



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.

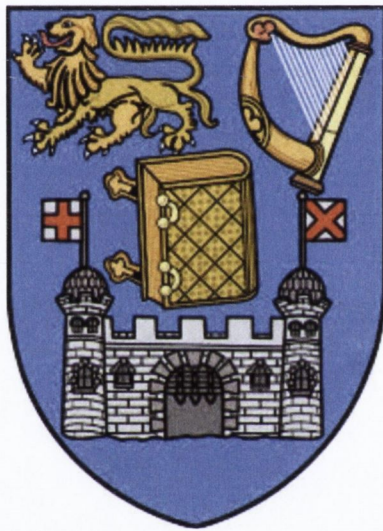
Access Agreement

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.

Genetic and molecular studies of rare variation in schizophrenia

Ciara Fahey



A thesis submitted to the University of Dublin for the degree of

Doctor of Philosophy

Department of Psychiatry

University of Dublin

Trinity College-March 2014

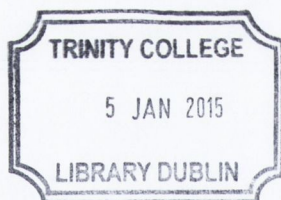
Declaration

University of Dublin Trinity College – March 2014

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, except where stated otherwise. 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.

Signed:

Carla Ahearn



Thesis 10437

Summary

Schizophrenia (SCZ) is a severe and debilitating disorder and it has a worldwide prevalence estimated at approximately 1%. The underlying pathogenesis of SCZ is poorly understood and there is a great need for diagnostic tools and new therapeutic interventions. It has a very strong genetic component but its genetic architecture remains to be elucidated.

In this study I sought to investigate rare variants associated with SCZ using a number of approaches. My first approach was to analyse a hexanucleotide repeat expansion in the *C9orf72* gene. This variant is associated with ALS/FTD, and clinical resemblance has been reported between frontal lobe syndrome and negative symptoms in SCZ. I applied this investigation to 1,271 Irish psychosis cases and 1,283 controls. I did not find any of the Irish Psychosis case control samples to have a hexanucleotide expansion with >30 repeats. I did however find seven samples with >22 repeats. Overall I saw no difference in the distribution of the mutation between case and control samples. The functional impact and importance of the repeat expansion at different lengths remains to be elucidated. In addition, I used available GWAS data to study the relationship between the founder haplotype and the repeat expansion in this large homogenous sample. My data supports the claim that carriers of the expansion arose from a single common founder, or at the very least that if the expansion arose multiple times, it has done so on the same haplotype background. This study was published in *Neurobiology of Aging*.

My second approach was to investigate a rare variant in the *PAK7* gene in collaboration with other members of my group. A gene-based analysis of large (>100kb), rare copy number variants (CNVs) conducted in the Wellcome Trust Case Control Consortium 2 (WTCCC2) SCZ sample, found an association between duplications at chr20p12.2 and SCZ ($P = 0.007$). I set out to use PCR and Sanger Sequencing to discover the specific breakpoints for this duplication. I identified the event as being a tandem duplication of 148,951bp at chr20:9,684,767-9,833,717 (hg18) with a 15bp sequence inserted between the first and second copies of the repeated sequence. This study was published in *Human Molecular*

Genetics. I then utilised UK10K data to follow up on my *PAK7* duplication analysis, and investigate if there were any exonic *PAK7* sequence variants present in this data set. I performed functional annotation of the genetic variants in the neurodevelopmental and obesity samples using Annovar. Eighty two variants were identified in the cases and controls. I then carried out a carrier-based association analysis on this data, but found no statistically significant differences in the rate of mutation carriers between cases and controls at *PAK7*.

My third approach to rare variant analysis was to use a new method namely High Resolution Melt (HRM) analysis. My initial goal was to test the sensitivity and specificity of the HRM method by using it to detect known mutations in HapMap samples, with the intention of then applying it to the screening of SCZ candidate genes in an Irish psychosis case and control sample. I applied HRM to the exons of two SCZ associated genes *PAK7* and *VIPR2*. I choose HapMap samples to test the efficiency of this method, as DNA sequence information for all of these samples is publicly available on online databases. Its overall lack of both a high sensitivity and a high specificity meant that it was not a method that would be useful for identifying rare variants.

Finally, I carried out a functional approach to rare variant analysis. A targeted exome study by my group identified an excess of LoF variants in an Irish psychosis case sample compared to controls. One of the LoF variants identified was a nonsense mutation at R156 in the *FYN* gene. *FYN* regulates the phosphorylation and trafficking of NMDA Receptors (NMDARs), acting as a point of convergence for many signalling pathways regulating NMDARs. My approach to investigate this variant was to design plasmids containing green fluorescent protein (GFP)-tagged wildtype and mutant versions of the gene and transfect them into neuroblastoma cell lines first to test the efficiency, and to then move on to primary cortical neurons from newborn mice. I did not see a significant difference in the expression of *FYN* in the WT compared to the Mut cells. I did however identify a significant difference between WT and Mut with regard to Phospho-AKT expression in the *FYN* Mut transfected cells ($p < 0.001$). AKT is a member of the kinase family, and plays a major regulatory role in cellular

function; a decrease in the phosphorylated version of AKT has been reported in first episode SCZ patients.

Acknowledgments

First and foremost I would like to thank Derek for giving me the opportunity to complete my PhD. Thank you so much for your incredible patience, kindness and understanding. You have been an absolute delight to work for, thank you so so much for everything.

To my unofficial second in command Emma, you have been an incredible pal to me since I started in the group, and with all my madness over the last few years, you have been there to listen to me go on and on and on! May the red wine drinking, nonsense talk and non-frequent runs in the morning continue for a long long time to come.

Elaine, thank you so much for all of your help, and for keeping your head, when I was losing mine with the GAIL. Ric, thank you so much for looking out for me, supplying me with lots of entertainment and tolerating my incredibly poor British accent! You are an incredibly kind nerd. To Paul for educating us all on what goes on down on the farm, and also being a very helpful nerd. To my wonderful office mates over the years, it has been a delight talking absolute nonsense with you all. To Siobhan for bringing in lovely cookies, and rambling in the corner. To Sarah for being a great pal, and a wonderful PhD-mate. To Daniela, Ines and Stefania for guiding me through the challenge that is functional analysis, and the rest of the Neuropsychiatric genetics group

Finally to my family: Jelly and Pa for being the bestest south side neighbours and OC drinking buddies a sister could have. To Aoibheann, Ian, Ruairi and, Cara for always looking after me, and hosting wonderful Seisúns in the Green House. Denise, Declan, Saoirse, Iarla, Sadhbh, Brian, Mag and Jessica Fletcher thank you all for your support. Finally Mam and Dad, thank you so much for everything over the years, and for treating your mature! student daughter to lovely meals out, wonderful holidays and spins in Connemara . I dedicate this thesis to you, the best parents a girl could have. Love ye lots.

Statement of Work

This work was undertaken by the author with the Neuropsychiatric Genetics Research Group, Department of Psychiatry and Institute of Molecular Medicine, Trinity College Dublin.

List of Abbreviations

AA	Amino Acid	DSM-IV	Diagnostic Statistical Manual IV
AAS	Amino Acid Substitution	DST	Dystonin
ADHD	Attention Deficit Hyperactivity Disorder	DZ	Dizygotic
AGP	Autism Genome Project	EB	Elution Buffer
AMP	Ampicillin	EEG	Electroencephalography
AMPA	2-amino-3-(3-hydroxy-5-methyl-isoxazol-4-yl) propanoic acid	EPS	Extra-Pyramidal Side-effects
ASD	Autism Spectrum Disorders	EtBr	Ethidium Bromide
AS-PCR	Allele Specific-Polymerase Chain Reaction	FAM	6-carboxyfluorescein
ATAV	Association Tests for Annotated Variants	FBS	Foetal Bovine Serum
BDNF	Brain-Derived Neurotrophic Factor	FIS	Functional Impact Score
BSA	Bovine Serum Albumin	FYN	Tyrosine-protein kinase FYN
BWA	Burrows-Wheeler Aligner	GAI	Illumina Genome Analyzer II
CAM	Cell Adhesion Molecule	GATK	Genome Analysis Toolkit
CDCV	Common Disease, Common Variant	GFP	Green Fluorescent Protein
cDNA	Complementary DNA	GO	Gene Ontology
ChIP-Seq	Chromatin Immunoprecipitation Sequencing	GRIN2B	Glutamate Receptor, Ionotropic, N-methyl D-aspartate 2B
Chr	Chromosome	GRIP1	Glutamate Receptor Interacting Protein 1
CNV	Copy Number Variant	GWAS	Genome-Wide Association Study
CVRV	Common disease, Rare Variant	H2O	Hydrogen Dioxide
DAPI	4',6-diamidino-2-phenylindole	HPRD	Human Protein Reference Database
ddNTPs	Dideoxy Nucleotide Triphosphate	HTG	Hypertriglyceridemia
DGV	Database of Genomic Variants	ICD-10	International Classification of Diseases-10
dH₂O	distilled Hydrogen Dioxide	ID	Intellectual Disability
DISC1	Disrupted-in-Schizophrenia 1	IDT	Integrated DNA Technologies
DMEM	Dulbecco's Modified Eagle Medium	IGV	Integrative Genomics Viewer
DNA	Deoxyribonucleic Acid	PBST	Phosphate-Buffered Saline-Triton
dNTP	Deoxyribonucleotide Triphosphate	PCR	Polymerase Chain Reaction

ISC	The International Schizophrenia Consortium	ORF	Open Reading Frame
LB	Lysogeny Broth	PP2	Polyphen 2
LD	Linkage Disequilibrium	PSD95	Post-Synaptic Density Protein 95
LMW	Low Molecular Weight	PSP	Post-Synaptic Proteome
LOD	Log of Odds	QC	Quality Control
LoF	Loss of Function	RNA	Ribonucleic Acid
MAF	Minor Allele Frequency	SAP	Shrimp Alkaline Phosphatase
MAP2	Microtubule-Associated Protein 2	SCID-P	Structured Clinical Interview for DSM-IV-Patient Edition
MECP2	Methyl CpG Binding Protein 2	SD	Standard Deviation
MHC	Major Histocompatibility Complex	SDS	Sequence Detection System
MR	Mental Retardation	SIFT	Scale-Invariant Feature Transform
MRI	Magnetic Resonance Imaging	SNP	Single Nucleotide Polymorphism
mRNA	Messenger Ribonucleic Acid	SNVs	Single Nucleotide Variant
MUT	Mutant	SOC	Super Optimal broth with Catabolite repression
MZ	Monozygotic	SCZ	Schizophrenia
NaOAc	Sodium Acetate	TAE	Tris Base, Acetic Acid and EDTA
NCBI	National Centre for Biotechnology Information	UCSC	University College, Santa Cruz
NEB	New England Biolabs	UTR	Untranslated Region
NGS	next generation Sequencing	WT	Wildtype
NLGN	Neurologin	XTERM	XTerminator
NMDA	N-methyl-D	ZNF804A	Zinc Finger Protein 804A
OR	Odds Ratio		

Publications arising from this thesis

Analysis of the hexanucleotide repeat expansion and founder haplotype at C9ORF72 in an Irish psychosis case-control sample.

Fahey C, Byrne S, McLaughlin R, Kenna K, Shatunov A, Donohoe G, Gill M, Al-Chalabi A, Bradley DG, Hardiman O, Corvin AP, Morris DW. *Neurobiol Aging*. 2014 35(6):1510.e1-5. PMID: 24411481.

An inherited duplication at the gene p21 Protein-Activated Kinase 7 (PAK7) is a risk factor for psychosis.

Derek W Morris PhD, Richard D Pearson PhD, Paul Cormican PhD, Elaine M Kenny PhD, Colm T O'Dushlaine PhD, Louis-Philippe Lemieux Perreault PhD, Eleni Giannoulatou PhD, Daniela Tropea PhD, Brion S Maher PhD, Brandon Wormley PhD, Eric Kelleher MD, **Ciara Fahey MSc**, Ines Molinos PhD, Stefania Bellini BSc, Matti Pirinen PhD, Amy Strange PhD, Colin Freeman PhD, Dawn L Thiselton PhD, Rachel L Elves PhD, Regina Regan PhD, Sean Ennis PhD, Timothy G Dinan MD PhD, Colm McDonald MD PhD, Kieran C Murphy MD PhD, Eadbhard O'Callaghan MD PhD, John L Waddington PhD, Dermot Walsh MD, Michael O'Donovan MD PhD, Detelina Grozeva MD, Nick Craddock MD PhD, Jennifer Stone PhD, Ed Scolnick PhD, Shaun Purcell PhD, Pamela Sklar MD PhD, Bradley Coe PhD, Evan E Eichler PhD, Roel Ophoff MD PhD, Jacobine Buizer PhD, Jin Szatkiewicz PhD, Christina Hultman PhD, Patrick Sullivan MD FRANZCP, Hugh Gurling MD PhD, Andrew McQuillin MD PhD, David St Clair MD PhD, Elliott Rees, George Kirov MD PhD, James Walters MD PhD, Douglas Blackwood MD PhD, Mandy Johnstone MD PhD, Gary Donohoe PhD, International Schizophrenia Consortium, SGENE+ Consortium, Francis A O'Neill MD PhD, Wellcome Trust Case Control Consortium 2, Kenneth S Kendler MD PhD, Michael Gill MD, Brien P Riley PhD, Chris C A Spencer PhD, Aiden Corvin MD PhD. *Hum Mol Genet*. 2014 Epub Jan 28. PMID: 24474471.

Contents

Declaration 2

Summary..... 3

Acknowledgments 6

Statement of Work 7

List of Abbreviations 8

Publications arising from this thesis 9

List of Figures 14

List of Tables 18

Introduction 20

 1.1 Genetic Variation 20

 1.2 Inheritance of complex diseases..... 20

 1.3 Schizophrenia Overview 22

 1.4 Neurobiology of schizophrenia 25

 1.5 Genetics of schizophrenia..... 26

 1.5.1 Mapping the genetics of complex diseases 27

 1.5.2 Linkage studies..... 27

 1.5.3 Linkage studies of schizophrenia 28

 1.6 Issues with genetic association studies 29

 1.6.1 Population stratification 29

 1.6.2 Power 30

 1.6.3 Pre-GWAS association studies of schizophrenia 31

 1.7 Genome-wide association studies in schizophrenia 31

 1.7.2 Copy Number Variation 34

 1.7.3 Next Generation Sequencing and Schizophrenia..... 36

1.8 My studies of rare variants involved in schizophrenia.....	42
1.8.1 Analysis of a hexanucleotide repeat in <i>C9ORF72</i> gene	42
1.8.2 Rare Variant Analysis of Schizophrenia Risk Genes	43
1.8.3 Investigation of a Loss of Function Variant in <i>FYN</i>	44
Chapter 2: Analysis of a hexanucleotide repeat expansion in the <i>C9ORF72</i> gene	46
2.1 Rare variation and the cross over between ALD/FTD and SCZ	47
2.2 Materials and Methods.....	49
2.2.1 Subjects.....	49
2.2.2 Genetic analysis	50
2.2.3 Haplotype analysis	52
2.4 Results.....	54
2.4.1 Genotyping of the repeat expansion	54
2.5 Discussion	64
Chapter 3: Rare Variant Validation: Investigation of Rare Variation in the <i>PAK7</i> gene	66
3.1 Studies of a duplication event in <i>PAK7</i>	67
3.2 Materials and Methods.....	69
3.2.1 <i>PAK7</i> breakpoint analysis.....	69
3.2.2 Gel Purification of PCR samples	70
3.3 Results.....	72
3.4 <i>PAK7</i> Association Analysis.....	73
3.4.1 UK10K Background	73
3.4.2 Variant Identification in UK10K data.....	74
3.4.3 Analysis of 82 variants in UK10K data using Annovar	76
3.4.4 Association analysis of <i>PAK7</i> UK10K data	78
3.5 Discussion	81
Chapter 4: Rare Variant Detection	

Testing the Specificity and Sensitivity of High Resolution Melt Analysis (HRM) for Rare Variant Detection.....	83
4.1 An alternative method of rare variant detection.....	84
4.2 HRM Methodology.....	84
4.2 Gene selection	85
4.2.1 <i>PAK7</i>	85
4.2.2 <i>VIPR2</i>	86
4.3 Materials and Methods.....	86
4.3.1 Samples.....	86
4.3.2 Primer Design.....	86
4.3.3 Sequencing of exonic fragments.....	89
4.4 Results.....	90
4.4.1 HRM Sensitivity Results	93
4.4.2 HRM Specificity Results.....	95
4.6 Discussion	97
Chapter 5: Investigation of a Loss of Function Variant in <i>FYN</i>	99
5.1 Identification of a rare variant in the <i>FYN</i> gene	100
5.2 My approach to further understanding the affect of this LoF variant.....	101
5.3 Plasmid Design.....	102
5.4 Materials and Methods.....	104
5.4.1 Plasmid Preparation.....	104
5.4.1.5 Plasmid Purification	108
5.4.2 Preparation of SH-SY5HY Neuroblastoma Cell Lines.....	110
5.4.3 Preparation of Mouse Cell Cultures.....	112
5.4.4 Transfection of Plasmids into Neuroblastoma cell lines	113
5.5 Immunohistochemistry.....	114

5.5.1 Western Blot to test specificity of the FYN antibody	115
5.5.1.1 Protein extraction from wild type and mutant <i>FYN</i> tissue	116
5.5.1.11 Western Blot Analysis to test the specificity of the FYN antibody.....	120
5.5.2 Confocal microscopy.....	121
5.6 Site Directed Mutagenesis	122
5.6.1 Site Directed PCR protocol.....	123
5.6.2 DPN1 digest.....	123
5.7 Results.....	124
5.7.1 Transfection of plasmids and subsequent Immunohistochemistry analysis	124
5.7.2 Transfection of the <i>FYN</i> WT plasmid into Neuroblastoma cell lines.....	127
5.8 Transfection and immunohistochemistry of Phospho-AKT in neuroblastoma cell lines	132
5.8.1 Immunohistochemistry of the transfected FYN WT plasmid	135
5.8.2 Immunohistochemistry of the transfected FYN Mut plasmid.....	137
5.8.3 Results of Phospho-AKT transfection.....	139
5.9 Transfection and Immunohistochemistry analysis in primary mouse neuronal cultures	140
5.9.1 Analysis of both FYN WT and FYN Mut plasmids in primary neuronal cultures ..	141
5.9.2 Results of FYN analysis on primary neuronal cultures.....	142
5.10 Site directed Mutagenesis	143
5.10.1 Transfection of a plasmid created via site directed mutagenesis into neuroblastoma cell lines.....	146
5.11 Discussion	147
Chapter 6: Discussion.....	150
References	159

List of Figures

Figure 2.1 Breakdown of samples included in the Psychosis Case group.

Figure 2.2: Electropherogram output for case sample MON109 with 28 repeats.

Figure 2.3: Electropherogram output for case sample RPG5076 with 27 repeats.

Figure 2.4: Electropherogram output for control Sample C0106 with 26 repeats.

Figure 2.5: Electropherogram output for Control Sample C0082 with 25 repeats.

Figure 2.6: Electropherogram output for control Sample C2025 with 24 repeats.

Figure 2.7: Histogram showing the repeat length in both case and control samples. Portion of the graph displaying results for repeats 16 – 28 is magnified in upper right of graph.

Figure 2.8: Histogram showing the relationship between repeat lengths and haplotype in cases only. Portion of the graph displaying results for repeats 16 – 28 is magnified in upper right of graph.

Figure 2.9: Histogram showing the relationship between repeat lengths and haplotype in controls only. Portion of the graph displaying results for repeats 16 – 28 is magnified in upper right of graph.

Figure 2.10: Histogram showing the relationship between repeat lengths and haplotype in both cases and controls. Portion of the graph displaying results for repeats 16 – 28 is magnified in upper right of graph.

Figure 3.1: Agarose gel image of PCR 400bp fragments and negative sample.

Figure 3.2: Capillary sequence of the unique region in the middle of the two copies of the tandem repeat sequence at the *PAK7* duplication. The first copy of the sequence ends at chr20:9,833,717(hg18) and the second copy starts with sequence from chr20:9,684,767(hg18), thereby defining the breakpoints. A 15bp sequence is present between the two copies.

Figure 3.3: Breakdown of Neurodevelopmental and obesity samples in the UK10K data.

Figure 3.4 Flowchart of breakdown of variants in the UK10K data set.

Figure 4.1 Melt curves of denatured double stranded DNA

Figure 4.2: Analysis of HRM data from VIPR2-Exon 3. Different genotypes for rs2270313 are highlighted in different colours. A/G=red, G/G= Blue. A. The aligned melt curve shows discrimination of the variants in their melt behaviour. B. The derivative melt curve allows me to set out pre and post-melt regions for more accurate analysis. C. The difference plot allows me to visualise small alterations in melt curves. D. The melt curve allows me to discriminate differences in the change in fluorescence for individual samples.

Figure 4.3: Derivative melt curves for hapmap samples NA12762 and NA12761 from Exon11 of the VIPR2 gene. Based on melt curve analysis, samples were subsequently analysed using Sanger sequencing and showed NA12762 to contain a rare variant (1KG_7_158824695) which has been reported in the 1000 genomes data.

Figure 5.1: Diagram of *FYN* protein with functional domains and site of variant under investigation.

Figure 5.2: Expression clone with CMV promoter. Diagram of the plasmids supplied by Genecopoeia. Plasmids contain a CMV promoter. The insert is located at site marked ORF (open reading frame) in green and is *GFP* tagged. There are two selection markers; neomycin and ampicillin.

Figure 5.3: Schematic of the plasmids used in transfection. 1. *GFP* plasmid 2. *FYN* WT (wildtype) plasmid tagged by *GFP*. 3. *FYN* MUT (mutant) plasmid tagged by *GFP*. The plasmid sizes are provided.

Figure 5.4: Western blot analysis to test the specificity of my *FYN* antibody.

Figure 5.5: 24hr Amaxa Transfection with *GFP* plasmid. A. Neuroblastoma transfected cell lines can be distinguished from non-transfected neurons upon observing staining with *GFP*. B. *FYN* antibody staining C. MAP2 staining indicates the presence of synapses. D. All three stains can be observed together.

Figure 5.6: 24hr Amaxa Transfection with WT plasmid. A. Neuroblastoma transfected cell lines can be distinguished from non-transfected neurons upon observing stain with *GFP*. B. *FYN* antibody staining C.MAP2 staining indicates the presence of synapses. D. All three stains can be observed together.

Figure 5.7: 24hr Amaxa Transfection with Mutant plasmid. A. Neuroblastoma transfected cell lines can be distinguished from non-transfected neurons upon observing staining with *GFP*. B. *FYN* antibody staining C.MAP2 staining indicates the presence of synapses. D. All three stains can be observed together.

Figure 5.8: Scatter plot showing the expression for both *GFP* and *FYN* WT plasmids, allowing me to differentiate transfected from non-transfected.

Figure 5.9: Chart showing the expression levels of both *GFP* (A) and *FYN* (B) for *FYN* Mut and *FYN* WT cells.

Figure 5.9: 24hr Amaxa Transfection with Mutant plasmid A. Neuroblastoma transfected cell lines can be distinguished from non-transfected neurons upon observing native *GFP*. B. Phospho-akt antibody staining C.MAP2 staining indicates the presence of synapses. D. All three stains can be observed together with DAPI.

Figure 5.10: Scatter plot showing the expression for both *GFP* and *FYN* Mutant plasmids, allowing me to differentiate transfected from non-transfected.

Figure 5.11: Overall comparison of *FYN* WT V'S *FYN* Mut transfection.

Figure 5.12: 24hr Amaxa Transfection with wt plasmid A. Neuroblastoma transfected cell lines can be distinguished from non-transfected neurons upon observing native *GFP*. B. Phospho-akt antibody staining C.MAP2 staining indicates the presence of synapses. D. All three stains can be observed together with DAPI.

Figure 5.13: Overall *GFP* and Phospho-AKT expression of both WT and Mutant transfected and non-transfected plasmids.

Figure 5.14: 24hr Amaxa Transfection on primary cultures with Mutant plasmid. A *GFP* antibody stain. B. *FYN* antibody staining C. MAP2 staining indicates the presence of synapses. D. All three stains can be observed together with DAPI.

Figure 5.15: Scatter plot showing the expression for both *GFP* and Phospho-AKT with the Mutant plasmids.

Figure 5.16: Results of Phospho-AKT transfection

Figure 5.17: 24hr Amaxa Transfection on primary cultures with WT plasmid. A *GFP* antibody stain. B. *FYN* antibody staining C. MAP2 staining indicates the presence of synapses. D. All three stains can be observed together with DAPI.

Figure 5.18: 24hr Amaxa Transfection on primary cultures with Mutant plasmid. A *GFP* antibody stain. B. *FYN* antibody staining C. MAP2 staining indicates the presence of synapses. D. All three stains can be observed together with DAPI.

Figure 5.19: Chart showing the mean expression values for *FYN* WT and *FYN* Mut in primary cultures. The transfection rate was significantly lower for primary cultures compared to cell lines. The number of cells available for analysis is indicated at the bottom of each column. It was also difficult to differentiate between transfected and non-transfected cells.

Figure 5.20 Sequence of wild type and site directed mutant plasmids after Site Directed Mutagenesis showing the successful introduction of the mutant allele.

Figure 5.21 24hr Amaxa Transfection on neuroblastoma cell lines with the site directed Mutant plasmid. A. native *GFP*, B. *FYN* antibody staining, C. MAP2 staining indicates the presence of synapses. D. All three stains can be observed together with DAPI.

List of Tables

- Table 1.1:** Criteria for diagnosis of schizophrenia according to the Diagnostics Statistics Manual Classification IV (DSM-IV).
- Table 1.2:** Criteria for diagnosis of schizophrenia according to the International Classification of Disease 10 (ICD-10).
- Table 2.1:** Genotyping Sample Preparation PCR.
- Table 2.2:** Primer sequence for the *C9ORF72* hexanucleotide repeat.
- Table 2.3** 57 SNPs with Haplotype data available from the GWAS samples
- Table 2.4:** Case samples with a repeat expansion of 20-28 where repeat number was confirmed by second genotyping assay.
- Table 2.5:** Control Samples with a repeat expansion number of 22-26 where repeat number was confirmed by second genotyping assay.
- Table 2.6:** Repeat lengths for both case and control samples.
- Table 2.7:** Haplotype results for case samples.
- Table 2.8:** Haplotype results for control samples.
- Table 2.9:** Haplotype results for both case and control samples.
- Table 3.1:** Primers for Breakpoint Analysis.
- Table 3.2:** Breakpoint PCR Components.
- Table 3.3:** Specificity of methods used to pinpoint the duplication breakpoint.
- Table 3.4:** Non-synonymous rare variants in the UK10K data.
- Table 3.5:** Association analysis 1: All rare exonic variants; n=29 variants.
- Table 3.6:** Association analysis 2: All rare non-syn variants; n=17.
- Table 3.7:** All rare non-syn damaging variants; n=6.

Table 4.1: Forward and reverse primers for Exonic regions of VIPR2 and *PAK7*.

Table 4.2: HRM Sample Preparation PCR.

Table 4.3: PCR Product Purification using ExoSAP-IT.

Table 4.4: HRM Sensitivity in Hapmap samples.

Table 4.5: HRM Specificity in HapMap samples.

Table 5.1: *FYN* colony screen primers

Table 5.2: PCR colony screen

Table 5.3: Example measurements obtained for transfected and non-transfected cells

Table 5.4: *FYN* LoF specific site directed mutagenesis primers

Table 5.5: Site Directed Mutagenesis PCR specifications

Introduction

1.1 Genetic Variation

Genetic variation in humans, either at the DNA or RNA level, directly brings about phenotypic diversity; from trait differences in behaviour and appearance to our susceptibility to disease. Phenotypic “traits” come in many forms, from those with clear patterns of inheritance that are determined by variation at just one base position or gene in the DNA sequence, to those whose inheritance is more complex involving the interaction of genetic risk variants, regulatory mechanisms and/or environmental influences (Lander et al., 2006). Whole genome sequencing has revealed that every individual carries on average four million genetic variations (Frazer et al., 2009), therefore elucidating which individual variant or which combination of variants is responsible for a particular phenotypic trait is an important challenge, particularly for the study of complex diseases.

As our knowledge of human genetic variation increases, progress has been made in applying this knowledge to develop insight into the molecular aetiology of complex diseases (Hindorff et al., 2009) such as Type 1 diabetes (Todd et al., 2007; WTCCC, 2007), inflammatory bowel disease (Duerr et al., 2006) and coeliac disease (Hunt et al., 2008). Psychiatric disorders, such as schizophrenia (SCZ), are substantially heritable and understanding the molecular aetiology involved is a significant research goal as current diagnostics and therapies are based on clinical assessment or have emerged from clinical observation rather than an understanding of underlying biology. However, this also presents substantial challenges (Burmeister, 1999).

1.2 Inheritance of complex diseases

Traits or diseases which are the result of the disruption of one gene, allele or biochemical pathway are known as single gene or Mendelian diseases and examples include cystic fibrosis and Huntington’s disease. Although the inheritance of some of these diseases are complicated by factors such as incomplete penetrance and varying age of onset, the pattern

of inheritance is still clear and follows strict dominant, and recessive or X-linked transmission patterns (Burmeister, 1999; Lander et al., 1994; Lander et al., 2006; Schork, 1997). With complex diseases inheritance is more complicated as they are caused by a combination of genetic and environmental risk factors. Disease liability is based upon a threshold whereby a certain combination of genetic and environmental factors can shift individuals towards or away from disease onset.

The number of risk alleles contributing to disease susceptibility, the frequency of the risk allele in the population, their effect size and whether or not they act additively or through interaction with one another can vary depending on the disease. There are two commonly proposed models; the common disease-common variant model (Lander, 1996; Reich et al., 2001), whereby the majority of genetic contribution to common disease is the result of multiple common variants of small effect with minor allele frequencies (MAF) $>5\%$, or the common disease-rare variant model whereby rare variants (MAF $<5\%$) with moderate to large effect contribute to disease susceptibility. There is evidence from the study of complex diseases to support both models and to indicate that these models are not mutually exclusive in that genetic variants of varying frequency and effect in the population (e.g. common and/or rare) contribute to disease liability. The number of variants contributing and the distribution of their effect sizes can also vary.

Additional complications such as incomplete penetrance, phenotypic heterogeneity, locus heterogeneity, allelic heterogeneity, epigenetic influences (e.g., imprinting), environmental influences and epistasis all make it difficult to elucidate the extent of the genetic contribution and underlying cause of these complex diseases (Burmeister, 1999; Lander et al., 1994; Schork, 1997).

1.3 Schizophrenia Overview

SCZ is a chronic, severe, and disabling mental disorder characterized by deficits in thought processes, perceptions, and emotional responsiveness. Symptoms typically emerge in early adulthood with the age of onset occurring between the ages of 15 to 25. One in 100 people will develop SCZ in their lifetime. At present there are no biological diagnostic tests, with diagnosis generally based on a thorough psychiatric interview. Early diagnosis of SCZ is very important, as it can decrease the impact of the disorder through tailored medication and therapy.

SCZ currently affects over 50,000 people in Ireland, and the economic cost of this is estimated at approximately €460 million per year, which breaks down as follows: direct costs of inpatient and outpatient care, medication and community services are €118 million, and indirect costs due to unemployment and high mortality are more than €343 million per year (Behan et al., 2008). Along with the economic costs, there are a number of social issues affecting sufferers of SCZ with many patients unable to live independently. Other debilitating social factors include unemployment, homelessness and increased risk of death due to accident, illness or suicide. The lifetime risk of suicide is 5-10% (Hor & Taylor, 2010). On average, individuals affected with SCZ die >20 years earlier than the general population (Tiihonen et al., 2009).

SCZ has been termed a complex disorder due to the fact it is understood to be caused by a combination of genetic and environmental factors. Environmental factors are both prenatal and perinatal, and include viral infections during pregnancy, famine, premature birth and advanced paternal age (Brown & Susser, 2008; Cannon, Jones, & Murray, 2002). These environmental factors however are not unique to SCZ. Individuals living in urbanised areas have been shown to be at a greater risk of the disorder (Krabbendam & van Os, 2005) and cannabis use has also been identified as a trigger for those with pre-existing liability to psychosis (Henquet et al., 2006).

It is widely believed that a complex combination of genetic factors and environmental influences are responsible for an individual's susceptibility to developing SCZ. The most strongly reported risk factor for both disorders is undoubtedly family history and epidemiological studies suggest that there may be genetic risk factors common to both disorders (Paul Lichtenstein et al., 2009). In the largest family study conducted to date, Lichtenstein *et al* (Paul Lichtenstein et al., 2009) found that offspring and siblings of individuals with SCZ were 9 times more likely than the general population to have SCZ.

SCZ symptoms can be divided into positive symptoms, negative symptoms and cognitive deficits. Positive symptoms include hallucinations, delusions, disorganised behaviour and thought disorders (Tables 1.1 and 1.2). The positive symptoms of SCZ have been found to overlap with other neuropsychiatric disorders such as bipolar disorder (BD) (Schulze et al., 2012). Negative symptoms include diminished characteristics such as social withdrawal, lack of empathy, neglect and loss of motivation. Diagnosis is based on confirmation of the key symptoms, and current treatments provide control rather than cure. Drug development is complicated by the heterogeneity of the disorders that likely includes a wide range of genetic and environmental risk factors.

The cognitive deficits associated with SCZ are among the earliest signs of the illness (Bowie, Reichenberg, Patterson, Heaton, & Harvey, 2006). Cognitive dysfunctions include verbal, visual and working memory; reasoning, problem solving and social cognition. Clinicians are currently using cognitive remediation therapy which involves repetitive practise leading to cognitive improvement. These methods have been used with SCZ sufferers in order to target higher-order thought processes such as attention and higher level processes, which in turn may lead to an improvement in overall cognition and function (Keefe et al., 2011).

Table 1.1: Criteria for diagnosis of schizophrenia according to the Diagnostics Statistics Manual Classification IV (DSM-IV)

Characteristic symptoms:	- Delusions - Hallucinations
Two (or more) of the following, each present for a significant portion of time during a 1-month period (or less if successfully treated)	- Disorganized speech - Grossly disorganized or catatonic behaviour - Negative symptoms
Social/occupational dysfunction:	For a significant portion of the time since the onset of the disturbance, one or more major areas of functioning such as work, interpersonal relations, or self-care are markedly below the level achieved prior to the onset (or when the onset is in childhood or adolescence, failure to achieve expected level of interpersonal, academic, or occupational achievement).
Duration:	Continuous signs of the disturbance persist for at least 6 months. This 6-month period must include at least 1 month of symptoms (or less if successfully treated) that meet Criterion A (i.e., active-phase symptoms) and may include periods of prodromal or residual symptoms. During these prodromal or residual periods, the signs of the disturbance may be manifested by only negative symptoms or two or more symptoms listed in Criterion A present in an attenuated form (e.g., odd beliefs, unusual perceptual experiences).
Schizoaffective and Mood Disorder exclusion:	Schizoaffective Disorder and Mood Disorder With Psychotic Features have been ruled out because either (1) no Major Depressive, Manic, or Mixed Episodes have occurred concurrently with the active phase symptoms; or (2) if mood episodes have occurred during active-phase symptoms, their total duration has been brief relative to the duration of the active and residual periods.
Substance/general medical condition exclusion:	The disturbance is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.
Relationship to a Pervasive Developmental Disorder:	If there is a history of Autistic Disorder or another Pervasive Developmental Disorder, the additional diagnosis of Schizophrenia is made only if prominent delusions or hallucinations are also present for at least a month (or less if successfully treated).

Table 1.2: Criteria for diagnosis of schizophrenia according to the International Classification of Disease 10 (ICD-10)

Diagnosis requires the presence of one symptom from section 1 or two symptoms from section 2, for a period of at least one month

1.	- Thought echo, insertion, withdrawal or broadcasting - Delusions of control, influence or passivity - Hallucinatory voices - Persistent delusions of other kinds which are culturally inappropriate or implausible.
2.	- Persistent hallucinations, accompanied by delusions without clear affective content, persistent over-valued ideas - Breaks of interpolations in the train of thought resulting in incoherence or irrelevant speech or neologisms - Catatonic behaviour such as excitement, posturing, negativism, mutism and stupor - Negative symptoms such as marked apathy, paucity of speech and blunting of emotional responses - Change in the overall quality of aspects of personal behaviour manifest as loss of interest, aimlessness, idleness, self-absorbed attitude and social withdrawal

1.4 Neurobiology of schizophrenia

SCZ is thought to be caused by alteration in brain structure and circuitry that can start during foetal development, and in turn is considered a neurodevelopmental disorder. Many of the genomic structural variants associated with SCZ implicate neurodevelopmental genes involved in neuronal proliferation, migration and synapse formation. Changes in neurotransmitter levels such as dopamine are thought to be associated with positive symptoms such as delusions and hallucinations. SCZ in its acute psychotic state has been found to be associated with an increase in dopamine synthesis, release and resting state synaptic dopamine concentrations (Guillin, Abi-Dargham, & Laruelle, 2007; Laruelle, 1998). Current treatments for SCZ are dopamine D2 receptor antagonists. Conventional

antipsychotics reduce the delusions and hallucinations; however they do not enhance cognitive deficits.

A second hypothesis involves glutamate and low activity of the N-methyl-D-aspartate (NMDA) receptor on γ -Aminobutyric acid (GABA) inhibitory interneurons in the prefrontal cortex. Glutamate is an abundant amino acid in the brain, and is a major excitatory neurotransmitter. It is not clear how NMDA receptors are involved in SCZ, however post mortem brain studies have shown a loss of GABA and key enzymes for glutamate biosynthesis (Coyle, 2006).

The neurodevelopmental model of SCZ has suggested an early insult, followed by a latent period through neural development, with the emergence of psychosis at the early adulthood stage. It has been suggested that this may be due to a lesion in early development, which may not manifest until later stage, and that it may influence pathways involved in excitatory and inhibitory neurons in the prefrontal cortex (Insel, 2010).

1.5 Genetics of schizophrenia

Like many complex diseases, much of the evidence that the cause of SCZ is rooted in genetics is based on family, twin and adoption studies of affected individuals. Family studies assess the risk for the disorder in the relatives of affected individuals but they cannot distinguish between genetics and environment. The leading risk factor for SCZ is having an affected first degree relative, which increases the risk ten-fold (Cardno & Gottesman, 2000). Studies of dizygotic twins have shown a concordance rate for SCZ of 17%, whereas in monozygotic (MZ) twins the concordance rate is 48%. Based on these data twin studies have found the heritability of SCZ to be 80 – 85%. Thus, genetic factors are central to the aetiology of the disorder. Another important area of study is discordant MZ twins, which have allowed investigators to study non-inherited factors. The risk of SCZ in the offspring of affected and unaffected MZ twins have been seen to be similar, which would imply that

unaffected MZ twins carry SCZ susceptibility genes (Gottesman, 1997). Genetic discoveries in rare psychiatric and neurological disorders have delivered changes in practice of clinical medicine (Zoghbi & Warren, 2010) and this encourages researchers to attempt to find the risk genes for SCZ in an effort to improve the diagnosis and treatment of the disorder.

1.5.1 Mapping the genetics of complex diseases

SNPs and structural variants have long been used in the mapping of susceptibility loci or genes to complex disorders and several methods and technologies exist for their identification. Constantly developing technology has allowed researchers to move from studying just one variant in one gene in a small sample of individuals, to analysis on a genome-wide level, surveying thousands of genes in thousands of individuals simultaneously. In fact one of the greatest challenges currently is handling the vast amounts of data that are being generated.

1.5.2 Linkage studies

Linkage studies analyse the co-segregation of DNA markers with disease (Altshuler, Daly, & Lander, 2008). The approach involves the genotyping of a panel of evenly spaced DNA markers (generally either microsatellites or SNPs), either in one large multi-generation family or in a number of smaller nuclear families to detect broad regions harbouring susceptibility loci. When using microsatellites, a marker density of one every 5-10 cM is optimal, requiring approximately 400 markers for a genome-wide study. SNPs are less informative than microsatellites, and so require increased marker density (1-2 SNPs every cM)(Evans & Cardon, 2004). For a genome-wide linkage screen, this density requires 5,000-10,000 SNPs be genotyped. For each marker a statistic is calculated to estimate the degree of linkage between markers and the disease state. The most common statistic used in linkage studies is the logarithm of odds (LOD) score. After identification of a linkage region of interest, the potential causative loci can be probed further by fine-mapping of the region.

Linkage studies have been used to successfully identify genetic variants for over 1,600 Mendelian diseases including cystic fibrosis and Huntington's disease (Botstein & Risch, 2003; Jimenez-Sanchez, Childs, & Valle, 2001). These types of studies were feverishly applied to complex diseases from the 1980's through to mid 2000's but with little success. The few successful examples include *NOD2* in Crohn's disease (Hugot et al., 2001) and *PTPN22* in Type 1 diabetes (Concannon et al., 2005). There are several reasons for this; linkage studies are successful in Mendelian diseases where the risk variants are rare and generally have a highly penetrant large effect thereby making them easier to detect in families using a panel of markers. When applied to complex diseases, where a combination of several genes or alleles of small effect along with environmental factors are involved, linkage studies are much less powerful and would require extremely large sample sizes and a more dense panel of markers (Hirschhorn & Daly, 2005; Risch & Merikangas, 1996), e.g. the disease susceptibility allele of *PPARG* in Type II diabetes would have required linkage analysis to be conducted in at least one million sibling pairs (Hirschhorn & Daly, 2005).

1.5.3 Linkage studies of schizophrenia

Genome-wide linkage studies in families with one or more individuals affected with SCZ have identified several regions of the genome potentially linked to the disorder including those at 6p24-p22 containing the *DNTBP1* (Dystrobrevin-binding protein 1) gene (Richard E. Straub et al., 1995), 8p22-p21 containing the gene *NRG1* (Neuregulin-1) (Hreinn Stefansson et al., 2002) and 13q14-q32 containing the genes *G30* and *DAOA* (D-amino acid oxidase activator) (Chumakov et al., 2002a; Lewis et al., 2003). Similarly in bipolar disorder several genes including *SLC6A4* (serotonin transporter) and *DRD4* (dopamine receptor D4) have been implicated from linkage peaks at 17q11 and 11p respectively (Serretti & Mandelli, 2008).

Replication of these findings in other samples was inconsistent and many loci did not reach genome-wide significance levels (E. S. Lander & Schork, 1994). Loci implicated in linkage studies have been identified on almost every chromosome for both disorders (Serretti &

Mandelli, 2008). In an attempt to increase the power of these individual studies, meta-analysis of multiple genome-wide scans have been performed.

Meta-analysis of schizophrenia genome-wide linkage studies found significant linkage on 8p, 13q, and 22q using data from 11 studies (Badner & Gershon, 2002), on 1q, 2q, 3p, 5q, 6p, 8p, 11q, 14p, 20q, and 22q using data from 20 studies (Lewis et al., 2003) and on 1q, 2q, 3q, 4q, 5q, 8p and 10q using data from 32 studies (Ng et al., 2008) . Only 8p was replicated across all three.

Technological developments that made it possible to assay large numbers of variants in an association study led to huge changes in how complex diseases were studied (Risch & Merikangas, 1996). Linkage studies in the majority of complex diseases (including both SCZ and BD) had been inconsistent and there was strong evidence that association studies (discussed below) were much more powerful at detecting common variants of small effects.

1.6 Issues with genetic association studies

Several factors must be considered when conducting association studies either on a single gene or genome-wide scale.

1.6.1 Population stratification

For any type of association study, homogenous genetic ancestry within and between cases and controls is key to prevent false positives and/or false negatives arising from differing allele and haplotype frequencies (Hirschhorn & Daly, 2005; Stranger et al., 2007; W. Y. S. Wang, Barratt, Clayton, & Todd, 2005). If the cases disproportionately represent a certain genetic population in comparison to the controls as a result of migratory history or differences in ethnic ancestry, then any SNP with allele frequencies differing between the cases and controls for these reasons will be found to be (incorrectly) associated with the

phenotype (Hirschhorn & Daly, 2005). A variety of methods have been proposed to identify and correct for population stratification in association studies with the most commonly used being genomic control. This method estimates the overall inflation of significant associations due to stratification and uses principle component analysis (PCA) for correction (Price et al., 2006; Price, Zaitlen, Reich, & Patterson, 2010). Using family based samples whereby association analysis measures the preferential transmission of alleles from heterozygous (unaffected) parents to affected offspring avoids issues with population stratification but naturally requires more extensive sample collection (Price et al., 2010).

