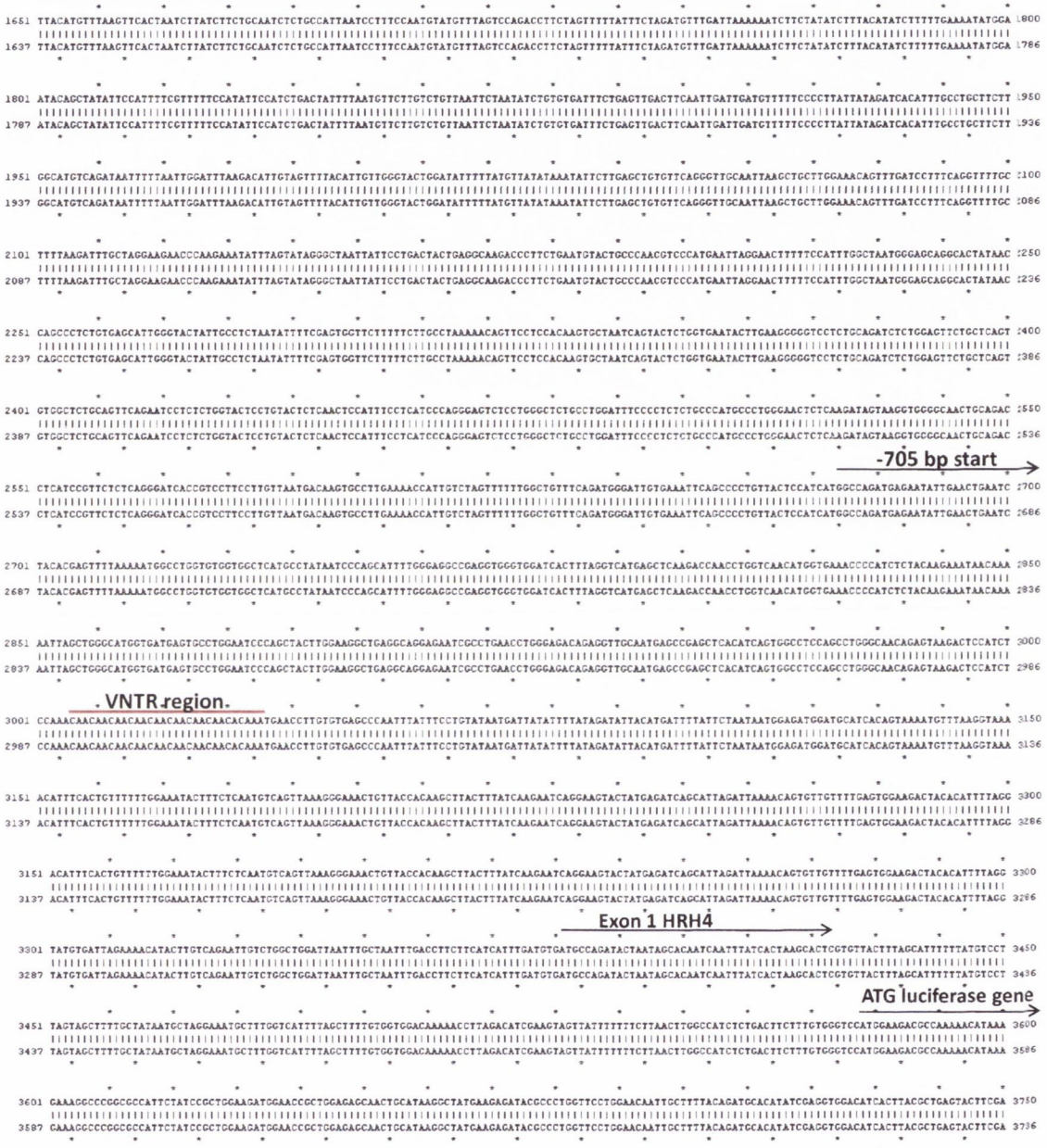
19RApGL3 aligned to proposed sequence



-3329 bp Ex1 10RpGL3 aligned to proposed sequence



-3329 bp Ex1 10RpGL3 aligned to proposed sequence (continued)



APPENDIX 4

Conferences Attended:

European Histamine Research Society XXXIX meeting – 13th – 17th July 2010, Durham UK.

