Bipolar Disorder Related Functional Variants in the Calcium Channel Gene Family.

Niamh Louise O'Brien

Division of Psychiatry

University College London,

Thesis submitted for the degree of

Doctor of Philosophy

UCL

I, **Niamh O'Brien** confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Bipolar disorder (BD) is a common highly heritable disorder. The calcium channel gene family has been widely implicated in BD aetiology and these genes include *CACNA1C,* and *CACNG4.* The association signal for *CACNA1C* with BD is located in the middle of the third intron of the gene. *CACNG4* encodes a transmembrane AMPA receptor regulator that is involved in trafficking AMPA receptors to the neuronal post-synaptic density.

High-resolution melting curve (HRM) analysis and whole genome sequencing (WGS) methods were used to identify functional variants in calcium channels genes in the UCL BD cohort. Variants that were predicted to impact gene regulation, transcription or to be damaging to protein structure were genotyped in the larger UCL BD and control cohort.

HRM analysis identified 26 calcium channel gene variants. These included two non-synonymous *CACNG4* variants that were associated with mental illness (rs371128228, $p=1.05x10^{-4}$, OR=4.39 and 17:65026851 (C/T), $p=5x10^{-4}$, OR=9.52). Fluorescent activated cell sorting analysis was used to determine the effect of rs371128228 on trafficking of GluR1 and GluR2 to the cell surface. This analysis demonstrated that the risk allele of rs371128228 significantly decreased cell surface trafficking of GluR1 ($p=0.026$) but no effect was observed on GluR2 trafficking.

WGS analysis of *CACNA1C* intron 3 identified two BD associated (p=0.015, OR=1.15) variants 105bp apart that were in complete LD. Both variants are predicted to create YY1 transcription factor binding sites. Luciferase reporter assays show a significant decrease in gene expression in the presence of both variants ($p=0.004$). Protein-DNA complex assays of the *CACNA1C* variants demonstrate increased nuclear proteins binding affinity for the variant alleles.

If these calcium channel variants are confirmed to be important risk factors for BD they could be used as markers for personalised treatment or in the identification of genetic subtypes of BD or other psychiatric illness.

Acknowledgements

As the end of my time working on this thesis draws to a close I am very aware of the amount of support and direction I have received over the course of my PhD. First and foremost I would like to thank my supervisor Dr.Andrew McQuillin for giving me the wonderful opportunity to carry out this work and PhD and for all of his advice and support throughout the years, no matter how silly the questions I may have asked were. I would like to thank Dr. Hugh Gurling for his introduction to the Molecular Psychiatry laboratory and support and advice during the first years of my PhD. I would like to extend my sincerest gratitude to Dr. Nicholas Bass, who after Dr. Gurling sadly passed away took over the mantle as my secondary supervisor and continued to support me through my PhD. Dr.Bass's advice and feedback during my write up and constant source of encouragement to keep going was invaluable.

My sincerest thanks go to The Bipolar Foundation: Equilibrium and the Fairclough family for providing the funding which has given me the opportunity to work in a field of great interest to me over the course of my PhD. Without this I would not have been able to work in the field of psychiatric genetics.

The next in line are my parents and family who have given me tremendous support throughout the course of my work. I can never thank them enough for the encouragement and motivation following my move to London and during the last number of years. They have always pushed me to strive for the best and have listened to me talk about my work non-stop, even when they may not have understood anything I was saying.

A very special thanks needs to go to both Alessia and Sally. Without their help I would not be the researcher I am today. They were always on hand no matter how small or large the questions I had about science were. I am greatly indebted to them for their support, advice and friendship during this time. I would like to thank all of my current and ex-colleagues of the Molecular Psychiatry lab, Michael, Yi, Jon, Alex, Kate, Giorgia and Mariam for their support.

I would like to thank everyone, especially Kleoniki, Chris, Jaspal and Niamh, who has listened to me complain about failed experiments and thesis writing. There have been many times when your friendship, support and ability to make me laugh have been the exact remedy I needed and you have never shown exasperation whether you felt it or not. A special thanks to Chris as well for helping me with the enormous task of proof-reading this thesis.

Last but not least I am sincerely grateful to my partner Gavriil for his constant and never ending support during my PhD. I would like to thank him for his cool calm attitude, his constant emotional support and motivation, the late night food runs and for never doubting my abilities.

Lastly, I would like to thank everyone again who has made this possible and made the journey extremely memorable and enjoyable.