1.6.2 Power

When conducting association studies another important consideration is the power of the study to detect a true association. The odds ratio (OR), MAF and size sample all contribute to this. SNPs with small (OR=1.1-1.5) to moderate (OR=1.5-2.0) effect sizes are difficult to detect and require large sample sizes (Spencer, Su, Donnelly, & Marchini, 2009; Stranger, Stahl, & Raj, 2010). This is particularly important but challenging in genome-wide tests of association given the numbers of SNPs being tested because p values must be corrected for the number of tests undertaken, e.g., a SNP with a MAF of 15% with an OR of 1.25 would require 6,000 cases and 6,000 controls to achieve a power of >80% for association at a genome-wide significance level (5×10^{-8}) (Stranger et al., 2010). The number of samples required to give sufficient power to reach a predetermined level of significance can be estimated using one of several genetic power calculators. The large number of statistical tests undertaken increase the risk of type I error (false positive association with disease). Lowering the significance level for association may reduce the number of type I errors, but it could increase the likelihood of type II error (false negative) because very large sample sizes will be required to detect highly significant associations with SNPs of small effect.

Another issue that must be considered is phenotype heterogeneity, which refers to variation within presentation of the disorder phenotype. Heterogeneity may weaken the magnitude of the effect and therefore overall power of the study to detect and identify

susceptibility genes (M. I. McCarthy et al., 2008). When dealing with large numbers of samples and hundreds of thousands of SNPs, ensuring the genotype data is robust is also extremely important in association studies. Problems with DNA extraction, genotyping methodologies between cases and controls and genotyping missingness can cause spurious associations. This has generally been remedied in GWAS by use of core technologies in centralized genomics facilities. (Anney et al., 2008). Most importantly, the gold standard of any association study is independent replication of significant findings in a sufficiently powered study.

1.6.3 Pre-GWAS association studies of schizophrenia

Association studies which have been performed for schizophrenia have generated evidence of association at more than 1,000 genes, the strongest candidates from both positional candidate gene studies, (i.e. combining results from both association and linkage) and from biological candidates (e.g. genes involved in dopamine, glutamate, or neurodevelopmental function) are dystrobrevin binding protein 1 (*DTNBP1*) (R. E. Straub et al., 2002), *NRG1* (H. Stefansson et al., 2002), *DAOA* (D-amino-acid oxidase) (Chumakov et al., 2002b) and *RGS4* (regulator of G-protein signalling 4) (Chowdari et al., 2002) and *DISC1* (Craddock, O'Donovan, & Owen, 2005; Gill, Donohoe, & Corvin, 2010; O'Donovan, Craddock, & Owen, 2009). However, a large association study of 14 SCZ candidate genes, including those mentioned above, in 1,870 cases and 2,002 controls failed to confirm any association with the disorder (Sanders et al., 2008). *DTNBP1*, *DAOA* and *DISC1* are also amongst the strongest candidate genes associated with bipolar disorder, along with *BDNF* (brain derived neurotrophic factor) and *SLC6A4* (serotonin transporter 4) (Craddock et al., 2005; Gill et al., 2010; Serretti & Mandelli, 2008).

1.7 Genome-wide association studies in schizophrenia

Genome-wide association studies (GWAS) involve analysing markers, predominantly single nucleotide polymorphisms (SNPs), in DNA samples from affected patients or cases and unaffected controls to find genetic variation that differs in frequency between cases and

controls and thus is associated with a particular disease. New genetic associations can then be used to develop strategies to detect, treat and prevent disease. GWAS studies have already been successful in identifying risk variants associated with many common traits and illnesses, e.g. type 2 diabetes (Frayling, 2007), Parkinsons Disease (Simon-Sanchez et al., 2009), Crohn's disease (Barrett et al., 2008) and prostate cancer (Thomas et al., 2008). A catalogue of published GWAS is maintained at <http://www.genome.gov/gwastudies/>.

GWAS analyses are based on the common disease-common variant hypothesis, i.e., the disorder is caused by a large number of common genetic variants (minor allele frequency (MAF) >1%) coupled with environmental factors (Lee et al., 2012). Since the first published GWAS of SCZ in 2008 (O'Donovan et al., 2008), there has been a steady flow of studies. The trend has been for sample numbers analysed together to increase over time, primarily driven by sharing of data through international collaboration. At the same time, the number of SNPs analysed has also increased, primarily driven by imputation (Marchini, Howie, Myers, McVean, & Donnelly, 2007) using dense reference panels such as 1000 Genomes (Abecasis et al., 2012). Common variants have been confirmed at more than 20 loci, although the effects on risk are modest (odds ratio (OR) = 1.1 – 1.24) (International Schizophrenia et al., 2009; O'Donovan et al., 2008; Ripke et al., 2011; J. Shi et al., 2009; Y. Shi et al., 2011; Stefansson et al., 2009; Yue et al., 2011).

Most effort is now channelled through the Psychiatric Genomics Consortium (PGC; <https://pgc.unc.edu/>), the purpose of which is to conduct meta-analyses of genome-wide genetic data for psychiatric disease. The PGC has focused on five critically-important disorders: Autism Spectrum Disorder (ASD), attention-deficit hyperactivity disorder (ADHD), BD, major depressive disorder and SCZ. The most recent large-scale GWAS of SCZ is described in a paper by Ripke et al (Ripke et al., 2013) where combining samples from a Swedish study, the PGC and a selection of independent samples to give <21,000 case and <38,000 control samples. They reported 22 loci that are genome-wide significant for SCZ and of these 13 are new. This study implicated a number of L-type voltage gated channel genes, involved in learning, memory and synaptic plasticity. A number of antipsychotics currently

prescribed to SCZ sufferers act on calcium channels. The calcium channel complex has also been associated with BD (Ferreira et al., 2008).

The region with the strongest association in the Ripke et al. study (Ripke et al., 2013) is the Major Histocompatibility Complex (MHC) region on chromosome 6. This region has consistently been identified in multiple GWAS of SCZ. This region covers a 4MB region on the short arm of chromosome 6, and contains more than 400 genes. The size of this region, its complexity, diversity, high density and high linkage disequilibrium (LD) make it a very complicated region to analyse. It has previously been associated with autoimmune and inflammatory disorders, both of which have increased rates in SCZ sufferers. Although its involvement has been implicated in both brain development and neuronal plasticity, the genetic association with SCZ is not understood. Five MHC regions have been reported to be associated with BD (Williams et al., 2011), but the evidence is not as strong as it is for SCZ.

An interesting locus to emerge from multiple GWAS is MIR137 (Ripke et al., 2011), (Ripke et al., 2013), a micro RNA whose target site has been shown to be significantly enriched amongst SCZ GWAS-associated genes. This story is not straight forward and highlights the complexity of fine-mapping risk loci to understand function. The associated SNP is actually 39kb upstream of MIR137 and is predicted by eQTL analysis to affect expression of the nearby DPYD gene, which itself has a MIR137 target site (Ripke et al., 2013). As the list of SCZ risk loci continues to grow, the challenge of linking the associated SNPs to gene function and from there to disease biology remains a major challenge.

In addition to pinpointing individual risk genes, GWAS have also advanced our knowledge of SCZ genetics in other ways. In a study by Lee et al (Lee et al., 2012) a number of conclusions were drawn about the causal variants involved in SCZ. Firstly that the genetic basis for SCZ was the same for males and females, secondly that variance is linearly related to the length of the chromosome and thirdly that a disproportionate amount of variation is attributed to a set of approximately 2,275 genes expressed in the Central Nervous System (CNS). GWAS

have allowed researchers to consider how many risk SNPs there may be in the human genome for SCZ and how much genetic risk they account for. It has been predicted that between 6,300 and 10,200 independent and mostly common SNPs could contribute to SCZ, and these account for at least 32% of the variance in liability (Ripke et al., 2013). The current emphasis is on clearly identifying the top 2,000 loci using larger samples, which hopefully will be sufficient to clearly identify the biological processes involved in risk. GWAS have provided empirical evidence of a genetic overlap between SCZ, BD and major depressive disorder (Lee et al., 2013). The genetic overlap between SCZ and BD backs up previous genetic epidemiological studies (P. Lichtenstein et al., 2009) but the overlap between SCZ and major depressive disorder was unexpected. These data challenge the traditional categorical diagnosis of psychiatric disorders because shared genes suggests shared biology between disorders.

Family studies have suggested shared genetic liabilities for SCZ, major depressive disorder (Tsuang), ASD and ADHD (Ronald, Simonoff, Kuntsi, Asherson, & Plomin, 2008), but the extent of these remain unclear. A correlation analysis can be carried out between these disorders using GWAS data to estimate genetic variance and liability in disease (Lee et al., 2013). This correlation was calculated by analysing SNPs between case-control sets for different disorders (r_{gSNP}). If a disorder showed a higher genetic similarity with another case group over their own control group it was termed a positive correlation. This analysis showed a high r_{gSNP} between SCZ and BD, a moderate one between SCZ and major depressive disorder and a low correlation between SCZ and ASD. It should also be noted however that misclassification could inflate correlation rates.

1.7.2 Copy Number Variation

Rare higher penetrant genetic mutations have also been identified that substantially increase risk of illness. These mutations take the form of copy number variants (CNVs), which are previously undetectable sub microscopic deletions, duplications and

rearrangements in the human genome, which can range from kilobases (kb) to megabases (Mb) in size.

There are a number of major mechanisms involved in the generation of structural variants. Non Homologous recombination (NAHR), which involves alignment and subsequent crossing over between 2 sites in the genome that share regions of sequence homology, is believed to be responsible for 95% of large deletions and duplications (Malhotra & Sebat, 2012). Non-homologous end joining (NHEJ) occurs due to aberrant repair of DNA double strand breaks and is thought to be the cause of 70.8% of all deletions. Fork stalling results in replication based genomic rearrangements induced by errors during the DNA replication process. L1 mediated retrotransposition, which are long interspersed nuclear elements, are responsible for 89.6% of small insertions (Malhotra & Sebat, 2012).

The involvement of rare variants conferring either a large or small effect on the pathogenesis of the disease has yet to be fully explored. In addition to CNVs and smaller indel mutations, rare single base mutations could contribute to individual cases of SCZ (Mitchell & Porteous, 2011). Findings from Walsh et al (Walsh et al., 2008) saw a three- fold increase of rare genetic variants in cases versus controls, and with an average of over 1,000 CNVs in the genome, this in turn could affect over 4 million bases.

In addition to 22q11.2 deletion syndrome identified pre-GWAS, new deleterious CNVs at chromosomes 1q21.1, 2p16.3, 3q29, 7q36.3, 15p13.3, 16p11.2 and 16p13.2 have been identified as new genomic disorders that increase SCZ risk (Levinson et al., 2011; "Rare chromosomal deletions and duplications increase risk of schizophrenia," 2008; Rujescu et al., 2009; Stefansson et al., 2008; Vacic et al., 2011), in some instances by as much as 20-fold (Vassos et al., 2010). Most loci represent large regions of recurrent genomic mutation implicating many genes, but others impact individual genes (e.g. 2p16.3 (*NRXN1*) (Rujescu et al., 2009) and 7q36.3 (*VIPR2*) (Vacic et al., 2011). CNV studies have thrown up a number of challenges when trying to link genetics to disease biology. Although CNVs confer a higher

risk for SCZ than common variants, they are neither necessary nor sufficient for the development of a neuropsychiatric condition. For CNVs that contain many genes, it is difficult to figure out which genes are important – one gene, some genes or all genes? But it is possible, as made evident by the identification of *KCTD13* as the likely risk gene at the 16p11.2 CNV (Golzio et al., 2012). Penetrance of these CNVs may be highly dependent on the functionality of other genes within the locus, or environmental factors and so these interactions require further investigation. Another complicating factor is the same CNVs being linked to multiple neuropsychiatric disorders, along with other developmental and neurological disorders.

The detection of CNVs may hold important clinical value, in particular for the use of genome-wide Chromosomal Microarrays (CMAs) for diagnostics. CNVs are now established as important risk factors for autism spectrum disorders and CMAs are now used for diagnosis of autism spectrum disorders (Miller et al., 2010).

1.7.3 Next Generation Sequencing and Schizophrenia

Next Generation Sequencing (NGS) has revolutionised genetic discovery by generating vast amounts of sequence data in a manner that is cheap, quick and accurate. Novel targeted approaches have further increased the volume of information obtained. Whole Exome Sequencing (WES) involves targeted capture of the entire exome, which generates a comprehensive list of variants within the coding region of the genome. Whole Genome Sequencing (WGS) on the other hand produces more complete and uniform sequence coverage, but also requires a substantial increase in both computational and analytical burden.

1.7.3.1 Whole Exome Studies

Whole exome studies have been used to investigate de novo mutations in targeted studies of small numbers of candidate genes in SCZ samples (Moens et al., 2011). Recent studies

have indicated a role for rare sequence variation in SCZ (Cooper et al., 2011). In a study carried out by Xu et al (Xu et al., 2011) they claimed that *de novo* germline mutations account for approximately 50% of sporadic SCZ cases, where individuals do not have a family history. Their study included 53 trios with no family history and observed 34 *de novo* point mutations, 19 of which were thought to be protein affecting. All *de novo* mutations were also absent in controls. In this study none of the variants had been seen in any previous studies of SCZ, and all mutations were found in different genes. A subsequent study by the same group, but with a greater sample size found mutations in genes with a higher expression in foetal life. The number of LoF events was also found to be 2.8 times greater in case than controls. They also made an observation that the higher the paternal age, the greater the number of *de novo* mutations in the offspring (Xu et al., 2012).

In a paper by Gulsener *et al* on *de novo* mutations in SCZ using a study design including the proband, parents and unaffected sibling, they found that damaging *de novo* mutations converged in a network of genes co-expressed in the dorsolateral and ventrolateral prefrontal cortex during fetal development (Gulsener et al., 2013). The prefrontal cortex matures late in brain development, which in turn is consistent with the age of onset of SCZ. Neuroimaging studies of SCZ have also shown anatomical and functional deficits in this region (Shepherd, Laurens, Matheson, Carr, & Green, 2012). Genes implicated in SCZ function in important processes in foetal brain development including axon guidance, neuronal cell mobility, synaptic function and chromosome remodelling (Gilman et al., 2012).

1.7.3.2 Functional Assessment of Genetic Associations

At the beginning of this chapter I stated that “Genetic variation in humans, either at the DNA or RNA level, directly brings about phenotypic diversity” and thus far I have discussed the types of variants involved and methods of detection. Equally as important as the identification of human variants is understanding their potential functional consequences. Genetic variation along with environmental and epigenetic factors contribute to natural variation in gene expression which in turn brings about different phenotypic traits (King &

Wilson, 1975). These inherited variations in gene expression influence morphology, physiology, behaviour and susceptibility to disease. Several studies have used family- and population-based linkage and association studies to identify regulatory variants and it is estimated that anywhere from 30%-60% of human genes are under the influence of regulatory polymorphisms known as expression quantitative trait loci or eQTLs (Cheung et al., 2011; Cookson, Liang, Abecasis, Moffatt, & Lathrop, 2009; Nica & Dermitzakis, 2008; Pastinen, Ge, & Hudson, 2006).

The huge technological advances over the past decade, particularly those of next generation sequencing, has not alone revolutionised how genetic variation can be identified and analysed but also how the functional aspect of our genomes can be studied. Analysis has progressed from the study of single genes, to hundreds of genes using microarrays and more recently the entire transcriptome using next generation sequencing. These tools have not alone provided huge insight into how the transcriptome functions in different cells and tissues under different conditions but also into how genetic variation affects gene expression which in turn contributes to disease phenotypes. Correlating how certain genotypes influence gene or transcript expression is one of the great challenges of genetic disease studies. It is essential in order to gain a greater understanding of the pathophysiology and potential therapeutic treatment of a given disease and to allow the prediction of disease risk (Montgomery & Dermitzakis, 2011).

Simply comparing gene expression levels of a particular gene between affected and unaffected individuals is not considered powerful enough to detect important differences and provides little information with regards the causative mechanisms (Cookson et al., 2009). Combining expression information with association data between a genetic variant and a trait has more potential. For example, if an allele has a higher frequency in cases compared to controls and is also associated with variance in gene expression, it is easier to establish causality. Mutations responsible for the majority of rare Mendelian diseases generally lie within the coding sequence of a gene and cause an amino acid change in the protein which inhibits normal functioning. In these cases, predicting the functional outcome

of the associated variant, whether it be missense, nonsense, or involved in splicing can be achieved using a variety of *in silico* tools (e.g., bioinformatics prediction tools developed to evaluate functional effects of coding SNPs such as Polyphen (Ramensky, Bork, & Sunyaev, 2002) or using molecular methods (e.g., gene reporter constructs, gene knockdown, expression analysis) and/or animal models.

There is greater difficulty with complex diseases whereby the majority of variants associated with complex diseases to date (~88%) are intronic or intergenic (Hindorff et al., 2009). This suggests that these variants have a regulatory function and contribute to disease susceptibility by influencing gene expression (Cookson et al., 2009; Emilsson et al., 2008). Again, several *in silico* tools exist for predicting the function of non-coding variants (e.g. Tfscore) however any findings must be validated biologically in order to infer true causation. Most associated loci are therefore commonly tested for potential association with gene expression.

1.7.3.3 Assessing variants identified from linkage and early association studies

The most commonly used methods for investigating potential functional disease associated loci involve treating the transcript levels of an associated gene as a quantitative trait and determining if the level of gene expression differs based on genotype or by testing for allele specific expression differences at a given associated SNP locus (Cheung et al., 2010; Cookson et al., 2009; Emilsson et al., 2008; Gilad, Rifkin, & Pritchard, 2008; Nica et al., 2011; Yan, Yuan, Velculescu, Vogelstein, & Kinzler, 2002). Successful eQTL and allelic expression imbalance studies have been conducted for several complex diseases, including psychosis, on variants associated through linkage and association studies e.g. *DTNBP1* and SCZ (Bray et al., 2008), e.g. *NRG1* and SCZ (Nicodemus et al., 2009), *HTRA1* and age related macular degeneration (Yang et al., 2006).

1.7.3.4 Microarrays and GWAS

If just one or a small number of genes are being tested, analysis of mRNA transcript expression for eQTL analysis can be effectively carried out by RT-PCR or by direct cDNA sequencing. However, the development of microarray technology revolutionised how gene expression could be studied, providing a practical and economical tool for studying the expression of a large number of genes simultaneously. In addition, it offers researchers an opportunity to combine gene expression data for large number of individuals with genotype data in order to interrogate the impact of genetic variation on gene expression, essentially a genome wide eQTL study.

Such studies were conducted on control population samples e.g. HapMap lymphoblastoid cell lines (Cheung et al., 2010; Morley et al., 2004), in order to assess how common variation affects expression and have also been applied to the study of complex diseases by integrating GWAS and gene expression data e.g. Moffatt et al (Moffatt et al., 2007) identified a series of strongly correlated SNPs on chromosome 17q23 associated with childhood asthma. The association region contained 19 genes, none of which had an evident disease role. Expression analysis on lymphoblastoid cell lines derived from the same families showed that the disease associated SNPs also explained ~29.5% of the variance in transcript levels of one of those 19 genes, *ORMDL3* (ORM1-like 3) which is now considered one of the best functional candidates for asthma. Similarly the gene based eQTL and allelic imbalance methodology can also be used to assess potential regulatory variants associated in GWAS e.g. *ZNF804A* and schizophrenia (Buonocore et al., 2010; Hill & Bray, 2011; Williams et al., 2010) and *PTGER4* in Crohn's disease (Libioulle et al., 2007). Several studies have recently shown that there is a significant excess of regulatory variants among disease associated loci (Nicolae et al., 2010; Richards et al., 2011).

1.7.3.5 Transcriptome sequencing

In recent years next generation sequencing of RNA (RNA-seq) has become the method of choice for the genome-wide characterisation of mRNA levels as it gives more accurate

resolution of individual exon or overall transcript levels than microarrays with the additional benefits of allowing novel gene and transcript discovery and characterisation of splice sites (Majewski & Pastinen, 2011; Marguerat & Bahler, 2010; Marioni, Mason, Mane, Stephens, & Gilad, 2008; Mortazavi, Williams, McCue, Schaeffer, & Wold, 2008; Z. Wang, Gerstein, & Snyder, 2009). It can also be used simultaneously for the analysis of eQTLs (Majewski & Pastinen, 2011) and allelic imbalance (Heap et al., 2010). Studies so far have focused on using this methodology to increase our understanding of the transcriptome (Cheung et al., 2011; Nica et al., 2011). Several studies combining genotype and expression data from the same individuals have been carried out with the intent of creating a genetic map of gene regulation in different tissues and in different populations (Montgomery et al., 2010; Pickrell et al., 2010), the extent of splicing (Lalonde et al., 2011; Pan, Shai, Lee, Frey, & Blencowe, 2008), percentage of regulated genes and extent of tissue specific expression (Nica et al., 2011). It is beginning to be applied to the study of disease e.g. cancer (Ding, Wendl, Koboldt, & Mardis, 2011), schizophrenia (Mudge et al., 2008) and Alzheimer's disease (Twine, Janitz, Wilkins, & Janitz, 2011) and will no doubt in the future offer huge insight into the functionality of associated variants.

In addition to eQTL and AEI studies, microarrays and next generation sequencing technologies can also be used to examine aspects of gene expression regulation on a transcriptome-wide level through the examination of transcription factors and RNA-binding proteins (Costa, Angelini, De Feis, & Ciccodicola, 2010; Park, 2009). Chromatin or RNA immunoprecipitation (ChIP or RIP) followed by array analysis or high-throughput sequencing allows the analysis of the binding targets of such master regulators of gene expression, several of which have been linked to complex diseases (Licatalosi et al., 2008; Robertson et al., 2007).

Clearly the methodology employed for the study of potential candidate genes in complex diseases has undergone huge advances in the past few years and is only set to continue. In psychiatric illness a particularly exciting development is induced pluripotent stem cells (IPS) which can be derived from patient cells and differentiated into neurons (Buxbaum & Sklar, 2011). This offers the opportunity for the in-depth analysis of the gene expression profiles

and processes that may underlie disease causation in relevant tissue. A preliminary study observed over 600 differentially expressed genes in IPS-derived neurons from SCZ patients compared to fibroblasts (Brennand et al., 2011).

How we study complex diseases is advancing at a rapid pace and presumably as sequencing costs continue to drop, studies of whole genomes combined with whole transcriptome data will start to become a possibility to offer unprecedented insights into disease biology.

1.8 My studies of rare variants involved in schizophrenia

My PhD studies have taken place at a time (2011 – 2014) when the field of SCZ genetics and psychiatric genetics in general, is experiencing a boom in terms of the quantity of genetic data being generated for both population- and family-based sample collections. GWAS studies of common variation have gathered pace over this time and in parallel, studies of rare variation have begun to scale-up, largely driven by NGS technologies. I focused my investigations in four areas of rare variant analysis.

1.8.1 Analysis of a hexanucleotide repeat in *C9ORF72* gene

My first study follows on from the phenomenon observed for several CNVs and different psychiatric and neurological disorders where there are several examples of rare mutations that are not just strictly associated with one disorder but instead are associated with risk of several disorders. At the time I was starting my studies, my group was approached by Prof. Orla Hardiman and colleagues from TCD and Beaumont Hospital in Dublin. They had just completed a study of a rare variant, specifically a hexanucleotide repeat expansion in the *C9ORF72* gene and had generated data that strongly supported initial reports that associated this mutation with amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). For my studies I applied this investigation to a large psychosis case-control sample.

ALS and FTD are both progressive neurodegenerative disorders and are believed to be at two ends of a disease continuum (Fecto & Siddique, 2011), (Lillo & Hodges, 2009). ALS is caused by degeneration of motor neurons, whereas FTD is a dementia caused by degeneration of neurons in the frontal and anterior temporal lobes leading to early behavioural and personality changes and language impairments. Clinical resemblance has been reported between frontal lobe syndrome and negative symptoms in SCZ (Stefansson et al., 2008), (Ziauddeen, Dibben, Kipps, Hodges, & McKenna, 2011) and people with younger onset FTD can be diagnosed with psychosis before being diagnosed with dementia (Velakoulis, Walterfang, Mocellin, Pantelis, & McLean, 2009). Evidence for potential genetic overlap between FTD and SCZ has also been reported (Momeni et al., 2010; Schoder et al., 2010). In relation to *C9ORF72*, a high incidence of psychotic symptoms has been reported in FTD cases that carry the repeat expansion (Snowden et al., 2012). These data led Huey et al to test for the *C9ORF72* repeat expansion in a schizophrenia patient sample without FTD (Huey et al., 2013). No repeat expansions were detected in this relatively small sample (n=192) so I asked a similar question but in a much more powerful sample of 1,271 Irish psychosis cases and 1,283 controls. In addition, I used available GWAS data to study the relationship between the founder haplotype and the repeat expansion in this large homogenous sample.

1.8.2 Rare Variant Analysis of Schizophrenia Risk Genes

My second project addressed the challenge of how to efficiently detect rare variants in candidate genes in large samples such as my psychosis case-control sample. Sequencing may appear to be the obvious solution to this but at the time of study NGS approaches were very expensive and only efficient when sequencing large numbers (e.g. hundreds) of candidate genes together.

I was given the opportunity to collaborate with other members of the group who were carrying out a study on duplications in the *PAK7* gene. This gene was selected because a CNV study led by our group identified it as a potential SCZ risk gene (Morris et al., 2014). I

sought to study this gene by mapping the specific breakpoints, and carry out an association analysis using data available to me and study rare coding variation by accessing sequence data from the UK 10K Genomes Project.

My third project addressed the challenge of how to efficiently detect rare variants in candidate genes in large samples such as my psychosis case-control sample. In my case I wanted to scan the psychosis case-control sample for rare coding variants in just a couple of genes that had emerged as strong candidate genes from CNV studies that directly implicated them in SCZ, namely *VIPR2* and *PAK7*.

VIPR2 (7q36.3) encodes the VPAC2 receptor which binds vasoactive intestinal peptide (VIP) in the brain. *VIPR2* helps to regulate the formation and activity of neurons in the brain and has been found to play important roles in behavioural processes in mice. In a study by Vacic et al, *VIPR2* has been strongly associated with SCZ, where microduplications with variable breakpoints within a 362kb region were detected in 29 of 8,290 cases and only 2 of 7,431 controls (Vacic et al., 2011). Subsequent analysis by Levinson et al (Levinson et al., 2011) and De-Jesus Hernandez et al (Nieratschker, Meyer-Lindenberg, & Witt, 2011) have confirmed this region as an area of interest with regard to SCZ. I also analysed the exons of the *PAK7* gene, a serine/threonine protein kinase which is predominantly expressed in the brain, and is thought to promote neurite development and growth. Ultimately, this method proved unreliable for mutation detection.

1.8.3 Investigation of a Loss of Function Variant in *FYN*

My final project focused on the challenge raised by the numerous NGS-based rare variant studies – with very large numbers of rare and potentially functional coding variants being detected, how we determine the functional impact of these variants at the protein level. For this I choose to look at a loss-of-function (LoF) variant in the *FYN* gene, identified by our group in a SCZ case using a targeted NGS study (Kenny et al., 2013). I was involved in this

study when working as a Research Assistant prior to starting my PhD studies. The mutation in this gene led to a stop codon and so I studied the specific effect of the mutation on the protein in both cell lines and primary mouse neuron cultures. I used a combination of transfection and immunohistochemistry to look at the expression of the wildtype and mutant using various antibody stains and confocal microscopy.

Chapter 2 describes my study of the hexanucleotide expansion at *C9ORF72* in the psychosis case-control sample. Chapter 3 describes my studies of rare variants in the *PAK7* gene. Chapter 4 describes my study of the High Resolution Melt Analysis method as a tool for mutation detection when tested on the *VIPR2* and *PAK7* genes. Chapter 5 describes my functional studies of the stop codon mutation in the *FYN* gene. Finally, Chapter 6 is a Discussion chapter where I reflect on my studies during my PhD and highlight how my areas of study will remain relevant within SCZ genetics research for a number of years to come.

Chapter 2

Analysis of a hexanucleotide repeat expansion in the

***C9orf72* gene**

2.1 Rare variation and the cross over between ALD/FTD and SCZ

Nucleotide repeat elements, including microsatellites and short tandem repeats, are common in eukaryotic genomes. Recently, a hexanucleotide repeat expansion (HRE), (GGGGCC)_n, in a noncoding region of *C9orf72* was linked to the neurodegenerative diseases amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). ALS is characterized by a loss of motor neurons, with 90% of ALS cases being sporadic and the other ~10% having a family history; the *C9orf72* HRE represents the most common genetic cause of both familial and sporadic ALS. FTD is characterized by degeneration of the frontal and temporal lobes of the brain and is the second most common type of dementia in people younger than 65. Again, the *C9orf72* HRE is one of the most common genetic causes of FTD. Increasing evidence suggests that ALS and FTD are two related diseases in a continuous clinical spectrum, and there is a possibility that the *C9orf72* HRE also contributes to Alzheimer's and Huntington's diseases.

Normal human *C9orf72* alleles have 2 to 25 intronic GGGGCC repeats, with the majority having fewer than eight repeats and more than half having two repeats. The expanded repeats associated with ALS/FTD are thought to have variable lengths, ranging from tens to thousands of hexanucleotide repeats, but correlations between the repeat lengths and clinical onset or progression have yet to be established. Although a molecular understanding of *C9orf72* HRE pathological phenotypes has begun to emerge, the mechanisms by which the GGGGCC repeat expansion causes ALS/FTD pathology is unknown and our understanding of nucleotide repeats in the context of human disease is still in its infancy.

In a previous study by Prof. Orla Hardiman's group in TCD and Beaumont Hospital found this expansion mutation in 41% of Irish familial ALS cases and in 5% of sporadic ALS cases (Byrne et al., 2012; Majounie et al., 2012). These findings are representative of data reported across different geographical locations, for example, the expansion mutation was found in 39.3% of familial ALS cases and in 7% of sporadic ALS cases from the USA, Europe, and Australia (Byrne et al., 2012; Majounie et al., 2012) The function of the *C9orf72* protein is

unknown, but the repeat expansion is thought to reduce expression, in turn leading to disease through haploinsufficiency (Renton et al., 2011), and nuclear foci containing the expanded repeat in RNA transcripts have been found in spinal cord neurons of ALS patients (DeJesus-Hernandez et al., 2011). The repeat sequence is generally reported to be pathogenic when >30 repeats are present and the number of repeats can range from 700-1600 copies in FTD/ALS sufferers (Renton et al., 2011). The repeat range in the unaffected population was initially reported to be not higher than 23 but repeat numbers >30 have since been reported for neurologically normal controls (Ferrari et al., 2012). Overall, the functional impact of repeats that range from 20-30 remains unclear and a reliable cut-off for the pathological repeat number is yet to be determined (Xi et al., 2012).

The repeat expansion mutation is believed to have arisen on a single founder haplotype. The first evidence of a founder haplotype was reported by Laaksovirta and colleagues (Laaksovirta et al., 2010) who detected a very strong association between a 42-SNP haplotype and familial ALS sample at the chromosome 9p21 locus in a Finnish sample. In a subsequent study (van der Zee et al., 2013), across a wider European population, this haplotype was shortened to 20 SNPs over a 140kb region. The most recent study by Smith et al (Smith et al., 2013) used next-generation sequencing to fine map the region and identify an 82-SNP founder haplotype (110kb) spanning the *C9orf72* locus on which the mutation arose.

ALS and FTD are both progressive neurodegenerative disorders and are believed to be at two ends of a disease continuum (Fecto & Siddique, 2011; Lillo & Hodges, 2009). ALS is caused by degeneration of motor neurons whereas FTD is a dementia caused by degeneration of neurons in the frontal and anterior temporal lobes leading to early behavioural and personality changes, and language impairments. Clinical resemblance has been reported between frontal lobe syndrome and negative symptoms in SCZ (Mok et al., 2012; Ziauddeen et al., 2011) and people with younger onset FTD can be diagnosed with psychosis before being diagnosed with dementia (Velakoulis et al., 2009) Evidence for

potential genetic overlap between FTD and SCZ has also been reported (Momeni et al., 2010; Schoder et al., 2010).

In relation to *C9orf72*, a high incidence of psychotic symptoms has been reported in FTD cases that carry the repeat expansion (Snowden et al., 2012). These data led Huey et al. to test for the *C9orf72* repeat expansion in a SCZ patient sample without FTD (Huey et al., 2013). No repeat expansions were detected in this relatively small sample (n=192) but I now ask a similar question but in a much more powerful sample of 1,271 Irish psychosis cases and 1,283 controls. In addition, I use available genome-wide association study (GWAS) data to study the relationship between the founder haplotype and the repeat expansion in this large homogeneous sample.

2.2 Materials and Methods

2.2.1 Subjects

Subjects were recruited through community and in-patient mental health facilities throughout the island of Ireland for genetic studies of psychotic disorders. All participants provided written informed consent and were interviewed using the Structured Clinical Interview for DSM-IV Axis 1 Diagnoses (SCID) [First et al, 2002]. Diagnosis of a major psychotic disorder was made by the consensus lifetime best estimate method using DSM-IV criteria with all available information – interview, the Operational Criteria Checklist for Psychotic Illness (OPCRIT) (McGuffin, Farmer, & Harvey, 1991), family or staff report, and chart review. All cases were over 18 years of age, of Irish origin (with 4 Irish grandparents) and had been screened to exclude substance-induced psychotic disorder or psychosis due to a general medical condition. The total case sample numbered 1,271 and included patients with a DSM-IV diagnosis of SCZ (n=742), bipolar affective disorder I (n=261), schizoaffective disorder (n=162) and a smaller numbers of patients with major depressive disorder, delusional disorder samples or other psychotic illness (n=106) (Figure 2.1).

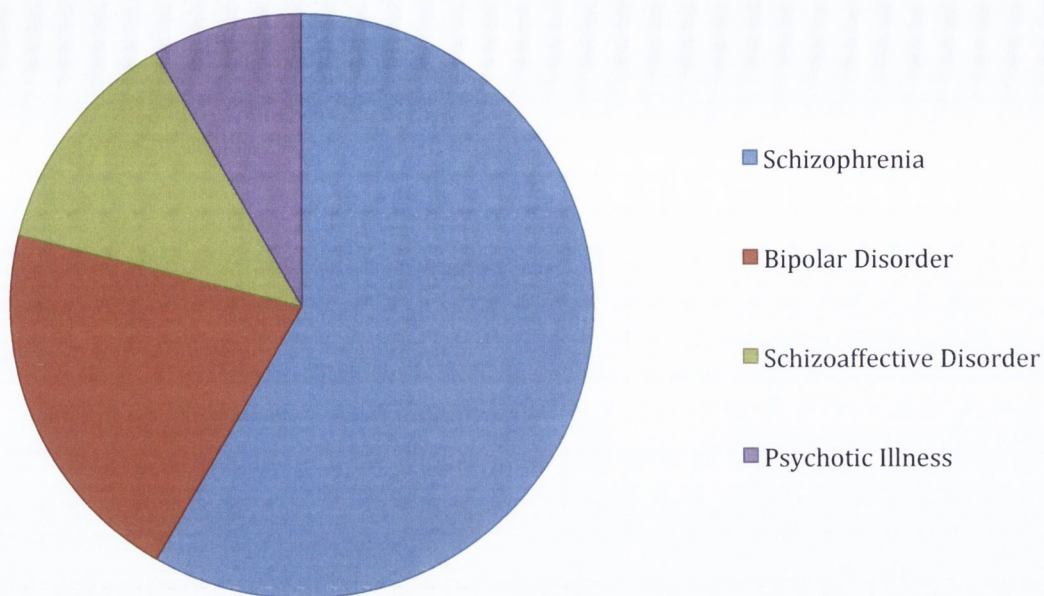


Figure 2.1 Breakdown of samples included in the Psychosis Case group.

Twenty eight case samples failed genotyping, giving a final number for analysis of 1,243. The control sample (n=1,283) was collected from the Irish Blood Transfusion Service and was made available for this study by the Trinity College Dublin Biobank. Participating individuals in this control collection gave written informed consent and met the same ethnicity criteria as cases. These controls were not specifically screened for psychiatric illness but were not taking medications, were not financially remunerated and unlikely to be over-represented for psychosis. This sample number included 33 duplicate samples, which all produced identical genotypes to the original samples. Removal of these 33 samples plus 16 samples that failed genotyping left a final control sample for analysis of 1,234.

2.2.2 Genetic analysis

Genotyping of the repeat expansion was carried out using reverse primed PCR as per Table 2.1 and analysis of all PCR products was carried out using the ABI PRISM Genetic Analyser

3130 and Genemapper 3.0 software (both Applied Biosystems). Primer sequences used for the PCR stage were obtained from Renton et al (Renton et al., 2011) (Table 2.2).

Table 2.1 : Genotyping Sample Preparation PCR	
Faststart PCR Master Mix (Roche)	14µl
7-Deaza-dGTP (Roche) (5nm)	5 µl
Q Solution (Qiagen) 1 x	5 µl
DMSO (Sigma-Aldrich) 7%	2 µl
MgCl2 (Qiagen) 0.9mM	1 µl
Primer Mix	2 µl
DNA (5nm)	9 µl
Total	38 µl

Table 2.2 : Primer sequence for the <i>C9ORF72</i> hexanucleotide repeat	
Forward	6-FAM/AGTCGCTAGAGGCGAAAGC
Reverse	TACGCATCCCAGTTTGAGACGGGGGCCGGGGCCGGGGCCGGGG
Anchor	TACGCATCCCAGTTTGAGACG

A touchdown PCR cycling programme was used, with the annealing temperature gradually lowered from 70°C to 56°C in 2°C increments, and a 3 minute extension time for each cycle. DNA samples were in 96-well plate format and all plates were run with 2 positive control DNA samples that have both previously been confirmed to contain 34+ repeats, and 2 negative control H₂O samples. Seven samples had a number of repeats >22 and all were repeated independently in Kings College London, giving the same result.

2.2.3 Haplotype analysis

A haplotype is a set of single-nucleotide polymorphisms (SNPs) on a single chromosome of a chromosome pair that are associated statistically. It is thought that these associations, and the identification of a few alleles of a haplotype sequence, can unambiguously identify all other polymorphic sites in its region. Such information is very valuable for investigating the genetics of common diseases, and has been investigated for the by the International HapMap Project.

The majority of samples used were previously analysed as part of a genome-wide association study (GWAS) of SCZ ("Genome-wide association study implicates HLA-C*01:02 as a risk factor at the major histocompatibility complex locus in schizophrenia," 2012). Genotype data was generated using the Affymetrix 6.0 SNP array and imputed using 1000 Genomes data (1000 Genomes Project Consortium) (Abecasis et al., 2012). Both repeat expansion genotype information and imputed GWAS data were available for 779 cases and 815 controls. Data on 333 SNPs spanning the founder haplotype region were available from the GWAS dataset. From this data, a set of 57 SNPs were extracted from this dataset that represented SNPs that capture the founder haplotype *C9orf72* (Smith et al., 2013)(Table 2.3) and this data was used to phase haplotypes using Plink (<http://pngu.mgh.harvard.edu/purcell/plink/>; (S. Purcell et al., 2007)). The SNPs that make up this haplotype are a collection of variants in high linkage disequilibrium (LD) over a region of 95kb that extends telomeric from intron 1 of the gene on the short arm of chromosome 9. LD was calculated using Haploview (<http://www.broadinstitute.org/haploview>; (Barrett, Fry, Maller, & Daly, 2005) (Figure 2.1). The numbering of the SNPs in Table 2.3 matches the numbering in Figure 2.1 and relates to the set of 333 GWAS SNPs available to me from this region of chromosome 9.

Table 2.3 57 SNPs with Haplotype data available from the GWAS samples

#	Name	Position	#	Name	Position
333	rs2453555	27553868	212	rs2814709	27503298
331	rs2453554	27551800	206	rs2484314	27501284
329	rs774359	27551049	205	rs2492817	27501151
322	rs774357	27549835	203	rs10812605	27500360
299	rs2453565	27541040	202	rs12554036	27499211
282	rs700791	27535960	199	rs1537712	27498489
280	rs812858	27535467	197	rs7036117	27497836
273	rs3849943	27533382	196	rs2246591	27497804
272	rs3849942	27533281	194	rs1319236	27497209
262	rs3849941	27528948	191	rs10738774	27496051
259	rs2477523	27526894	190	rs2492818	27495639
258	rs2814707	27526397	176	rs2492812	27489627
245	rs2453552	27517512	171	rs10967965	27487988
244	rs4879540	27517479	166	rs10812604	27486661
241	rs700782	27516047	157	rs9969832	27483063
240	rs10967973	27515751	156	rs2764331	27482934
238	rs3849939	27515221	147	rs7019351	27481262
234	rs700795	27510508	146	rs9969707	27481104
227	rs774351	27506640	145	rs7046653	27480967
226	rs774352	27506590	143	rs2492815	27480296
224	rs774353	27506436	142	rs868856	27479251
221	rs774354	27505967	139	rs868857	27478906
220	rs774355	27505794	138	rs17779457	27478092
219	rs814030	27505194	136	rs7041409	27477221
218	rs4609281	27504962	134	rs12115670	27476172
217	rs2783009	27504584	133	rs895021	27474911
216	rs2440619	27504323	130	rs895022	27474371
214	rs2814708	27503716	122	rs4879515	27472235
			93	rs10511816	27458461

2.4 Results

2.4.1 Genotyping of the repeat expansion

Sequencing of DNA involves a dye terminator reaction, to synthesize an array of DNA fragments of differing lengths, all complementary to the original DNA sequence of interest and ending in fluorescently labeled terminator nucleotides, and gel electrophoresis, to separate the fluorescently labeled DNA fragments by size. Within the sequencing machine itself, dye terminator reactions are loaded into capillaries (one reaction per tube) containing polyacrylamide gel. As gel electrophoresis progresses, the DNA fragments in each tube migrate down the capillaries, the smaller traveling faster than the larger; this orders the fragments so that the smallest fragment exits first. As each fragment exits the capillary tube, it is hit with a laser beam that excites the fluorescent dye attached to its terminator nucleotide. A camera captures an image of this fluorescence. The images captured by the camera are converted to a readable form called an electropherogram, from which the sequence of the DNA of interest can then be determined. The electropherogram is a graphical representation of data received from a sequencing machine and is also known as a trace.

Of the 1,243 case and 1,234 control samples analysed, none were found to have a hexanucleotide expansion with >30 repeats. I found seven samples with >22 repeats (Tables 3.1 & 3.2). Case sample MON109 was found to contain 28 repeats (Figure 2.2), and RPG5076 contained 27 repeats (Figure 2.3). Of the 1,234 control sample CO106 was found to contain 26 repeats (Figure 2.4), samples C0082 was found to contain 25 repeats (Figure 2.5), and sample C2025 was found to contain 24 repeats (Figure 2.6). It is interesting that the two samples with the longest number of repeats are both case samples but overall there is no evidence of association between the repeat expansion and SCZ using any definition of the pathological repeat number between 20 and 30.

Table 2.4 Case samples with a repeat expansion of 20-28 where repeat number was confirmed by second genotyping assay		
Cases	No of Repeats	Diagnosis
MON109	28	Schizophrenia
RPG5076	27	Schizophrenia
GASP4006	21	Bipolar Disorder
RPG1035	21	MDD Psychosis
RPG2035	21	Schizophrenia
GASP3213	20	Schizoaffective Disorder
RPG2117	20	Schizophrenia

Table 2.5 Control Samples with a repeat expansion number of 22-26 where repeat number was confirmed by second genotyping assay	
Controls	No of Repeats
C0106	26
C0082	25
C2025	24
C0362	23
C0508	23
C0129	22
C1613	22

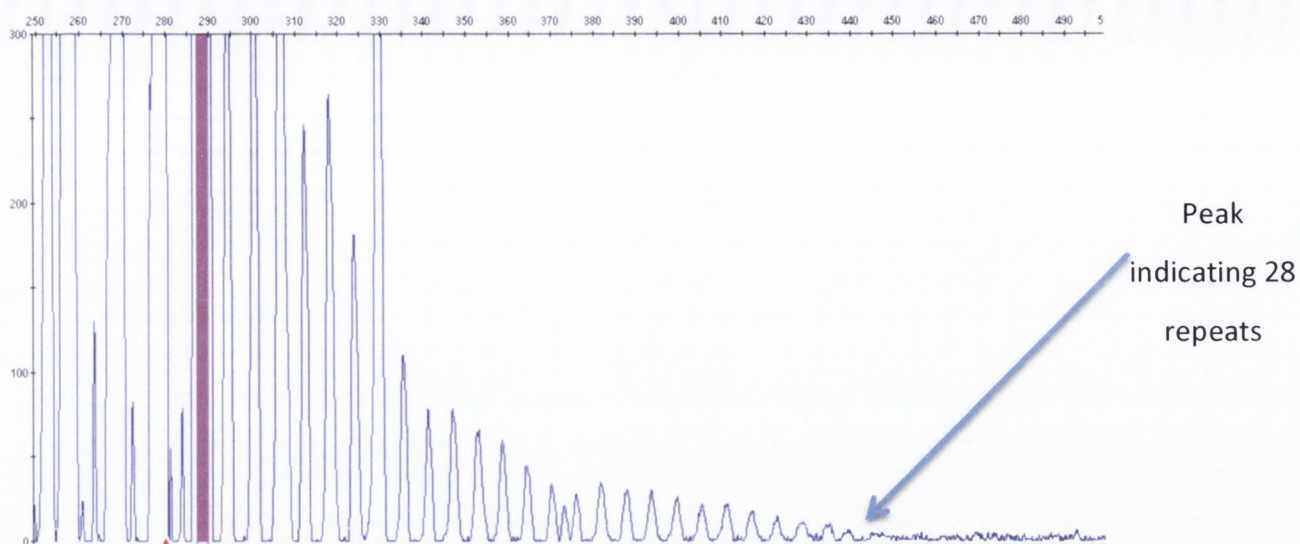


Figure 2.2: Electropherogram output for case sample **MON109** with 28 repeats

(Each peak in the electropherogram represents one hexanucleotide repeat).