Neuroscience Ireland Annual meeting – 2nd-3rd Sept 2010, University College Dublin, Ireland

Translational Medicine 2010: Discovery to Delivery – 7th October 2010, Trinity College Dublin, Ireland

13th Annual Meeting of the Institute of Molecular Medicine – 12th-13th November 2010, Trinity Centre for Health Sciences at St. James' Hospital, Dublin, Ireland.

COST Action BM0806 Management Committee Meeting and Working Groups Meeting – 28th–29th March 2011, Krakow, Poland.

Royal Academy of Medicine in Ireland, Section of Biomedical Sciences Annual Meeting 2011 – 22nd June 2011, Trinity College Dublin, Ireland

Presentation:

Micallef S, Quinn E, Morris D, Sasse A. Linkage Disequilibrium Mapping on the Human Histamine 4 Receptor Gene. All Ireland School of Pharmacy – 33rd Research Seminar – 18th–19th April 2011, Royal College of Surgeons Ireland, Dublin, Ireland (*poster*).

Powe L, Jung M, Mohsin A, Micallef S, Sasse A. Investigating a Rare Single Nucleotide Polymorphism in the Human H4 Receptor Gene. 2nd Year Medicine Research Poster Symposium – 2nd June 2011, School of Medicine, Trinity College Dublin, Ireland (*poster*).

Micallef S, Quinn E, Morris D, Sasse A. Genetic polymorphism in human histamine 4 receptor gene and schizophrenia. Neuroscience Ireland Annual Meeting – 1st-2nd September 2011, National University of Ireland, Maynooth, Ireland (*poster*).

Micallef S, Quinn E, Morris D, Sasse A. Genetic polymorphism in human histamine 4 receptor gene and schizophrenia. Neuroscience 2011, Society of Neuroscience – 12th- 16th November 2011, Washington DC Convention Centre, USA (*poster*).

Micallef S, Quinn E, Morris D, Sasse A. Genetic polymorphism in human histamine 4 receptor gene and schizophrenia. TCD postgraduate research symposium – January 2012, Trinity College Dublin, Ireland (*Oral*).

Micallef, S., Quinn, E., Morris, D., Sasse, A. Investigating the effect of a triplet repeat expansion on the human histamine 4 receptor gene. All Ireland School of Pharmacy 34th Research Seminar, University College Cork, Ireland, 2nd – 3rd April 2012 (*poster*).

Micallef S, Quinn E, Morris D, Sasse A. Microsatellite repeats expansions in human histamine 4 receptor gene. EHRS meeting 2012 – 2nd - 5th May, Belfast, Northern Ireland. (*2nd Poster prize award and Early stage researcher award*)

Micallef, S., Krause, J., Sasse, A. Characterisation of the promoter region of the human histamine 4 receptor. All Ireland School of Pharmacy 35th Research Seminar, University of Ulster, Northern Ireland, 25-26th March 2013 (*poster*).

Publications:

Micallef, S., Quinn, E., Morris, D., Sasse, A. Microsatellite Repeat Expansions in Human Histamine (4) Receptor Gene.Joint COST BM0806/41st EHRS Meeting, Belfast, UK, 2-5 May 2012. Inflam. Research (2012) 61(2), Suppl. 2, S52.

Micallef S, Stark H, Sasse A. Polymorphisms and genetic linkage of histamine receptors. Life Sci 2013; **93**(15)**:** 487-494.

Micallef, S. and Sasse, A., Identification of the active promoter of human histamine H4 receptor. *In preparation.*

Micallef, S. Quinn, E., Morris, D. and Sasse, A., Microsatellite repeat expansions determine HRH4 promoter activity. *In preparation.*