Table of Contents

ABSTRACT 3

ACKNOWLEDGEMENTS 4

TABLE OF CONTENTS 6

LIST OF FIGURES 10

LIST OF TABLES 13

LIST OF ABBREVIATIONS 16

CHAPTER 1 INTRODUCTION 21

1.1 BIPOLAR DISORDER 21

1.1.1 OVERVIEW 21

1.1.2 HISTORY AND CLASSIFICATION 24

1.1.3 DIAGNOSTIC CRITERIA 26

1.1.4 BIPOLAR DISORDER SUBTYPES 30

1.1.5 BIOCHEMICAL PATHWAYS IMPLICATED IN BIPOLAR DISORDER 36

1.1.6 TREATMENT OF BIPOLAR DISORDER 40

1.2 GENETICS OF BIPOLAR DISORDER 46

1.2.1 HERITABILITY OF BIPOLAR DISORDER 46

1.2.2 MOLECULAR GENETIC STUDIES OF BIPOLAR DISORDER 50

1.3 CALCIUM CHANNEL GENES 81

1.3.1 INTRODUCTION 81

1.3.2 GENETIC VARIATION IN CALCIUM CHANNEL GENES 95

1.4 INTRONIC GENETIC VARIATION 99

1.4.1 INTRODUCTION 99

1.4.2 ENCODE DATA 100

1.4.3 UTILISING DATA SETS SUCH AS ENCODE TO INVESTIGATE NGS 103

CHAPTER 2 AIMS OF THESIS 104

CHAPTER 3 MATERIAL AND METHODS 105

3.1 GENERAL METHODS 105

3.1.1 UCL RESEARCH SUBJECTS 105

3.1.2 GENOMIC DNA EXTRACTIONS 107

3.1.3 DNA QUANTIFICATIONS 110

3.1.4 Polymerase Chain Reaction 111
3.1.5 Agarose Gel Electrophoresis 114
3.1.6 Mammalian Cell Culture 116
3.2 Methods for results Chapter 4 and 5: High-resolution melting curve analysis in Calcium Channel genes 120
3.2.1 Research Samples 120
3.2.2 High-resolution Melting 120
3.2.3 Purification of samples 121
3.2.4 Sanger Sequencing 123
3.2.5 Variant Selection for KASPar genotyping 126
3.2.6 SNV Genotyping with KASPar 127
3.2.7 Plasmid Construction 133
3.2.8 Cellular Transfections 147
3.2.9 Flow Cytometry 147
3.3 Methods for results chapter 5: Investigation of variants using Next Generation Sequencing Analysis 149
3.3.1 Research Samples for Next Generation Sequencing 149
3.3.2 Next Generation Sequencing and Analysis 149
3.3.3 Variant selection 151
3.3.4 KASPar Genotyping 152
3.3.5 Statistical Analysis 152
3.3.6 Plasmid Construction 153
3.3.7 Luciferase Assay 158
3.3.8 Electromobility Shift Assay 160
3.3.9 EMSA Supershift Assay 164
CHAPTER 4 HIGH-RESOLUTION MELTING CURVE ANALYSIS OF REGULATORY REGIONS IN CALCIUM CHANNEL GENES 165
4.1 Introduction 165
4.2 Hypothesis 169
4.3 Aims 169
4.4 HRM methods for screening *CACNA1C*, *CACNA1D*, and *CACNB3* 170
4.5 Results 171
4.5.1 HRM Screening of Genes 171
4.5.2 Investigation of the Effects of Variants 172
4.5.3 Genotyping of Variants from HRM 175
4.6 Discussion 176