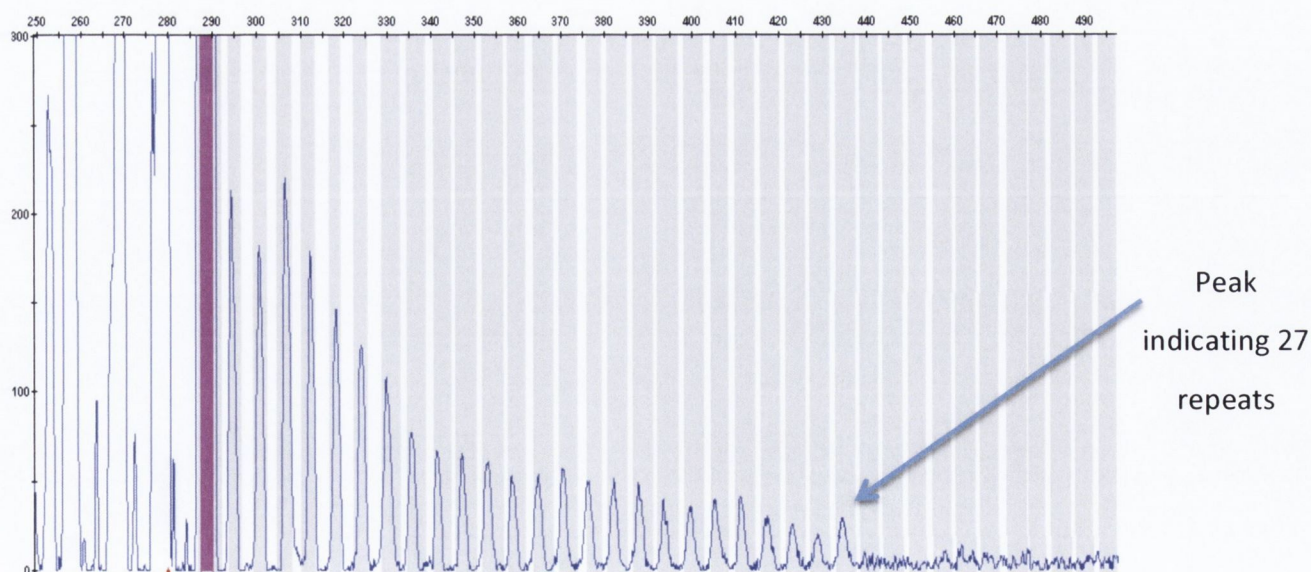


Figure 2.3: Electropherogram output for case sample **RPG5076** with 27 repeats

(Each peak in the electropherogram represents one hexanucleotide repeat).

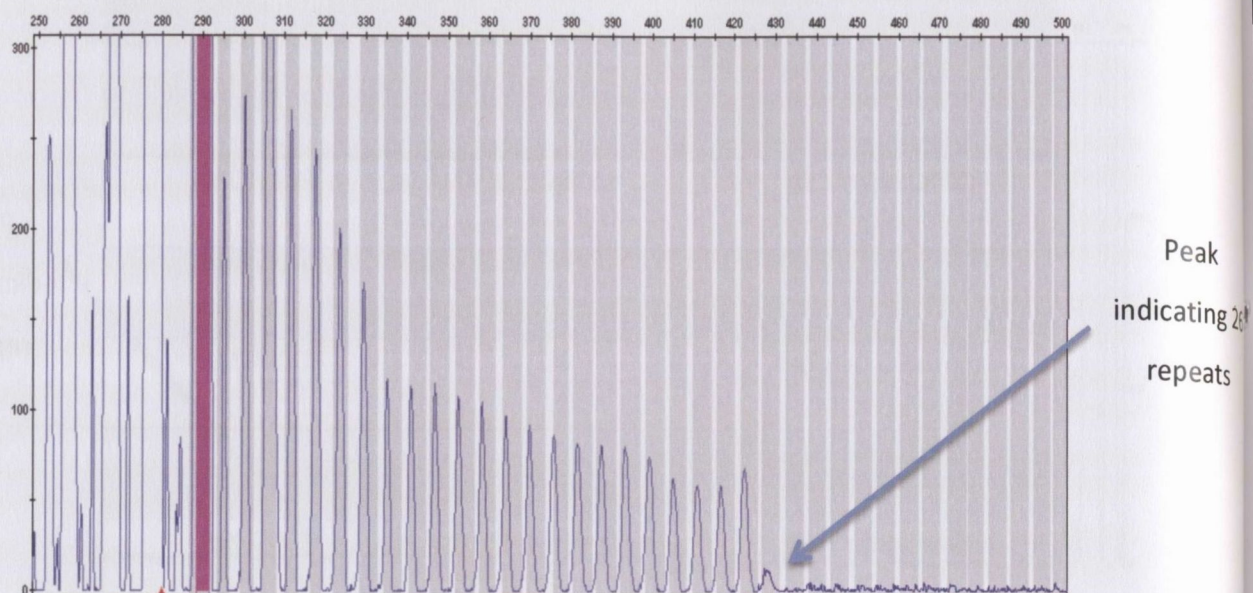


Figure 2.4: Electropherogram output for control Sample **C0106** with 26 repeats(Each peak in the electropherogram represents one hexanucleotide repeat).

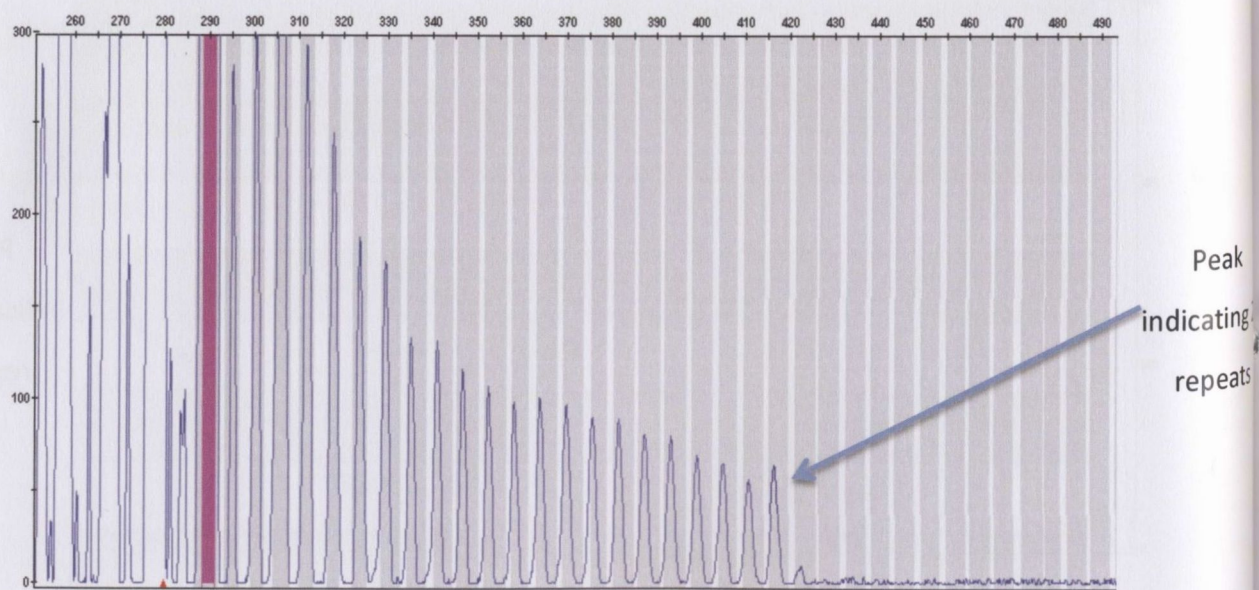


Figure 2.5:Electropherogram output for Control Sample **C0082** with 25 repeats.
(Each peak in the electropherogram represents one hexanucleotide repeat).

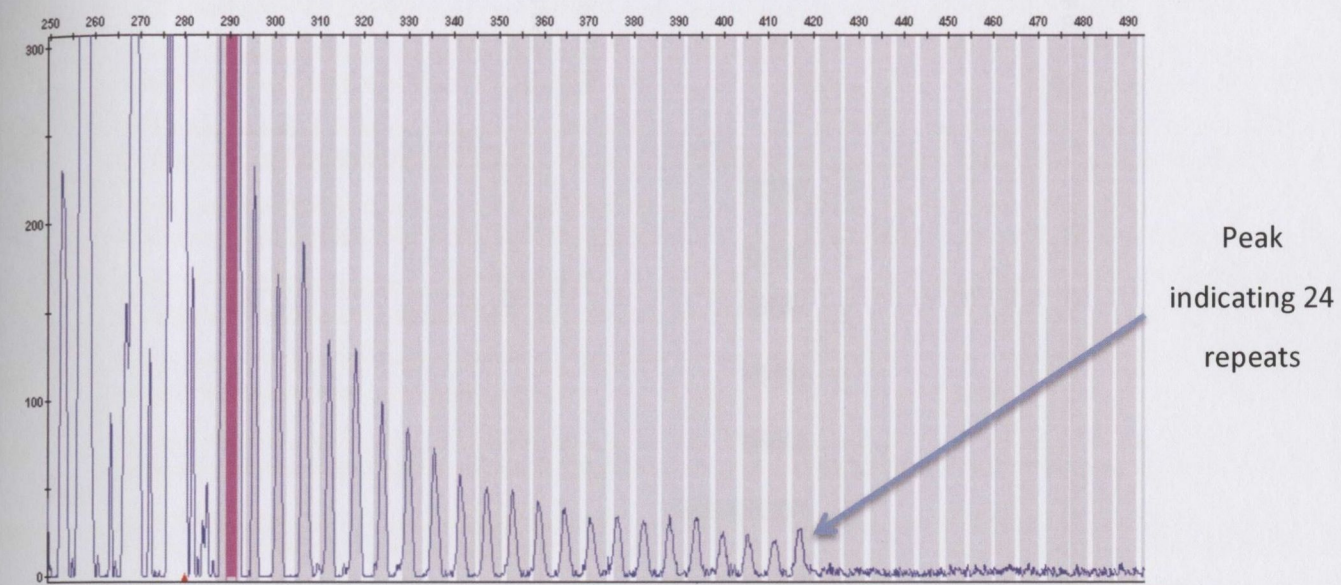


Figure 2.6: Electropherogram output for control Sample **C2025** with 24 repeats (Each peak in the electropherogram represents one hexanucleotide repeat).

Table 2.6 Repeat lengths for both case and control samples																											
Repeat No.	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Cases (n:1,243)	399	44	197	104	25	245	18	79	37	21	16	14	9	7	5	6	6	4	2	3	0	0	0	0	0	1	1
Controls (n:1,234)	415	42	187	106	22	236	11	70	42	28	20	18	4	10	3	7	2	1	1	2	2	2	1	1	1	0	0

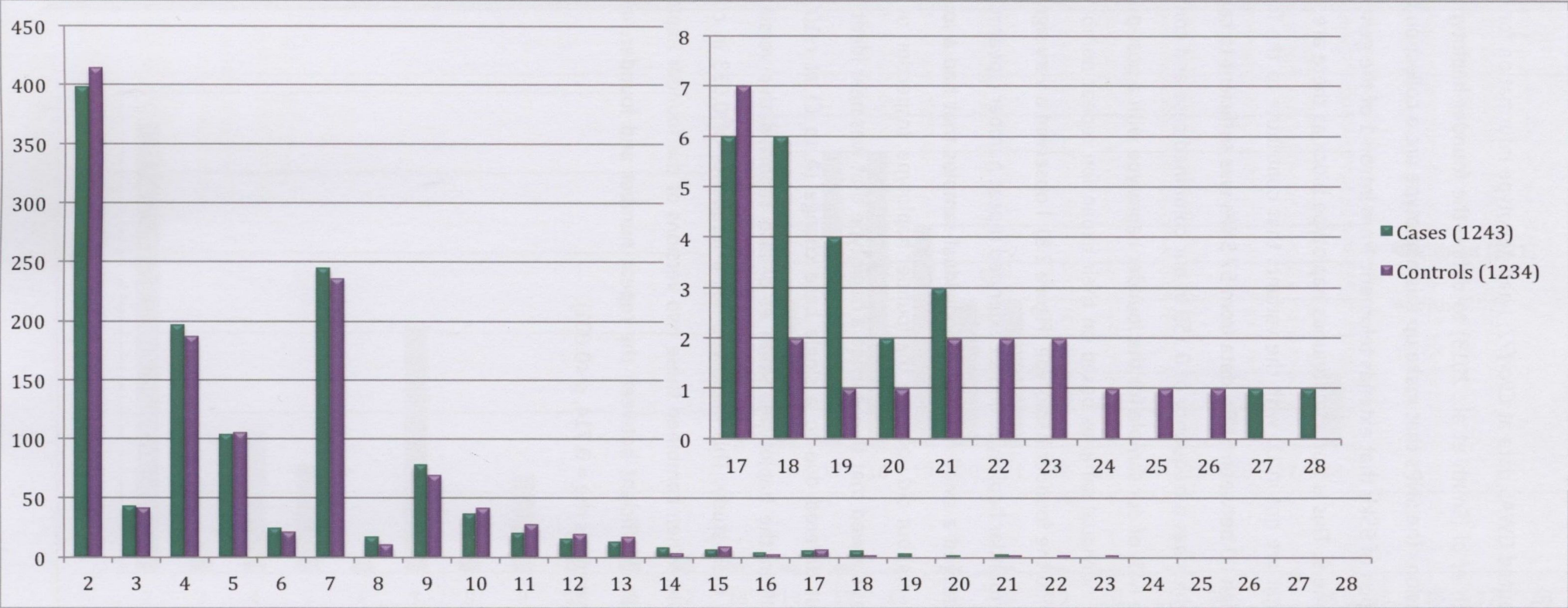


Figure 2.7 showing the repeat length in both case and control samples. Portion of the graph displaying results for repeats 16 – 28 is magnified in upper right of graph.

From the imputed GWAS data at *C9orf72*, I had genotype information for 57 of the 82 SNPs used by Smith *et al* (Smith *et al.*, 2013) to define the founder haplotype that carries the repeat expansion. The SNPs that make up this haplotype are a collection of variants in high LD over a region of 95kb that extends telomeric from intron 1 of the gene on the short arm of chromosome 9. This is not a continuous haplotype block as there are many SNPs within this region that are not in LD with the variants that contribute to the founder haplotype. Due to the high LD between SNPs, data from 57 SNPs was sufficient to capture the founder haplotype, which has a frequency of 0.159 in our combined case and control sample. I was able to phase 64% of our sample for the founder haplotype with a post-probability of >0.8 . I next plotted individual samples based on their expansion repeat number and whether or not they carried the founder haplotype (Figure 2.8). I observed a very significant correlation between the founder haplotype and an extended repeat number (spearman's $\rho = 0.685$, $p < 0.001$). I observed a small number of individual samples that had a long repeat number (e.g. >16 repeats) but did not carry the founder haplotype. Inspection of the haplotypes in these samples showed that they carried a haplotype that was near identical to the founder haplotype but differed due to a single base change (A to C) at rs10511816, the most telomeric SNP in the haplotype. Smith *et al* also reported this version of the founder haplotype in their study. This haplotype has a frequency of 0.033 in combined case and control sample. When combined these two versions of the founder haplotype for analysis, the correlation coefficient between the repeat number and founder haplotype was even stronger (spearman's $\rho = 0.714$, $p < 0.001$).

Table 2.7 Haplotype results for case samples

No. of repeats	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Without (521)	279	18	85	62	10	36	1	18	4	1	0	1	4	1	1	0	0	0	0	0	-	-	-	0	0	0	0
With (257)	2	2	4	5	2	119	7	43	21	13	12	5	1	4	1	4	4	3	2	2	-	-	-	0	0	0	1

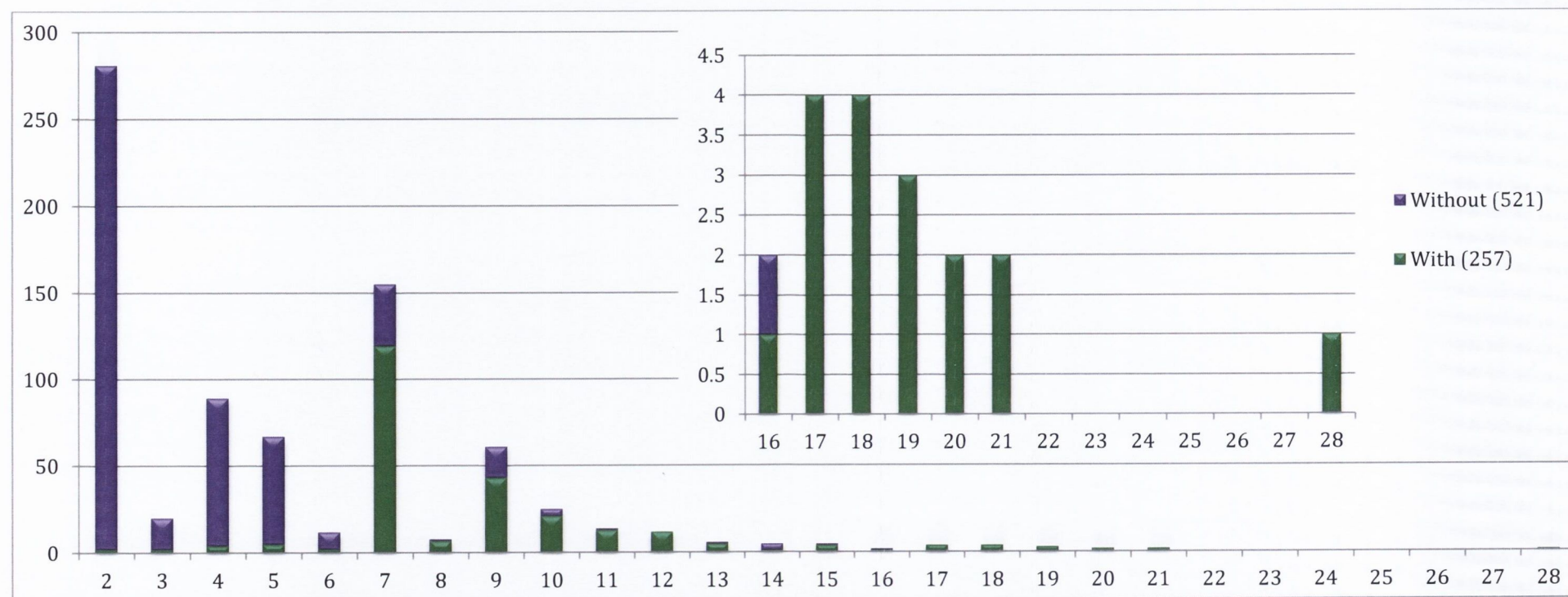


Figure 2.8: Histogram showing the relationship between repeat lengths and haplotype in cases only. Portion of the graph displaying results for repeats 16 – 28 is magnified in upper right of graph.

Table 2.8 Haplotype results for control samples

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Without (563)	311	20	78	69	11	38	2	16	9	2	1	2	-	2	-	1	-	-	-	-	-	-	-	-	-	-	-
With (253)	2	5	4	4	-	115	5	35	20	19	13	8	4	4	3	5	2	1	-	2	1	1	-	1	-	-	-

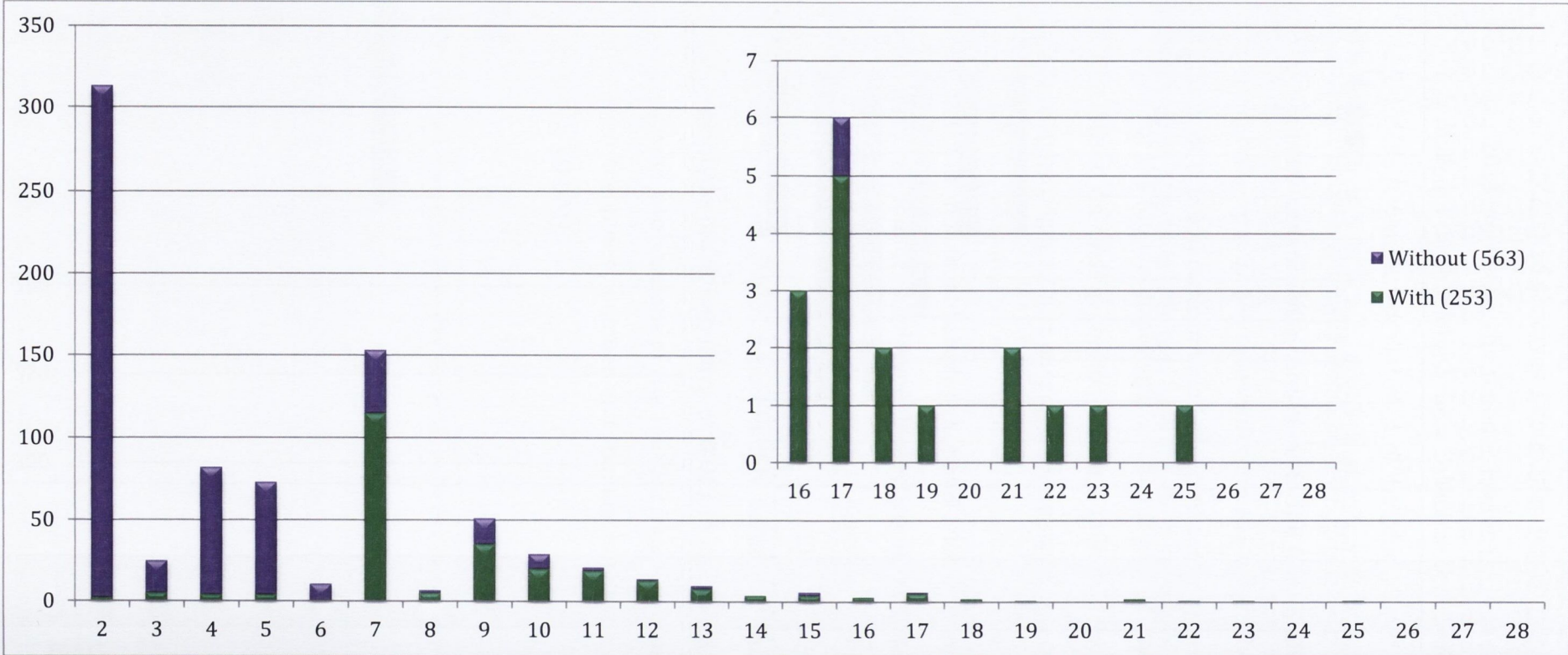


Figure 2.9 Histogram showing the relationship between repeat lengths and haplotype in controls only. Portion of the graph displaying results for repeats 16 – 28 is magnified in upper right of graph.

Table 2.9 Haplotype results for both case and control samples

No. of rpts	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Without (1082)	590	38	163	131	21	75	3	31	13	3	1	3	4	3	1	1	0	0	0	0	0	1	-	0	-	-	0
With (512)	4	7	8	9	2	233	12	81	41	32	25	13	5	8	4	9	6	4	2	4	1	0	-	1	-	-	1

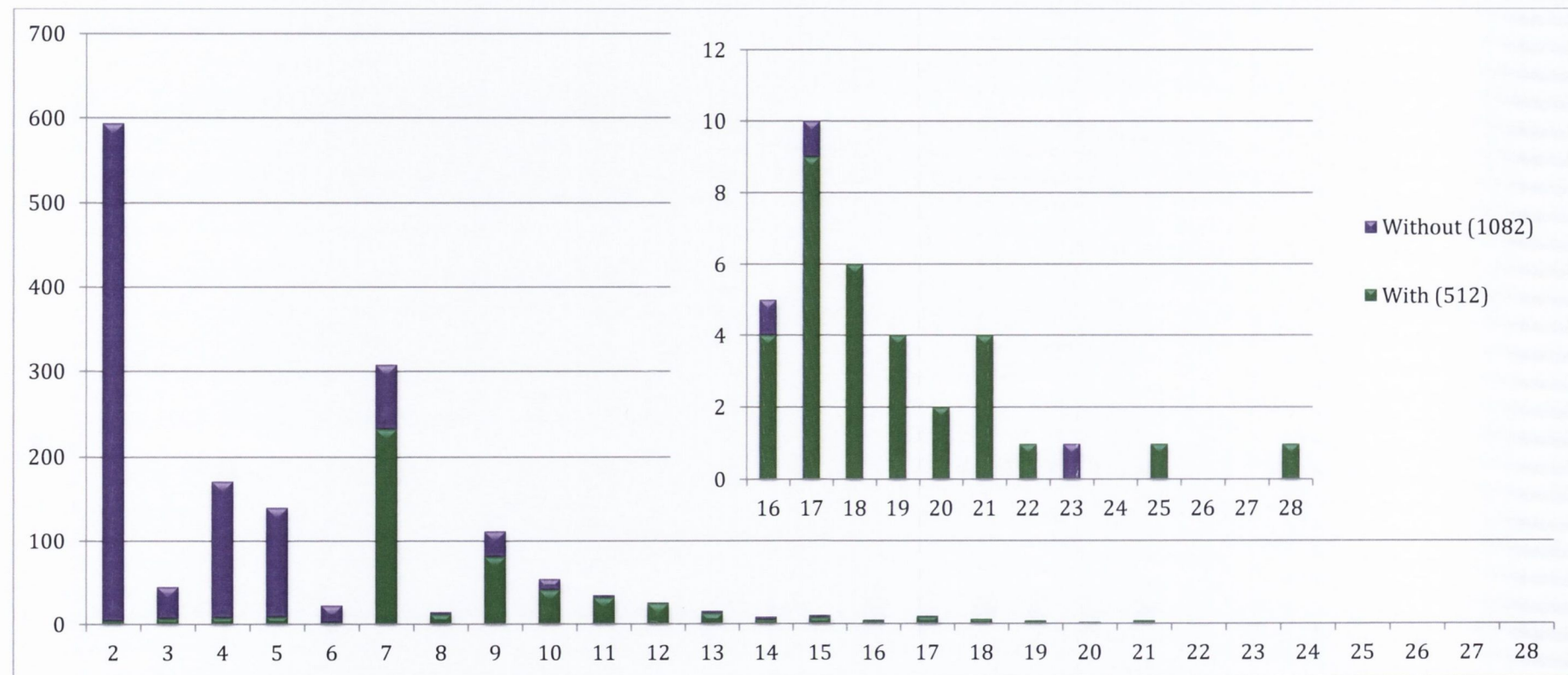


Figure 2.10 Histogram showing the relationship between repeat lengths and haplotype in both cases and controls. Portion of the graph displaying results for repeats 16 – 28 is magnified in upper right of graph.

2.5 Discussion

The hexanucleotide repeat expansion 'GGGGCC' at the *C9orf72* gene has been strongly linked with ALS and FTD. There is some evidence for clinical and genetic overlap between FTD and SCZ. In a study of *C9orf72*, the presence of psychosis in FTD patients dramatically increased the odds that these individuals carried the repeat expansion mutation (Snowden et al., 2012). A previous study reported no carriers of the pathogenic-length repeat expansion in small sample of 192 schizophrenia patients (Huey et al., 2013). Here, I genotyped the repeat expansion at *C9orf72* in a much larger Irish psychosis case-control sample. I found no carriers of >30 repeats, the putative start of the pathogenic range for this mutation. I found seven samples with >22 repeats, 2 SCZ cases and 5 controls. The case samples had the longest number of repeats in our samples (27 and 28 repeats), but overall I saw no difference in the distribution of the mutation between case and control sample. Reports have shown a high level of intermediate copy numbers (>20–30+) in a number of Parkinson's disease patients (Nuytemans et al., 2013). The functional impact and importance of the repeat expansion at different lengths remains to be elucidated. In my control sample, 0.8% of individuals carried 20 or more repeats, whereas in a similar study carried out in the UK population, the *C9orf72* repeat expansion was seen in 0.15% of the control population (Beck et al., 2013)

I utilized available GWAS data for my sample to investigate the founder haplotype on which this repeat expansion is reported to have arisen (Smith et al., 2013). My data supports the claim that carriers of the expansion arose from a single common founder, or at the very least that if the expansion arose multiple times, it has done so on the same haplotype background. In the Irish population, two versions of the founder haplotype are evident and are differentiated from each other by the SNP rs10511816. Based on my data, the frequency of the founder haplotype in the Irish population is 0.192. All samples that carry 17 or more repeats also carry a version of the founder haplotype and overall there is a very strong correlation between repeat length and the founder haplotype. Of the 621 samples carrying >6 repeats, 78% carry the founder haplotype. The largest number of repeats carried by an individual sample that does not carry a copy of the founder haplotype is 16 repeats. The

founder haplotype is only detected in 3% of 973 samples that have between 2 and 6 repeats.

In summary, I find no strong evidence to support a role for the repeat expansion at *C9orf72* in psychosis. Greater understanding of the functional impact of repeat lengths <30 may help in the interpretation of my data. I do present further strong support to back the claim that this repeat expansion arose in a single common founder. My study has now been published in *Neurobiology of Aging* (Fahey et al., 2013).

Chapter 3

Rare Variant Validation

Investigation of Rare Variation in the *PAK7* gene

3.1 Studies of a duplication event in PAK7

The emerging genetic architecture of SCZ includes a spectrum of risk variation from rare mutations of large effect, to common risk variants of small effect (odds ratio (OR) <1.15) which collectively account for at least 25% of susceptibility. Through genome-wide association study (GWAS) and subsequent meta-analysis, more than 20 independent common loci have been confirmed, but there are likely to be many more. A much smaller number of rare mutations of moderate or large effect have been identified, but these will be particularly important in facilitating dissection of the risk phenotype in model systems.

The list of rare, highly penetrant schizophrenia mutations (OR= 2-30) includes recurrent *de novo* copy number variants (CNVs) involving deletions or duplications of large, genomic regions (>100kb), but also the accumulation of different CNV events at specific loci implicating single genes (*NRXN1* [MIM 600565] and *VIPR2* [MIM 601970]). Almost all of the confirmed CNVs are also risk factors for other psychiatric or developmental phenotypes (e.g. intellectual disability, attention deficit hyperactivity disorder (ADHD), autism). This is in keeping with epidemiological and GWAS data supporting shared genetic liability between schizophrenia and other psychiatric disorders, in particular bipolar disorder. As current findings explain a modest proportion of total SCZ susceptibility, expanding the number of risk mutations will be important in understanding molecular etiology but also the relationships between these clinical disorders.

In a study carried out by my group we reported a gene-based analysis of large (>100kb), rare (<1% minor allele frequency (MAF)) CNVs in the Wellcome Trust Case Control Consortium 2 (WTCCC2) schizophrenia sample of 1,564 cases and 1,748 controls, all from Ireland. We found evidence of association with duplications at 20p12.2 and further support in a large independent European sample. All of the carriers were from the British Isles and an extended analysis including 32,911 Irish/UK subjects provided further association support. By haplotype analysis and sequencing we showed that the CNV is not the result of repeated

de novo events and is an inherited risk mutation potentially inherited from a single European ancestor.

The p21-activated kinases (PAKs) are a family of serine/threonine protein kinases, which are regulated by the Rho family of small G proteins and are involved in multiple intracellular signaling pathways. Six PAK genes are expressed in human and based on their regulatory functions are classified into Group I (*PAK 1-3*) and Group II (*PAK4-6*) members. Group I PAKs are activated by RAC-PAK signaling to promote axon connectivity, and synapse formation, in the developing brain in a pathway regulated by another SCZ risk gene *DISC1* [MIM:60521]. Aberrant synaptic network development and maintenance represents a plausible molecular mechanism for psychosis susceptibility and mutations involving other Group I PAKs (*PAK2* (the 3q29 microdeletion syndrome locus [MIM 609425]) and *PAK3* [MIM:300558]) are known to be associated with neurodevelopmental syndromes characterized by psychosis. *PAK7* is a brain-specific isoform within the cell it is localized to filipodia, where it has been shown to promote the induction of neurite outgrowth, filopodium formation and synaptic vesicle trafficking. The Group II PAKs, including *PAK7*, differ structurally from Group I members and their mechanism of activation requires clarification.

A gene-based analysis of large (>100kb), rare copy number variants (CNVs) had been conducted in the Wellcome Trust Case Control Consortium 2 (WTCCC2) SCZ sample of 1,564 cases and 1,748 controls all from Ireland, and was further extended to include an additional 5,196 UK controls. Association was found between duplications at chr20p12.2 and SCZ ($P = 0.007$) and evidence of replication came from large independent European SCZ study ($P=0.052$) and UK bipolar disorder case-control cohorts ($P = 0.047$). A combined analysis of Irish/UK subjects including additional psychosis cases (SCZ and BD) identified 22 carriers in 11,707 cases and 10 carriers in 21,204 controls (meta-analysis CMH $P=2 \times 10^{-4}$ (odds ratio (OR)=11.3, 95% CI=3.7, ∞)).

In my studies I investigated a duplication event in nineteen of the 22 cases and 8 of the 10 controls in this study. This ‘common’ event amongst the carriers was a tandem ~149kb duplication overlapping the *PAK7* gene. Haplotype analysis indicated that the duplication was in LD with local haplotypes ($P=2.5 \times 10^{-21}$), indicative of a single ancestral duplication event (Morris et al., 2014). The potential that this variant was a single ancestral mutation was interesting and could be further supported if the exact same breakpoints were found in multiple duplication carriers. I then utilised the UK10K resource to follow up on my *PAK7* duplication analysis, and investigate if there were any exonic *PAK7* sequence variants present in this data set.

3.2 Materials and Methods

3.2.1 *PAK7* breakpoint analysis

Using an Agilent custom designed comparative genomic hybridization (CGH) microarray with high probe density in the *PAK7* region, a duplication event was confirmed in one individual from our Irish psychosis case control sample and the breakpoints identified by CGH (chr20:9,684,902-9,833,151) mapped closely to those predicted by SNP array analyses (chr20:9,685,413-9,831,947). Based on these estimated breakpoints and assuming that this was a tandem duplication that is not inverted, I attempted to sequence from one copy of the duplication into the second copy in order to identify the precise breakpoints. I designed a number of primers around the target region of varying lengths to narrow down the breakpoints (Table 3.1). After testing the selection of primers, I found the 400bp product size was ideal to capture the duplication breakpoints for PCR analysis.

Table 3.1 Primers for <i>PAK7</i> Breakpoint Analysis	
Product Size: 1281bp	For: CAAATCACAAAGGGATCCTGA Rev: GGGCTAGGTGTGTTCCCATATA
Product Size: 851bp	For: TCTCTGTTGGATGGAGCTTCT Rev: CGATGTAAAAAGACACAAGAGAAA
Product Size: 400bp	For: GAAGGTAGGTATGGCCAGGA Rev: TTA CTGGATTTGGGGCATGT

Table 3.2 Breakpoint PCR Components	
10X buffer(containing 15mM MgCl2:Qiagen)	1.5µl
Forward Primer (25µM)	1.5µl
Reverse Primer (25µM)	1.5µl
dNTPs (2.5mM of each dNTP: NEB)	1.5µl
Taq (5U/µl; Qiagen)	0.15µl
dH ₂ O	4.35µl
Total	10.5µl

3.2.2 Gel Purification of PCR samples

A 2% agarose (Sigma Aldrich) gel was prepared in a volume of 50ml with distilled water and TAE. The mix was heated for 1min until dissolved. 3.5µl ethidium bromide (EtBr) was added after the TAE-agarose had cooled. 3µl low molecular weight ladder and 8ul of DNA containing 3ul of loading dye was added to separate wells of the gel. The gel was then run for 40 mins at 100 Volts.

I set out to use PCR and Sanger Sequencing to discover the specific breakpoints for this duplication. Using a forward primer positioned at the putative end of the first copy of the duplication and a reverse primer positioned at the putative start of the second copy of the duplication, I successfully PCR-amplified this unique region in a carrier sample. Samples were run on a 2% agarose gel, and samples positive for the duplication were initially identified by the presence of a band at 400bp. The PCR samples which did not amplify identified non-carrier samples which can be seen in Figure 3.1.

Table 3.3 Specificity of methods used to pinpoint the duplication breakpoint	
GWAS	Chr20:9,685,413-9,831,947 (hg18)
aCGH	Chr20:9,684,902-9,833,151 (hg18)
Sanger sequencing	Chr20:9,684,767-9,833,717 (hg18)

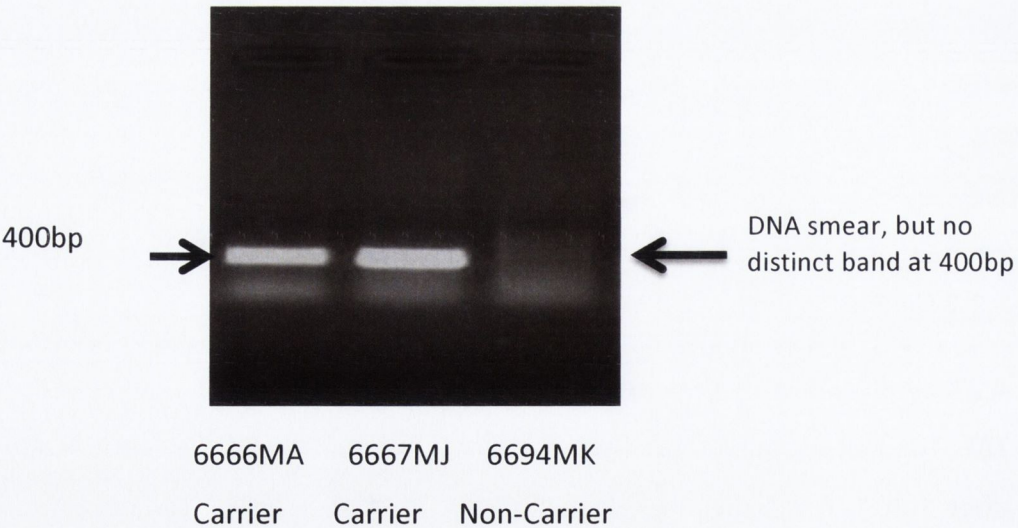


Figure 3.1 Agarose gel image of PCR 400bp fragments for Carrier (positive) and Non-Carrier (negative) samples.

3.3 Results

Having pinpointed the region where the duplication was likely to occur, I then used Sanger Sequencing to look at the specific sequence of the region. I identified the event as being a tandem duplication of 148,951bp at chr20:9,684,767-9,833,717(hg18) with a 15bp sequence inserted between the first and second copies of the repeated sequence (Figure 3.2).

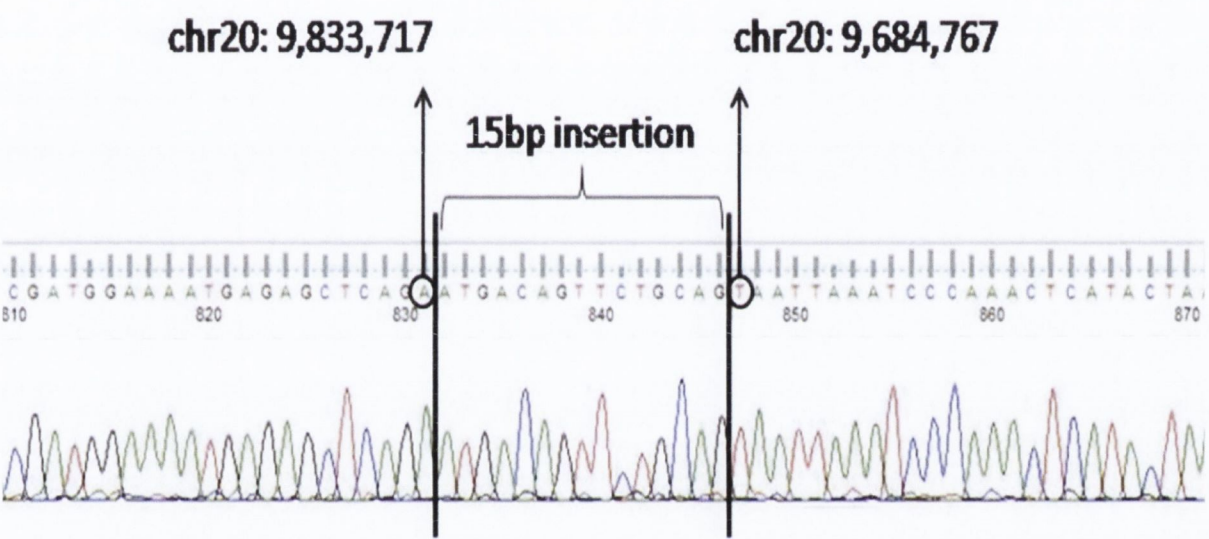


Figure 3.2 Capillary sequence of the unique region in the middle of the two copies of the tandem repeat sequence at the *PAK7* duplication. The first copy of the sequence ends at chr20:9,833,717(hg18) and the second copy starts with sequence from chr20:9,684,767(hg18), thereby defining the breakpoints. A 15bp sequence is present between the two copies.

By sequencing 5 other Irish carriers and 2 UK carriers, I confirmed that they all shared exactly the same breakpoints with the same 15bp sequence inserted between the two copies of the repeat sequence, suggesting that this represents the same duplication event. Thus, by haplotype analysis and sequencing, we established that the association is driven by the approximately 149 kb duplication, with a likely common European ancestor.

3.4 PAK7 Association Analysis

3.4.1 UK10K Background

UK10K (<http://www.uk10k.org/>) is a project to allow researchers to understand the link between rare genetic variants and human disease. The aim of the project is to study the genetic code of 10,000 people in detail, and create a genotype/phenotype resource, which will permit future human genetic research. It is a collaborative effort between multiple research centres, including from the Neuropsychiatric Genetics Research Group in TCD (Prof. Louise Gallagher). The project design involves studying and comparing the DNA of 4,000 control individuals, to identify changes that have no evident effect on phenotypes. The second approach is to look at the exomes of 6,000 people with extreme health problems, and to then compare the genetic changes between the two groups. Three thousand of these whole exome samples are individuals with neurodevelopmental disorders. Although the control sample has whole genome sequence data, these data are low X coverage sequencing for use as pooled data meaning that it is not possible to identify variants at an individual level, as is possible for the exome data for the different case groups. Thus, I decided to use the obesity case group as a control group to compare against the neurodevelopmental disorders case group. The obesity samples were used as a control set as the data was available to me, however ideally I would have used a control population that was randomly selected and more representative of the general population. Another reason why this control group was not ideal is that a higher risk of metabolic disorders has been reported in SCZ sufferers (Henderson, 2005).

The neurodevelopmental disorders case group is predominantly made up of SCZ and autism spectrum disorder samples. In brief, autism spectrum disorders (ASDs) include autism, Asperger's syndrome and pervasive developmental disorder. They have an onset in childhood and are characterized by impairments in social interaction and communication and a pattern of repetitive behaviour and restricted interests (Lord et al., 2000; Lord, Rutter, & Le Couteur, 1994). Prototypical ASD is diagnosed in 15–20 per 10,000 children (Fombonne, 2009), with broader ASD affecting between 60 and 100 in 10,000 (Baron-Cohen et al., 2009; Fernell & Gillberg, 2010). Treatments for ASD include behavioural interventions

and the use of psychotropic medications to treat comorbid conditions, but core symptoms persist.