CHAPTER 5 HIGH RESOLUTION MELTING CURVE ANALYSIS OF *CACNG4* 179

5.1 INTRODUCTION 179

5.2 HYPOTHESIS 181

5.3 AIMS 183

5.4 HRM SCREENING OF *CACNG4* 183

5.5 RESULTS 185

5.5.1 HRM SCREENING OF GENES 185

5.5.2 INVESTIGATION OF VARIANT EFFECTS 185

5.5.3 GENOTYPING OF VARIANTS DETECTED BY HRM 187

5.5.4 EFFECT OF RS371128228 ON AMPA-R TRAFFICKING 193

5.6 DISCUSSION 198

CHAPTER 6 INVESTIGATION OF VARIANTS IN CALCIUM CHANNEL GENES USING NEXT GENERATION SEQUENCING DATA 205

6.1 EXONIC REGIONS 205

6.1.1 INTRODUCTION 205

6.1.2 HYPOTHESIS 206

6.1.3 AIM 206

6.1.4 RESULTS 206

6.1.5 DISCUSSION 213

6.2 INTRONIC VARIANTS 215

6.2.1 INTRODUCTION 215

6.2.2 AIMS 217

6.2.3 HYPOTHESIS 217

6.2.4 RESULTS 218

6.2.5 DISCUSSION 237

CHAPTER 7 GENERAL DISCUSSION 241

7.1 STUDY LIMITATIONS AND POTENTIAL CONFOUNDS 245

7.2 FUTURE DIRECTIONS 246

REFERENCES 248

APPENDIX I 274

APPENDIX II 281

II.I DNA EXTRACTION FROM BLOOD SAMPLES 281

II.II AGROSE GELS 281

II.III CELL CULTURE 282

II.IV TRANSIENT TRANSFECTION 282
II.V PURIFICATION OF PCR PRODUCTS 282
II.VI MINI AND MIDI PREP 283
II.VII EMSA 283
APPENDIX III 285
APPENDIX IV 288
PUBLICATIONS 290

List of Figures

Figure 1. Common symptoms of Bipolar Disorder. 23
Figure 2. Weighted mean for discharge diagnoses in 6 psychiatric treatment centres in the United States between 1972-1988 29
Figure 3. Mood changes in BD-I and BD-II over time. 31
Figure 4. Effect of Lithium and Valproate on the inositol synthesis pathway. 42
Figure 5. Estimates of the lifetime risk of developing bipolar disorder. 47
Figure 6. Heritability estimates for a number of Psychiatric Illnesses. 50
Figure 7. Relationship of associated loci in psychiatric illness and sample size 68
Figure 8. Concept of Next Generation Sequencing through parallel sequencing of genomic DNA. 72
Figure 9 Timeline of progress in psychiatric genetics from 1989-2013. 74
Figure 10. Families of genes highlighted from pathway analysis in 200 WGS BD patients. 79
Figure 11. Formation of the Voltage-Gated Calcium Channel in the cell membrane. 82
Figure 12. Mechanisms by which neuronal activity generates increased levels of intercellular and nuclear calcium transients. 85
Figure 13. Processes activated following the influx of Ca^{2+} ions. 86
Figure 14. Activation of gene transcription by the influx of calcium through plasma membrane calcium channels 87
Figure 15. TARP Phylogenetic tree 93
Figure 16. Trafficking of AMPA-R to the postsynaptic density through TARP interaction. 94
Figure 17. Structure of the Alpha Pore of VDCC. 96
Figure 18. Primer 3 results for primers in the intronic region of CACNA1C. 113
Figure 19. Neubauer-Improved Haemocytometer. 117
Figure 20. HRM melt curve analysis. 122
Figure 21. BigDye reaction for Terminator Cycle Sequencing. 124
Figure 22. Staden Trace display of a *CACNG4* variant. 126
Figure 23. Diagrammatic representation of the mechanisms underlying KASPar Chemistry 129

Figure 24. KASPar genotyping for *CACNA1C.* 131
Figure 25. Vector sequence for pCMV6-XL4. 134
Figure 26. Vector Sequence for pTAG-RFP. 134
Figure 27. pGEM-T Vector 136
Figure 28. GluR1Elution 136
Figure 29 PCR of *E.coli* single colonies for LIC insert. 141
Figure 30. Circular map of the pCI-SEP-GluR1 vector. 145
Figure 31. Circular vector map of the pCI-SEP-GluR2 vector. 145
Figure 32. From Samples to variant genotyping. A stepwise depiction of the experiments involved in whole genome sequencing 150
Figure 33. pJET1.2 Blunt End Cloning Vector 153
Figure 34. pGL3-Basic Vector 157
Figure 35. pRL-SV40 *Renilla* Vector 157
Figure 36. Regional association plot of CACNB3. 167
Figure 37. Regional association plot of *CACNA1C*. 168
Figure 38. Regional association plot of *CACNA1D* 168
Figure 39. Exome sequencing coverage of regulatory regions in *CACNA1D*, *CACNA1C* and *CACNB3*. 170
Figure 40. TESS predictions for 12:49212493 (G/C) in the *CACNB3* promoter. 173
Figure 41. TESS predictions for 12:2162217 (G/A) and 12:2162570 (C/G) in the *CACNA1C* promoter. 174
Figure 42. TARP γ2 and γ4 protein homology. 181
Figure 43. Location of the associated variant from case-case analysis between *CACNG4* and *CACNG5.* 182
Figure 44. Regional association plot of *CACNG4*. 182
Figure 45. An example of primer design for exon 4 and part of the 3'UTR of *CACNG4*. 184
Figure 46. Amino acid changes in *CACNG4.* 188
Figure 47. TARP γ4 structure 188
Figure 48. Histogram of the Fluorescence of SEP for *CACNG4*/GluR1. 194
Figure 49. Histogram of the Fluorescence of SEP for *CACNG4/GluR2* 195
Figure 50. Fluorescence intensity of SEP in HEK293 cells. 197
Figure 51. Linkage disequilibrium between *CACNG5* and *CACNG4*. 200