SCZ and ASD share some clinical features such as cognitive impairment and deficits in social functioning (Craddock & Owen, 2010) and further support for biological overlap between the disorders comes from epidemiological (Sullivan et al., 2012) and neuroimaging studies (Al-Sarraj et al., 2011). The most recent evidence for shared aetiology comes from genetic studies, especially studies of rare copy number variants (CNVs) (Malhotra & Sebat, 2012)). There is substantial heterogeneity at these CNV sites in terms of type (deletion or duplication), penetrance and size, and these CNV loci are associated with multiple other neuropsychiatric, developmental and neurological phenotypes (Cooper et al., 2011; Malhotra & Sebat, 2012). However, in certain instances, mutations in SCZ and ASD cases only impact a single gene such as deletions at NRXN1 suggesting a potential risk mechanism involving synapse function (Glessner et al., 2009; Kim et al., 2008; Kirov et al., 2008; Marshall et al., 2008; Reichelt, Rodgers, & Clapcote, 2012; Rujescu et al., 2009; Stefansson et al., 2008; Szatmari et al., 2007; Vrijenhoek et al., 2008). Where SCZ and ASD have been combined for CNV (Guilmatre et al., 2009) or sequencing (Myers et al., 2011) analysis, the data supports shared biological pathways for the disorders in synaptogenesis and glutamate neurotransmission. Based on the emerging evidence that SCZ and ASD share common pathogenic mechanisms, studies such as the UK10K project have now combined the two disorders for analysis.

3.4.2 Variant Identification in UK10K data

I utilised this resource to follow up on my *PAK7* duplication analysis, and investigate if there were any exonic *PAK7* sequence variants present in this data set. In collaboration with Dr. Paul Cormican, I began with 2,598 neurodevelopmental exomes (<http://www.uk10k.org/studies/neurodevelopment.html>) and removed 620 due to relatedness leaving 1,978 cases. Relatedness or the coefficient of relationship is a measure of the degree of consanguinity (or biological relationship) between two individuals. The

coefficient of relationship ("r") between two individuals B and C is obtained by a summation of coefficients calculated for every line by which they are connected to their common ancestors. Each such line connects the two individuals via a common ancestor, passing through no individual which is not a common ancestor more than once.

For the obesity controls we started with 1,208 samples and removed 91 due to relatedness leaving 1,117 samples for analysis. Relatedness was estimated using the PLINK Identity-By-Descent (IBD) method. Samples were removed if they had a kinship coefficient of less than 0.0156, i.e., were closer than second cousins. Final analysis generated 82 variants that were present in the *PAK7* gene in the combined neurodevelopmental disorder cases and obesity controls in the UK10K data. A breakdown of these numbers can be seen in Figure 3.3.

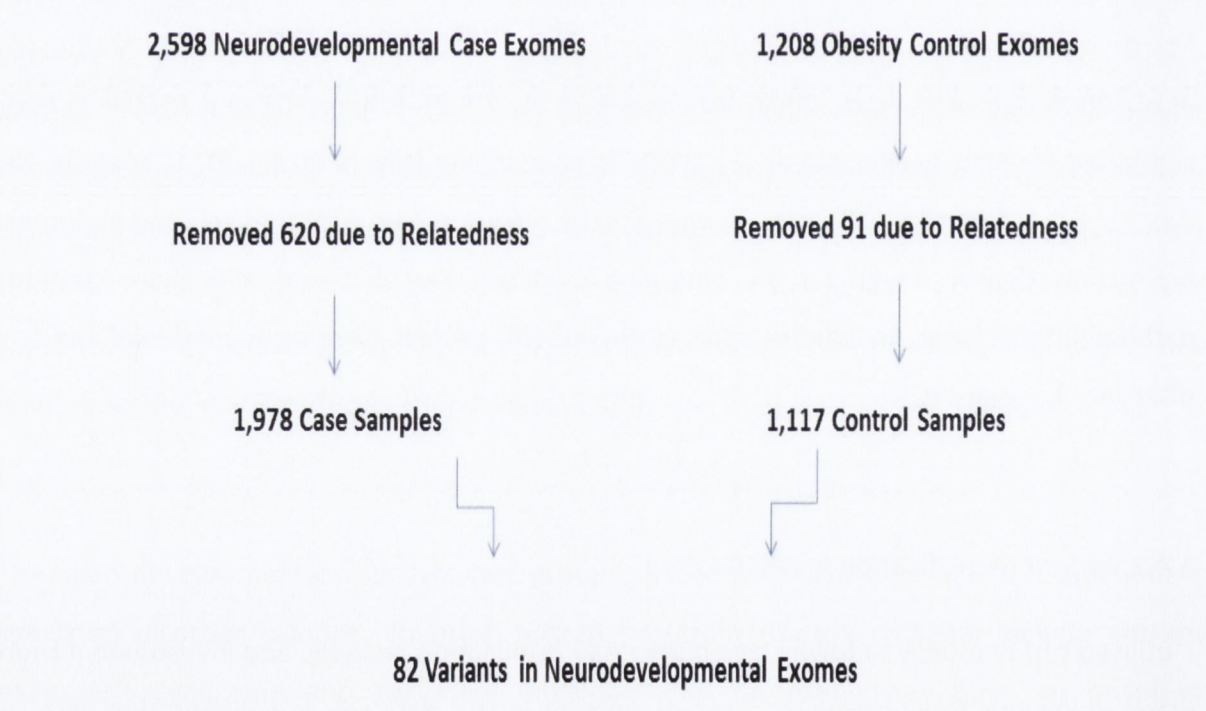


Figure 3.3: Breakdown of numbers of genetic variants detected at *PAK7* in the neurodevelopmental and obesity samples in the UK10K data.

3.4.3 Analysis of 82 variants in UK10K data using Annovar

I performed functional annotation of the genetic variants in the neurodevelopmental and obesity samples using Annovar (<http://www.openbioinformatics.org/annovar/>). By inputting a list of variants with chromosome start and stop positions, reference nucleotides and observed nucleotides, it can identify SNPs, INDELs or CNVs that cause coding changes to the amino acid sequence. It also provides information on the potential functional importance of the amino acid changing or non-synonymous variants. This result is based on the Sorting Intolerant From Tolerant (SIFT) (<http://sift.jcvi.org/>) and Polymorphism Phenotyping (PolyPhen) (<http://genetics.bwh.harvard.edu/pph2/>) scores. SIFT and PolyPhen are both programmes that predict whether an amino acid substitution affects protein function based on sequence homology and the physical properties of the amino acids.

Using Annovar for the 82 variants I observed, the output was as follows: 49 non-coding variants, 15 synonymous mutations and 18 non-synonymous mutations. The 49 non-coding and 15 synonymous mutations were filtered out leaving 18 not amino acid changing variants. One common non-synonymous variant was next removed because my focus was on rare variants (frequency < 0.01). This then left me with 17 non-synonymous variants, of which 6 turned out to be potentially damaging rare non-synonymous mutations based on SIFT and PolyPhen scores. The breakdown of this data can be seen in flowchart format in Figure 3.4.

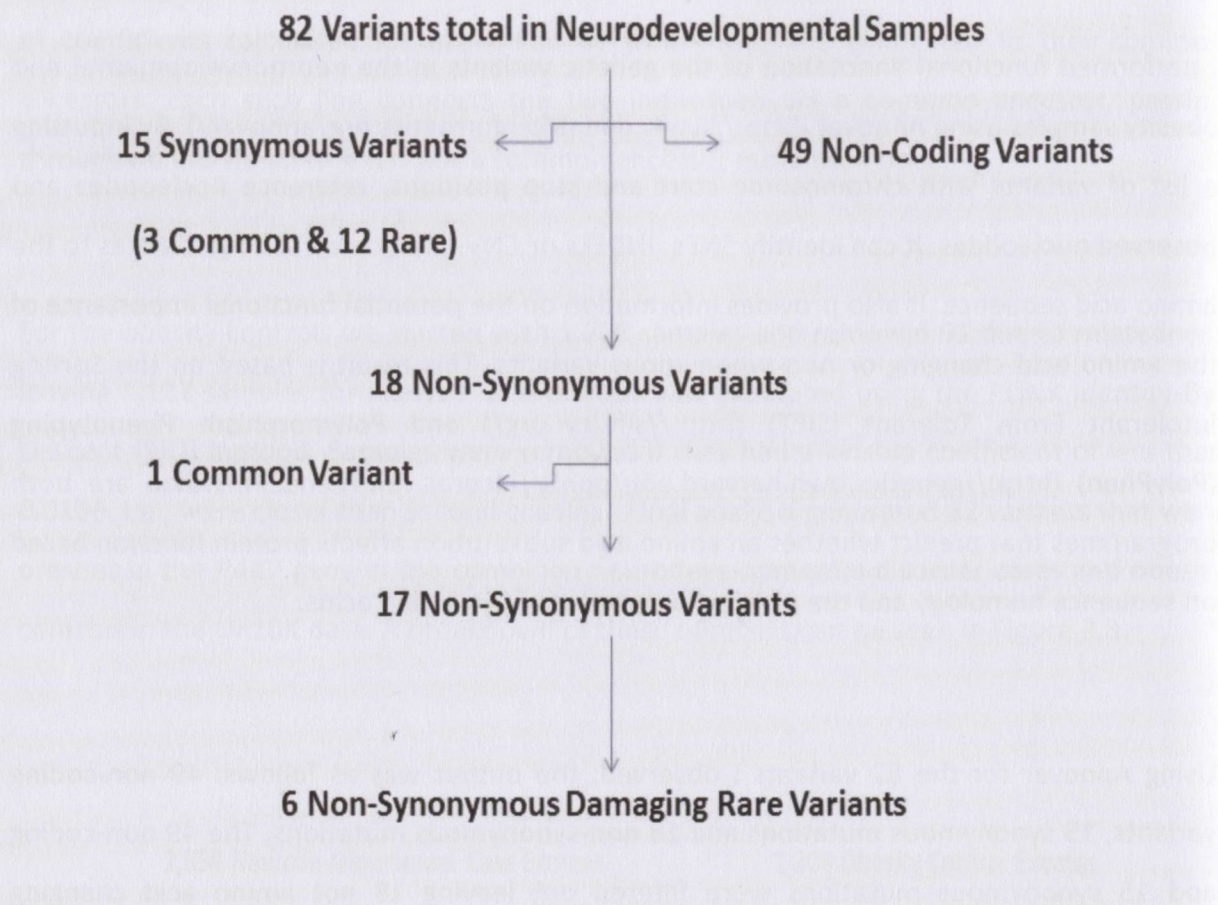


Figure 3.4: Flowchart of variants detected in the *PAK7* gene in UK10K data.

Information on the 6 non-synonymous variants identified in the dataset can be seen in Table 3.4. For each variant detected, I have given the number of affected (AFF) and unaffected (UNAFF) individuals that were heterozygous in the data set, the reference allele (Ref) and observed allele (Obs), the exonic function, amino acid change and SIFT and PolyPhen predictions for this SNP. For SIFT analysis a rating of “D” and “T” are defined as either damaging or tolerated, and in the case of PolyPhen the “P” , “D” and “B” are termed possibly damaging, damaging and benign. I classified a variant as damaging if it was scored as D by SIFT, or P or D by PolyPhen.

Table 3.4 Non-synonymous rare variants in the UK10K data.

SNP	AFF	UNAFF	Ref	Obs	Exonic Func.	AA Change	SIFT	PolyPhen
Var-20-9520135	1	0	G	A	Non-synonymous	Leu-Phe	D	P
Var-20-9520248	1	0	C	T	Non-synonymous	Arg-Gln	T	P
Var-20-9561177	1	0	T	A	Non-synonymous	His-Leu	T	D
Var-20-9523285	0	1	G	A	Non-synonymous	Arg-Val	D	P
Var-20-9624917	0	1	C	A	Non-synonymous	Arg-Ser	D	D
Var-20-9560814	10	3	T	C	Non-synonymous	Glu-Gly	D	D

3.4.4 Association analysis of PAK7 UK10K data

I next wanted to undertake a carrier-based association analysis to investigate if the overall frequency of cases that carry rare exonic mutations was significantly higher compared to controls samples. Overall I found 29 different rare coding variants (17 non-synonymous and 12 synonymous), which were carried by 61 cases samples and 32 controls. I then constructed a 2x2 contingency Table to look at in the frequency of mutation carriers between cases and controls. I calculated the Chi-Square to compare observed data with data I would expect, the confidence interval for the level of uncertainty around the measure of effect, the odds ratios; a relative measure of risk and P-value; the statistical significance. These tests were applied to each data set to look at the association between the 2 groups. For my first analysis of the 29 rare exonic variants, I found that there was no difference in the rate of case versus control mutation carriers. The data for this analysis can be seen in Table 3.5. The frequency of mutation carriers in the case and control samples is also detailed in the last column of the table.

Table 3.5 Association analysis 1: All rare exonic variants; n=29 variants				
	Mut Carriers	Mut Non-Carriers	Total	(Mut Carriers)
Cases	61	1917	1978	0.031
Controls	32	1085	1117	0.029
Chi-square = 0.118, p=0.732, OR = 1.08, 95CI = 0.6, 1.7				

Secondly, I wanted to look at just the rare non-synonymous variants in the data set. For this investigation I had narrowed down my analysis to 17 variants, and these were present in 50 case samples and 29 controls (Table 3.6). Once again when Chi square, odds ratio and P values were calculated for these, I did not find a statistically significant difference between cases and controls.

Table 3.6 Association analysis 2: All rare non-syn variants; n=17				
	Mut Carriers	Mut Non-Carriers	Total	(Mut Carriers)
Cases	50	1928	1978	0.025
Controls	29	1088	1117	0.026
chi-square = 0.013, p=0.906, OR = 0.973, 95CI = 0.6, 1.6				

Finally, I examined the 6 non-synonymous damaging variants from my Annovar output. I found 13 carriers in the case samples and only 5 carriers in the controls (Table 3.7). The frequency of mutation carriers is nearly double for the cases compared to controls, but given the small sample numbers involved, the data is not statistically significant (p=0.624).

Table 3.7 All rare non-syn damaging variants; n=6				
	Mut Carriers	Mut Non-Carriers	Total	(Mut Carriers)
Cases	13	1965	1978	0.007
Controls	5	1112	1117	0.004
chi-square = 0.542, p=0.624, OR = 1.471, 95CI = 0.49, 4.7				

3.5 Discussion

A combined analysis of Irish/UK subjects including additional psychosis cases (SCZ and BD) identified 22 carriers in 11,707 cases and 10 carriers in 21,204 controls (meta-analysis CMH $P=2 \times 10^{-4}$ (odds ratio (OR)=11.3, 95% CI=3.7, ∞)) found an association between duplications at chr20p12.2 and SCZ ($P = 0.007$). Nineteen of the 22 cases and 8 of the 10 controls carried duplications starting at 9.68Mb with similar breakpoints across samples.

I was given an opportunity to collaborate with the group on this detailed study of duplications present in the *PAK7* gene in a number of SCZ patients. By sequencing I was able to show that the CNV at *PAK7* is not the result of a recurrent de novo event but is an inherited mutation potentially from a single European ancestor. The mutation involves a tandem duplication of 148,951bp at chr20:9,684,767-9,833,717 (hg18) overlapping the *PAK7* gene. The risk duplication identified in this study overlaps the first two exons, enhancer and promoters regions in the known transcripts of *PAK7*. Because *PAK7* is exclusively brain expressed the impact of the duplication on gene expression and the genetic mechanisms involved are yet to be determined. Further clinical and genetic assessment of patient carriers, their families and carriers identified in other disease or normal populations will be important in quantifying the penetrance and range of phenotypic expression of the *PAK7* duplication.

I performed functional annotation of the genetic variants in the neurodevelopmental and obesity samples using Annovar. Eighty two variants were identified in the cases and controls, on which I carried out a carrier-based association analysis, but found no statistically significant differences in the rate of mutation carriers between cases and controls. Interestingly, looking at the rare non-synonymous variants, I observed that the frequency of mutation carriers is nearly double for the cases compared to controls but this was not statistically significant. One reason for this could be that the sample number does not give me sufficient statistical power to detect association in this dataset. Overall, the

UK10K dataset did not provide support for rare genetic variants at *PAK7* leading to increased risk of neurodevelopmental disorders.

Chapter 4

Rare Variant Detection

Testing the Specificity and Sensitivity of High Resolution

Melt Analysis (HRM) for Rare Variant Detection

4.1 An alternative method of rare variant detection

The goal of genetic research in SCZ is to identify the genetic loci that confer risk in order to gain insight into the as yet, relatively understood pathophysiology and ultimately improve diagnosis and treatment and reduce mortality and morbidity. The methodologies and technologies used to investigate the genetic basis of these and other such complex illness are advancing at a rapid pace and continuously offering new insights into how the genome can confer susceptibility to disease. NGS technology is a research tool to detect rare variants that may be associated with a rare genetic disease. Different research approaches can be employed to distinguish causative from non causative rare variants. However NGS is still an expensive method of rare variant detection, especially with small sample numbers, and genes under investigation. My approach in this section of my thesis was to look at the potential of High Resolution Melt (HRM) analysis to identify rare variants, which would be further investigated using Sanger sequencing.

4.2 HRM Methodology

The appeal of HRM analysis screening from Applied Biosystems (<http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/real-time-pcr/high-resolution-melting-hrm.html>) as a tool for mutation detection is its potential for rapid, high-throughput, inexpensive and sensitive post-PCR analysis. The methodology is based on using dsDNA-binding dyes and real-time PCR instrumentation. PCR is followed by using saturating concentrations of a fluorescent dye which specifically intercalates with double stranded DNA, the denaturation behaviour of the amplified region can then be analysed.

The amplicon DNA is heated from 50-95°C, the dye dissociates from the double stranded DNA as it denatures into single stranded DNA and thus the melting can be monitored using the decrease in fluorescence (Figure 4.1).

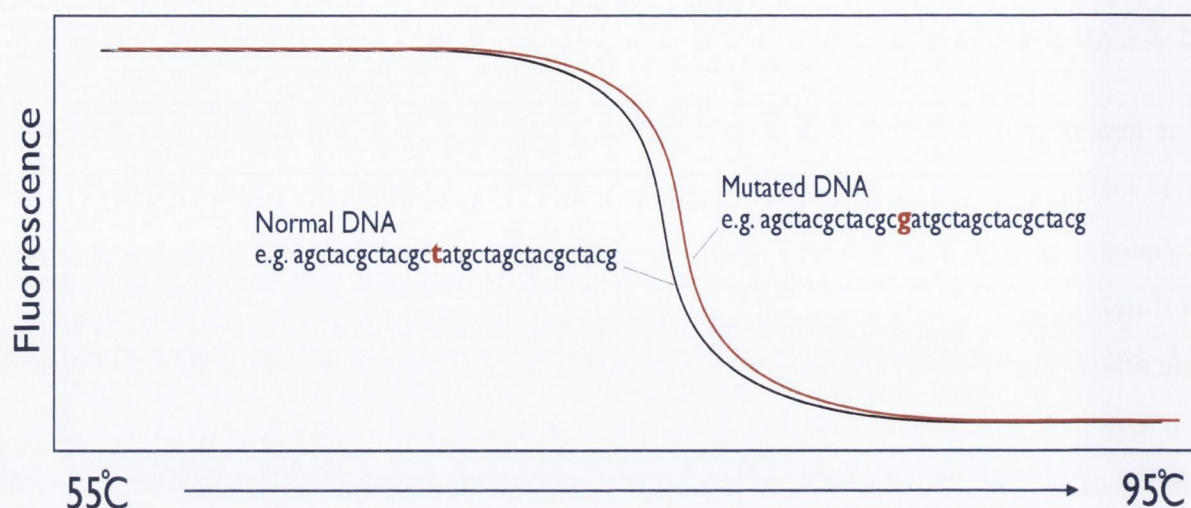


Figure 4.1 Melt curves of denatured double stranded DNA.

Observations in the resulting melting curve are dependent on a number of factors including the sequence composition and length, GC content and strand complementarity. HRM enables investigation of genetic variation including SNP genotyping, DNA methylation, heterozygosity screening and mutation scanning. HRM is also a closed tube method, which allows for minimal downstream processing and decreased contamination. I used this method to investigate its efficiency at detecting rare variants in two target SCZ-associated genes. If this method was successful at identification, I would also confirm using Sanger Sequencing analysis.

4.2 Gene selection

4.2.1 *PAK7*

HRM was applied to the exons of the *PAK7* gene, which was selected following evidence of association between duplications at chr20p12.2 and SCZ ($P = 0.007$). (Morris et al., 2014) (see Chapter 3).

4.2.2 VIPR2

VIPR2 (7q36.3) encodes the VPAC2 receptor which binds vasoactive intestinal peptide (VIP) in the brain. *VIPR2* helps to regulate the formation and activity of neurons in the brain and has been found to play important roles in behavioural processes in mice (Vacic et al., 2011). In a SCZ study carried out by Vacic *et al*, microduplications with variable breakpoints were identified within a 362kb region, and were detected in 29 of 8,290 cases compared with 2 of 7,431 control samples (Vacic et al., 2011). All were located within 89kb upstream of the *VIPR2* gene. Subsequent analysis by Levinson *et al* (Levinson et al., 2011) and De-Jesus Hernandez *et al* (DeJesus-Hernandez et al., 2011) have confirmed this region as an area of interest for SCZ.

4.3 Materials and Methods

4.3.1 Samples

HRM analysis was carried out on 78 CEPH samples, which represent trios with ancestry from northern and western Europe. These samples were chosen due to the fact that genotype data is freely available for them from both the HapMap (<http://hapmap.ncbi.nlm.nih.gov/>) and 1,000 Genomes (<http://www.1000genomes.org/>) websites, allowing me to cross-check my HRM results with available online data. By using this method of analysis and choosing not to set controls for each sample, I was able to look at each fragment without prior knowledge of the common variants. My primary interest in testing this method was to assess its viability for detection of rare variants rather than genotyping common polymorphisms.

4.3.2 Primer Design

The first step was the careful design of primers for the coding sequence of each of my genes of interest (Table 4.1). *VIPR2* and *PAK7* exons were determined using the NCBI CCDS database (July 2011) (<http://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi>), and a total of 11 amplicons were designed for the *VIPR2* gene, and 5 for the *PAK7* gene. Primers were

designed using Primer 3 software (<http://primer3.ut.ee/>). I was unable to design primers for *VIPR2* Exon1 and 13 and Exon 6 of *PAK7*. Amplicons with a target length of <350bp were chosen to ensure sensitive detection of sequence variants. Other primer specifications included a length <20bp, a primer melt temperature of between 58-62°C and GC content was between 30-80% (Table 4.1). Quality checks were carried out for each primer set, including a blast search and standard PCR on 2 HapMap samples in duplicate to test the robustness of my chosen primers.

Table 4.1: Forward and reverse primers for Exonic regions of <i>VIPR2</i> and <i>PAK7</i> .				
Gene	Exon	Forward Primer	Reverse Primer	Size (bp)
<i>VIPR2</i>	2	CCTGAACCACAGCTGAGTGA	CTGTGAGGTTTGGGGTTGTT	300
	3	GAACTCAACCGACGGACAGT	CCTGCCTCACTCTGCTTCTC	188
	4	AGTGGAACCTCACCACCGTA	GGCCAGTTTTTCATGGACTTG	198
	5	CGCCTAGAAGGCCTAAATGG	GCTGGTCACCCTCAGCTTAC	268
	6	AGACAGTGGGACCCTGTGGT	CCCCTCCAATTCCTCTCTCT	242
	7	AAAATGTTTGCGTTGTTTGG	TCTCCTCTGTTTGCGGTTTC	291
	8	TGCAGTCTGGTCATGGACTC	GTAACCAGGCGTCTCTCTGC	121
	9	TGTCTGCCCAAGTGTGTGAT	GCGAAAGTGGTTGGTGGTAA	180
	10	GGGGTCTCCTGTCTCCACAC	GGGATCCACGTGTCTGTTTC	232
	11	GAAGAGATGGTCAGGCAGGA	GCTGGCTCATCTCCTTCTGT	290
	12	GCATCTCAGGAAGCAAGTGA	TCCTGCCTGACCATCTCTTC	182
<i>PAK7</i>	1	GCTTTGCCACCTACACGAAT	ATTTTGGGGACTGGGTTAGG	296
	2	GCTGGATGTGGTAAGCCAGT	CCAGTAGGGTGTGCCTAACA	335
	3	AGAACCAATGGGTCCCTGAG	TTGCATTCTTTTCTTTTCA	266
	4	GCCCATGACAGACAGAGTGA	TGCCTCTCTTGAGCATTTCA	267
	5	AGCACCTCGAATTCTCTGGA	TGGCCACAAGAGAACAAAAA	274

DNA quality and quantity were verified, and a starting concentration of 20ng/μl was used. HRM analysis, and subsequent melt curve analysis was carried out on 78 HapMap samples for each of the *VIPR2* and *PAK7* fragments. HRM PCR components can be see in Table 4.2.

Table 4.2: HRM Sample Preparation PCR	
Meltdoctor HRM mastermix	10μl
Forward primer (5μM)	1.2μl
Reverse Primer (5μM)	1.2μl
H ₂ O	5.6 μl
DNA	2 μl
Total	20 μl

PCR cycling conditions were as follows 95°C for 10mins, followed by 40 cycles of 95°C for 15secs and 60°C for 1min. Subsequent melt curve conditions were Denaturation 95°C for 10secs, Annealing 60°C for 1min, HRM at 95°C for 15secs and final annealing at 60°C for 15secs. Pre and post-melt temperatures were set as close as possible to the melting transition region of each derivative melt curve. Melting data was also normalised, temperature shifted and displayed as derivative curves.

Analysis was performed using the aligned melt curve which shows discrimination of the variants in their melt behaviour. Analysis was also carried out using the derivative melt curve which allows me to set out pre and post-melt regions for more accurate analysis. Difference and melt plots were also analysed for each fragment for all samples which allowed me to visualise small alterations in melt curves and discriminate differences in the change in fluorescence for individual samples.

I did not use the stringency of a known wild type, as my purpose in using this method of analysis was to identify rare variants and positive controls would not be routinely available.

4.3.3 Sequencing of exonic fragments

Sanger sequencing was carried out on DNA samples taken directly from the HRM analysis plate at a concentration of 1ng/ul. All samples were purified with Shrimp Alkaline Phosphatase (SAP, USB Corp) and Exonuclease 1 (Exo1, New England Biolabs) the plate was then run at 37°C for 30mins, followed by 80°C for 15mins. The sequencing reaction was then carried out using 10µl DNA, 2µl Big Dye Terminator v1.1, 1µl forward primer, 4µl H₂O and 1µl 5 x Buffer. Samples were denatured at 96°C for 1 min, followed by 25 cycles of 96°C for 10secs, 50°C for 3secs and 60°C for 75secs. Purification of the DNA sample was then carried out using 45µl SAM solution (Applied Biosystems) and 10µl Big Dye X Terminator (Applied Biosystems) (Table 4.3), which were vortexed for 30mins and spun at 1000g for 2mins. Samples were heated to 37°C for 1 hour followed by 80°C for 20 mins. Samples were subsequently run on the Genetic Analyser 3130 (Applied Biosystems).

Table 4.3: PCR Product Purification using ExoSAP-IT	
PCR product	3µl
Reaction components	3µl
SAP	1µl
Exo 1	0.2µl
Dilution Buffer	1µl
H ₂ O	2µl
Total	10.2 µl

4.4 Results

Mutation detection was carried out as follows: Dr. Morris and I independently undertook visual inspection of all plots and curves produced by HRM analysis for each amplicon with no prior knowledge of SNP genotypes for any common variants present. This data was combined with the automatic calls by the HRM to generate a list of samples for each amplicon that were likely to contain a variant and required follow-up sequencing analysis, to identify and confirm the presence of a variant.

Before HRM curves are plotted, the raw data is normalized. Melt curves are normally plotted with fluorescence on the Y axis and temperature on the X axis. Since different genetic sequences melt at slightly different rates, they can be viewed, compared and analysed using the curves generated for each sample. Figure 4.2 gives an example output from the HRM software where the rs2270313 SNP in the *VIPR2*-Exon 3 amplicon has been detected in a HapMap sample.

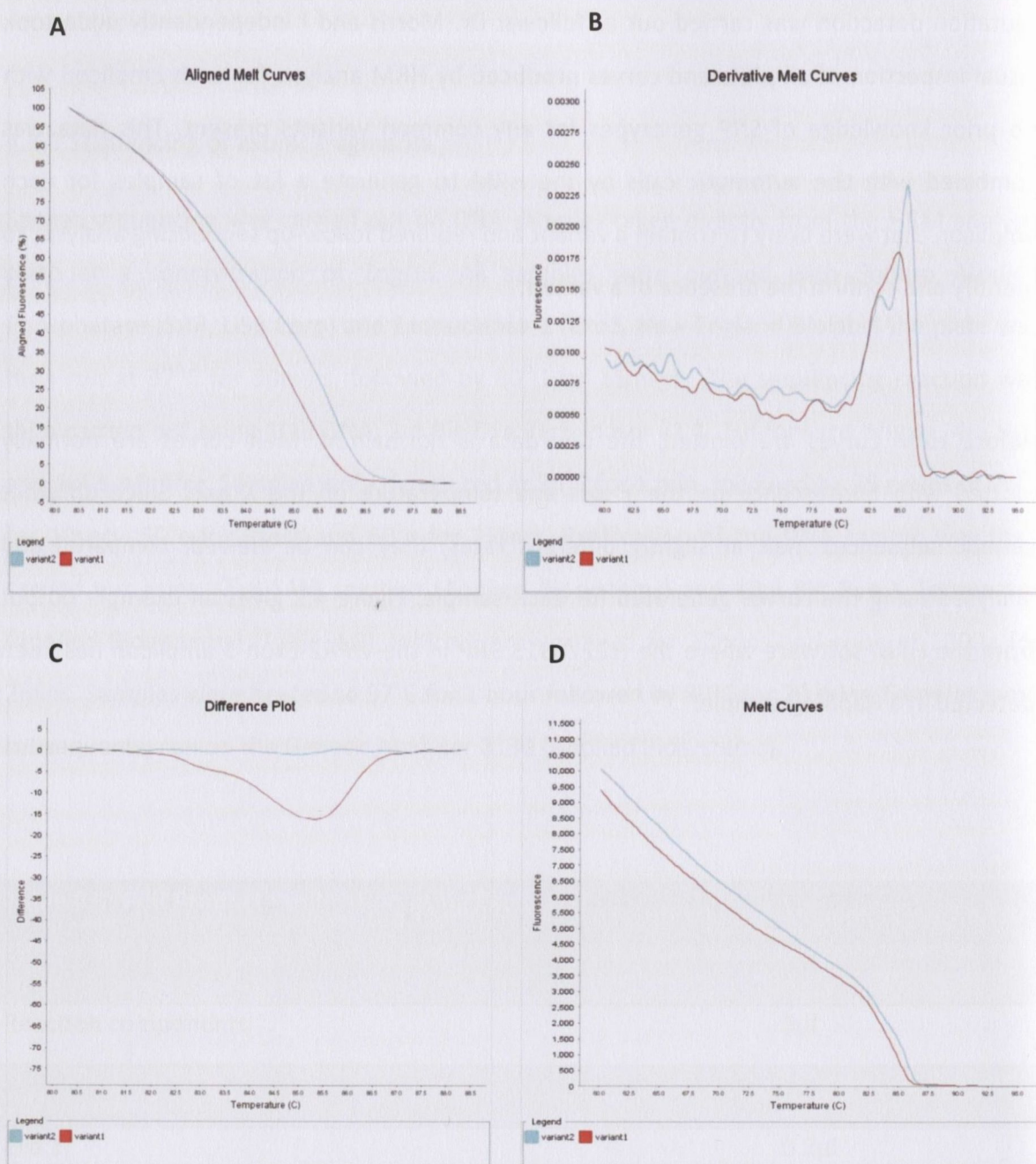


Figure 4.2 Analysis of HRM data from VIPR2-Exon 3. Different genotypes for rs2270313 are highlighted in different colours. A/G=red, G/G= Blue. **A.** The aligned melt curve shows discrimination of the variants in their melt behaviour. **B.** The derivative melt curve allows me to set out pre and post-melt regions for more accurate analysis. **C.** The difference plot allows me to visualise small alterations in melt curves. **D.** The melt curve allows me to discriminate differences in the change in fluorescence for individual samples.

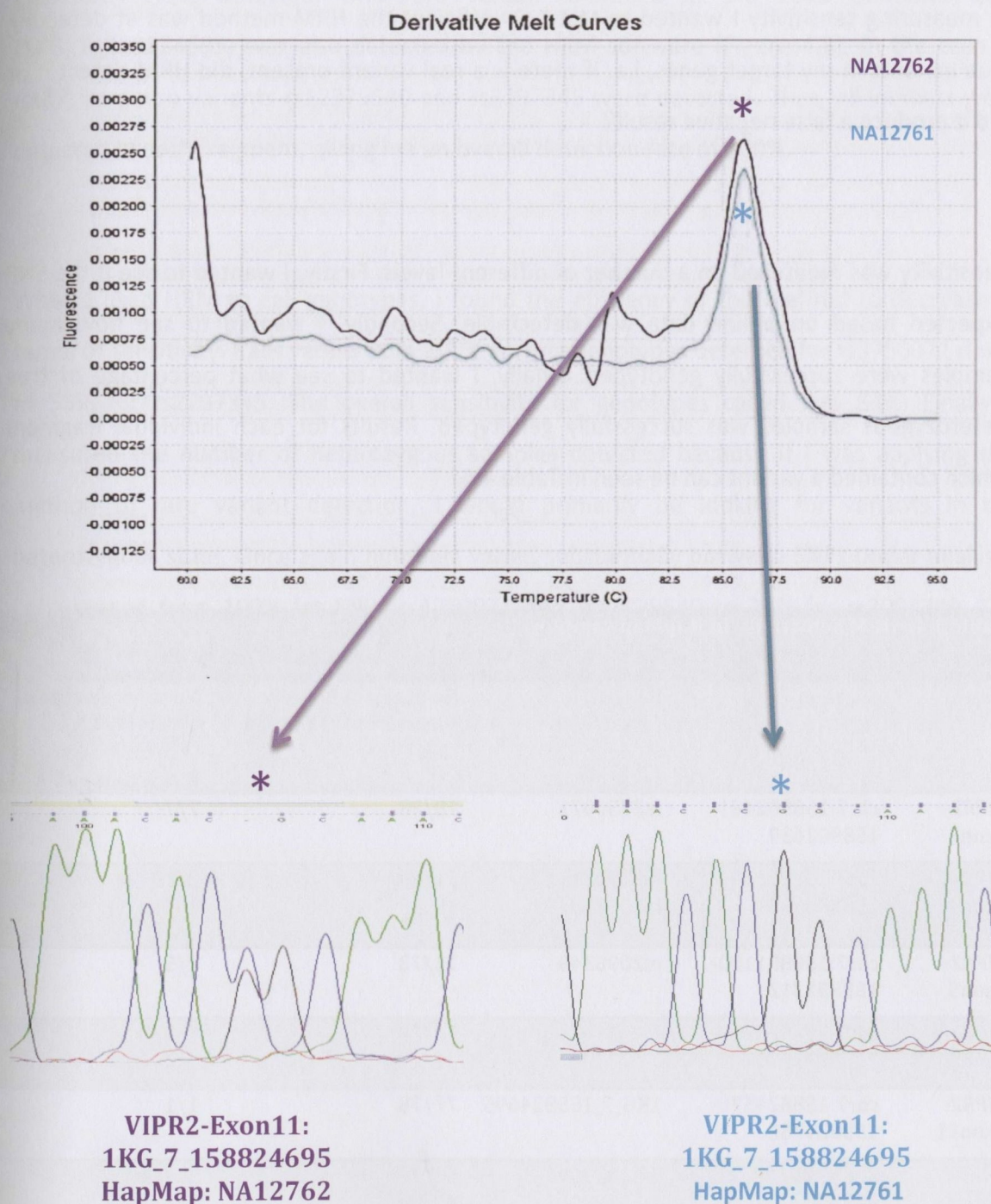


Figure 4.3: Derivative melt curves for HapMap samples NA12762 and NA12761 from Exon 11 of the *VIPR2* gene. Based on melt curve analysis, samples were subsequently analysed using Sanger sequencing and showed NA12762 to contain a rare variant (1KG_7_158824695) which has been reported in the 1000 Genomes data.

4.4.1 HRM Sensitivity Results

In measuring sensitivity I wanted to test how efficient the HRM method was at detecting true variants in my target genes, i.e. if there is a real variant present, did HRM detect it or did it produce a false negative result?

Sensitivity was measured on a number of different levels. Firstly, I wanted to see if the SNP expected based on online data was detectable. Secondly, I wanted to see how many samples were successfully genotyped. Finally, I wanted to see what percentage of true heterozygous samples was successfully genotyped. Results for each individual fragment which contained a variant can be seen in Table 4.4.

Table 4.4: HRM Sensitivity in HapMap samples

Gene	Location	Variant	No of genotypes called (#correct calls / #total calls)	No of heterozygous genotypes called (#correct calls / #total calls)
VIPR2-Exon3	chr7:158902481-158902639	rs3750071	78/78	7/7
VIPR2-Exon4	chr7:158896409-158896595	rs2270313 rs17837874	51/78 78/78	8/29 10/10
VIPR2-Exon5	chr7:158851100-158851312	rs2098349	72/78	1/5
VIPR2-Exon9	chr7:158827215-158827387	rs2071624	56/78	1/4
VIPR2-Exon11	chr7:158824570-158824782	1KG_7_158824695	77/78	1/1
PAK7-Exon 1	Chr20:9520082-9520327	rs12329540 rs6039495	62/78 77/78	21/21 1/1
PAK7-Exon 5	Chr20:9543502-9543737	rs2297345	42/78	6/11
Overall Sensitivity			84%	68%

All common variants in the *VIPR2* fragments rs3750071, rs2270313, rs2098349 and rs2071624 were detected. One rare variant identified in the 1000 Genomes data (1KG_7_158824695) was also detected by the HRM software (Figure 4.3). In the case of *PAK7*, common variants rs12329540 and rs2297345 were detected. Thus, all variants were detected in each fragment, giving me an overall detection rate of 100%.

When I used HRM to call genotypes, I found the efficiency of the method to decrease in terms of sensitivity. Calls varied from 100% correct genotypes detected for rs3750071 down to 53% for rs2297345. The overall sensitivity for genotypes called was 84%. Finally, I measured the number of heterozygous samples detected because if I was applying this method to rare variant detection, I would primarily be looking for variants in the heterozygous state. Once again numbers varied substantially between SNPs under analysis. Calls ranged from 100% for SNPs rs3750071 and rs12329540 to 20% for rs2098349. The overall sensitivity for heterozygous genotypes across all fragments was 68%.

4.4.2 HRM Specificity Results

In measuring specificity I wanted to test how accurate the HRM method was at detecting amplicons that did contain true variants, i.e. if the HRM suggested a variant was present, was that correct or had it produced a false positive result?

Unlike the sensitivity analysis that focused on the amplicons that contained known variants in the HapMap samples, the specificity analysis included all amplicons that had been successfully amplified. In order to test the specificity of the HRM methodology, I differentiated the calls into ones which varied only slightly from the normal plot and called them ‘Possibles’, and samples which varied more substantially which were termed ‘Probables’. For the purpose of this study I carried all Possible and Probable samples through for Sanger Sequencing to confirm or not the presence of a variant, in order to test the specificity of the calls made.

The results from this analysis are presented in Table 4.5. Here I am presenting the specificity results for sample amplicons considered as ‘Probable’ carriers of a variant. Ten amplicons produced different HRM profiles that suggested the ‘Probable’ instance of a variant. In some cases, e.g. *VIPR2*-Exon2, there was just one different type of profile. In other cases, e.g. *PAK7*-Exon 1, there were multiple different profiles. For each different profile, a sample that produced that profile was sequenced to identify the variant. As can be seen in the third column of Table 4.5, this was performed with varying degrees of success. For example, for *VIPR2*-Exon2, the one ‘Probable’ profile did not yield a true variant. For *PAK7*-Exon 1, only one of the four ‘Probable’ profiles produced a true variant. Thus, the remaining three profiles represented false positive results. I calculated an overall specificity by determining what percentage of ‘Probable’ profiles identified a true variant. The result of just 20% indicates a very high false positive rate.

Table 4.5 HRM Specificity in HapMap samples			
Gene:Exon	Amplicon size (bp)	No of Different 'Probable' Profiles	Number of Variants detected with Sanger Sequencing
VIPR2-Exon 2	300	1	0/1
VIPR2-Exon 3	188	2	1/2
VIPR2-Exon 4	198	3	1/3
VIPR2-Exon 6	242	2	0/2
VIPR2-Exon 7	291	2	0/2
VIPR2-Exon 11	290	1	0/1
PAK7-Exon 1	296	4	1/4
PAK7-Exon 3	266	1	0/1
PAK7-Exon 4	267	3	0/3
PAK7-Exon 5	274	1	1/1
Overall Specificity			20%

4.6 Discussion

High resolution melt analysis is a method which potentially offers the user a post PCR method to identify variants in nucleic acid sequences. The mutation scanning ability of this method scans DNA samples for minor genetic variations as candidates for further direct sequencing. Low reagent consumption, with little waste, and also only requires a PCR volume of 20ul. In theory it is a simple, fast workflow with no additional instrumentation required after PCR amplification. Another benefit of it is the low sample consumption: following HRM analysis, the PCR amplicon can be used directly in a Sanger sequencing reaction.

I chose this method because in theory it had the potential to be both fast and more cost effective than other methods of variant analysis such as sequencing for small numbers of genes. The method's promise of low reagent consumption, a simple fast workflow and low sample consumption made it an ideal candidate for rare variant analysis. I choose HapMap samples to test the efficiency of this method, as DNA sequence information for all of these samples is publicly available on online databases. My aim was to test the method's efficiency and cost-effectiveness before applying it to my large case-control sample.

I decided to define the efficiency of this method using two criteria. Firstly, was the method sensitive enough, i.e. did the method detect all variants that were present or did it miss some (false negatives). A high rate of false negatives would comprise the suitability of the method for a rare variant detection study. Secondly, was the method specific enough, i.e. did the method only identify amplicons that contained true variants or did it identify many amplicons that when sequenced were found not to contain variants. A high rate of false positives would result in a lot of extra unnecessary and costly lab work and again would comprise the methods efficiency. Based on the HRM analysis carried out I decided not to carry the method through to the psychosis case-control samples. Its overall lack of both a high sensitivity and a high specificity meant that it was not a method that would be useful

for identifying rare variants. Although the method was able to detect common variants, and in one case a rare variant from 1000 Genomes data, only 68% of heterozygous genotypes were called correctly, and the HRM calls led to a substantial number of false positive samples being brought forward for subsequent sequencing.

HRM was chosen because it promised fast and sensitive SNP detection, however this was not true for my set of experiments. There are a number of reasons why this may have happened. The Applied Biosystems website recommends designing primers which capture 120-300bp amplicons. The amplicons that I designed ranged from 121-335bp with an average of 246bp. Therefore, amplicon size is not a likely explanation for the poor performance of this method in my hands. This method also recommended running at least 3 replicates of each PCR reaction. However this was not feasible if I were to use this method on a large-scale project. To have included a known variant may also have assisted in detection of mutations, but since I was using this method for rare variant detection, this was also not possible.

Probably the most important limitation of HRM analysis is that it may not reveal all the sequence variations in a DNA fragment. There is still a strong empirical component to the determination of which base changes are detected by HRM. Base differences that do not change the %G+C content are inherently more difficult to detect than those that do, and the probability that a base change is detectable by HRM is inversely correlated with DNA fragment length. In our experience and that of others, detecting single base changes in sequences of >200 bp is frequently problematic. The inherent simplicity of HRM analyses would suggest that HRM analyses should be highly reproducible. However, reproducibility can be poor if there are variations in initial nucleic acid template.

Chapter 5

Investigation of a Loss of Function Variant in *FYN*

5.1 Identification of a rare variant in the *FYN* gene

The backdrop to my work in this chapter is a targeted sequencing study undertaken by my group, and in which I was involved (Kenny et al., 2013). SCZ and ASD are complex neurodevelopmental disorders that may share an underlying pathology suggested by shared genetic risk variants. Our group sequenced the exonic regions of 215 genes in 147 ASD cases, 273 SCZ cases and 287 controls, to identify rare risk mutations. Genes were primarily selected for their function in the synapse and were categorised as: 1) Neurexin and Neuroligin Interacting Proteins, 2) Postsynaptic Glutamate Receptor Complexes, 3) Neural Cell Adhesion Molecules, 4) *DISC1* and Interactors, and 5) Functional and Positional Candidates. Thirty-one novel loss-of-function (LoF) variants that are predicted to severely disrupt protein-coding sequence were detected amongst 2,861 rare variants. We found an excess of LoF variants in the combined cases compared to controls ($p=0.02$). This effect was stronger when analysis was limited to singleton LoF variants ($p=0.0007$) and the excess was present in both SCZ ($p=0.002$) and ASD ($p=0.001$).

Understanding the impact of these LoF variants requires functional investigation. As it is not feasible to investigate all variants, just one was selected; a nonsense mutation in *FYN* at chr6: 112025283 in a SCZ patient.