Figure 52. Normalised molecular abundance of GluR's and CACNG proteins for the inner core architecture of GluR's in different brain regions 203
Figure 53. Linkage Disequilibrium in the third intron of *CACNA1C.* 216
Figure 54. PROMO transcription factor binding site predictions for CACNA1C intronic variants 223
Figure 55. mRNA secondary structures for rs79398153 and rs116947827 227
Figure 56. mRNA centroid structural predictions of variants rs79398153 and rs116947827. 228
Figure 57. mRNA secondary structures for rs79398153 and rs116947827. 228
Figure 58. Relative luciferase activity reporter gene assays for the experimental constructs in HEK293 cells. 231
Figure 59. Electomobility shift assay for rs79398153. 234
Figure 60. Electromobility shift assay for rs116947827. 235
Figure 61. Super shift EMSA assay of *CACNA1C* variants with a YY1 specfic antibody. 236

List of Tables

Table 1. Summary of all the GWAS studies of BD to date....58
Table 2. Next generation sequencing studies conducted on Bipolar Disorder....75
Table 3. Calcium channel genes. 83
Table 4. Classification of the alpha pores of the calcium channel family....90
Table 5. Histone Markers in the Genome 102
Table 6. Standardised PCR reactions mixes for optimisations. 114
Table 7. PCR conditions used for primer optimisation....114
Table 8. Optimisation condition for HRM assays. 122
Table 9. Big Dye v3.2 Reaction Mix for sequencing....124
Table 10. Big Dye v3.1 Thermocycling conditions....124
Table 11. Allele Mix for KASPar Assay....130
Table 12. Optimisation Conditions for KASPar Genotyping 130
Table 13. Thermocycling conditions for KASPar Genotyping....132
Table 14. PCR conditions for LIC cloning. 139
Table 15. *pfx* primer optimisation conditions. 139
Table 16. PCR reaction mixture for site directed mutagenesis 147
Table 17. Thermocycling conditions for site directed mutagenesis....147
Table 18. *Pfx* cycling conditions....154
Table 19. Components of the Dual Luciferase Reporter Assay Kit....159
Table 20. EMSA Conditions....163
Table 21. EMSA supershift reaction mixture. 164
Table 22. HRM variants from analysis in *CACNA1C, CACNA1D* and *CACNB3*....172
Table 23. Genotyping results for calcium channel variants detected by HRM. 176
Table 24. Variants found using HRM analysis in *CACNG4*....186
Table 25. Polyphen2 and SIFT prediction effects for non-synonymous variants in *CACNG4.* 186
Table 26. Results of SNVs genotyped from *CACNG*4. 190
Table 27. Total variant counts for rs371128228 and 17:65026851 (C/T) in *CACNG4.* 191
Table 28. Total case control sample size for rs371128228 and 17:65026851 (C/T) *CACNG4* variants. 192

Table 29. Combined analysis of psychosis group versus Healthy Controls for *CACNAG4* variants..... 192
Table 30. Mean Fluorescence from *CACNG4*/AMPA-R experiments. 196
Table 31. Non-synonymous SNVs in CACNA1S and CACNA1D from whole genome sequencing of 99 BD samples. 209
Table 32. Non-synonymous SNVs in *CACNA2D4, CACNB1* and *CACNB2* from whole genome sequencing of 99 BD samples. 210
Table 33. Variants from WGS for follow up genotyping. 211
Table 34. Genotyping results for calcium channel variants detected by WGS..... 212
Table 35. *CACNA1C* intronic variants matching study criteria..... 220
Table 36. Pairwise linkage disequilibrium analysis between the BD associated SNP reported here and previous GWAS SNPs..... 220
Table 37. Tests of association with *CACNA1C* intronic variants with BD. 221
Table 38. Allele frequencies from 1000G dataset for rs79398153 and rs116947827. 224
Table 39. TFBS predicted by PROMO when both rs79398153 and rs116947827 are present. 226
Table 40. Human mRNA from brain regions from the UCSC genome browser from the third intron of *CACNA1C*. 227
Table 41. Student T-test for changes in luciferase expression with *CACNA1C* intronic variants compared to the wild type control..... 230
Table 42. ICD-10 criteria for categorising a manic episode [7]. 275
Table 43. ICD-10 criteria for categorising a depressive episode [7]. 276
Table 44. ICD-10 criteria for categorising a depressive episode [1]. 277
Table 45. ICD-10 classifications of Bipolar Disorder. 278
Table 46. ICD-10 classifications of Bipolar Disorder. 279
Table 47. ICD-10 classification of mixed affective episode 280
Table 48. Primers used for HRM analysis of *CACNA1C* 285
Table 49. Primers used for HRM analysis of *CACNA1D*..... 285
Table 50. Primers used for HRM analysis of *CACNB3*..... 285
Table 51. Primers used for HRM analysis of *CACNG4*..... 286
Table 52.Primers for Genotyping *CACNA1C* promoter SNP1 287
Table 53. Primers for Genotyping variants from the fourth exon of *CACNG4*..... 287

Table 54. Primers for Genotyping *CACNA1C* intronic variants identified in WGS data..... 288
Table 55. Primers for Genotyping *CACNA2D4* and *CACNA1D* intronic variants identified in WGS data..... 289

List of Abbreviations

5-HT	5-Hydroxy Tryptamine Serotonin
5-HTT	5-Hydroxy Tryptamine Transporter
AID	α Interacting Domain
ADHD	Attention Deficiency Hyperactivity Disorder
ADS	Alcohol Dependence Syndrome
ADS-I	Alcohol Dependence Syndrome Type 1
ADS-II	Alcohol Dependence Syndrome Type 2
AF	Allele Frequency
AMPA	α-Amino-3-Hydroxy-5-Methyl-4-Isoxazolepropionic Acid
AMPA-R	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor
ANK3	Ankyrin 3
APA	American Psychiatric Association
APS	Ammonium Sulphate
ASC	Autism Sequencing Collaboration
ASD	Autism Spectrum Disorder
ASMAD	Amish Study of Major Affective Disorders
ATP6V1G3	ATPase, H+ Transporting, Lysosomal V1 gene
BCL-2	B-cell CLL/lymphoma-2 gene
BD	Bipolar Disorder
BD I	Bipolar I Disorder
BD II	Bipolar II Disorder
BDNF	brain derived neurotrophic factor gene
BID	B Interacting Domain
BIC	Bayesian Information Criterion
BiGs	Bipolar Genome Study
BrS	Brugada Syndrome
BSA	Bovine Serum Albumin
Ca^{2+}	Calcium ion
CACNA1C	Calcium Channel, Voltage-Dependent, L Type, A 1C Subunit
CACNA1D	Calcium Channel, Voltage-Dependent, L Type, D 1D Subunit
CACNA1S	Calcium Channel, Voltage-Dependent, L Type, S 1D Subunit
CACNA2D4	Calcium Channel, Voltage-Dependent, A2/Δ Subunit 4
CACNB3	Calcium Channel, Voltage-Dependent, B 3 Subunit
CACNG2	Calcium Channel, Voltage-Dependent, Γ Subunit 2
CACNG4	Calcium Channel, Voltage-Dependent, Γ Subunit 4
CACNG5	Calcium Channel, Voltage-Dependent, Γ Subunit 5
CAN	Ca^{2+}-Activated Non-Selective Cation Channels
CBP	CREB-binding protein
CBT	Cognitive Behavioural Therapy
ChIP-Seq	Chromatin Immunoprecipitation Followed By Sequencing
cM	Centimorgans