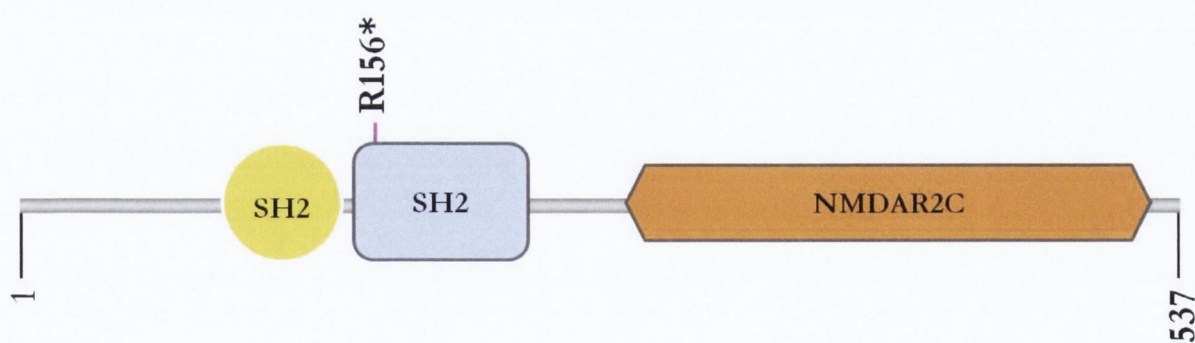


Figure 5.1 Diagram of *FYN* protein showing functional domains and the site of variant under investigation.

This nonsense mutation was identified at R156 in the *FYN* gene (Figure 5.1). *FYN* regulates the phosphorylation and trafficking of NMDA Receptors (NMDARs), acting as a point of convergence for many signalling pathways regulating NMDARs (Hennah et al., 2009). *FYN*-mutant mice exhibit blunting of long term potentiation and impaired spatial learning, plus other neurological defects including un-coordinated hippocampal architecture and reduced neural CAM-dependent neurite outgrowth (Beggs, Soriano, & Maness, 1994; Grant et al., 1992).

Increased expression of *FYN* has been reported in the prefrontal cortex of SCZ patients compared to controls (S. E. McCarthy et al., 2009). Hattori *et al* reported reduced expression of *FYN* in SCZ in a study of platelet samples from SCZ patients, first-degree relatives and controls (Hattori et al., 2009). Altered *NRG1* (Neuregulin 1)-*erbB4* signalling has been suggested to contribute to *NMDA* hypofunction in SCZ (Hahn et al., 2006) and *NRG1-erbB4* signalling stimulates *NR2B* phosphorylation through the activation of *FYN* (Bjarnadottir et al., 2007). Studies using *FYN*-deficient mice support a role for *FYN* in the induction of epilepsy (Cain, Grant, Saucier, Hargreaves, & Kandel, 1995) and data from *FYN*-transgenic mice indicates that this may be mediated via phosphorylation of *NR2B* (Kojima et al., 1997).

5.2 My approach to further understanding the affect of this LoF variant

My approach to investigate this variant was to design plasmids containing green fluorescent protein (*GFP*)-tagged wildtype and mutant versions of the gene and transfect them into neuroblastoma cell lines first to test the efficiency, and to then move on to primary cortical neurons from newborn mice. Immunohistochemistry and confocal microscopy could then be employed to determine the transfection efficiency and localisation of the wildtype and mutant proteins, using antibodies specific to *FYN* and Phospho-akt

5.3 Plasmid Design

Plasmids containing the wildtype (WT) and mutant (Mut) versions of *FYN* were ordered from Genecopoeia (www.genecopoeia.com). cDNA sequences of the WT and Mut versions of the human transcript were provided to the company to aid design. For the WT plasmid, the full coding sequence was provided. For the Mut plasmid, the sequence up to, but not including the mutant stop codon was provided. The plasmids were designed with *GFP* located upstream of the coding DNA inserts so that the resultant WT and Mut proteins would be co-expressed with *GFP*. The expression of *GFP* is used as a marker of successful transfection of the plasmid.

The plasmids contain a very strong promoter, human cytomegalovirus (CMV) promoter, which maximises the expression of genes within the plasmids. Each plasmid has an ampicillin (AMP) gene, allowing selection of transformed E.coli cells in carbenicillin containing media (Figure 5.2). A third plasmid with just *GFP* was also ordered. This plasmid acts as a control to determine the effect of transfection on neurons. A schematic of all three plasmids can be found in Figure 5.3. The insert is located at site marked ORF (open reading frame) in green and is *GFP* tagged.

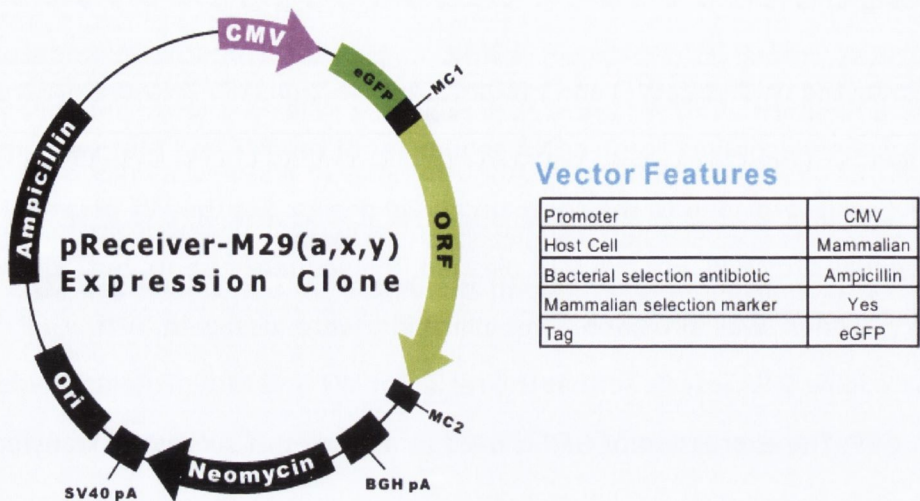


Figure 5.2: Expression clone with CMV promoter. Plasmids contain a CMV promoter. The insert is located at site marked ORF (open reading frame) in green and is *GFP* tagged. There are two selection markers; neomycin and ampicillin.

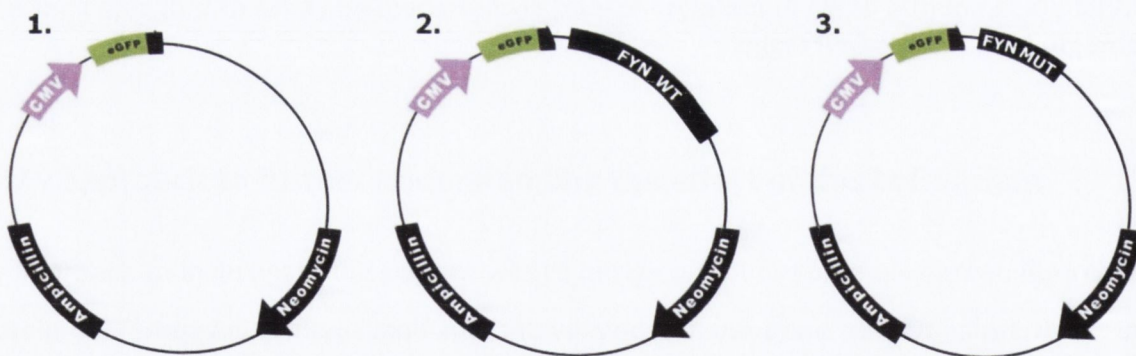


Figure 5.3: Schematic of the plasmids used in transfection 1. *GFP* plasmid (**6.5KB**) 2. *FYN* WT (wildtype) plasmid tagged by *GFP* (**7.5KB**). 3. *FYN* MUT (mutant) plasmid tagged by *GFP*.(**6.5KB**)

5.4 Materials and Methods

5.4.1 Plasmid Preparation

As limited plasmid stocks are provided, it is necessary to increase the amount of plasmid using a technique known as transformation, which allows the introduction of foreign DNA into bacterial cells. Plasmids were transformed into competent E.coli cells (JM109 Promega), using a 'heat-shock' protocol as outlined below. The plasmid is mixed with the competent cells, cooled on ice and then subjected to a 'heat shock', which disrupts the cell surface allowing entry of the plasmid. Not all bacterial cells will be successfully transformed thus a selection marker is required. In this case the plasmids contain an ampicillin gene so only successfully transformed cells will grow in media or agar containing carbenicillin (ampicillin confers resistance to carbenicillin). The transformed cells are allowed to multiply and the plasmids are then purified from the cells. This is achieved as outlined below.

5.4.1.1 Plasmid Transformation

1. Label and chill eppendorf tubes (1 per insert plasmid, one positive control and one negative control).
2. Warm the S.O.C. medium (Super Optimal broth with Catabolite) repression Media to 37°C in a water bath and set another water bath to 42°C.
3. Remove JM109 cells from the -80°C freezer and thaw on ice for 5 mins: enough for 50µl per plasmid insert transformation and 25µl for each of the positive and negative control.
4. Once thawed, pipette 50µl into the plasmid insert tubes and 25µl into the control tubes.
5. Make a 15ng/µl dilution of each plasmid with nuclease free water in new eppendorf tubes.
6. Add 1µl of the plasmid dilutions to the corresponding tubes containing 50µl JM109 cells. For the positive control add 1µl of the control plasmid provided with the JM109 cells and for the negative control add 1µl H₂O.

7. Heat-shock the cells for 45 secs at 42°C in a water bath. Immediately place the tubes on ice for 2 mins.
8. Add 400µl warm (37°C) S.O.C. medium to each tube (prepared as per section 2.3.12.4).
9. Incubate for 1hr at 37°C in a water bath. Allow LB agar plates (prepared as per section 2.3.12.4) to reach room temperature.
10. For the insert plasmids inoculate plates with 200µl, 100µl, 50µl and 25µl of the 400µl transformation media, spreading evenly across the plate with a plate spreader (this number of dilutions is performed in the initial transformation to determine the dilution that gives an adequate number of well-spaced colonies, just two dilutions are used in future transformations).
11. For the positive and negative control inoculate one agar plate each with 50µl of the 400µl transformation media. Discard the remaining media.
12. Incubate the plates in a 37°C oven overnight.

5.4.1.2 Post Transformation

Presence of colonies on the agar plates indicates successful transformation of plasmids into the bacterial cells. These colonies are selected and grown up in media as below.

1. Warm S.O.C media to 37°C in a water bath.
2. Ensure the positive control plate has colony growth and the negative control does not.
3. Add 200µl warm S.O.C media to six wells per plasmid of a 24-well cell culture plate (Nunclon. Fisher Scientific TKT-190-010Y).
4. Select four transformants and add to the wells of the culture plate using an inoculation loop.
5. Incubate at 150 rpm in a shaking incubator at 37°C for 8 hours.

5.4.1.3 Colony Screen

A colony screen must be performed to ensure that the bacteria cells do contain the plasmids and that the plasmids have the correct insert. Primers were designed using primer3. The

maximum amplicon size was set to 900bp and they overlapped by a minimum of 70bp to ensure near complete coverage of the insert sequence. PCR cycling conditions were as follows 95°C for 10mins, followed by 40 cycles of 95°C for 15secs and 58°C for 1min. For the *FYN* wildtype, 3 primer sets were designed, one of which amplified the entire shorter *FYN* mutant insert (Table 5.1).

A PCR master mix was set up for each separate primer set, with adequate volumes for the number of colonies being screened. A pipette tip is dipped into the broth and then into the PCR reaction. This is sufficient as too much broth will impair the PCR reaction. PCR is then performed as per Table 5.2 for the wildtype plasmid; PCR reactions are set up for all 3 primer sets. For the mutant plasmid, a PCR reaction is set up for primer set 1 that covers the plasmid and also with primer set 2, which should not amplify. This is a control measure to ensure that the plasmid contains the mutant sequence.

Table 5.1 <i>FYN</i> colony screen primers			
Amplicon Name	FWD Sequence	Rev Sequence	Amplified in
Set 1	ATGGGCTGTGTGCAATGTAA	GCCAAGTTTTCCAAAGTACCA	WT/MUT
Set 2	CAGCAATTATGTGGCTCCAG	GCTCGATGTAAGCCATTCT	WT only
Set 3	CATCGTCACCGAGTATATGAACA	GGCTCTGTCGCGGTAAAGTA	WT only

Table 5.2 PCR colony screen	
Transformed Plasmid	<0.5µl
10X buffer(containing 15mM MgCl2:Qiagen)	1µl
Forward Primer (25µM)	1µl
Reverse Primer (25µM)	1µl
dNTPs (2.5mM of each dNTP: NEB)	1µl
Taq (5U/µl; Qiagen)	0.1µl
dH ₂ O	5.5µl
Total	10µl

Following PCR, the products are analysed on a gel to ensure that they have a band of the correct size for all 3 primers sets for the wildtype insert and just for primer set 1 for the mutant insert.

5.4.1.4 Growth of Positive Colonies

Once it is confirmed that the colonies contain the plasmids with the correct insert, they are grown up to attain greater levels of the plasmids. This is achieved as described below:

1. Warm S.O.C media to 37°C in a water bath.
2. Add 200ml warm media to a conical flask (2 per plasmid).
3. Select two positive cultures with the cleanest and brightest bands on the gel for each plasmid.
4. Take the 200µl media from the starter culture of these colonies and add to the labelled conical flasks. To the remaining positive cultures add 500µl media (to be stored as a back-up). Incubate all at 100-150 rpm in a shaking incubator at 37°C overnight.

5.4.1.5 Plasmid Purification

Once large quantities of transformed cells are attained, it is necessary to purify the plasmids from the cells. This is achieved using the Promega midi-prep kit. This system purifies plasmid DNA using a silica membrane column and a variety of wash buffers which remove protein, RNA and endotoxin contaminant from the plasmid DNA. This is performed with the following steps:

1. Remove the conical flasks from the shaking incubator.
2. Pour 200ml of culture to four 50ml falcon tubes.
3. Pellet cells at 4,000g for 10 mins. Discard the supernatant.
4. Resuspend one cell pellet thoroughly in 12ml of Cell Resuspension Solution by pipetting and transfer to new 50ml tubes and so on until all four pellets have been resuspended and combined in 12ml of Cell Resuspension Solution in one 50ml tube.
5. Add 12ml of Cell Lysis Solution. Invert gently 3–5 times to mix and incubate for 3 mins at room temperature.
6. Add 12ml of Neutralization Solution. Invert gently 10–15 times to mix. Does not over mix as this may cause the centrifuged lysate to be cloudy which can cause clogging of the columns.
7. Centrifuge the lysate at 7,000g for 30 mins at room temperature using a fixed angle rotor. If the lysate is not sufficiently clear add to a fresh 50ml tube and centrifuge for another 5-10 mins.
8. Assemble the blue PureYield™ Clearing Column and white PureYield™ Maxi Binding Column in a stack, with the clearing column on top. Place this column stack on the vacuum manifold.
9. Pour one half of the lysate into the blue PureYield™ Clearing Column.
10. Apply maximum vacuum until the lysate has passed through both the clearing and binding columns. Then add the remaining lysate and maintain vacuum until the liquid has cleared both columns.
11. Slowly release the vacuum. Then remove and discard the blue PureYield™ Clearing Column, leaving the PureYield™ Maxi Binding Column on the vacuum manifold.
12. Add 5ml of Endotoxin Removal Wash to the PureYield Maxi Binding Column, apply a vacuum and allow the solution to be pulled through the column.

13. Add 20ml of Column Wash (ensure the recommended volume of ethanol has been pre-added) to the binding column, and allow the vacuum to pull the solution through the column.
14. Dry the membrane by applying a vacuum for 8-10 mins.
15. Remove the PureYield™ Maxi Binding Column from the vacuum manifold, and tap the tip of the column on a paper towel.
16. Place a 1.5ml microcentrifuge tube into the base of the Eluator™ Vacuum Elution Device.
17. Assemble the Eluator™ Vacuum Elution Device, and insert the DNA binding column into the device and when secure place the assembly onto a vacuum manifold.
18. Add 500µl of Nuclease-Free Water to the DNA binding membrane in the binding column. Wait for 1 min. Apply maximum vacuum for 1 min and remove the microcentrifuge tube.

5.4.1.6 Check Plasmid Quality

The quality of DNA in the purified plasmids must be checked by Nanodrop (Thermo Scientific) and by agarose gel electrophoresis:

1. Quantify the plasmids using the Nanodrop and check the 260/280 quality ratios.
2. Make a 1% 60ml agarose gel with 3µl EtBr.
3. Make a 10µl ~10ng/µl dilution of each purified plasmid. Add 2µl 6X loading dye and load to gel.
4. Add Hind III marker, which yields bands of 2 - 23kb.
5. Run for 40mins at 100Volts.

5.4.1.7 Concentration of Plasmid DNA by Ethanol Precipitation

If the concentration of the purified plasmids is lower than that required for transfection, it is necessary to precipitate the DNA. This is achieved by ethanol precipitation as below:

1. Prepare 3M Sodium acetate (NaOAc) solution.
2. Add 11µl NaOAc for every 100µl DNA.

3. Add 2 X DNA volume of ice cold ethanol.
4. Precipitate for at least 30 mins at -20°C.
5. Centrifuge for 10 mins at 14,000g.
6. Pour off the ethanol and wash the sides of the tubes with ice cold 70% ethanol.
7. Centrifuge for 10 mins at 14,000g.
8. Pour off the ethanol and allow to air dry until all the ethanol has evaporated.
9. Resuspend in an adequate volume of TE.

5.4.1.8 Sequence Plasmid Inserts and adjacent plasmid sequence

Once it is determined that the plasmid is of adequate concentration and quality (based on 260/280 value being in the range of 1.8-2.2), the insert must be sequenced to ensure the correct insert is present and no errors were introduced during transformation. PCR products from the colony screen were sent to Source Bioscience for sequencing to ensure the plasmids contain the correct sequence. The resulting sequences were aligned to reference cDNA using T-Coffee (www.tcoffee.org/). Primers were also designed to incorporate the adjoining sequence to the insert, and samples were sequenced to ensure no variation in sequence.

5.4.2 Preparation of SH-SY5HY Neuroblastoma Cell Lines

Preparation of SH-SY5HY neuroblastoma cell lines were carried out on a newly ordered batch and all experiments were carried out on fresh stocks, with a maximum passage number of 25. Initial preparation was carried out as per manufacturer's guidelines. Cells were grown using DMEM/F12 medium containing 2mM glutamine, 2mM penicillin/streptomycin and 10% foetal bovine serum (FBS) and stored at 37°C in 5% CO₂. The cells were split every 3-4 days as per protocol.

1. Empty media from the confluent flask and wash the cells with 5ml warm (37°C) PBS.
2. Dilute trypsin to 1x concentration using PBS

3. Remove PBS from cells and add 5ml trypsin. Incubate at 37°C until the cells have detached (5 mins).
4. Add 5mls warm media to halt the trypsin.
5. Centrifuge for 2mins at ~1200rpm
6. Pour off supernatant and resuspend in 10mls fresh media.
7. Count cells –if required
8. Split cells at an appropriate dilution into new culture flasks/plates.

It is important that you do not culture SH-SY5Ys beyond P30 or if an overt phenotype change has been observed (cells may change shape and become less 'neurone like').

5.4.2.1 Preparation of Cell lines for Karyotyping

Karyotyping of the SH-SY5HY Neuroblastoma Cell Line was carried out to ensure that it was suitable for further experimentation. After multiple passaging of cell lines, chromosomal rearrangements can occur, and so I sought to karyotype this cell line to ensure its suitability.

1. Cell lines were split the day before harvesting.
2. Flasks were grown to 75% confluency
3. 10 mg/mL colchicin was added to each flask and left overnight
4. Cells were washed with 1%PBS and trypsinise.
5. Cells were spun at 2,000 rpm for 2mins.
5. PBS was removed and pellet was resuspended in 500 µl of 0.56% KCl., and left at RT for 15 min.
6. Cells were spun at 2,000 rpm for 2mins.
7. KCl was removed and the pellet was resuspended in cold MeOH:Acetic Acid 3:1 and let sit on ice for 10 min.
8. Cells were spun at 2,000 rpm for 2mins.
9. Supernatant was removed, and pellet was resuspended in 5ml fixative (1:3) Methanol:Acetic acid.
10. 1 drop of suspension was placed on each slide.
11. Metaphase spreads were imaged on a light microscope.

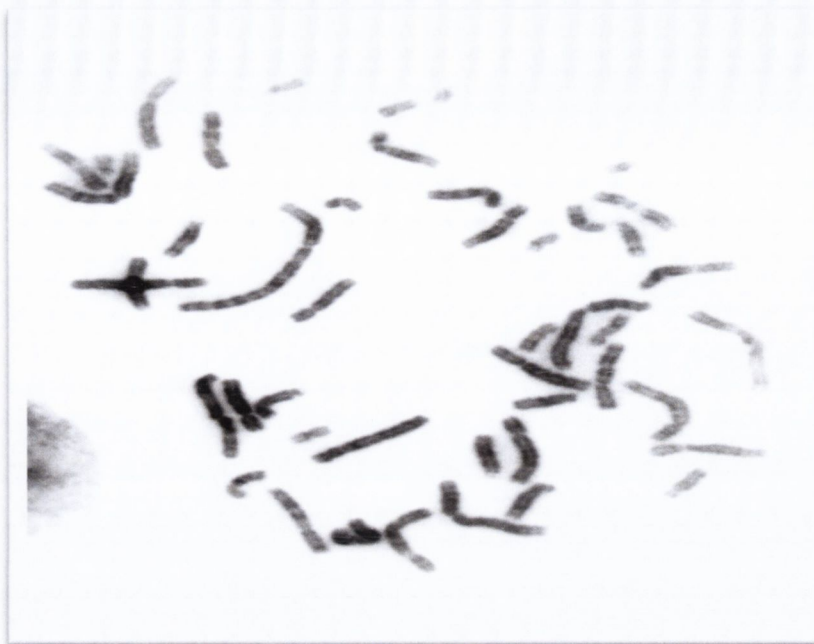


Figure 5.4: Metaphase spread of the SY-5HY neuroblastoma cell line. (modal number = 47; the cells possess a unique marker comprised of a chromosome 1 with a complex insertion of an additional copy of a 1q segment into the long arm, resulting in trisomy of 1q)

5.4.3 Preparation of Mouse Cell Cultures

Preparation of mouse cell primary cultures from mice primary cortical neurons was performed by Ines Molinos in our group as below. All pipettes, tubes, tips, surgical tools and Petri dishes were UV irradiated for 20 mins before use.

1. Cold DPBS (Dulbecco's Phosphate-Buffered Saline) (Invitrogen 14190-094) was added to each Petri dishes.
2. The pups head was cut and placed into a Petri dish.
3. In a new Petri dish with DPBS the head the skull was cut longitudinally
4. The skull was removed and brain removed.
5. The whole brain was placed in a clean Petri dish with DPBS.
6. 1µl DNase I and 20µl 10% warmed Trypsin-EDTA (Invitrogen R-001-100) was added and samples were left in the incubator at 37°C for 10 mins.

7. The reaction was blocked by adding double the volume (440µl) of DMEM (Dulbecco's Modified Eagle Medium) media + 10 % FBS (Fetal Bovine Serum; section 2.7.6.5).
8. The tissue was dissociated by pipetting up and down no more than 15 times with a P1000 pipette tip.
9. The sample was left to settle for 5-10 mins. The supernatant was transferred (~550µl) to a clean eppendorf tube.
10. The sample was spun at 800 rpm for 5 mins at 10°C.
11. Supernatant was removed without disturbing the pellet.
12. The pellet was resuspended in the same volume of DMEM + 10% FBS equal to the volume of supernatant removed.
13. 100µl diluted sample was placed on each coverslip (Sparks, MIC3366). and incubated for 2 hours in a 5% CO₂ incubator at 37°C.
14. 400µl warmed Neurobasal media and supplements were added and returned to the incubator overnight.
15. The next day 200µl media and add 200µl of 7µM feeding media are added.
16. Media is changed every two days

5.4.4 Transfection of Plasmids into Neuroblastoma cell lines

Transfection is a technique that allows the introduction of foreign DNA into eukaryotic cells. Transfection of *FYN* wildtype, *FYN* mutant and *GFP* plasmids were performed. This was performed using the Amaxa nucleofactor system and the nucleofection kit for mice neurons (Amaxa® Mouse Neuron Nucleofector Kit VAPG-1001) and for neuroblastoma cell line (Amaxa® Cell Line Nucleofector Kit V) as below. Control cells were also processed for transfection, but in the absence of the plasmid.

1. Add 200µl feeding media to number of wells required of a 24-well cell culture plate and equilibrate to 37°C in a 5% CO₂ incubator. Place a poly-lysine coated cover slide into each well.
2. Add the nucleofector solution to the nucleofector supplement and mix by pipetting. This can be stored for up to three months at 4°C. Equilibrate to room temperature before use.

3. For each transfection add 3µg plasmid into a sterile 1.5ml Eppendorf tube.
4. Once the number of mouse prefrontal cortex neuronal cells per sample is known, take the volume needed to have 6-700 cells.
5. Spin the cells at 80g for 5 mins. Carefully remove as much of the supernatant as possible.
6. Resuspend the pellet in 100µl of pre-warmed nucleofector solution/supplement mixture.
7. Transfer the neuronal suspension into the 1.5ml Eppendorf tube containing the plasmid and mix by gently flicking the tube for approximately 20 secs.
8. Transfer 100µl of cell/DNA mixture into a nucleofection cuvette and place it into the cuvette holder avoiding air bubbles.
9. Place into the Amaxa device and select programme G-05.
10. Immediately after nucleofection, carefully add 300µl pre-warmed DMEM (Sigma D5796)-10%FBS (section 2.7.6.5) to minimize shearing forces when retrieving the cells from the cuvette.
11. Use a plastic pipette supplied in the nucleofection kit to collect the cells from the cuvette and seed them into an equilibrated 24 well plate containing feeding media.
12. Incubate in a 5% CO₂ incubator at 37°C.
13. After 2-4 hours carefully replace medium with 500µl fresh feeding media to remove cellular debris.
14. Gene expression is often detectable after 6-8 hours and can be observed up to 12-14 days after Nucleofection.
15. After 24 hours, remove half the volume of media and add the same volume as that removed of 7 AraC feeding media.
16. Change media once or twice a week (use 3.5 M AraC Feeding Media).

5.5 Immunohistochemistry

Immunohistochemistry is the process of detecting a known antigen within cells or tissues by taking advantage of the specificity of antibody-antigen interactions. The antigen is the protein of interest, in this case *FYN* protein, and other biological markers. Fluorescently labelled antibodies can be produced that are specific to the antigen/protein of interest. In

order to increase specificity, an indirect two-step process is employed. An unlabelled primary antibody binds to the antigen. Then a secondary fluorescently labelled antibody is employed which is specific to the primary antibody. Upon binding of the secondary antibody to the primary antibody, fluorescence at a particular wavelength is emitted, which can be detected by confocal microscopy. This allows determination of the distribution and localization of the wildtype and mutant versions of *FYN*, and also for markers of dendrites and of synaptic function.

The level of non-specific background staining is low by this approach. Also, it is a very sensitive method as the signal is amplified due to the binding of many secondary antibodies to each primary antibody. A number of antibodies were used; *FYN* (marker of *FYN* protein), *GFP* (marker of *GFP* protein, an indicator of successful transfection), MAP2 (microtubule associated protein 2; acts as a dendritic marker) and phospho-akt. DAPI is also used which stains nuclei. Only four colours are detectable by confocal microscopy at one time, thus as DAPI (stains blue) and *GFP* (stains green) must be used in each slide to indicate the cellular location and type, *GFP* must be used to indicate transfection, and a combination of any other two antibodies can be used provided the colours are different.

5.5.1 Western Blot to test specificity of the *FYN* antibody

In order to test the specificity of the *FYN* antibody, both WT and *FYN*-KO spleen tissue was obtained from Dr. Elena Tarabra from Albert Einstein College of Medicine, New York, USA. I wanted to test the specificity of the antibody which corresponds to amino acid 35-51 (TPQHYPSFGVTSIPNYN) of human *FYN*, a region present in both WT and the truncated Mut plasmids.

5.5.1.1 Protein extraction from wild type and mutant *FYN* tissue

FYN WT and mutant tissue was sourced from Dr. Elena Tarabra from Albert Einstein College of Medicine. Protein was then extracted from both spleen and heart tissue to test for specificity of the *FYN* antibody by western blot analysis as follows:

1. Lysis buffer was prepared containing both ColP and NP40
2. 500 µl of lysis buffer was added to each tissue sample of heart and spleen of both wild type and mutant
3. Samples were then homogenised in lysis buffer
4. Sonication was then carried out on the sample for 3 minutes
5. Samples were placed on ice for 30 minutes
6. Samples were spun at 13,000rpm for 20 minutes at 4°C
7. Supernatant was then removed
8. Quantification of the samples was carried out using the nanodrop.
9. Samples were stored at -80°C until required

5.5.1.2 Sample preparation

1. Protein concentration was calculated and diluted to a total volume of 12.5µl in synapto buffer.
2. Loading dye was prepared by mixing 4X LDS sample buffer and 10X reducing agent with dH₂O, and kept at room temperature
3. Aliquots of the WT and MUT samples were recovered from the -80°C.
4. Samples were kept on ice during preparation. Samples were then spun down
5. Samples and loading dye were then placed on a heating plate at 100°C for 5 minutes.
6. The comb was then removed from the stacking gel, making sure not to damage any well.
7. The running apparatus was then mounted by positioning the gel glasses with the open side facing the middle chamber
8. The middle chamber was filled with 1X running buffer, and the external chamber was filled to the dashed line level.
9. Samples were spun and loaded with a molecular weight marker.

10. The power was set at 200V and 150mA and run for approx. 1.5hrs.

5.5.1.3 Running Gel

1. APS (ammonium persulfate) was made up fresh by adding 0.1g to 1ml of dH₂O and was mixed by vortexing.
2. 10ml of both the running and stacking gel were made up according to recipe below
3. 7mls of the gel was then poured in the gel apparatus avoiding bubbles and covered with water.
4. The gel was left to polymerize for at least 45 mins under the hood.
5. A 5% stacking gel was prepared in a 15ml tube (see recipe).
6. The water was removed from the top of the running gel
7. Residual water was removed with blotting filter paper
8. The stacking gel was poured avoiding bubbles
9. The comb was set and left to polymerize for 30 minutes

5.5.1.4 Blotting

1. TBE 1X and 15ml of methanol were prepared
2. Ten rectangles of filter paper and a piece of PVDF membrane were prepared for each gel.
3. The blotting sandwich was placed into the semi-dry blotting apparatus. Five pieces of filter paper soaked in TBE1X buffer, membrane (activated at least 20 seconds in methanol and rinsed in TBE1X buffer, the gel rinsed in TBE1X (after opening the 2 glasses and removing the stacking gel with a scalpel), the other five pieces of filter paper soaked in TBE1X.
4. Power supply was set at 150V and 150mA and run for approx. 1hr.
5. 5%Milk-TBS-Tween 0.1% for the blocking step and antibody dilutions was prepared

5.5.1.5 Primary antibody incubation

1. At the end of the blotting, the membrane was rinsed in TBS-T 0.5%, and incubated in red Ponceau solution for 10 minutes.
2. After 10 minutes the membrane was washed with deionised water.
3. The membrane was incubated in 5% milk-TBS-T 0.1% at RT
4. 5% milk-TBS-T 0.1% was prepared
5. Primary antibodies were prepared with 5% BSA 5% milk with the specific primary antibody (AbI) dilutions. The AbI dilution will depend on several factors: amount of total protein loaded for each well, amount of target antigen in the total protein extract, affinity and specificity of the AbI for the target antigen.
6. The membrane was cut at the proper molecular weight and incubated with 6ml of AbI solution overnight.

5.5.1.6 Secondary Antibody application

1. The membrane was washed 3 times/5mins each at RT with PBS
2. Secondary antibody (AbII) solution was prepared by diluting an HRP-conjugated or biotinylated AbII in 5% milk-TBS-T 0.1%. (The AbII) concentration will depend on the AbI concentration and on AbII properties)
3. The membrane was moved into a clean box containing the AbII solution.
4. The membrane was incubated for 60mins at RT

5.5.1.7 Amplification

1. The membrane was washed 3 times 5 mins each at RT
2. At the end of the third wash, the membrane was moved into a clean box containing the amplification mix and incubated for 60mins at RT.

5.5.1.8 Development

1. At the end of the incubation with AbII HRP-conjugated or Amplification Mix, the membrane was washed 3 times for 5 mins
2. During the last wash, a transparent plastic bag and an aliquot of ECL (same amount of both solutions for commercial ECL were prepared.
3. The membrane was drained and an aliquot of ECL was applied to the membrane for 1-2 mins,
4. The membrane is then placed between two pieces of transparent plastic.
5. The membrane is acquired quickly to avoid bleaching of the membrane.

5.5.1.9 Primary antibody staining

1. Prepare heat inactivated goat serum (Sigma G9023) in 1ml aliquots by heating to 56°C for 30 mins in a water bath, flicking every 5 mins. Store unused aliquots at -20°C.
2. Wash each well of the 24-well culture plate containing transfected cells with PBS (Phosphate-Buffered Saline) 3 times with a 5 min incubation at 250 rpm on a shaker in between each wash.
3. Block the non-specific binding by adding 400 µl 10% Goat serum in 0.1% PBS-T (Phosphate-Buffered Saline-Triton; 2.7.6.6) to each well and incubate for 1 hr at room temperature at 250 rpm.
4. Prepare dilutions of the primary antibodies in 1% Goat serum in 0.1% PBS-T, in the ratio as specified beside each one:
 - Rabbit anti-*FYN* antibody, Millipore, 06-133, 1:1000
 - Goat anti-*GFP* antibody, ABCAM, ab5450, 1:500
 - Mouse anti-MAP2 antibody, Millipore, MAB3418, 1:200
 - Rabbit anti-Phospho-AKT antibody, Cell Signalling, SER473, 1:2000

5.5.1.10 Secondary antibody staining

Wash each well of the 24-well culture plate containing transfected cells with PBS 3 times with a 5 minute incubation at 250 rpm on a shaker in between each wash.

1. Prepare dilutions of the secondary antibodies in 1% Goat serum in 0.1% PBS-T, according to the ratio specified beside each one:
 - Goat anti-rabbit Cy3, Invitrogen A10520, 1:500.
 - Donkey anti-goat Dylight 488, Jackson Immuno Research, 705-485-003, 1:500.
 - Goat anti-mouse Dylight 649, Jackson Immuno Research, 115-495-003, 1:500.
 - Goat anti-rabbit Dylight 488, Jackson Immuno Research, 115-485-003, 1:500.
 - Goat anti-guinea pig Cy3, Jackson Immuno Research, 106-165-003, 1:500
2. Add 350µl of the secondary antibody in 0.05% PBS-T to each well and incubate for 2 hrs, covered with aluminium foil.
3. Wash each well 3 times with PBS for 5 mins each with incubation at 250 rpm.
4. Mount the slides with Vectashield Hard set with DAPI (Vector H-1500). Place the slides in a slide tray and store at 4°C until the mounting media hardens. Once the mounting media is hard, seal the coverslips with nail polish.

5.5.1.11 Western Blot Analysis to test the specificity of the FYN antibody

Western blot analysis was carried out as per section 5.5. I was able to visualise that the antibody I was using for further immunohistochemistry analysis was specific to the *FYN* protein with a thick band in the WT at 59kDa (Figure 5.4). A faint band can be seen in the KO spleen sample, but this is likely to be due to non-specific binding. There are numerous well documented reasons for non-specific binding in Western blot analysis, but the presence of a large band in the WT at 59kDa supports the specificity of the antibody for my studies.

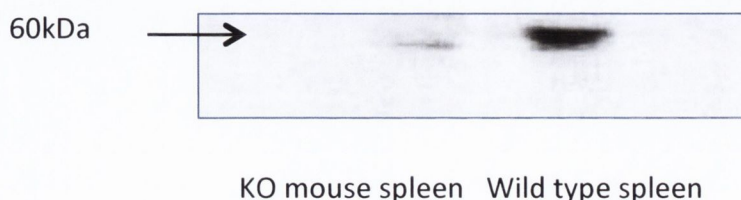


Figure 5.4: Western blot analysis to test the specificity of the *FYN* antibody.

5.5.2 Confocal microscopy

A Zeiss LSM 510 laser module confocal microscope is used for analysis with the LSM 5 software (Zeiss). Image analysis is performed using Image J (rsb.info.nih.gov/ij/). I first carried out transfection and immunochemistry on neuroblastoma cell lines. The purpose of this primary analysis was the following: 1) To see if the method was viable to bring forward to primary neuronal mouse cultures, 2) To determine if there were differences in expression of *GFP* to differentiate transfected from non-transfected cells, 3) To see if transfection of the *FYN* Mut plasmid results in a different level of *FYN* expression to *FYN* WT using Image J software. All experiments were carried out in parallel. All three plasmids were transfected using the Amaxa system on the same flask of harvested cells, at passage number 5. All images were captured at the same time using confocal microscopy, with all settings identical for each set of images. I then utilized image J software to measure the expressions in 50 cells chosen at random for each experiment.

5.5.2.1 Image Analysis

Measurements were calculated by obtaining the area and intensity for each image. The intensity signal was then divided by the area, and the background signal was then subtracted to give a final measurement of expression (Corvin et al., 2012; Ripke et al., 2013). As a further method of quality control, I deemed a cell to be transfected only if the expression levels were more than 4 times the strength of the background signal. Table 5.1 gives an example of the measurements of *GFP* expression generated for each cell. I carried the same method of analysis through to all future experiments.

Table 5.3: Example measurements obtained for transfected and non-transfected cells

	Area	Intensity	Intensity/Area	Expression	Transfected/ Non-Transfected
Cell 1	2158	113424	52.56	49.01	Transfected
Background	2158	7668	3.55		
Cell 2	1310	8267	6.31	2.66	Non-Transfected
Background	1310	4788	3.65		

5.6 Site Directed Mutagenesis

Site directed mutagenesis is a method used to create specific changes in the DNA sequence of a gene. It requires the synthesis of a short DNA primer. This synthetic primer contains the desired mutation and is complementary to the template DNA around the mutation site so it can hybridize with the DNA in the gene of interest.

Table 5.4 *FYN* LoF specific site directed mutagenesis primers

Forward:	5' GTACTTTGGAAAACCTTGGCTGAAAAGATGCTGAGCGAC 3'
Reverse:	5' GTCGCTCAGCATCTTTTCAGCCAAGTTTCCAAAGTAC 3'

Table 5.5 Site Directed Mutagenesis PCR specifications	
Phusion DNA pol	25 µl
Template (Original wild type plasmid)	25ng/ µl
FWD 10uM	1 µl
REV 10uM	1 µl
ddH2O	up to 50 µl

5.6.1 Site Directed PCR protocol

The PCR cycle used was 98°C for 30 secs, followed by 12 cycles of 98°C for 30 secs , 55°C for 1 min and 68°C for one min.

5.6.2 DPN1 digest

1. 0.5 µl of dpn1 (20U/ µl) was added to one half of each PCR product (or 0.25 µl as split in two)
2. The other half of PCR product was not treated with DPN1
3. Samples were mixed thoroughly by pipetting and spun for 1 min in a microcentrifuge
4. The samples were Incubated at 37°C for 1hr in a water bath

5.7 Results

5.7.1 Transfection of plasmids and subsequent Immunohistochemistry analysis

3 plasmids (*GFP*, *FYN* WT & *FYN* Mut) were transfected into neuroblastoma cell lines and immunohistochemistry was performed. Three markers were used in parallel in the first set of analyses: *GFP*, *FYN* and MAP2. The *GFP* signal, which can be seen in Figure 5.6 acts as an indicator of transfected neurons as each plasmid contains a *GFP* gene. MAP2 (Microtubule Associated Protein) is primarily expressed in neurons and so it can be used to distinguish neurons from glia.

As a control for all experiments, cells were also processed for transfection, but in the absence of any plasmid. Subsequent immunohistochemistry was carried out in an identical fashion to the plasmid transfections with antibodies specific to *GFP*, *FYN* and MAP2. I found that no signal appears for *GFP* or *FYN* but with a strong signal for MAP2, which is a stain for neuronal cells (Figure 5.5.) The absence of a strong signal of native *FYN* in both the negative control and *GFP* images is due to the microscope settings. In order to optimise the detection of the transfected plasmid, the imaging conditions failed to show the endogenous *FYN*.

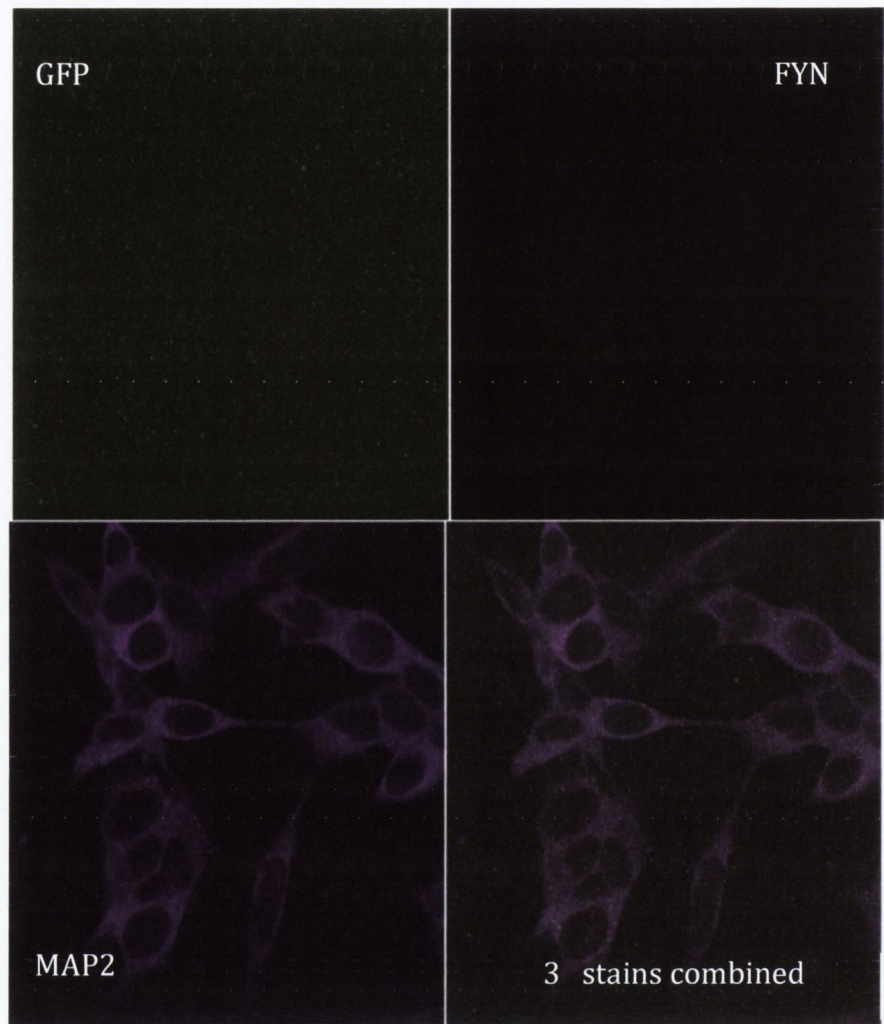


Figure 5.5 24hr Amaxa Transfection Control . Transfection was carried out with no plasmid. Subsequent Immunohistochemistry showed no signals for GFP, and FYN. Staining for MAP2 of neuronal cells was present.