CNS	Central Nervous System
CNV	Copy Number Variants
COMT	Catechol-o-Methyltransderase
CREB	cyclic-AMP- response element (CRE)-binding protein
D_1-type	Dopamine type 1 receptor
DAT	Dopamine Transporter
DAOA	D-amino acid oxidase
dbSNP	Single Nucleotide Polymorphism Database
dbGAP	Database of Genotypes and Phenotypes
dCTP	Deoxycytidine Triphosphate
DGKH	Diacylglycerol Kinase Eta
dGTP	Deoxyguanosine Triphosphate
DHSs	DNase-Hypersensitivity Site
DLR™	Dual-Luciferase® Reporter
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl Sulfoxide
dNTP	Deoxynucleoside Triphosphate
DRE	downstream response element
DREAM	downstream response element (DRE)-antagonist modulator
DSM	Diagnostic and Statistical Manual of Mental Disorders
DTI	Diffusion Tensor Imaging
DTT	Dithiothreitol
DZ	Dyzygotic
ECACC	European Collection Of Animal Cell Cultures
EDTA	Ethylenediaminetetraacetic Acid
EGA	European Genome-phenome archive
EGFR	Epidermal Growth Factor receptor
EMSA	Electromobility Shift Assay
ENCODE	Encyclopaedia Of DNA Elements
EMSA	Electromobility Shift Assay
EPSP	Excitatory Postsynaptic Potentials
ER	Endoplasmic Reticulum
ESP	Exome Sequencing Project
EsR	Estrogen Receptor
EtBr	Ethidium Bromide
EUR	European
EuroPoPP-MH	European profile of prevention and promotion of mental health study
EVS	Exome Variant Server
EXaC	Exome Aggregation Consortium
FBS	Foetal Bovine Serum
FIN	Finnish population
fMRI	Functional Magnetic Resonance
G-6-PO_4	Glucose-6-Phosphate
GABRB1	GABA A receptor β1 subunit gene
GAIN	Genetic Association Information Network
GBR	British in England and Scotland
GC	Genomic Control
GERP	Genomic Evolutionary Rate Profiling
GFP	Green Fluorescent Protein

GM12878	Immortalized B Lymphocyte Epstein-Barr Line
GO	Gene Ontology
GPCR	G protein coupled receptors
GRP78	glucose related protein
GRM7	Glutamate Receptor, Metabotropic 7 gene
GSK	GlaxoSmithKline
GSMA	Genome Scan Meta-Analysis
GRIA1	Glutamate Receptor 1
GWAS	Genome-wide Association Study
H2A.Z	Variant of H2A Histone Protein
H3k4me1	Histone 3 Mono Methylated Lysine 4
H3k4me2	Histone 3 Di Methylated Lysine 4
H3k4me3	Histone 3 Tri Methylated Lysine 4
H3k9ac	Histone 3 Lysine 9 Acetylation
H3k9me1	Histone 3 Mono Methylated Lysine 9
H3k9me3	Histone 3 Tri Methylated Lysine 9
H3k27me3	Histone 3 Tri Methylated Lysine 27
H3k36me3	Histone 3 Tri Methylated Lysine 36
H3k79me2	Histone 3 Di Methylated Lysine 79
H4k20me1	Histone 4 Mono Methylated Lysine 20
HEPES	2-[4-(2-hydroxyethyl) piperazin-1-yl] ethanesulfonic acid
hESC	Human Embryonic Stem Cells
HEK293	Human Embryonic Kidney 293 transformed cell line
HGP	Human Genome Project
HM	Hypomania
HPA	Hypothalamic-Pituitary-Adrenal Axis
HRM	High-resolution Melting
HWE	Hardy-Weinberg Equilibrium
IBS	Iberian population in Spain
ICD10	International Statistical Classification Of Diseases And Related Health Problems 10th Revision
IMPase	Inositol Monophosphatase
INRICH	interval-based enrichment analysis
$InsP_3$	Inositol 1,4,5-Triphosphate
IPA	Ingenuity Pathway Analysis software
iPS	Induced Pluripotent stem cell
IPT	Interpersonal Therapy
ITIH1	heavy chain 1 of the Inter-α trypsin inhibitor gene
K^+	Potassium Ion
K652	Human Myeloid Leukemia Cell Line
KASPar	KBiosciences Competitive Allele-Specific PCR
KCNC2	Potassium voltage-gated channel 2 gene
KEGG	Kyoto Encyclopedia of Genes and Genomes
LAR II	Luciferase Assay Reagent II
LB	Lysogeny Broth
LD	Linkage Disequilibrium
LIC	Ligase Independent Cloning
LTP	Long Term Potentiation
MAF	Minor Allele Frequency
MAOA	Monoamine Oxidase A gene