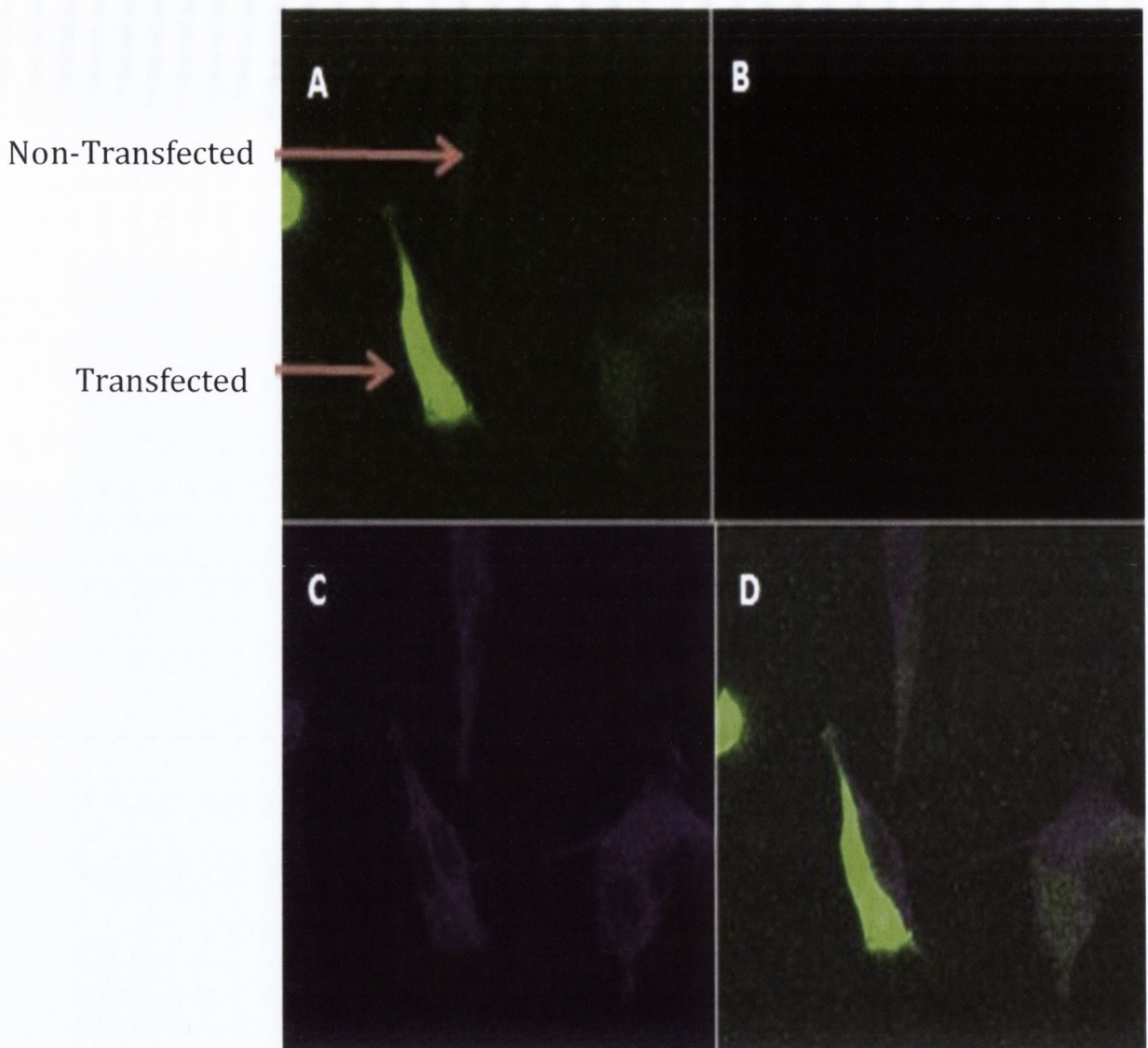


Figure 5.6 24hr Amaxa Transfection with *GFP* plasmid. **A.** Neuroblastoma transfected cell lines can be distinguished from non-transfected neurons upon observing staining with *GFP*. **B.** *FYN* antibody staining **C.**MAP2 staining indicates the presence of synapses. **D.** All stains can be observed together.

5.7.2 Transfection of the *FYN* WT plasmid into Neuroblastoma cell lines

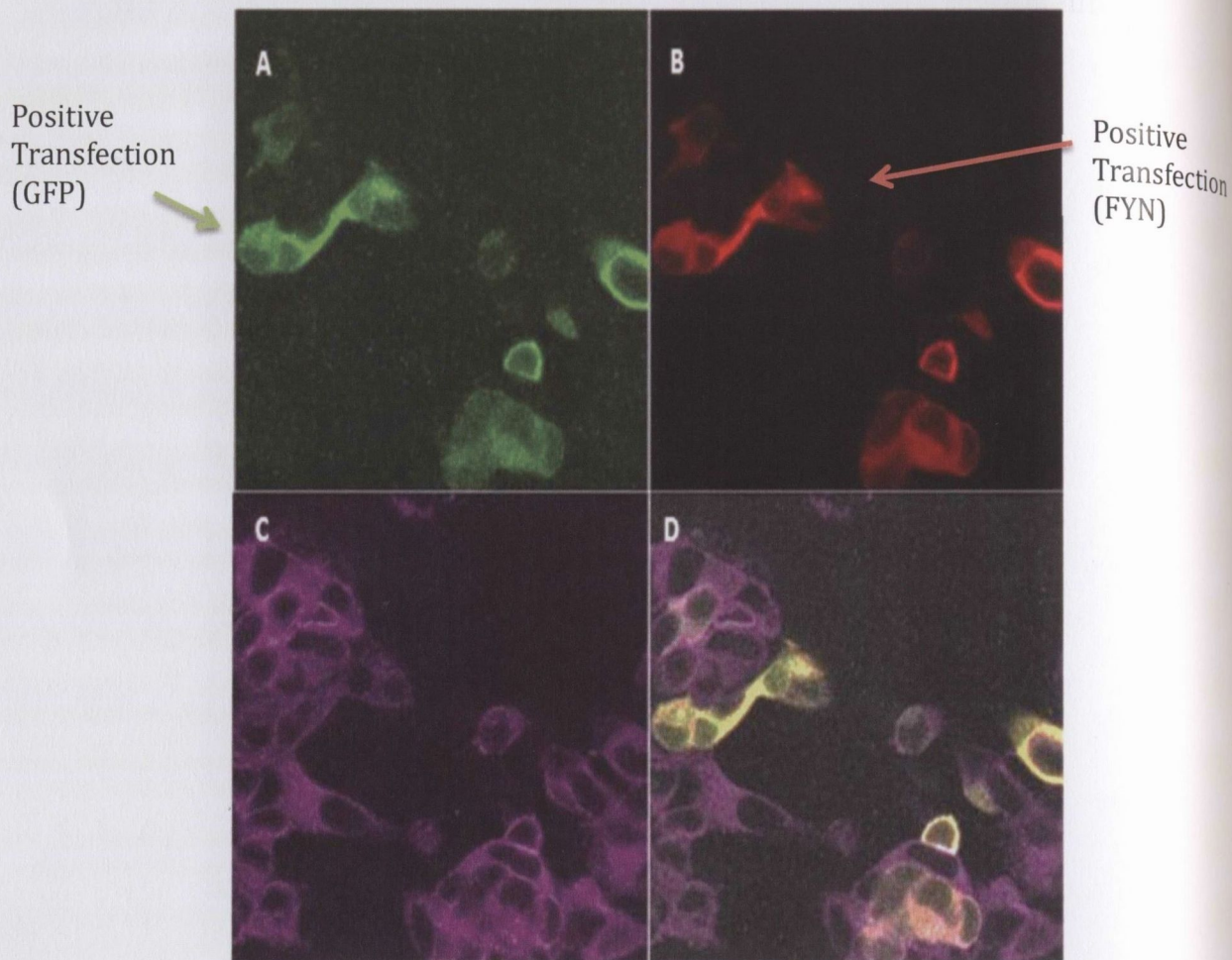


Figure 5.7 24hr Amaxa Transfection with *FYN* WT plasmid. A. Neuroblastoma transfected cell lines can be distinguished from non-transfected neurons upon observing stain with *GFP*. B. *FYN* antibody staining C. MAP2 staining indicates the presence of synapses. D. All three stains can be observed together.

Transfection with the *FYN* WT plasmid was successful (Figure 5.7). A strong *FYN* signal can be seen, which overlaps specifically with the signal for *GFP*, allowing me to confirm that both *GFP* and *FYN* signals can be used as indicators of transfection. Since these were also carried out in the exact same way as the negative control, and *GFP* transfection, and neither of these had a strong *FYN* signal, I know that the signal is true *FYN* signal. The scatter plot containing the individual results for each of the 50 cells counted in this transfection experiment Figure 5.8.

Measurements were calculated by obtaining the area and intensity for each image. The intensity signal was then divided by the area, and the background signal was then subtracted to give a final measurement of expression (Corvin et al., 2012; Ripke et al., 2013). As a further method of quality control, I deemed a cell to be transfected only if the expression levels were more than 4 times the strength of the background signal. Table 5.1 gives an example of the measurements of *GFP* expression generated for each cell. I carried the same method of analysis through to all future experiments.

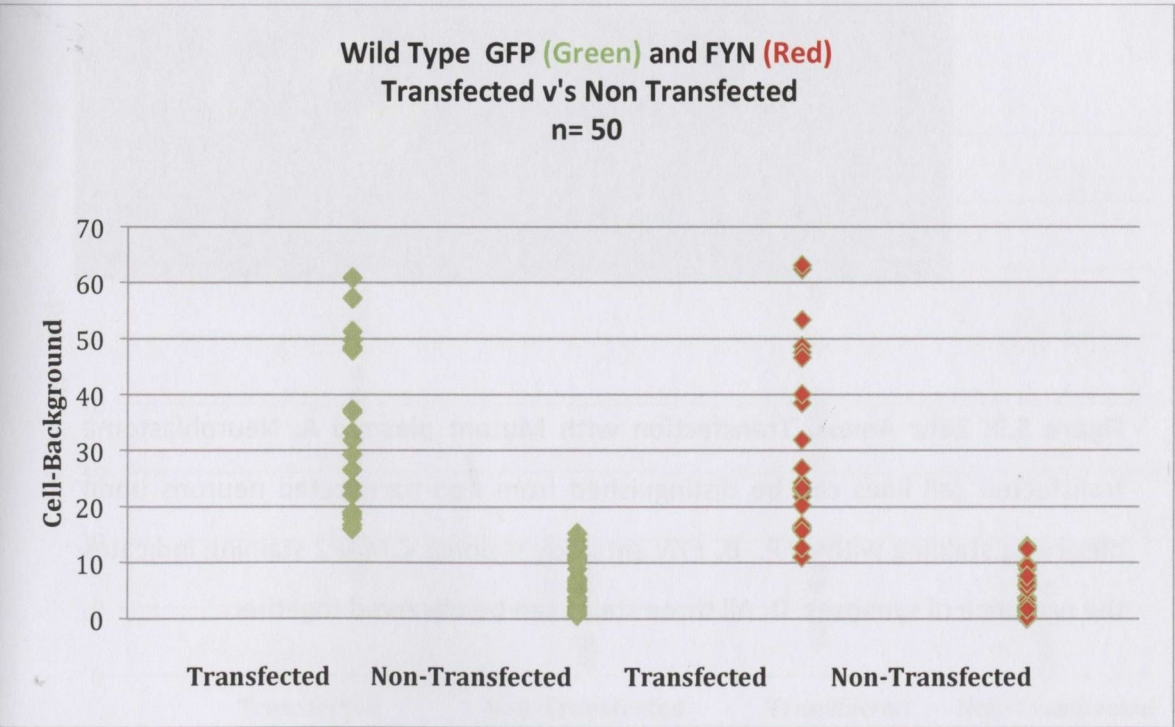


Figure 5.8: Scatter plot showing the expression for both *GFP* and *FYN* WT plasmids, allowing me to differentiate transfected from non-transfected.

5.7.3 Transfection of the *FYN* Mut plasmid into Neuroblastoma cell lines

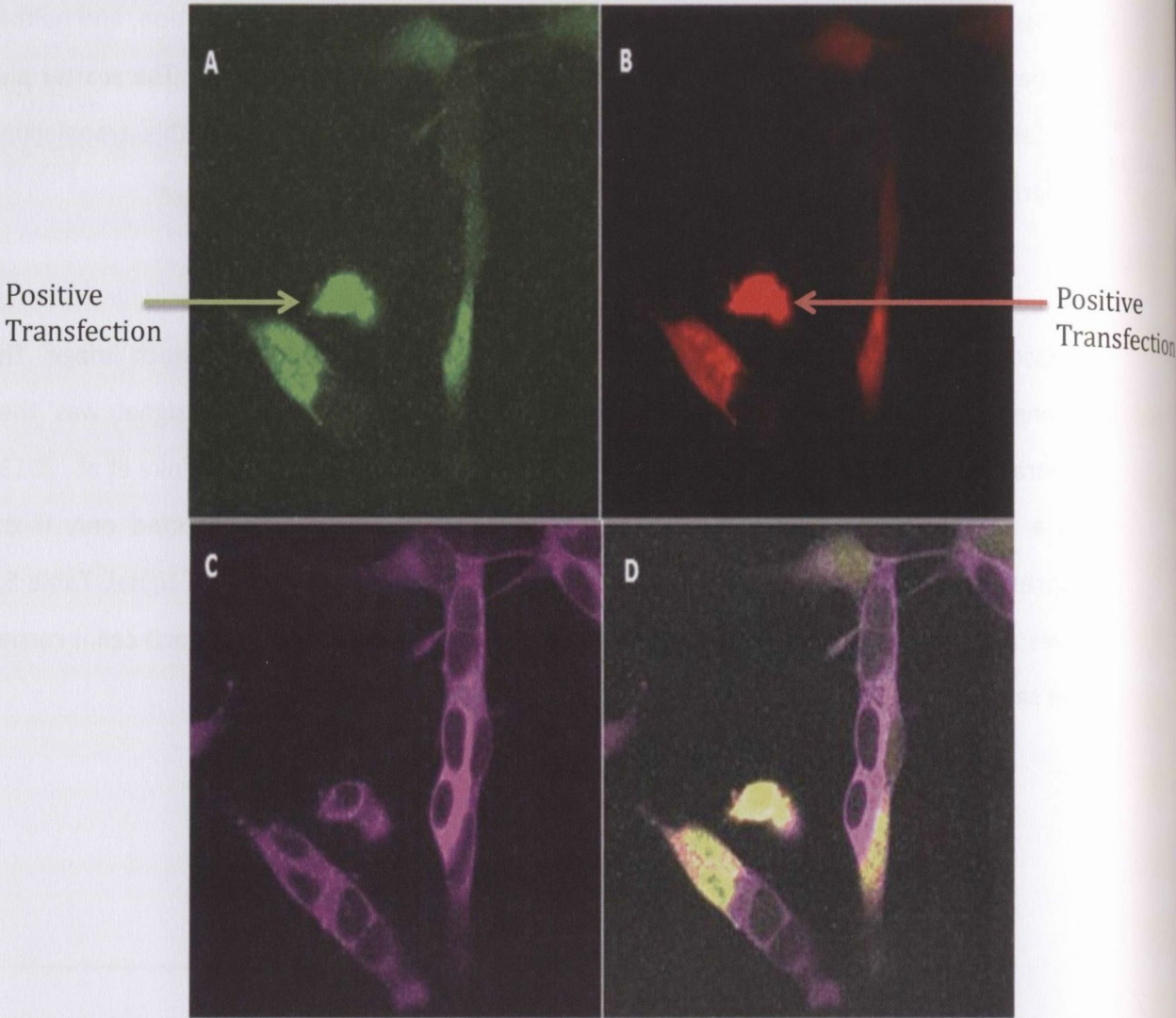


Figure 5.9: 24hr Amaxa Transfection with Mutant plasmid **A.** Neuroblastoma transfected cell lines can be distinguished from non-transfected neurons upon observing staining with *GFP*. **B.** *FYN* antibody staining **C.**MAP2 staining indicates the presence of synapses. **D.** All three stains can be observed together.

Transfection with the Mut plasmid was successful. I was able to clearly differentiate transfected from non-transfected cells. A clear *FYN* signal is present, which overlaps specifically with the signal for *GFP*. Fifty cells were chosen at random for each slide, and measurements were carried out using Image J software.

Measurements once again were calculated by obtaining the area and intensity for each image. The intensity signal was then divided by the area, and the background signal was then subtracted to give a final measurement of expression (Corvin et al., 2012; Ripke et al., 2013). As a further method of quality control, I deemed a cell to be transfected only if the expression levels were more than 4 times the strength of the background signal. Table 5.1 gives an example of the measurements of *GFP* expression generated for each cell. I carried the same method of analysis through to all future experiments.

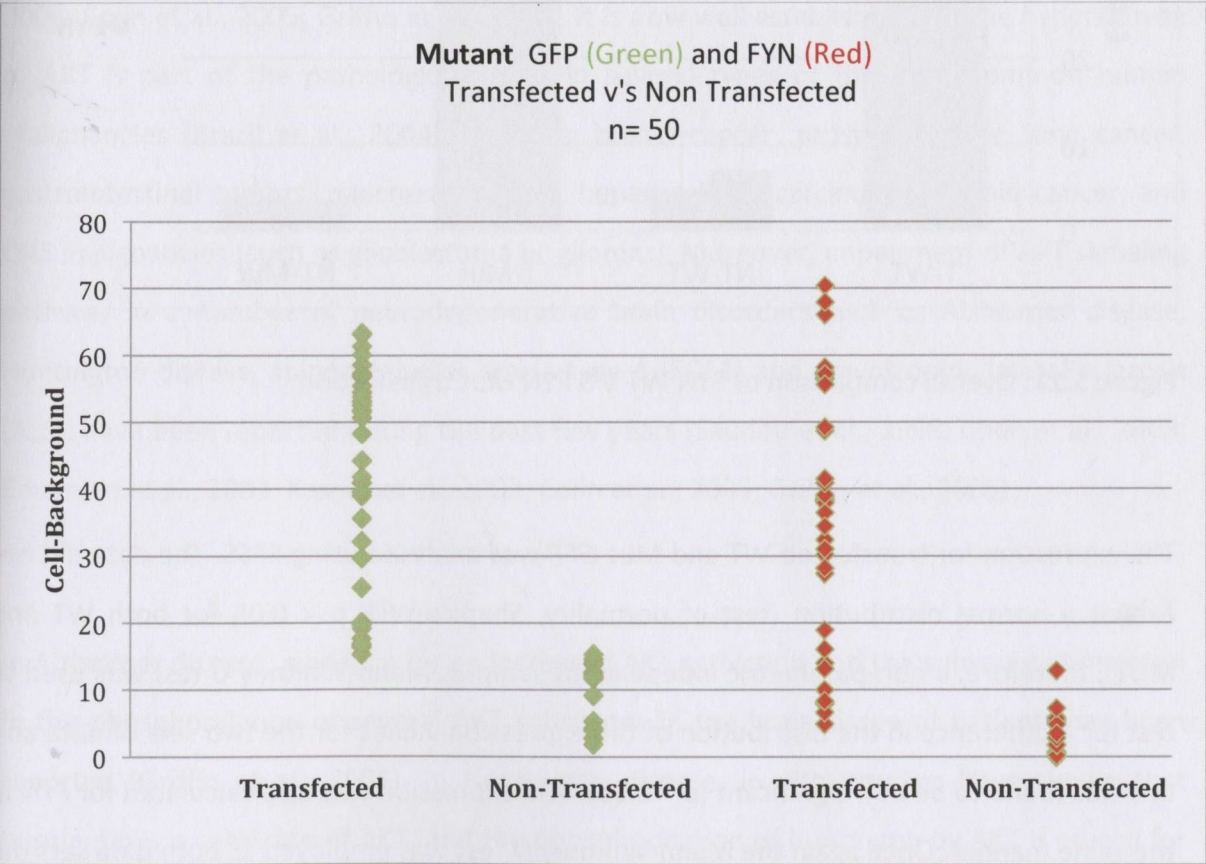


Figure 5.10: Scatter plot showing the expression for both *GFP* and *FYN* Mutant plasmids, allowing me to differentiate transfected from non-transfected.

5.7.4 Comparison of FYN WT V'S FYN Mut transfection

In summary, I did not see a significant difference in the expression of *FYN* in the WT compared to the Mut cells. Although the rate of transfection was higher for the Mut, the expression levels only varied slightly between the WT and Mut transfected cells. From this set of experiments, it would appear that the truncated protein does not lead to a decrease in *FYN* expression. I next wanted to investigate if there was a difference in function between WT and Mut transfected cells.

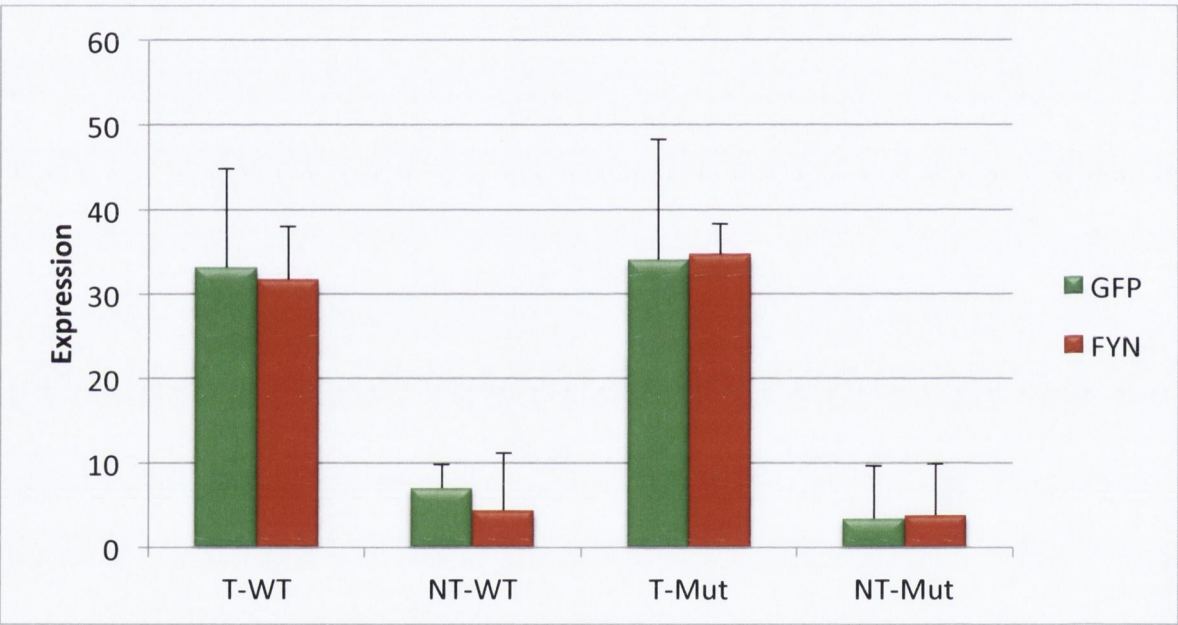


Figure 5.11: Overall comparison of FYN WT V'S FYN Mut transfection

The expression for transfected WT and Mut *GFP* was analysed using SPSS. The data did not exhibit a normal distribution (test of normality, Shapiro-Wilk $p < 0.05$ for both WT and MUT), therefore, a non-parametric independent samples Mann-Whitney U test was used to test for a difference in the distribution of the expression values for the two sets of data and this was found to be non-significant ($p = 0.85$). The expression was also calculated for *FYN* in the same manner. Once again the Mann-Whitney U test was employed as both data sets did not show a normal distribution (Shapiro-Wilk $p < 0.05$), and the expression values for the two sets of data were found to be non-significant ($p = 0.63$).

5.8 Transfection and immunohistochemistry of Phospho-AKT in neuroblastoma cell lines

AKT is a member of the kinase family, and plays a major regulatory role in cellular function. Levels have also been found to increase during brain development. Impairment of the AKT signalling pathway has been found in a number of neurodegenerative brain disorders such as Alzheimer's disease (Saudou, Finkbeiner, Devys, & Greenberg, 1998), Huntington's disease (Chen et al., 2003), and ALS (Kaspar, Llado, Sherkat, Rothstein, & Gage, 2003). AKT also plays a basic role in neuronal cell size and survival and can be used as a marker of neuronal cell function (Brazil, Yang, & Hemmings, 2004).

Deregulation of AKT signaling pathway is directly associated with some of the most prevalent and incurable human diseases (Blain and Massague, 2002; Liang et al., 2002; Shin et al., 2002; Viglietto et al., 2002; Chen et al., 2003; Emamian et al., 2003, 2004; Brazil et al., 2004; Colin et al., 2005; Griffin et al., 2005). It is now well established that the hyperactivity of AKT is part of the pathologic process in several types of the most common human malignancies (Brazil et al., 2004), including breast cancer, prostate cancer, lung cancer, gastrointestinal tumors, pancreatic cancer, hepatocellular carcinoma, thyroid cancer, and CNS malignancies (such as glioblastoma or gliomas). Moreover, impairment of AKT signaling pathway in a number of neurodegenerative brain disorders such as Alzheimer disease, Huntington disease, spinocerebellar ataxia type 1 (SCA1) and amyotrophic lateral sclerosis (ALS), have been reported during the past few years (Saudou et al., 1998; Chen et al., 2003; Emamian et al., 2003; Kaspar et al., 2003; Colin et al., 2005; Griffin et al., 2005).

In Alzheimer disease, evidence for an increased AKT activation and the subsequent increase in the phosphorylation of several AKT substrates in the brain tissue of patients has been reported (Griffin et al., 2005). In Huntington disease, in vitro studies have shown that huntingtin is a substrate of AKT, and the phosphorylation of huntingtin by AKT is crucial for mediating the neuroprotective effects of IGF-1. Furthermore, changes in AKT signaling in an animal model of Huntington disease, as well as in patients with Huntington disease, are also

reported (Saudou et al., 1998; Colin et al., 2005). In SCA1, Chen et al. (2003) reports a regulatory role of AKT in the association of ataxin-1 with 14-3-3 through the phosphorylation of Ser-776 of ataxin-1, which modulates the neurotoxicity of ataxin-1 (Emamian et al., 2003). In ALS, IGF-1 stimulates the activity of AKT in the spinal cord and prolongs the lifespan of the mouse model of ALS, by increasing the survival of motor neurons (Kaspar et al., 2003).

Since the initial report in 2004 (Emamian et al., 2004), many other studies have been published supporting that impaired AKT signaling plays an important role in the pathogenesis of SCZ. This study and several subsequent ones uncovered association between AKT1 genetic variants and SCZ in case/control samples and family cohorts (Emamian et al., 2004; Ikeda et al., 2004; Schwab et al., 2005; Bajestan et al., 2006). The original study reported a significant association between SCZ and an AKT1 haplotype in U.S. families (Caucasian of Northern European origin). The genetic association of AKT1 with schizophrenia was later confirmed in European (Schwab et al., 2005), Japanese (Ikeda et al., 2004), and Irish (Thiselton et al., 2008) populations. Taken together, despite the presence a negative report in a case-control Japanese population (Ohtsuki et al., 2004), it seems that the genetic association of AKT with SCZ is not limited to a specific genetic background and is reproducible in populations with different backgrounds.

The main advantage of studies on signaling molecules in complex human disorders with unknown etiology, such as SCZ and ASD, is their potential for discovering the abnormal function of proteins involved in the pathogenesis regardless the actual causative factor, whether it is genetic or environment or the interplay between them. Despite the fact that SCZ is complex and extremely difficult to understand at the molecular level, dissecting the role of signaling molecules such as AKT or GSK3 in such a complex disorder has obvious therapeutic implications.

AKT1 protein level is significantly reduced in different tissues from patients with SCZ. Lymphocytes from individuals with SCZ expressed 68% less AKT1 relative to control subjects (Emamian et al., 2004). More importantly, AKT1 was reduced in the hippocampus and frontal cortex in postmortem brain samples from those with SCZ relative to comparison subjects. This decrease in protein level in the brain was specific to AKT1 isoform and AKT2 and AKT3 levels were unaffected (Emamian et al., 2004). The activity of major AKT1 targets such as GSK3 was also found to be altered in individuals with SCZ (Emamian et al., 2004). In my study Phospho-Akt (Ser473) antibody was used to detect levels of phosphorylated AKT, a decrease in the phosphorylated version of AKT has been reported in first episode SCZ patients (Szamosi, Kelemen, & Keri, 2012).

5.8.1 Immunohistochemistry analysis of Phospho akt on transfected FYN WT plasmid in neuroblastoma cell lines.

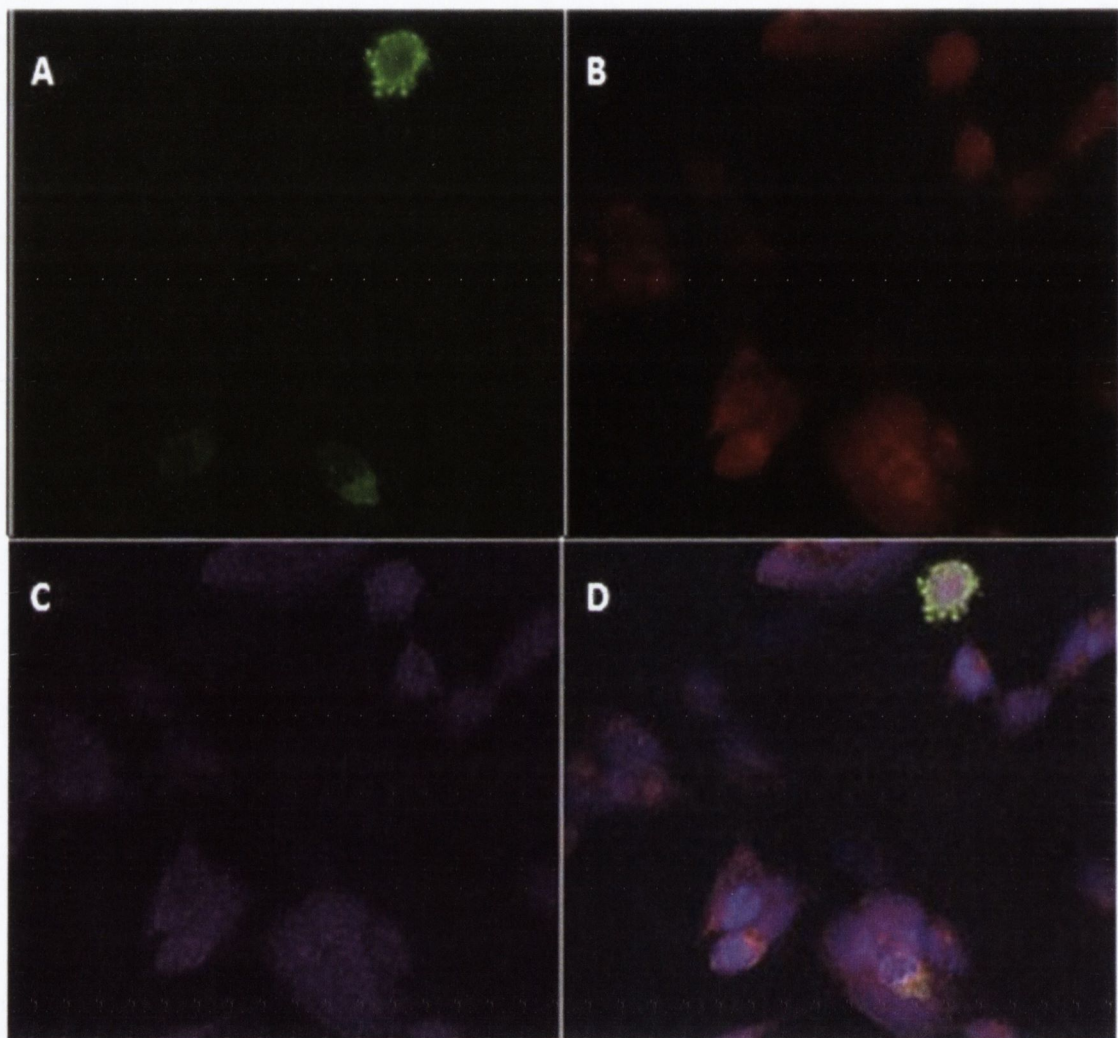


Figure 5.12: 24hr Amaxa Transfection with WT plasmid and subsequent Phospho AKT Immunohistochemistry analysis. A. Neuroblastoma transfected cell lines can be distinguished from non-transfected neurons upon observing native *GFP*. B. Phospho-akt antibody staining C. MAP2 staining indicates the presence of synapses. D. All three stains can be observed together with DAPI.

FYN WT and Mut plasmids were transfected into neuroblastoma cell lines as per section 5.4.4 and immunohistochemistry was performed as per section 5.5. As a control, cells were also processed for transfection, but in the absence of any plasmid. Three markers were used in parallel in the first set of analyses: *GFP*, *FYN* and Phospho-AKT. I did not transfect the *GFP* plasmid on this occasion, as I only wanted to compare the WT and Mut cells.

Transfection of the *FYN* WT plasmid was successful. I was able to clearly differentiate transfected from non- transfected cells using *GFP* as an indicator. Once again 50 cells were chosen at random for each slide, and measurements were carried out using Image J software.

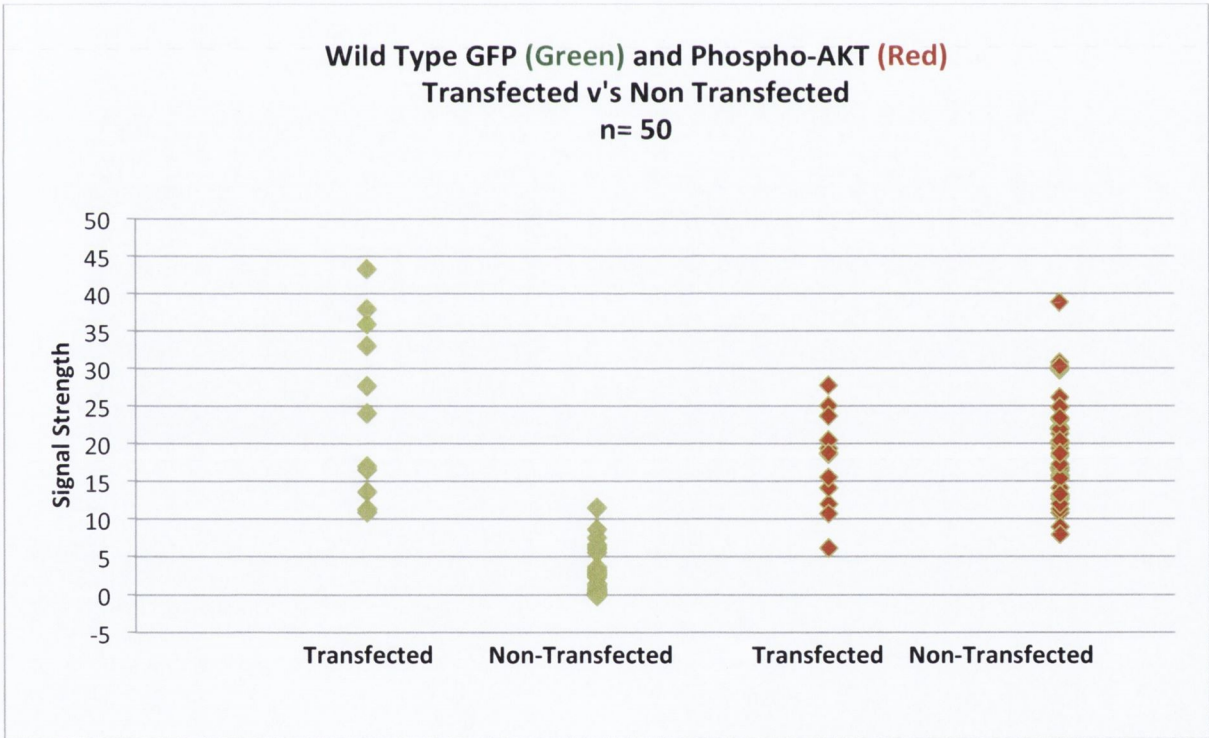


Figure 5.13: Scatter plot showing the expression for both *GFP* and Phospho-AKT from WT transfected plasmids.

5.8.2 Immunohistochemistry analysis of Phospho akt on transfected FYN Mut plasmid in neuroblastoma cell lines.

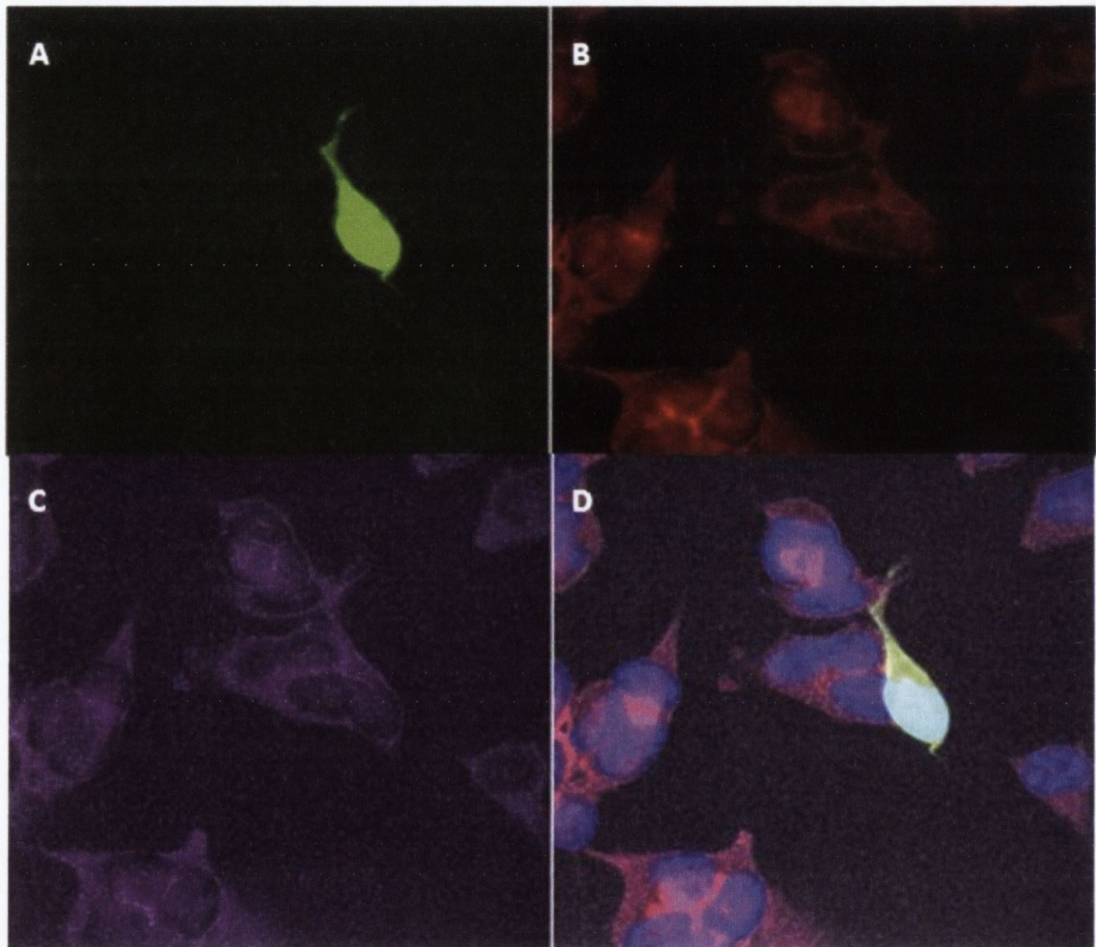


Figure 5.14 24hr Amaxa Transfection with Mutant plasmid **A.** Neuroblastoma transfected cell lines can be distinguished from non-transfected neurons upon observing native *GFP*. **B.** Phospho-akt antibody staining **C.**MAP2 staining indicates the presence of synapses. **D.** All three stains can be observed together with DAPI.

Transfection with the Mut plasmid was successful. I was able to clearly differentiate transfected from non- transfected cells using *GFP* as an indicator. Clear *GFP* and Phospo-AKT signals were present. All specific values for the analysis of *FYN* Mut can be found in Appendix 1.6 and 1.7. Fifty cells were chosen at random for each slide, and measurements were carried out using Image J software.

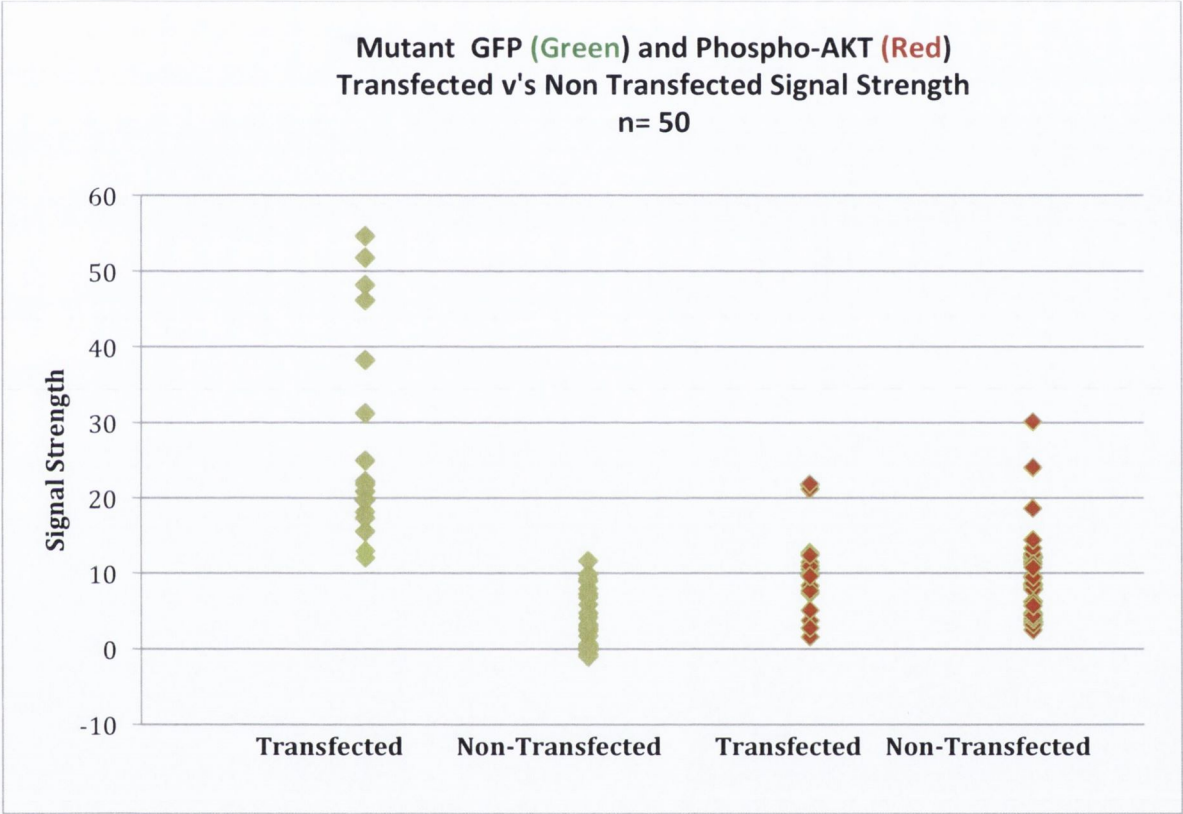


Figure 5.15: Scatter plot showing the expression for both *GFP* and Phosho-AKT with the Mutant plasmids.

5.8.3 Overall results of Phospho-AKT transfection

I carried out a transfection of *GFP*, *FYN* WT and *FYN* Mut plasmids on neuroblastoma cell lines. All plasmids were transfected in an identical fashion, and subsequent immunohistochemistry was carried out with antibodies specific to Phospho-AKT and MAP2. Once again I found a higher rate of transfection for the *FYN* Mut plasmid (Figure 5.11).

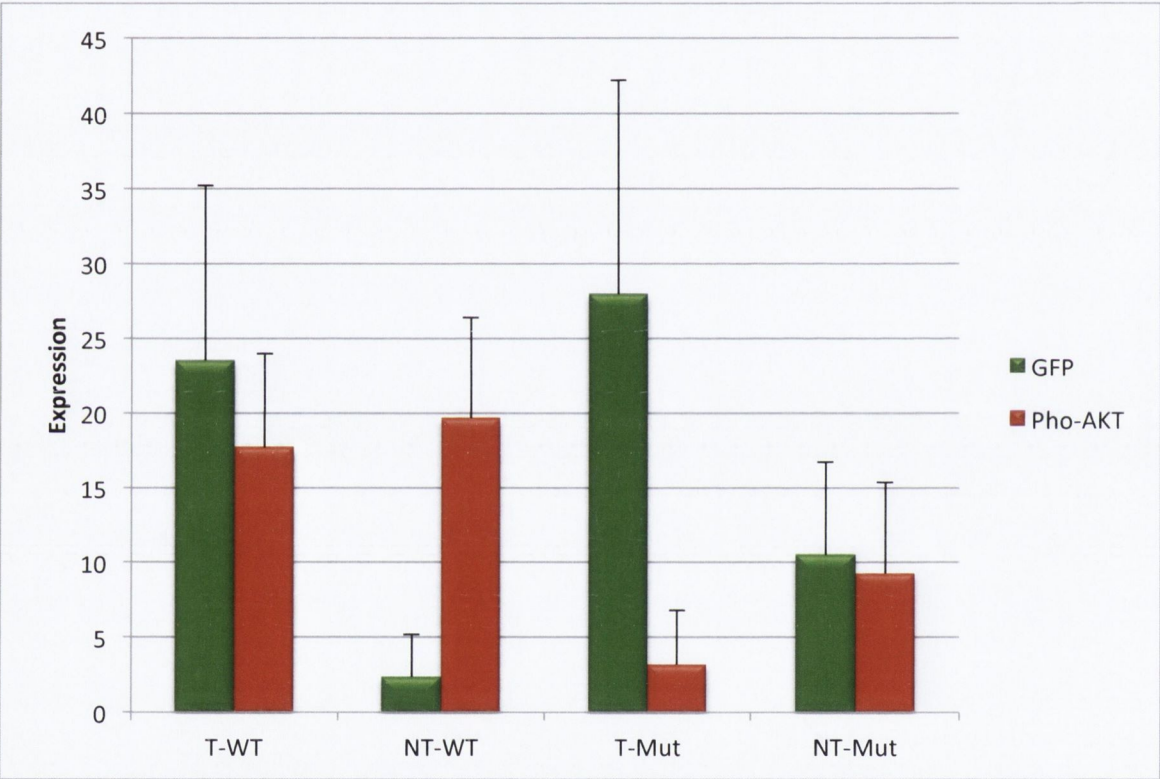


Figure 5.16: Overall Results of Phospho-AKT transfection showing the difference in expression of Transfected v's Non-Transfected in both WT and Mut plasmids

The data for transfected WT and Mut Phospho-AKT expression was analysed in a similar fashion. The data for the WT did exhibit a normal distribution (test of normality, Shapiro-Wilk $p = 0.996$) but the data for the Mut did not exhibit a normal distribution (test of normality, Shapiro-Wilk $p = 0.002$). Therefore, non-parametric independent samples Mann-Whitney U test was used once again to test for a difference in the distribution of the

expression values for the two sets of data. This analysis did identify a significant difference between WT and Mut with less Phospho-AKT expression in the *FYN* Mut transfected cells ($p < 0.001$). These data suggest that the presence of the mutant *FYN* protein impacts on neuronal cell function.

Similar analysis was carried out for the non-transfected cells, and I detected a significant difference between WT and Mut for both *GFP* and Phospho-AKT. For *GFP*, expression was higher in Mut cells compared to WT ($p=0.082$). For Phospho-AKT, expression was higher in WT cells compared to Mut ($p=0.429$).

5.9 Transfection and Immunohistochemistry analysis in primary mouse neuronal cultures

After successful transfections on neuroblastoma cell lines, I wanted to see if I could replicate the experiments in mouse primary neuronal cultures. Although cell line use is standard in most research labs, primary cultures are considered a superior model of *in vivo* behaviour, and are more representative of functioning brain tissue. Primary cultures are derived directly from excised tissue and are extremely labour intensive. Primary cultures also involve sacrificing animals, and so all of the tissue culture for this set of experiments was carried out by Dr. Ines Molinos.

Once the primary cultures were harvested, I was able to carry out an Amaxa transfection in a similar fashion to the neuroblastoma cell lines. Unlike cell lines however, the expression of primary cultures is best analysed after 7 days (Corvin et al., 2012) the benefit of this was the potential to allow me to see any changes in the morphology of the cells. Subsequent immunohistochemistry was carried out in an identical manner to the neuroblastoma cell lines.

5.9.1 Analysis of both FYN WT and FYN Mut plasmids in primary neuronal cultures

FYN WT and *FYN* Mut plasmids were transfected into cortical neuronal cells from newborn mice as per section 2.3.11 and immunohistochemistry was performed as per section 2.3.12.

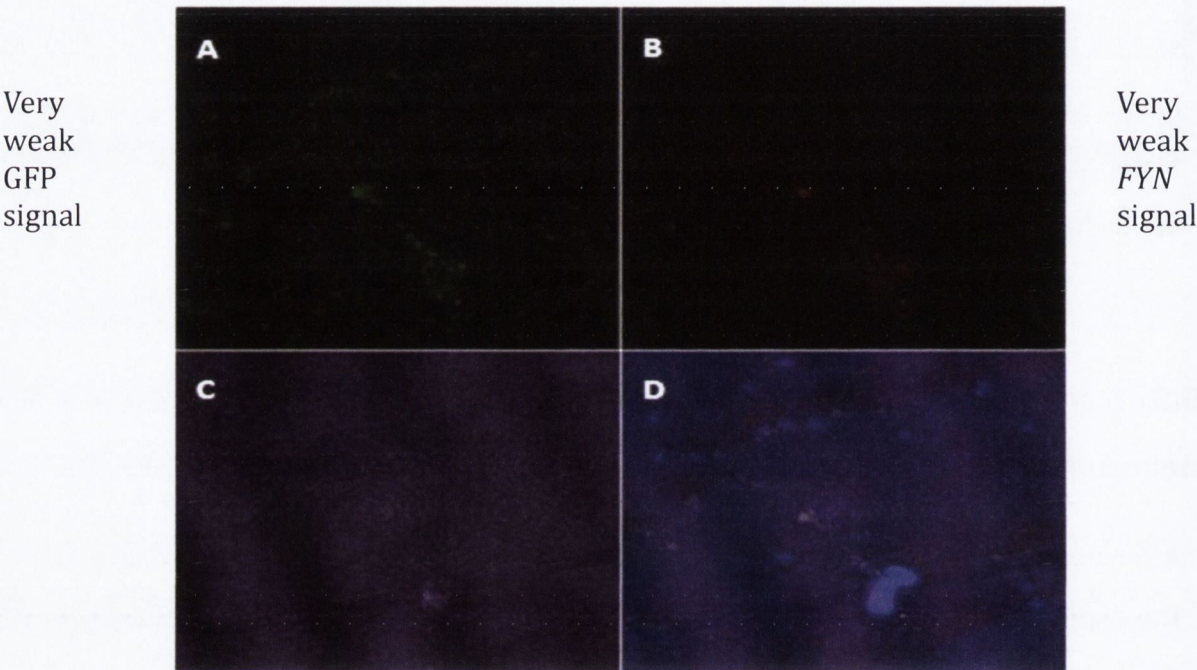


Figure 5.17 24hr Amaxa Transfection on primary cultures with WT plasmid. **A** *GFP* antibody stain. **B.** *FYN* antibody staining **C.**MAP2 staining indicates the presence of synapses. **D.** All three stains can be observed together with DAPI.

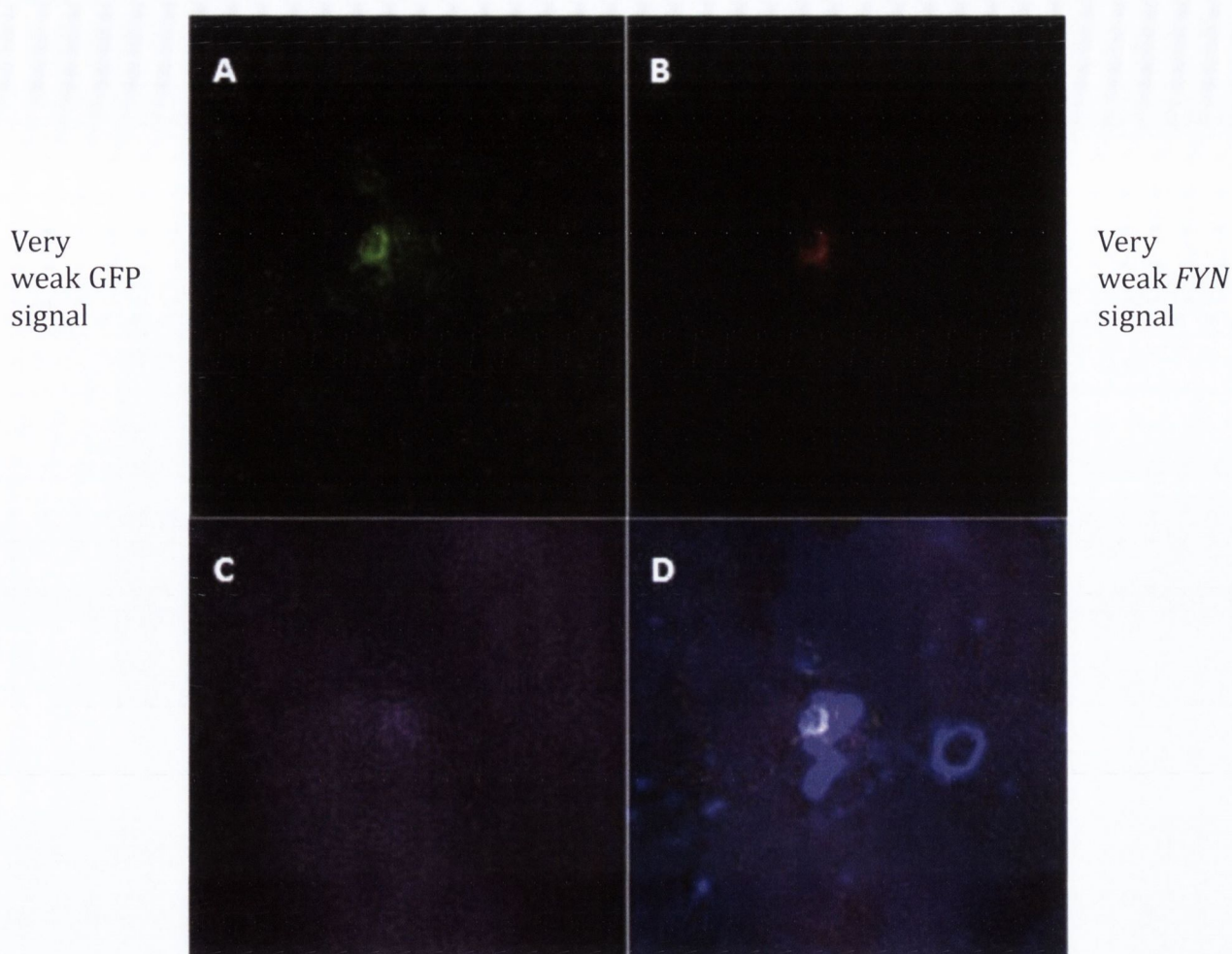


Figure 5.18 24hr Amaxa Transfection on primary cultures with Mutant plasmid. A *GFP* antibody stain. B. *FYN* antibody staining C. MAP2 staining indicates the presence of synapses. D. All three stains can be observed together with DAPI.

5.9.2 Results of *FYN* analysis on primary neuronal cultures

After successful transfection of neuroblastoma cell lines, I sought to use this same method in primary cortical mouse neurons, to see if it had a similar effect. Unfortunately the transfection efficiency was very poor, which can be seen in Figures 5.12 and 5.13. For example, only 2 transfected cells were identified for the *FYN* WT (Figure 5.14), and thus the data was too noisy to attempt analysis and draw conclusions. Obtaining a substantial number of transfected cells from the primary cultures proved very difficult, in turn hampering my ability to carry out extensive studies as had been possible for the cell lines. This experiment was repeated using the Amaxa system, but did not prove fruitful.

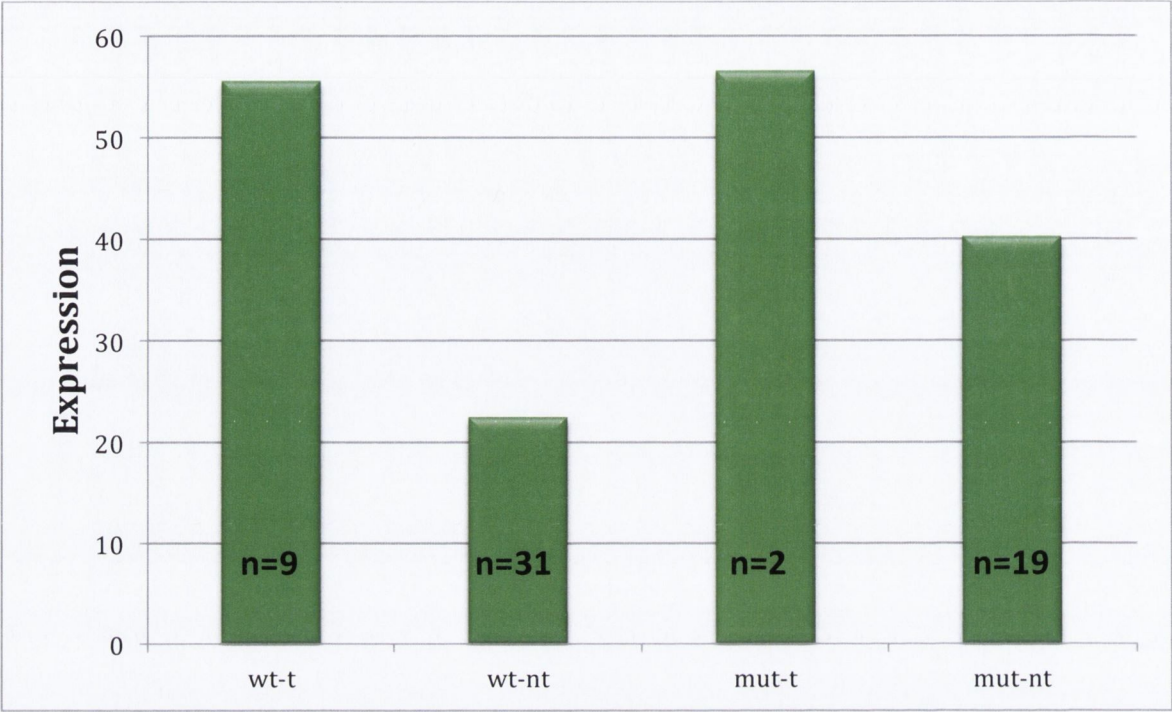


Figure 5.19: Chart showing the mean expression values for *FYN* WT and *FYN* Mut in primary cultures. The transfection rate was significantly lower for primary cultures compared to cell lines. The number of cells available for analysis is indicated at the bottom of each column. It was also difficult to differentiate between transfected and non-transfected cells.

5.10 Site directed Mutagenesis

Site directed mutagenesis is a method used to create specific changes in the DNA sequence of a gene. It requires the synthesis of a short DNA primer. This synthetic primer contains the desired mutation and is complementary to the template DNA around the mutation site so it can hybridize with the DNA in the gene of interest. The mutation in my experiments is a single base change (a point mutation) in the *FYN* gene. The single-stranded primer is extended using a DNA polymerase, which copies the rest of the gene.

A “hotstart” enzyme is desirable, as the proof-reading capability of most of these enzymes may otherwise degrade the primers during setup. Following the reaction, the product is digested with DpnI. This digest is crucial. DpnI only cleaves at methylated sites, so it chews up the template plasmid but not the PCR product. DpnI will cleave the sequence G^{Me}ATC where ^{Me}A means that the adenylate nucleotide is methylated. DpnI will not cleave the unmethylated sequence GATC. When we apply an oligonucleotide primer to this template and extend it, the new DNA that is synthesized in vitro will be unmethylated, therefore a marked difference is created between the parental strand (methylated) and the mutagenized strand (unmethylated). The copied gene contains the mutated site (unmethylated) and the parental strand (methylated), and they are then introduced into a host cell as a vector and cloned. Primers were designed using primerX (<http://bioinformatics.org/primerx/>).

I tested the site directed mutagenesis technique because it could significantly reduce the cost of investigation of variants from sequencing studies, particularly where multiple variants within the same gene may be available for study. By using this approach of transfecting plasmids containing the gene of interest into neuronal cells, it would mean only having to purchase a wildtype version of the plasmid, and the mutant version could be developed using site directed mutagenesis. Furthermore, it would be possible to introduce other interesting mutations in the gene of interest.

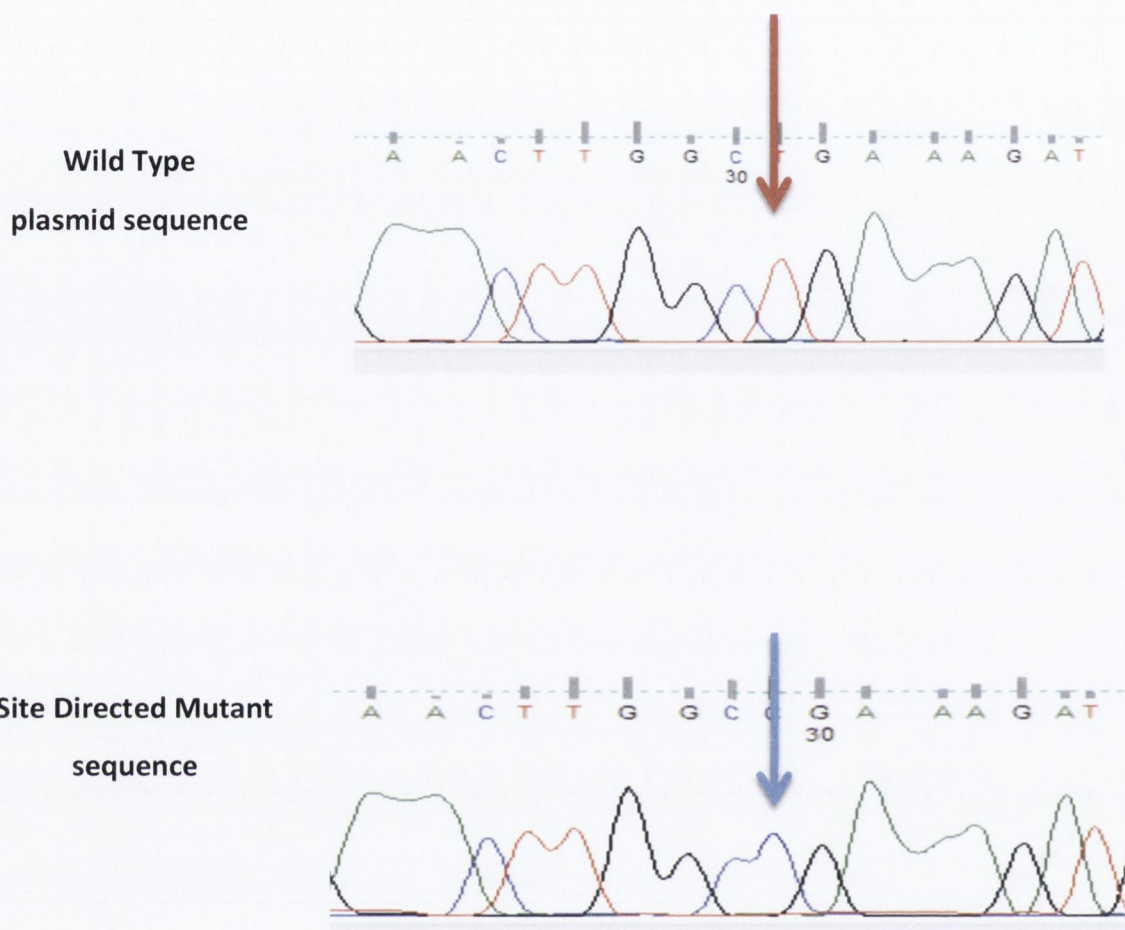


Figure 5.20 Capillary Sequence output for both the wild type and site directed mutant plasmids, showing the successful introduction of the mutant allele.

5.10.1 Transfection of a plasmid created via site directed mutagenesis into neuroblastoma cell lines

After creating a site directed version of the mutant plasmid, I wanted to see if it would transfect efficiently. For this experiment I choose to transfect the plasmid into neuroblastoma cell lines, as they turned out to be easier to work with than primary cell cultures.

Transfected
SDM plasmid
with a very
strong GFP
signal

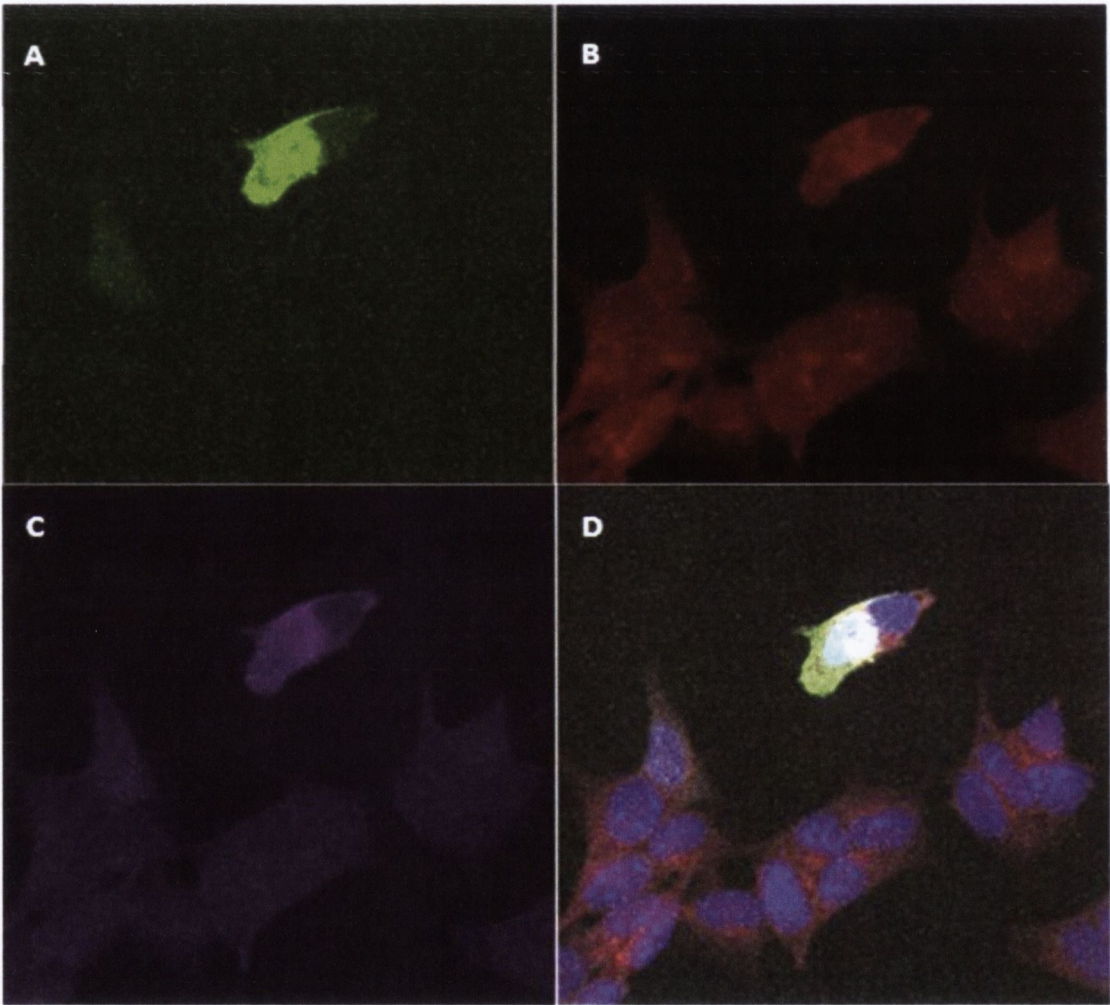


Figure 5.21 24hr Amaxa Transfection on neuroblastoma cell lines with the site directed Mutant plasmid. A. native *GFP*, B. *FYN* antibody staining, C. MAP2 staining indicates the presence of synapses. D. All three stains can be observed together with DAPI.

The result of transfection showed a very bright signal, and a transfection rate similar to that of the original mutant plasmid supplied by Genecopia (Figure 5.16). Due to time constraints, I was not able to duplicate or follow up on this area of study further. I have however standardised an efficient method of introducing mutations into WT plasmids for future investigations into this LoF variant in my lab.

Given more time, I would have liked to research the SDM plasmid in more detail. The transfection rate for this SDM plasmid was very high, and gave a very bright signal. There are many areas I would have liked to look into, to see what caused this change, including the methylation profile of the SDM mutant, or perhaps something in the DsPI digest protocol may have altered it in some way, to make it easier to transfect. I consulted with others that had carried out the SDM protocol in other labs, but could find no specific reason for this greater transfection rate, and signal strength.

5.11 Discussion

Elucidation of the impact of LoF variants identified from sequencing studies requires functional investigation. I selected a LoF variant for investigation based on 1) prior evidence of its role in synaptic function and 2) the small gene size which should make the task of transfecting a plasmid containing the gene into both neuroblastoma cell lines and mouse neurons a less difficult task. This nonsense mutation in the *FYN* gene results in truncation of more than half of the protein.

Before performing transfection and subsequent immunohistochemistry, I carried out a number of steps to ensure the specificity of my analysis. I karyotyped the neuroblastoma cell line to ensure there were no large structural abnormalities present, which are the first abnormality to be generated in passaged cell lines. I sequenced both the WT and Mut plasmids to ensure that they contained the correct inserts, and also carried out Western Blot analysis of the *FYN* antibody, to ensure its specificity to the *FYN* gene.

Confocal microscopy image analysis following transfection with the WT and Mut version of *FYN* and with a control *GFP* plasmid, revealed successful transfection of all plasmids. Overall the rate of transfection was greater for the mutant plasmid, which could potentially be attributed to its smaller size. For the two groups of successfully transfected cells (WT and Mut), an equal level of *FYN* expression was observed, indicating that there was no difference in the level of *FYN* expression between WT and Mut plasmids.

Despite *FYN* expression being the same between WT and Mut plasmids, I next wanted to investigate if there was a difference in neuronal cell function for cells transfected with WT and Mut plasmids. I used the expression of AKT as measured by Phospho-AKT as a measure of this phenotype. Confocal microscopy image analysis following transfection with the WT and Mut version of *FYN* and with a control *GFP* plasmid, revealed successful transfection of all plasmids. For the two groups of successfully transfected cells (WT and Mut), the analysis did identify a difference between WT and Mut with significantly less Phospho-AKT expression in the *FYN* Mut transfected cells ($p < 0.001$). This was a very interesting result highlighting the potential functional impact of a rare mutation identified in a SCZ patient from our sample. These experiments also found significant differences for *GFP* and Phospho-AKT expression between WT and Mut for non-transfected cells. Due to time constraints I was unable to repeat this experiment to confirm the decrease in expression of Phospho-AKT in the Mut transfected cells, and to further test data from the non-transfected cells. An ambition post-PhD is to see that this work can be completed and work towards a publication.

Although transfection of *GFP*, *FYN* WT and *FYN* Mut was achieved in primary mouse cultures, it did not show a high level of transfection, and in turn it was possible to do a formal analysis of *FYN* or Phospho-AKT expression due to low numbers of transfected cells. There are a number of reasons why the transfection process was not as successful in primary cultures. For example, they require a 7-day turnaround, which can lead to overcrowding of cells. When this method proved inadequate I undertook a second method

of transfection using lipofectamine (not described here) a cationic-lipid transfections reagent, however this method also proved unsuccessful.

Site directed mutagenesis proved a successful method of vector modification. I was able to introduce a single base change difference in my WT plasmid to create an identical copy of my mutant plasmid. This was then successfully transfected into cell lines. This method can now be utilised for future studies in the lab.

With the use of functional studies come a large number of caveats. I sought to minimise all variables throughout this set of experiments where I could, such as transfecting at an early passage, carrying out transfections on the same day in an identical manner and with the same antibody mix for every immunohistochemistry experiment on each day. However, the variability of these methods at all stages of the experiment could potentially lead to varying levels of transfection, expression and morphology. It is also impossible to know how many plasmids are being transfected into each cell, therefore even though I could classify cells as transfected, these cells could contain differing levels of functioning plasmid. Cells may also be imaged on different planes, which in turn could mean a cell which appears untransfected in one plane, could be transfected on a different plane.

Chapter 6

Discussion

Schizophrenia is a severe and debilitating disorder and symptoms can be dividing into three broad categories; positive symptoms (hallucinations, delusions), negative symptoms (apathy, social withdrawal, disorganised speech) and cognitive deficits (attention, working memory and problem solving). It is a complex neurodevelopmental disorders and its underlying pathogenesis is poorly understood. It has a very strong genetic component but its genetic architecture remains to be elucidated.

In comparison to other complex diseases, the limited understanding of the pathophysiology has rendered the development of effective treatments and even diagnosis extremely difficult. Knowledge of the pathophysiology of a disease usually involves detailed examination of gene expression and other biological changes in the affected tissue or organ. In psychiatric illnesses however this is made difficult by the fact that the organ involved, the brain, is largely inaccessible for molecular analysis. Even with huge developments in neuroscience, imaging and electrophysiology tools, the brain remains difficult to study and characterise. Therefore there is much to learn by pursuing the genetic cause, attempting to identify the genetic loci involved, and using the information to decipher the convoluted genetic vulnerabilities which lead to disease causation.

The human genome is composed of approximately 3 billion nucleotides and is highly variable among individuals. DNA sequence variants range in size from small single base pair changes to large variations in copy number of sections of the genome. Variants can be neutral, which is most often the case, they can be beneficial, which contributes to evolution or they can be harmful and contribute to disease. The impact of a variant depends not only on the type of variant it is but where it resides relative to functional elements of the genome.

At the time when my PhD studies began neither linkage studies using large multi-affected families, nor candidate gene association studies had identified many robust findings. The availability of genomic technologies which allow GWAS changed genetic studies of SCZ, as it

did for many other common phenotypes. GWAS examine millions of loci across the genome to identify common variants of small effect associated with disease, and have become the most popular method of genetic studies of complex disorders. They rely on the premise that at least a proportion of the underlying heritability of the disorder under study is due to common variants of small effect. Since the first published GWAS of SCZ in 2008 (O'Donovan et al., 2008), there has been a steady flow of studies. The trend has been for sample numbers analysed together to increase over time, primarily driven by sharing of data through international collaboration. At the same time, the number of SNPs analysed has also increased, primarily driven by imputation (Marchini et al., 2007) using dense reference panels such as 1000 Genomes (Abecasis et al., 2012).

In the largest GWAS of SCZ published to date by Ripke et al (Ripke et al., 2013) they combined samples from a Swedish study, the PGC and a selection of independent samples to give 21,000 case and 38,000 control samples. They reported 22 loci that are genome-wide significant for SCZ and of these 13 are new. This study implicated a number of L-type voltage gated channel genes, involved in learning, memory and synaptic plasticity. Calcium channel genes have also been implicated in the genetic etiology of BD, once again highlighting the genetic overlap between these disorders. This study estimated that 8,300 mostly common SNPs (95% credible interval of 6,300-10,200) contribute to the etiology of SCZ, which collectively accounts for 32% of the overall variance. The study also proposes targeting the identification of the top 2,000 risk loci, and from this, identify the biological pathways that are fundamental to disease risk.

GWAS studies have also revealed the extent to which SCZ shares genetic variation with other psychiatric disorders. For example the overlap in heritability attributed to common variation between SCZ and BD is estimated at approximately 15%. In addition to this the overlap between SCZ and major depressive disorder was found to be approximately 9% and with ASD approximately 3% (Lee et al., 2013).

While GWAS is primarily focused on the study of common variation, it also has the capacity to study rare variation in the form of copy number variants (CNVs). CNVs contribute an important genetic component to the heterogeneity and etiology of SCZ. On average there are >1000 CNVs in the genome and their relative contribution to genomic variation could be greater than SNPs given their capacity to directly impact more of the genome. With regard to SCZ studies, risk CNVs have been confirmed at a number of loci throughout the genome (Malhotra & Sebat, 2012) and have been found to impact genes involved in neuronal function (Walsh et al., 2008), synaptic activity (Malhotra et al., 2011) and neurodevelopment (Walsh et al., 2008). *De novo* CNVs identified in SCZ trios have also been found to be significantly enriched in components of the postsynaptic NMDA receptor (NMDAR) and Activity-Regulated Cytoskeleton-Associated Protein (ARC) signalling complexes (Kirov et al., 2012). As with common SNPs, for rare SCZ-associated CNVs we observe that the same CNVs are associated with multiple neuropsychiatric disorders, along with other developmental and neurological disorders (Malhotra et al., 2011). Given the heterogeneity in these studies at both the level of the CNVs (type, size, penetrance) and the multiple phenotypes involved, many challenges still remain to elucidate the causative relationship between genes affected by CNVs and disease phenotype.

Next-generation sequencing technologies have more recently allowed the search for rare variants to extend beyond large CNVs to much smaller variants including single nucleotide changes. In the most recent study by Purcell et al (S. M. Purcell et al., 2014) used exome sequencing to study rare disruptive variants in a large set of genes with a higher likelihood of having a role in SCZ based on existing genetic evidence. The study focused on 2,500 genes in a data set of 2,536 cases and 2,543 controls and discovered a polygenic burden of ultra-rare nonsense mutations which increases the risk for SCZ. Once again genes were notably related to calcium channels and ARC complex.

The shared genetic etiology between neurodevelopmental disorders was investigated in the largest and most recent exome sequencing study of *de novo* mutations in SCZ. In this study they focused on the biological processes involved in SCZ, the crossover between SCZ and

other neurodevelopmental disorders and the relevance of *de novo* mutations in SCZ, to establish if there was a higher than expected rate of mutations affecting protein sequences (Fromer et al., 2014). They showed enrichment of nonsynonymous *de novo* mutations in a number of genes, indicating that a proportion of mutations are pathogenic for SCZ. This study also observed an enrichment for mutations involved in both the NMDAR and ARC complexes. ARC is a protein coding gene required for consolidation of synaptic plasticity as well as formation of long-term memory. It also plays a role in regulation of cell morphology and cytoskeletal organization. An overlap of genes with *de novo* mutations were also identified between SCZ, ASD and ID highlighting the crossover between these disorders. Results from this study support the suggestion that understanding of the genetic architecture of these disorders might be better achieved if research were to integrate finding across multiple disorders.

Given the emerging relevance of rare variation to SCZ, I sought to investigate rare variants using a number of approaches. With increased numbers of reports detailing genetic overlap between SCZ and other disorders including neurological disorders, I became involved in a project that followed up a study by Prof. Orla Hardiman's group, who found a hexanucleotide expansion mutation in 41% of Irish familial ALS cases and in 5% of sporadic ALS cases. This study strongly supported initial reports that associated this mutation with ALS and FTD. Evidence for potential genetic overlap between FTD and schizophrenia has been reported. Clinical resemblance has been reported between frontal lobe syndrome and negative symptoms in schizophrenia (Mok et al., 2012; Ziauddeen et al., 2011)(Mok et al., 2012; Ziauddeen et al., 2011) people with younger onset FTD can be diagnosed with psychosis before being diagnosed with dementia (Velakoulis et al., 2009) .

The repeat expansion 'GGGGCC' in the intronic region of the *C9ORF72* gene is thought to reduce gene expression, in turn leading to disease through haploinsufficiency. The repeat sequence is generally reported to be pathogenic when >30 repeats. Of the 1,243 case and 1,234 control samples that I analysed, none were found to have a hexanucleotide expansion with >30 repeats. I was also able to use GWAS data to study the relationship between the

founder haplotype and the *C9ORF72* repeat expansion in this large homogeneous sample. My data supports the claim that carriers of the expansion arose from a single common founder, or at the very least that if the expansion arose multiple times, it has done so on the same haplotype background.

I did find seven samples with >22 repeats. The functional impact and importance of the repeat expansion at different lengths remains to be elucidated. There have been conflicting reports with regard to the importance of the intermediate repeat length (20<30). In a study by Rutherford et al (Rutherford et al., 2012), they found no association between the normal range (<30) and the disease phenotype, whereas in a study by van der Zee et al (van der Zee et al., 2013) a significant loss of function was reported with increasing numbers of the normal repeat expansion. An area I would have liked to investigate further would have been the family history of the patients in my analysis to see if there were any reports of FTD or ALS in family members. Another avenue would have been to look at the repeat number in phenotyped family members of the case samples carrying the repeat, to see if there is any evidence of linkage between the repeat and SCZ as it's possible that an intermediate repeat length could have some functional impact that contributes to risk of the disorder.

Although NGS is the gold standard for genetic analysis, it is still relatively expensive, and so there is currently a blind spot for smaller-scale and less expensive methods of rare variant detection for researchers interested to screen a small number of genes in 1000's of samples for follow-up studies of genome-wide discovery studies. One method that does claim to allow for rapid, high-throughput, inexpensive and sensitive mutation detection is HRM. I sought to test this methods efficiency by analysing HapMap samples, as DNA sequence information for all of these samples is publicly available on online databases. Based on my analysis, I decided not to carry the method through to the psychosis case-control samples, due to its overall lack of both sensitivity and specificity. Although the method was able to detect common variants, and in one case a rare variant from 1000 Genomes data, the HRM produced a number of false negative results and also produced calls led to a substantial number of false positive samples being brought forward for subsequent sequencing.

At the same time as I was testing the HRM method on the *PAK7* gene, I was given the opportunity to collaborate directly in a project within our group investigating its association with SCZ. A combined analysis of Irish/UK subjects with SCZ and BD identified 22 carriers in 11,707 cases and 10 carriers in 21,204 controls of a duplication starting at 9.68Mb with similar breakpoints across all samples. I successfully PCR-amplified this unique region in a carrier sample, and confirmed that they all shared exactly the same breakpoints with the same 15bp sequence inserted between the two copies of the repeat sequence, suggesting that this represents the same duplication event. To follow up on this interesting discovery, I had wanted to screen the *PAK7* gene for rare coding variants that may also increase risk of SCZ in my psychosis case-control sample but this was not possible because of the poor performance of the HRM method. An alternative was to look at exome sequencing data from UK10K which includes a sample collection of individuals with neurodevelopmental disorders (SCZ and ASD).

I performed functional annotation of the genetic variants in the case and control samples using Annovar. I identified 82 variants, and carried out a carrier-based association analysis on the data. I found no statistically significant differences in the rate of mutation carriers between cases and controls in the UK10K data set. In my final analysis of the 6 non-synonymous damaging variants identified, I found 13 carriers in the case samples and only 5 carriers in the controls. Although the frequency of mutation carriers is nearly double for the cases compared to controls, given the small sample numbers involved, the data was not statistically significant. My studies do highlight the need for large sample sizes for rare variant association analysis. These results then led me to investigate the gene further by utilising the Exome Variant Server database (EVS) (<http://evs.gs.washington.edu/EVS/>). This project allows researchers to access exome sequencing data for richly phenotyped samples from studies of heart, lung and blood disorders. I analysed the data for *PAK7* to see if any of the rare and damaging non-synonymous variants identified in my case-control sample were present in this dataset. I found only one present, which was the more “common” variant rs140079492 and it was present in two samples. Based on the Polyphen scores recorded for variants in the EVS, 19 possibly or probably damaging rare damaging non-synonymous variants and 1 frameshift variant are reported in the *PAK7* gene in 6,500 samples. The

UK10K and the EVS data both highlight the fact that many rare variants exist and establishing association with any phenotype will be major challenge for human genetic studies when the function and pathogenicity of rare variants is largely unknown.

Finally, I used a functional approach to rare variant analysis. Understanding the impact of LoF variants requires a functional approach to measure the effect of a mutation on the properties and function of genes and gene products. With this in mind I sought to investigate a nonsense mutation in the *FYN* gene identified by our group in a targeted sequencing study (Kenny et al., 2013). My approach was to design plasmids containing (*GFP*)-tagged WT and Mut versions of the gene and transfect them into neuroblastoma cell lines and primary cultures. This was then followed with immunohistochemistry to compare the expression levels of both *FYN* and Phospho-AKT in WT and Mut transfected and non-transfected cells.

The 3 plasmids (*GFP*, *FYN* WT & *FYN* Mut) were transfected into neuroblastoma cell lines and primary cultures. Unfortunately the transfection rate was extremely low in the primary cultures. However I did find a high transfection rate for both the *GFP* and *FYN* Mut plasmids in the neuroblastoma cell lines. I did not see a significant difference between the expression of *FYN* in the WT compared to the Mut cells. However I did see a significant difference in Phospho-AKT expression between WT and Mut cells. I found that the Mut transfected cells had a lower level of expression than the WT ($p < 0.001$). I also carried out Site Directed Mutagenesis on the WT *FYN* plasmid, introducing a single base change difference to create an identical copy of my mutant plasmid. This was then successfully transfected into cell lines. This method can now be utilised for future studies in the lab. If time had permitted I would have liked to study the primary neuronal cultures further, as this would allow me to look at a selection of morphological factors, such as dendrite length and number of synapses in the WT and Mut cells.

Even with sophisticated technology, we are still some way from figuring out the specifics of the genetic etiology of SCZ: what genes are involved, how they incur risk and whether or not certain mutations are associated with the different subtypes. SCZ is a highly heterogeneous disorder and with each investigation into its causes, we realize just how deep its complexities are. Even following GWAS, sequencing studies of rare variants are just part of the beginning of discovery; many risk genes still need to be identified and their products and functions elucidated before the biological basis of this devastating disorder can be understood.

References

- Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., Handsaker, R. E., . . . McVean, G. A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422), 56-65. doi: 10.1038/nature11632
- Al-Sarraj, S., King, A., Troakes, C., Smith, B., Maekawa, S., Bodi, I., . . . Shaw, C. E. (2011). p62 positive, TDP-43 negative, neuronal cytoplasmic and intranuclear inclusions in the cerebellum and hippocampus define the pathology of C9orf72-linked FTL and MND/ALS. *Acta Neuropathol*, 122(6), 691-702. doi: 10.1007/s00401-011-0911-2
- Altshuler, D., Daly, M. J., & Lander, E. S. (2008). Genetic Mapping in Human Disease *Science* (Vol. 322, pp. 881-888).
- Anney, R. J., Kenny, E., O'Dushlaine, C. T., Lasky-Su, J., Franke, B., Morris, D. W., . . . Gill, M. (2008). Non-random error in genotype calling procedures: implications for family-based and case-control genome-wide association studies. *Am J Med Genet B Neuropsychiatr Genet*, 147B(8), 1379-1386.
- Badner, J. A., & Gershon, E. S. (2002). Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Mol Psychiatry*, 7(4), 405-411.
- Baron-Cohen, S., Scott, F. J., Allison, C., Williams, J., Bolton, P., Matthews, F. E., & Brayne, C. (2009). Prevalence of autism-spectrum conditions: UK school-based population study. *Br J Psychiatry*, 194(6), 500-509. doi: 10.1192/bjp.bp.108.059345
- Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 21(2), 263-265. doi: 10.1093/bioinformatics/bth457

- Barrett, J. C., Hansoul, S., Nicolae, D. L., Cho, J. H., Duerr, R. H., Rioux, J. D., . . . Daly, M. J. (2008). Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet*, 40(8), 955-962. doi: 10.1038/ng.175
- Beck, J., Poulter, M., Hensman, D., Rohrer, J. D., Mahoney, C. J., Adamson, G., . . . Mead, S. (2013). Large C9orf72 hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population. *Am J Hum Genet*, 92(3), 345-353. doi: 10.1016/j.ajhg.2013.01.011
- Beggs, H. E., Soriano, P., & Maness, P. F. (1994). NCAM-dependent neurite outgrowth is inhibited in neurons from Fyn-minus mice. *J Cell Biol*, 127(3), 825-833.
- Behan, C., Kennelly, B., & O'CALLAGHAN, E. (2008). The economic cost of schizophrenia in Ireland: a cost of illness study. *Irish journal of psychological medicine*, 25(3).
- Bjarnadottir, M., Misner, D. L., Haverfield-Gross, S., Bruun, S., Helgason, V. G., Stefansson, H., . . . Andresson, T. (2007). Neuregulin1 (NRG1) signaling through Fyn modulates NMDA receptor phosphorylation: differential synaptic function in NRG1+/- knock-outs compared with wild-type mice. *J Neurosci*, 27(17), 4519-4529. doi: 10.1523/jneurosci.4314-06.2007
- Botstein, D., & Risch, N. (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet*, 33 Suppl, 228-237.
- Bowie, C. R., Reichenberg, A., Patterson, T. L., Heaton, R. K., & Harvey, P. D. (2006). Determinants of real-world functional performance in schizophrenia subjects: correlations with cognition, functional capacity, and symptoms. *Am J Psychiatry*, 163(3), 418-425. doi: 10.1176/appi.ajp.163.3.418

- Bray, N. J., Holmans, P. A., van den Bree, M. B., Jones, L., Elliston, L. A., Hughes, G., . . . O'Donovan, M. C. (2008). Cis- and trans- loci influence expression of the schizophrenia susceptibility gene DTNBP1 *Hum Mol Genet.* (Vol. 17, pp. 1169-1174).
- Brazil, D. P., Yang, Z. Z., & Hemmings, B. A. (2004). Advances in protein kinase B signalling: AKTion on multiple fronts. *Trends Biochem Sci*, 29(5), 233-242. doi: 10.1016/j.tibs.2004.03.006
- Brennand, K. J., Simone, A., Jou, J., Gelboin-Burkhart, C., Tran, N., Sangar, S., . . . Gage, F. H. (2011). Modelling schizophrenia using human induced pluripotent stem cells. *Nature*, 473(7346), 221-225.
- Brown, A. S., & Susser, E. S. (2008). Prenatal nutritional deficiency and risk of adult schizophrenia. *Schizophr Bull*, 34(6), 1054-1063. doi: 10.1093/schbul/sbn096
- Buonocore, F., Hill, M. J., Campbell, C. D., Oladimeji, P. B., Jeffries, A. R., Troakes, C., . . . Bray, N. J. (2010). Effects of cis-regulatory variation differ across regions of the adult human brain *Hum Mol Genet.* (Vol. 19, pp. 4490-4496).
- Burmeister, M. (1999). Basic concepts in the study of diseases with complex genetics. *Biol Psychiatry*, 45(5), 522-532.
- Buxbaum, Joseph D., & Sklar, P. (2011). Human Induced Pluripotent Stem Cells: A New Model for Schizophrenia? *Cell Stem Cell*, 8(5), 461-462.
- Byrne, S., Elamin, M., Bede, P., Shatunov, A., Walsh, C., Corr, B., . . . Hardiman, O. (2012). Cognitive and clinical characteristics of patients with amyotrophic lateral sclerosis carrying a C9orf72 repeat expansion: a population-based cohort study. *Lancet Neurol.* doi: 10.1016/s1474-4422(12)70014-5
- Cain, D. P., Grant, S. G., Saucier, D., Hargreaves, E. L., & Kandel, E. R. (1995). Fyn tyrosine kinase is required for normal amygdala kindling. *Epilepsy Res*, 22(2), 107-114.

- Cannon, M., Jones, P. B., & Murray, R. M. (2002). Obstetric complications and schizophrenia: historical and meta-analytic review. *Am J Psychiatry*, 159(7), 1080-1092.
- Cardno, A. G., & Gottesman, II. (2000). Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. *Am J Med Genet*, 97(1), 12-17.
- Chen, H. K., Fernandez-Funez, P., Acevedo, S. F., Lam, Y. C., Kaytor, M. D., Fernandez, M. H., . . . Zoghbi, H. Y. (2003). Interaction of Akt-phosphorylated ataxin-1 with 14-3-3 mediates neurodegeneration in spinocerebellar ataxia type 1. *Cell*, 113(4), 457-468.
- Cheung, V. G., Nayak, R. R., Wang, I. X., Elwyn, S., Cousins, S. M., Morley, M., & Spielman, R. S. (2010). Polymorphic Cis and Trans Regulation of Human Gene Expression. *PLoS Biol*, 8(9), e1000480.
- Cheung, V. G., Nayak, R. R., Wang, I. X., Elwyn, S., Cousins, S. M., Morley, M., & Spielman, R. S. (2011). Polymorphic Cis- and Trans- Regulation of Human Gene Expression. *PLoS Biol*, 8(9), e1000480.
- Chowdari, K. V., Mirnics, K., Semwal, P., Wood, J., Lawrence, E., Bhatia, T., . . . Nimgaonkar, V. L. (2002). Association and linkage analyses of RGS4 polymorphisms in schizophrenia *Hum Mol Genet*. (Vol. 11, pp. 1373-1380).
- Chumakov, I., Blumenfeld, M., Guerassimenko, O., Cavarec, L., Palicio, M., Abderrahim, H., . . . Cohen, D. (2002a). Genetic and physiological data implicating the new human gene G72 and the gene for d-amino acid oxidase in schizophrenia *Proc Natl Acad Sci U S A*. (Vol. 99, pp. 13675-13680).
- Chumakov, I., Blumenfeld, M., Guerassimenko, O., Cavarec, L., Palicio, M., Abderrahim, H., . . . Cohen, D. (2002b). Genetic and physiological data implicating the new human

- gene G72 and the gene for d-amino acid oxidase in schizophrenia (Vol. 99, pp. 13675-13680).
- Concannon, P., Erlich, H. A., Julier, C., Morahan, G., Nerup, J. r., Pociot, F., . . . the Type 1 Diabetes Genetics, C. (2005). Type 1 Diabetes (Vol. 54, pp. 2995-3001).
- Cookson, W., Liang, L., Abecasis, G., Moffatt, M., & Lathrop, M. (2009). Mapping complex disease traits with global gene expression. *Nat Rev Genet*, 10(3), 184-194.
- Cooper, G. M., Coe, B. P., Girirajan, S., Rosenfeld, J. A., Vu, T. H., Baker, C., . . . Eichler, E. E. (2011). A copy number variation morbidity map of developmental delay. *Nat Genet*, 43(9), 838-846. doi: 10.1038/ng.909
- Corvin, A. P., Molinos, I., Little, G., Donohoe, G., Gill, M., Morris, D. W., & Tropea, D. (2012). Insulin-like growth factor 1 (IGF1) and its active peptide (1-3)IGF1 enhance the expression of synaptic markers in neuronal circuits through different cellular mechanisms. *Neurosci Lett*, 520(1), 51-56. doi: 10.1016/j.neulet.2012.05.029
- Costa, V., Angelini, C., De Feis, I., & Ciccodicola, A. (2010). Uncovering the complexity of transcriptomes with RNA-Seq. *J Biomed Biotechnol*, 2010, 853916.
- Coyle, J. T. (2006). Glutamate and schizophrenia: beyond the dopamine hypothesis. *Cell Mol Neurobiol*, 26(4-6), 365-384. doi: 10.1007/s10571-006-9062-8
- Craddock, N., O'Donovan, M. C., & Owen, M. J. (2005). The genetics of schizophrenia and bipolar disorder: dissecting psychosis. *J Med Genet*, 42(3), 193-204.
- Craddock, N., & Owen, M. J. (2010). Data and clinical utility should be the drivers of changes to psychiatric classification. *Br J Psychiatry*, 197(2), 158; author reply 158-159. doi: 10.1192/bjp.197.2.158
- DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., . . . Rademakers, R. (2011). Expanded GGGGCC hexanucleotide repeat in

noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*, 72(2), 245-256. doi: 10.1016/j.neuron.2011.09.011

Ding, L., Wendl, M. C., Koboldt, D. C., & Mardis, E. R. (2011). Analysis of next-generation genomic data in cancer: accomplishments and challenges. *Hum Mol Genet*, 19(R2), R188-196.

Duerr, R. H., Taylor, K. D., Brant, S. R., Rioux, J. D., Silverberg, M. S., Daly, M. J., . . . Cho, J. H. (2006). A Genome-Wide Association Study Identifies IL23R as an Inflammatory Bowel Disease Gene *Science* (Vol. 314, pp. 1461-1463).

Emamian, E. S., Karayiorgou, M., & Gogos, J. A. (2004). Decreased phosphorylation of NMDA receptor type 1 at serine 897 in brains of patients with Schizophrenia. *J Neurosci*, 24(7), 1561-1564. doi: 10.1523/jneurosci.4650-03.2004

Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A. S., Zink, F., Zhu, J., . . . Stefansson, K. (2008). Genetics of gene expression and its effect on disease. *Nature*, 452(7186), 423-428.

Evans, D. M., & Cardon, L. R. (2004). Guidelines for Genotyping in Genomewide Linkage Studies: Single-Nucleotide-Polymorphism Maps Versus Microsatellite Maps. *The American Journal of Human Genetics*, 75(4), 687-692.

Fahey, C., Byrne, S., McLaughlin, R., Kenna, K., Shatunov, A., Donohoe, G., . . . Morris, D. W. (2013). Analysis of the hexanucleotide repeat expansion and founder haplotype at C9ORF72 in an Irish psychosis case-control sample. *Neurobiol Aging*. doi: 10.1016/j.neurobiolaging.2013.12.003

Fecto, F., & Siddique, T. (2011). Making connections: pathology and genetics link amyotrophic lateral sclerosis with frontotemporal lobe dementia. *J Mol Neurosci*, 45(3), 663-675. doi: 10.1007/s12031-011-9637-9

- Fernell, E., & Gillberg, C. (2010). Autism spectrum disorder diagnoses in Stockholm preschoolers. *Res Dev Disabil*, 31(3), 680-685. doi: 10.1016/j.ridd.2010.01.007
- Ferrari, R., Mok, K., Moreno, J. H., Cosentino, S., Goldman, J., Pietrini, P., . . . Momeni, P. (2012). Screening for C9ORF72 repeat expansion in FTLD. *Neurobiol Aging*. doi: 10.1016/j.neurobiolaging.2012.02.017
- Ferreira, M. A., O'Donovan, M. C., Meng, Y. A., Jones, I. R., Ruderfer, D. M., Jones, L., . . . Craddock, N. (2008). Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet*, 40(9), 1056-1058. doi: 10.1038/ng.209
- Fombonne, E. (2009). Epidemiology of pervasive developmental disorders. *Pediatr Res*, 65(6), 591-598. doi: 10.1203/PDR.0b013e31819e7203
- Frayling, T. M. (2007). Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nat Rev Genet*, 8(9), 657-662. doi: 10.1038/nrg2178
- Frazer, K. A., Murray, S. S., Schork, N. J., & Topol, E. J. (2009). Human genetic variation and its contribution to complex traits. *Nat Rev Genet*, 10(4), 241-251.
- Genome-wide association study implicates HLA-C*01:02 as a risk factor at the major histocompatibility complex locus in schizophrenia. (2012). *Biol Psychiatry*, 72(8), 620-628. doi: 10.1016/j.biopsych.2012.05.035
- Gilad, Y., Rifkin, S. A., & Pritchard, J. K. (2008). Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet*, 24(8), 408-415.
- Gill, M., Donohoe, G., & Corvin, A. (2010). What have the genomics ever done for the psychoses? *Psychol Med*, 40(4), 529-540.

- Gilman, S. R., Chang, J., Xu, B., Bawa, T. S., Gogos, J. A., Karayiorgou, M., & Vitkup, D. (2012). Diverse types of genetic variation converge on functional gene networks involved in schizophrenia. *Nat Neurosci*, 15(12), 1723-1728. doi: 10.1038/nn.3261
- Glessner, J. T., Wang, K., Cai, G., Korvatska, O., Kim, C. E., Wood, S., . . . Hakonarson, H. (2009). Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*, 459(7246), 569-573. doi: 10.1038/nature07953
- Golzio, C., Willer, J., Talkowski, M. E., Oh, E. C., Taniguchi, Y., Jacquemont, S., . . . Katsanis, N. (2012). KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature*, 485(7398), 363-367. doi: 10.1038/nature11091
- Gottesman, II. (1997). Twins: en route to QTLs for cognition. *Science*, 276(5318), 1522-1523.
- Grant, S. G., O'Dell, T. J., Karl, K. A., Stein, P. L., Soriano, P., & Kandel, E. R. (1992). Impaired long-term potentiation, spatial learning, and hippocampal development in *fyn* mutant mice. *Science*, 258(5090), 1903-1910.
- Guillin, O., Abi-Dargham, A., & Laruelle, M. (2007). Neurobiology of dopamine in schizophrenia. *Int Rev Neurobiol*, 78, 1-39. doi: 10.1016/s0074-7742(06)78001-1
- Guilmatre, A., Dubourg, C., Mosca, A. L., Legallic, S., Goldenberg, A., Drouin-Garraud, V., . . . Campion, D. (2009). Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation. *Arch Gen Psychiatry*, 66(9), 947-956. doi: 10.1001/archgenpsychiatry.2009.80
- Gulsuner, S., Walsh, T., Watts, A. C., Lee, M. K., Thornton, A. M., Casadei, S., . . . McClellan, J. M. (2013). Spatial and temporal mapping of de novo mutations in schizophrenia to a

- fetal prefrontal cortical network. *Cell*, 154(3), 518-529. doi: 10.1016/j.cell.2013.06.049
- Hahn, C. G., Wang, H. Y., Cho, D. S., Talbot, K., Gur, R. E., Berrettini, W. H., . . . Arnold, S. E. (2006). Altered neuregulin 1-erbB4 signaling contributes to NMDA receptor hypofunction in schizophrenia. *Nat Med*, 12(7), 824-828. doi: 10.1038/nm1418
- Hattori, K., Fukuzako, H., Hashiguchi, T., Hamada, S., Murata, Y., Isosaka, T., . . . Yagi, T. (2009). Decreased expression of Fyn protein and disbalanced alternative splicing patterns in platelets from patients with schizophrenia. *Psychiatry Res*, 168(2), 119-128. doi: 10.1016/j.psychres.2008.04.014
- Heap, G. A., Yang, J. H. M., Downes, K., Healy, B. C., Hunt, K. A., Bockett, N., . . . Plagnol, V. (2010). Genome-wide analysis of allelic expression imbalance in human primary cells by high-throughput transcriptome resequencing *Hum Mol Genet.* (Vol. 19, pp. 122-134).
- Henderson, D. C. (2005). Schizophrenia and comorbid metabolic disorders. *J Clin Psychiatry*, 66 Suppl 6, 11-20.
- Hennah, W., Thomson, P., McQuillin, A., Bass, N., Loukola, A., Anjorin, A., . . . Porteous, D. (2009). DISC1 association, heterogeneity and interplay in schizophrenia and bipolar disorder. *Mol Psychiatry*, 14(9), 865-873. doi: 10.1038/mp.2008.22
- Henquet, C., Rosa, A., Krabbendam, L., Papiol, S., Fananas, L., Drukker, M., . . . van Os, J. (2006). An experimental study of catechol-o-methyltransferase Val158Met moderation of delta-9-tetrahydrocannabinol-induced effects on psychosis and cognition. *Neuropsychopharmacology*, 31(12), 2748-2757. doi: 10.1038/sj.npp.1301197

- Hill, M. J., & Bray, N. J. (2011). Allelic differences in nuclear protein binding at a genome-wide significant risk variant for schizophrenia in ZNF804A. *Mol Psychiatry*.
- Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., & Manolio, T. A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits *Proc Natl Acad Sci U S A*. (Vol. 106, pp. 9362-9367).
- Hirschhorn, J. N., & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*, 6(2), 95-108.
- Hor, K., & Taylor, M. (2010). Suicide and schizophrenia: a systematic review of rates and risk factors. *J Psychopharmacol*, 24(4 Suppl), 81-90. doi: 10.1177/1359786810385490
- Huey, E. D., Nagy, P. L., Rodriguez-Murillo, L., Manoochchri, M., Goldman, J., Lieberman, J., . . . Mayeux, R. (2013). C9ORF72 repeat expansions not detected in a group of patients with schizophrenia. *Neurobiol Aging*, 34(4), 1309 e1309-1310. doi: 10.1016/j.neurobiolaging.2012.08.011
- Hugot, J.-P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J.-P., Belaiche, J., . . . Thomas, G. (2001). Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature*, 411(6837), 599-603.
- Hunt, K. A., Zhernakova, A., Turner, G., Heap, G. A. R., Franke, L., Bruinenberg, M., . . . van Heel, D. A. (2008). Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet*, 40(4), 395-402.
- Insel, T. R. (2010). Rethinking schizophrenia. *Nature*, 468(7321), 187-193. doi: 10.1038/nature09552
- International Schizophrenia, C., Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., . . . Sklar, P. (2009). Common polygenic variation contributes to

- risk of schizophrenia and bipolar disorder. *Nature*, 460(7256), 748-752. doi: 10.1038/nature08185
- Jimenez-Sanchez, G., Childs, B., & Valle, D. (2001). Human disease genes. *Nature*, 409(6822), 853-855.
- Kaspar, B. K., Llado, J., Sherkat, N., Rothstein, J. D., & Gage, F. H. (2003). Retrograde viral delivery of IGF-1 prolongs survival in a mouse ALS model. *Science*, 301(5634), 839-842. doi: 10.1126/science.1086137
- Keefe, R. S., Vinogradov, S., Medalia, A., Silverstein, S. M., Bell, M. D., Dickinson, D., . . . Stroup, T. S. (2011). Report from the working group conference on multisite trial design for cognitive remediation in schizophrenia. *Schizophr Bull*, 37(5), 1057-1065. doi: 10.1093/schbul/sbq010
- Kenny, E. M., Cormican, P., Furlong, S., Heron, E., Kenny, G., Fahey, C., . . . Morris, D. W. (2013). Excess of rare novel loss-of-function variants in synaptic genes in schizophrenia and autism spectrum disorders. *Mol Psychiatry*. doi: 10.1038/mp.2013.127
- Kim, H. G., Kishikawa, S., Higgins, A. W., Seong, I. S., Donovan, D. J., Shen, Y., . . . Gusella, J. F. (2008). Disruption of neurexin 1 associated with autism spectrum disorder. *Am J Hum Genet*, 82(1), 199-207. doi: 10.1016/j.ajhg.2007.09.011
- King, M. C., & Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees *Science* (Vol. 188, pp. 107-116).
- Kirov, G., Gumus, D., Chen, W., Norton, N., Georgieva, L., Sari, M., . . . Ullmann, R. (2008). Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Hum Mol Genet*, 17(3), 458-465. doi: 10.1093/hmg/ddm323

- Kirov, G., Pocklington, A. J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., . . . Owen, M. J. (2012). De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol Psychiatry*, 17(2), 142-153. doi: 10.1038/mp.2011.154
- Kojima, N., Wang, J., Mansuy, I. M., Grant, S. G., Mayford, M., & Kandel, E. R. (1997). Rescuing impairment of long-term potentiation in fyn-deficient mice by introducing Fyn transgene. *Proc Natl Acad Sci U S A*, 94(9), 4761-4765.
- Krabbendam, L., & van Os, J. (2005). Schizophrenia and urbanicity: a major environmental influence--conditional on genetic risk. *Schizophr Bull*, 31(4), 795-799. doi: 10.1093/schbul/sbi060
- Laaksovirta, H., Peuralinna, T., Schymick, J. C., Scholz, S. W., Lai, S. L., Myllykangas, L., . . . Traynor, B. J. (2010). Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. *Lancet Neurol*, 9(10), 978-985. doi: 10.1016/s1474-4422(10)70184-8
- Lalonde, E., Ha, K. C., Wang, Z., Bemmo, A., Kleinman, C. L., Kwan, T., . . . Majewski, J. (2011). RNA sequencing reveals the role of splicing polymorphisms in regulating human gene expression. *Genome Res*, 21(4), 545-554.
- Lander, E. S. (1996). The New Genomics: Global Views of Biology *Science* (Vol. 274, pp. 536-539).
- Lander, E. S., & Schork, N. J. (1994). Genetic dissection of complex traits. *Science*, 265(5181), 2037-2048.
- Lander, E. S., & Schork, N. J. (2006). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet.* (Vol. 4, pp. 442-458).

- Laruelle, M. (1998). Imaging dopamine transmission in schizophrenia. A review and meta-analysis. *Q J Nucl Med*, 42(3), 211-221.
- Lee, S. H., DeCandia, T. R., Ripke, S., Yang, J., Sullivan, P. F., Goddard, M. E., . . . Wray, N. R. (2012). Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet*, 44(3), 247-250. doi: 10.1038/ng.1108
- Lee, S. H., Ripke, S., Neale, B. M., Faraone, S. V., Purcell, S. M., Perlis, R. H., . . . Wray, N. R. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet*, 45(9), 984-994. doi: 10.1038/ng.2711
- Levinson, D. F., Duan, J., Oh, S., Wang, K., Sanders, A. R., Shi, J., . . . Gejman, P. V. (2011). Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. *Am J Psychiatry*, 168(3), 302-316. doi: 10.1176/appi.ajp.2010.10060876
- Lewis, C. M., Levinson, D. F., Wise, L. H., DeLisi, L. E., Straub, R. E., Hovatta, I., . . . Helgason, T. (2003). Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet*, 73(1), 34-48.
- Libioulle, C. c., Louis, E., Hansoul, S., Sandor, C., Farnir, F. d. r., Franchimont, D., . . . Georges, M. (2007). Novel Crohn Disease Locus Identified by Genome-Wide Association Maps to a Gene Desert on 5p13.1 and Modulates Expression of <italic>PTGER4</italic>. *PLoS Genet*, 3(4), e58.
- Licatalosi, D. D., Mele, A., Fak, J. J., Ule, J., Kayikci, M., Chi, S. W., . . . Darnell, R. B. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*, 456(7221), 464-469.
- Lichtenstein, P., Yip, B. H., Bjork, C., Pawitan, Y., Cannon, T. D., Sullivan, P. F., & Hultman, C. M. (2009). Common genetic determinants of schizophrenia and bipolar disorder in

- Swedish families: a population-based study. *Lancet*, 373(9659), 234-239. doi: 10.1016/s0140-6736(09)60072-6
- Lichtenstein, P., Yip, B. H., Björk, C., Pawitan, Y., Cannon, T. D., Sullivan, P. F., & Hultman, C. M. (2009). Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *The Lancet*, 373(9659), 234-239.
- Lillo, P., & Hodges, J. R. (2009). Frontotemporal dementia and motor neurone disease: overlapping clinic-pathological disorders. *J Clin Neurosci*, 16(9), 1131-1135. doi: 10.1016/j.jocn.2009.03.005
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., . . . Rutter, M. (2000). The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord*, 30(3), 205-223.
- Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord*, 24(5), 659-685.
- Majewski, J., & Pastinen, T. (2011). The study of eQTL variations by RNA-seq: from SNPs to phenotypes. *Trends in Genetics*, 27(2), 72-79.
- Majounie, E., Renton, A. E., Mok, K., Dopper, E. G., Waite, A., Rollinson, S., . . . Traynor, B. J. (2012). Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. *Lancet Neurol*, 11(4), 323-330. doi: 10.1016/s1474-4422(12)70043-1
- Malhotra, D., McCarthy, S., Michaelson, J. J., Vacic, V., Burdick, K. E., Yoon, S., . . . Sebat, J. (2011). High frequencies of de novo CNVs in bipolar disorder and schizophrenia. *Neuron*, 72(6), 951-963. doi: 10.1016/j.neuron.2011.11.007

- Malhotra, D., & Sebat, J. (2012). CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell*, 148(6), 1223-1241. doi: 10.1016/j.cell.2012.02.039
- Marchini, J., Howie, B., Myers, S., McVean, G., & Donnelly, P. (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*, 39(7), 906-913. doi: 10.1038/ng2088
- Marguerat, S., & Bahler, J. (2010). RNA-seq: from technology to biology. *Cell Mol Life Sci*, 67(4), 569-579.
- Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M., & Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res*, 18(9), 1509-1517.
- Marshall, C. R., Noor, A., Vincent, J. B., Lionel, A. C., Feuk, L., Skaug, J., . . . Scherer, S. W. (2008). Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet*, 82(2), 477-488. doi: 10.1016/j.ajhg.2007.12.009
- McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. A., & Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*, 9(5), 356-369.
- McCarthy, S. E., Makarov, V., Kirov, G., Addington, A. M., McClellan, J., Yoon, S., . . . Sebat, J. (2009). Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet*, 41(11), 1223-1227. doi: 10.1038/ng.474
- McGuffin, P., Farmer, A., & Harvey, I. (1991). A polydiagnostic application of operational criteria in studies of psychotic illness. Development and reliability of the OPCRIT system. *Arch Gen Psychiatry*, 48(8), 764-770.
- Miller, D. T., Adam, M. P., Aradhya, S., Biesecker, L. G., Brothman, A. R., Carter, N. P., . . . Ledbetter, D. H. (2010). Consensus statement: chromosomal microarray is a first-tier

- clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet*, 86(5), 749-764. doi: 10.1016/j.ajhg.2010.04.006
- Mitchell, K. J., & Porteous, D. J. (2011). Rethinking the genetic architecture of schizophrenia. *Psychol Med*, 41(1), 19-32. doi: 10.1017/S003329171000070X
- Moens, L. N., De Rijk, P., Reumers, J., Van den Bossche, M. J., Glassee, W., De Zutter, S., . . . Del-Favero, J. (2011). Sequencing of DISC1 pathway genes reveals increased burden of rare missense variants in schizophrenia patients from a northern Swedish population. *PLoS One*, 6(8), e23450. doi: 10.1371/journal.pone.0023450
- Moffatt, M. F., Kabisch, M., Liang, L., Dixon, A. L., Strachan, D., Heath, S., . . . Cookson, W. O. C. (2007). Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature*, 448(7152), 470-473.
- Mok, K. Y., Koutsis, G., Schottlaender, L. V., Polke, J., Panas, M., & Houlden, H. (2012). High frequency of the expanded C9ORF72 hexanucleotide repeat in familial and sporadic Greek ALS patients. *Neurobiol Aging*, 33(8), 1851 e1851-1855. doi: 10.1016/j.neurobiolaging.2012.02.021
- Momeni, P., Wickremaratchi, M. M., Bell, J., Arnold, R., Beer, R., Hardy, J., . . . Morris, H. R. (2010). Familial early onset frontotemporal dementia caused by a novel S356T MAPT mutation, initially diagnosed as schizophrenia. *Clin Neurol Neurosurg*, 112(10), 917-920. doi: 10.1016/j.clineuro.2010.07.015
- Montgomery, S. B., & Dermitzakis, E. T. (2011). From expression QTLs to personalized transcriptomics. *Nat Rev Genet*, 12(4), 277-282.
- Montgomery, S. B., Sammeth, M., Gutierrez-Arcelus, M., Lach, R. P., Ingle, C., Nisbett, J., . . . Dermitzakis, E. T. (2010). Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature*, 464(7289), 773-777.

- Morley, M., Molony, C. M., Weber, T. M., Devlin, J. L., Ewens, K. G., Spielman, R. S., & Cheung, V. G. (2004). Genetic analysis of genome-wide variation in human gene expression. *Nature*, 430(7001), 743-747.
- Morris, D. W., Pearson, R. D., Cormican, P., Kenny, E. M., O'Dushlaine, C. T., Lemieux Perreault, L. P., . . . Corvin, A. (2014). An inherited duplication at the gene p21 Protein-Activated Kinase 7 (PAK7) is a risk factor for psychosis. *Hum Mol Genet*. doi: 10.1093/hmg/ddu025
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Meth*, 5(7), 621-628.
- Mudge, J., Miller, N. A., Khrebtukova, I., Lindquist, I. E., May, G. D., Huntley, J. J., . . . Kingsmore, S. F. (2008). Genomic Convergence Analysis of Schizophrenia: mRNA Sequencing Reveals Altered Synaptic Vesicular Transport in Post-Mortem Cerebellum. *PLoS ONE*, 3(11), e3625.
- Myers, R. A., Casals, F., Gauthier, J., Hamdan, F. F., Keebler, J., Boyko, A. R., . . . Awadalla, P. (2011). A population genetic approach to mapping neurological disorder genes using deep resequencing. *PLoS Genet*, 7(2), e1001318. doi: 10.1371/journal.pgen.1001318
- Ng, M. Y. M., Levinson, D. F., Faraone, S. V., Suarez, B. K., DeLisi, L. E., Arinami, T., . . . Lewis, C. M. (2008). Meta-analysis of 32 genome-wide linkage studies of schizophrenia. *Mol Psychiatry*, 14(8), 774-785.
- Nica, A. C., & Dermitzakis, E. T. (2008). Using gene expression to investigate the genetic basis of complex disorders. *Human Molecular Genetics*, ddn285.
- Nica, A. C., Parts, L., Glass, D., Nisbet, J., Barrett, A., Sekowska, M., . . . The Mu, T. C. (2011). The Architecture of Gene Regulatory Variation across Multiple Human Tissues: The MuTHER Study. *PLoS Genet*, 7(2), e1002003.

- Nicodemus, K. K., Law, A. J., Luna, A., Vakkalanka, R., Straub, R. E., Kleinman, J. E., & Weinberger, D. R. (2009). A 5[prime] promoter region SNP in NRG1 is associated with schizophrenia risk and type III isoform expression. *Mol Psychiatry*, 14(8), 741-743.
- Nicolae, D. L., Gamazon, E., Zhang, W., Duan, S., Dolan, M. E., & Cox, N. J. (2010). Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet*, 6(4), e1000888.
- Nieratschker, V., Meyer-Lindenberg, A., & Witt, S. H. (2011). Genome-wide investigation of rare structural variants identifies VIPR2 as a new candidate gene for schizophrenia. *Expert Rev Neurother*, 11(7), 937-941. doi: 10.1586/ern.11.84
- Nuytemans, K., Bademci, G., Kohli, M. M., Beecham, G. W., Wang, L., Young, J. I., . . . Vance, J. M. (2013). C9ORF72 Intermediate Repeat Copies Are a Significant Risk Factor for Parkinson Disease. *Ann Hum Genet*. doi: 10.1111/ahg.12033
- O'Donovan, M. C., Craddock, N., Norton, N., Williams, H., Peirce, T., Moskvina, V., . . . Molecular Genetics of Schizophrenia, C. (2008). Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet*, 40(9), 1053-1055. doi: 10.1038/ng.201
- O'Donovan, M., Craddock, N., & Owen, M. (2009). Genetics of psychosis; insights from views across the genome. *Human Genetics*, 126(1), 3-12.
- Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*, 40(12), 1413-1415.
- Park, P. J. (2009). ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*, 10(10), 669-680.

- Pastinen, T., Ge, B., & Hudson, T. J. (2006). Influence of human genome polymorphism on gene expression *Hum Mol Genet* (Vol. 15, pp. R9-16).
- Pickrell, J. K., Marioni, J. C., Pai, A. A., Degner, J. F., Engelhardt, B. E., Nkadori, E., . . . Pritchard, J. K. (2010). Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature*, *464*(7289), 768-772.
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*, *38*(8), 904-909.
- Price, A. L., Zaitlen, N. A., Reich, D., & Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nat Rev Genet*, *11*(7), 459-463.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, *81*(3), 559-575. doi: 10.1086/519795
- Purcell, S. M., Moran, J. L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., . . . Sklar, P. (2014). A polygenic burden of rare disruptive mutations in schizophrenia. *Nature*, *506*(7487), 185-190. doi: 10.1038/nature12975
- Ramensky, V., Bork, P., & Sunyaev, S. (2002). Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.* (Vol. 30, pp. 3894-3900).
- Rare chromosomal deletions and duplications increase risk of schizophrenia. (2008). *Nature*, *455*(7210), 237-241. doi: 10.1038/nature07239
- Reich, D. E., & Lander, E. S. (2001). On the allelic spectrum of human disease. *Trends in Genetics*, *17*(9), 502-510.

- Reichelt, A. C., Rodgers, R. J., & Clapcote, S. J. (2012). The role of neurexins in schizophrenia and autistic spectrum disorder. *Neuropharmacology*, 62(3), 1519-1526. doi: 10.1016/j.neuropharm.2011.01.024
- Renton, A. E., Majounie, E., Waite, A., Simon-Sanchez, J., Rollinson, S., Gibbs, J. R., . . . Traynor, B. J. (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron*, 72(2), 257-268. doi: 10.1016/j.neuron.2011.09.010
- Richards, A. L., Jones, L., Moskvina, V., Kirov, G., Gejman, P. V., Levinson, D. F., . . . O'Donovan, M. C. (2011). Schizophrenia susceptibility alleles are enriched for alleles that affect gene expression in adult human brain. *Mol Psychiatry*.
- Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J. L., Kahler, A. K., Akterin, S., . . . Sullivan, P. F. (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet*. doi: 10.1038/ng.2742
- Ripke, S., Sanders, A. R., Kendler, K. S., Levinson, D. F., Sklar, P., Holmans, P. A., . . . Gejman, P. V. (2011). Genome-wide association study identifies five new schizophrenia loci. *Nat Genet*, 43(10), 969-976. doi: 10.1038/ng.940
- Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*, 273(5281), 1516-1517.
- Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., . . . Jones, S. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Meth*, 4(8), 651-657.
- Ronald, A., Simonoff, E., Kuntsi, J., Asherson, P., & Plomin, R. (2008). Evidence for overlapping genetic influences on autistic and ADHD behaviours in a community twin

- sample. *J Child Psychol Psychiatry*, 49(5), 535-542. doi: 10.1111/j.1469-7610.2007.01857.x
- Rujescu, D., Ingason, A., Cichon, S., Pietilainen, O. P., Barnes, M. R., Toulopoulou, T., . . . Collier, D. A. (2009). Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet*, 18(5), 988-996. doi: 10.1093/hmg/ddn351
- Rutherford, N. J., Heckman, M. G., Dejesus-Hernandez, M., Baker, M. C., Soto-Ortolaza, A. I., Rayaprolu, S., . . . Rademakers, R. (2012). Length of normal alleles of C9ORF72 GGGGCC repeat do not influence disease phenotype. *Neurobiol Aging*, 33(12), 2950 e2955-2957. doi: 10.1016/j.neurobiolaging.2012.07.005
- Sanders, A. R., Duan, J., Levinson, D. F., Shi, J., He, D., Hou, C., . . . Gejman, P. V. (2008). No Significant Association of 14 Candidate Genes With Schizophrenia in a Large European Ancestry Sample: Implications for Psychiatric Genetics *Am J Psychiatry*. (Vol. 165, pp. 497-506).
- Saudou, F., Finkbeiner, S., Devys, D., & Greenberg, M. E. (1998). Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. *Cell*, 95(1), 55-66.
- Schoder, D., Hannequin, D., Martinaud, O., Opolczynski, G., Guyant-Marechal, L., Le Ber, I., & Campion, D. (2010). Morbid risk for schizophrenia in first-degree relatives of people with frontotemporal dementia. *Br J Psychiatry*, 197(1), 28-35. doi: 10.1192/bjp.bp.109.068981
- Schork, Nicholas J. (1997). Genetics of Complex Disease . Approaches, Problems, and Solutions *Am J Respir Crit Care Med*. (Vol. 156, pp. S103-109).

- Schulze, T. G., Akula, N., Breuer, R., Steele, J., Nalls, M. A., Singleton, A. B., . . . McMahon, F. J. (2012). Molecular genetic overlap in bipolar disorder, schizophrenia, and major depressive disorder. *World J Biol Psychiatry*. doi: 10.3109/15622975.2012.662282
- Serretti, A., & Mandelli, L. (2008). The genetics of bipolar disorder: genome /`hot regions,/' genes, new potential candidates and future directions. *Mol Psychiatry*, 13(8), 742-771.
- Shepherd, A. M., Laurens, K. R., Matheson, S. L., Carr, V. J., & Green, M. J. (2012). Systematic meta-review and quality assessment of the structural brain alterations in schizophrenia. *Neurosci Biobehav Rev*, 36(4), 1342-1356. doi: 10.1016/j.neubiorev.2011.12.015
- Shi, J., Levinson, D. F., Duan, J., Sanders, A. R., Zheng, Y., Pe'er, I., . . . Gejman, P. V. (2009). Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature*, 460(7256), 753-757. doi: 10.1038/nature08192
- Shi, Y., Li, Z., Xu, Q., Wang, T., Li, T., Shen, J., . . . He, L. (2011). Common variants on 8p12 and 1q24.2 confer risk of schizophrenia. *Nat Genet*, 43(12), 1224-1227. doi: 10.1038/ng.980
- Simon-Sanchez, J., Schulte, C., Bras, J. M., Sharma, M., Gibbs, J. R., Berg, D., . . . Gasser, T. (2009). Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat Genet*, 41(12), 1308-1312. doi: 10.1038/ng.487
- Smith, B. N., Newhouse, S., Shatunov, A., Vance, C., Topp, S., Johnson, L., . . . Shaw, C. E. (2013). The C9ORF72 expansion mutation is a common cause of ALS+/-FTD in Europe and has a single founder. *Eur J Hum Genet*, 21(1), 102-108. doi: 10.1038/ejhg.2012.98

- Snowden, J. S., Rollinson, S., Thompson, J. C., Harris, J. M., Stopford, C. L., Richardson, A. M., . . . Pickering-Brown, S. M. (2012). Distinct clinical and pathological characteristics of frontotemporal dementia associated with C9ORF72 mutations. *Brain*, 135(Pt 3), 693-708. doi: 10.1093/brain/awr355
- Spencer, C. C. A., Su, Z., Donnelly, P., & Marchini, J. (2009). Designing Genome-Wide Association Studies: Sample Size, Power, Imputation, and the Choice of Genotyping Chip. *PLoS Genet*, 5(5), e1000477.
- Stefansson, H., Ophoff, R. A., Steinberg, S., Andreassen, O. A., Cichon, S., Rujescu, D., . . . Collier, D. A. (2009). Common variants conferring risk of schizophrenia. *Nature*, 460(7256), 744-747. doi: 10.1038/nature08186
- Stefansson, H., Petursson, H., Sigurdsson, E., Steinthorsdottir, V., Bjornsdottir, S., Sigmundsson, T., . . . Stefansson, K. (2002). Neuregulin 1 and Susceptibility to Schizophrenia. *The American Journal of Human Genetics*, 71(4), 877-892.
- Stefansson, H., Rujescu, D., Cichon, S., Pietilainen, O. P., Ingason, A., Steinberg, S., . . . Stefansson, K. (2008). Large recurrent microdeletions associated with schizophrenia. *Nature*, 455(7210), 232-236. doi: 10.1038/nature07229
- Stefansson, H., Sigurdsson, E., Steinthorsdottir, V., Bjornsdottir, S., Sigmundsson, T., Ghosh, S., . . . Stefansson, K. (2002). Neuregulin 1 and susceptibility to schizophrenia. *Am J Hum Genet*, 71(4), 877-892.
- Stranger, B. E., Forrest, M. S., Dunning, M., Ingle, C. E., Beazley, C., Thorne, N., . . . Dermitzakis, E. T. (2007). Relative Impact of Nucleotide and Copy Number Variation on Gene Expression Phenotypes *Science*. (Vol. 315, pp. 848-853).

- Stranger, B. E., Stahl, E. A., & Raj, T. (2010). Progress and Promise of Genome-Wide Association Studies for Human Complex Trait Genetics *Genetics*. (Vol. 187, pp. 367-383).
- Straub, R. E., Jiang, Y., MacLean, C. J., Ma, Y., Webb, B. T., Myakishev, M. V., . . . Kendler, K. S. (2002). Genetic variation in the 6p22.3 gene DTNBP1, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia. *Am J Hum Genet*, 71(2), 337-348.
- Straub, R. E., MacLean, C. J., O'Neill, F. A., Burke, J., Murphy, B., Duke, F., . . . Kendler, K. S. (1995). A potential vulnerability locus for schizophrenia on chromosome 6p24-22: evidence for genetic heterogeneity. *Nat Genet*, 11(3), 287-293.
- Sullivan, P. F., Magnusson, C., Reichenberg, A., Boman, M., Dalman, C., Davidson, M., . . . Lichtenstein, P. (2012). Family History of Schizophrenia and Bipolar Disorder as Risk Factors for Autism Family History of Psychosis as Risk Factor for ASD. *Arch Gen Psychiatry*, 1-5. doi: 10.1001/archgenpsychiatry.2012.730
- Szamosi, A., Kelemen, O., & Keri, S. (2012). Hippocampal volume and the AKT signaling system in first-episode schizophrenia. *J Psychiatr Res*, 46(3), 279-284. doi: 10.1016/j.jpsychires.2011.12.005
- Szatmari, P., Paterson, A. D., Zwaigenbaum, L., Roberts, W., Brian, J., Liu, X. Q., . . . Meyer, K. J. (2007). Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet*, 39(3), 319-328. doi: 10.1038/ng1985
- Thiselton, D. L., Vladimirov, V. I., Kuo, P. H., McClay, J., Wormley, B., Fanous, A., . . . Riley, B. P. (2008). AKT1 is associated with schizophrenia across multiple symptom dimensions in the Irish study of high density schizophrenia families. *Biol Psychiatry*, 63(5), 449-457. doi: 10.1016/j.biopsych.2007.06.005

- Thomas, G., Jacobs, K. B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., . . . Chanock, S. J. (2008). Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet*, 40(3), 310-315. doi: 10.1038/ng.91
- Tiihonen, J., Lonnqvist, J., Wahlbeck, K., Klaukka, T., Niskanen, L., Tanskanen, A., & Haukka, J. (2009). 11-year follow-up of mortality in patients with schizophrenia: a population-based cohort study (FIN11 study). *Lancet*, 374(9690), 620-627. doi: 10.1016/s0140-6736(09)60742-x
- Todd, J. A., Walker, N. M., Cooper, J. D., Smyth, D. J., Downes, K., Plagnol, V., . . . Clayton, D. G. (2007). Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet*, 39(7), 857-864.
- Tsuang, M. F., S. The Genetics of Mood Disorders (Johns Hopkins University Press, Baltimore, MD, 1990).
- Twine, N. A., Janitz, K., Wilkins, M. R., & Janitz, M. (2011). Whole Transcriptome Sequencing Reveals Gene Expression and Splicing Differences in Brain Regions Affected by Alzheimer's Disease. *PLoS ONE*, 6(1), e16266.
- Vacic, V., McCarthy, S., Malhotra, D., Murray, F., Chou, H. H., Peoples, A., . . . Sebat, J. (2011). Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. *Nature*, 471(7339), 499-503. doi: 10.1038/nature09884
- van der Zee, J., Gijselinck, I., Dillen, L., Van Langenhove, T., Theuns, J., Engelborghs, S., . . . Cavallaro, T. (2013). A pan-European study of the C9orf72 repeat associated with FTLT: geographic prevalence, genomic instability, and intermediate repeats. *Hum Mutat*, 34(2), 363-373. doi: 10.1002/humu.22244

- Vassos, E., Collier, D. A., Holden, S., Patch, C., Rujescu, D., St Clair, D., & Lewis, C. M. (2010). Penetrance for copy number variants associated with schizophrenia. *Hum Mol Genet*, 19(17), 3477-3481. doi: 10.1093/hmg/ddq259
- Velakoulis, D., Walterfang, M., Mocellin, R., Pantelis, C., & McLean, C. (2009). Frontotemporal dementia presenting as schizophrenia-like psychosis in young people: clinicopathological series and review of cases. *Br J Psychiatry*, 194(4), 298-305. doi: 10.1192/bjp.bp.108.057034
- Vrijenhoek, T., Buizer-Voskamp, J. E., van der Stelt, I., Strengman, E., Sabatti, C., Geurts van Kessel, A., . . . Veltman, J. A. (2008). Recurrent CNVs disrupt three candidate genes in schizophrenia patients. *Am J Hum Genet*, 83(4), 504-510. doi: 10.1016/j.ajhg.2008.09.011
- Walsh, T., McClellan, J. M., McCarthy, S. E., Addington, A. M., Pierce, S. B., Cooper, G. M., . . . Sebat, J. (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*, 320(5875), 539-543. doi: 10.1126/science.1155174
- Wang, W. Y. S., Barratt, B. J., Clayton, D. G., & Todd, J. A. (2005). Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet*, 6(2), 109-118.
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 10(1), 57-63.
- Williams, H. J., Craddock, N., Russo, G., Hamshere, M. L., Moskvina, V., Dwyer, S., . . . O'Donovan, M. C. (2011). Most genome-wide significant susceptibility loci for schizophrenia and bipolar disorder reported to date cross-traditional diagnostic boundaries. *Hum Mol Genet*, 20(2), 387-391. doi: 10.1093/hmg/ddq471

- Williams, H. J., Norton, N., Dwyer, S., Moskvina, V., Nikolov, I., Carroll, L., . . . O'Donovan, M. C. (2010). Fine mapping of ZNF804A and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder. *Mol Psychiatry*.
- WTCCC. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145), 661-678.
- Xi, Z., Zinman, L., Grinberg, Y., Moreno, D., Sato, C., Bilbao, J. M., . . . Rogaeva, E. (2012). Investigation of c9orf72 in 4 neurodegenerative disorders. *Arch Neurol*, 69(12), 1583-1590. doi: 10.1001/archneurol.2012.2016
- Xu, B., Ionita-Laza, I., Roos, J. L., Boone, B., Woodrick, S., Sun, Y., . . . Karayiorgou, M. (2012). De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat Genet*, 44(12), 1365-1369. doi: 10.1038/ng.2446
- Xu, B., Roos, J. L., Dexheimer, P., Boone, B., Plummer, B., Levy, S., . . . Karayiorgou, M. (2011). Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat Genet*, 43(9), 864-868. doi: 10.1038/ng.902
- Yan, H., Yuan, W., Velculescu, V. E., Vogelstein, B., & Kinzler, K. W. (2002). Allelic Variation in Human Gene Expression *Science* (Vol. 297, pp. 1143).
- Yang, Z., Camp, N. J., Sun, H., Tong, Z., Gibbs, D., Cameron, D. J., . . . Zhang, K. (2006). A Variant of the HTRA1 Gene Increases Susceptibility to Age-Related Macular Degeneration *Science* (Vol. 314, pp. 992-993).
- Yue, W. H., Wang, H. F., Sun, L. D., Tang, F. L., Liu, Z. H., Zhang, H. X., . . . Zhang, D. (2011). Genome-wide association study identifies a susceptibility locus for schizophrenia in Han Chinese at 11p11.2. *Nat Genet*, 43(12), 1228-1231. doi: 10.1038/ng.979

- Ziauddeen, H., Dibben, C., Kipps, C., Hodges, J. R., & McKenna, P. J. (2011). Negative schizophrenic symptoms and the frontal lobe syndrome: one and the same? *Eur Arch Psychiatry Clin Neurosci*, 261(1), 59-67. doi: 10.1007/s00406-010-0133-y
- Zoghbi, H. Y., & Warren, S. T. (2010). Neurogenetics: advancing the "next-generation" of brain research. *Neuron*, 68(2), 165-173. doi: 10.1016/j.neuron.2010.10.015