MAOI	Monoamine Oxidase Inhibitors
MAPK	Mitogen-Activated Protein Kinase
MDD	Manic Depressive Disorder
MFE	Minimum Free Energy
$MgCl_2$	Magnesium Chloride
MooDS	Systematic Investigation of the Molecular Causes of Major Mood Disorders and Schizophrenia
MLP	Minus Log *P*
MRI	Magnetic Resonance Imaging
MSP	Multiple Scan Probability
MYO5B	Ca^{2+} sensitive motor gene
MZ	Monozygotic
Na^+	Sodium ion
NCAN	Neurocan gene
NCS	Neuronal Calcium Sensors
NEK7	Never in mitosis-related gene 7
NGS	Next Generation Sequencing
$(NH_4)_2SO_4$	Ammonium Sulphate
NICE	National Institute for Health and Care Excellence
NIMH	National Institute Of Mental Health
NMDA	N-methyl-D-aspartate
NMDAR	N-methyl-D-aspartate receptors
NRGR	NIMH Repository and Genomics Resources
ODZ4	Homolog Of The Drosophila Pair-Rule Gene Ten-M
OPCRIT	Operational Criteria Checklist
OR	Odds Ratio
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PDZ	PSD-95/DLG/ZO-1 Binding motif
PGC	Psychiatric Genomics Consortium
PISA	Proteins, Interfaces, Structures and Assemblies database
PLB	Passive Lysis Buffer
PLC	Phospholipase C
PSD	Post-synaptic density
PSI-BLAST	Position-Specific Iterated Basic Local Alignment Search Tool
$Ptdins4,5P_2$	Phosphatidylinositol 4,5-Bisphosphate
RDC	Research Diagnostic Criteria
RDoC	Research Domain Criteria
RE	Recycling Endosomes
RFLP	Restriction Fragment Length Polymorphisms
RT	Room Temperature
SADS-L	Schizophrenia and Affective Disorder Schedule
SCZ	Schizophrenia
SDC	Sudden Cardiac Death
SEP	Super-Ecliptic pHluorin
SERT	Serotonin Reuptake Transporter
SERCA	Sarcoendoplasmic Recticulum Calcium ATPase

SIFT	Sort Intolerant From Tolerant
SKAT	Sequence Kernel Association Test
SLC6A4	Solute Carrier Family 6 Member 4 gene
SMR	Standardised Mortality Rate
sMRI	Structural Magnetic Resonance
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
SSAGA-I	Semi-Structured Assessment for the Genetics of Alcoholism
SSRI	Selective serotonin reuptake inhibitors
STEP-BD	Systematic Treatment Enhancement Program for Bipolar Disorder
SYN3	Synapsin-3 gene
TARP	Transmembrane AMPA Receptors Regulators
TBE	Tris-Borate-EDTA
TE	Tris-EDTA
TEMED	Tetramethylethylenediamine
TESS	Transcription Element Search System
TF	Transcription Factor
TFBS	Transcription Factor Binding Site
TS	Timothy Syndrome
TSI	Toscani in Italy
TSS	Transcriptional Start site
TSPAN8	Tetraspanin Eight gene
UCSC	University Of Santa Cruz California
UK-COGA	United Kingdom Collaborative Study on the Genetics of Alcoholism
UPD	Unipolar Depression
UTR	Untranslated Region
VDCC	Voltage Dependant Calcium Channels
VPA	Sodium Valproate
VWA	Von Willebrand factor-A domain
WGS	Whole Genome Sequencing
WTCCC	Wellcome Trust Case Control Consortium
YY1	Yin Yang I
1000G	1000 Genomes

Chapter 1 Introduction

1.1 Bipolar Disorder

1.1.1 Overview

Over the past decade there has been an effort to accurately estimate the frequency of mental disorders and their global burden. Calculations by the World Health Organisation (WHO) have identified BD as one of the top ten leading causes of work lost due to mental disability [1]. The European profile of prevention and promotion of mental health study (EuroPoPP-MH) states that there is a three-fold increase in the number of days lost at work if an individual suffers from a mental health disorder compared to a healthy individual [2]. Therefore mental disorders should be considered a high priority health challenge for the 21st century [3, 4]. Wittchen *et al.* state in their 2010 report, entitled 'The size and burden of mental disorders and other disorders of the brain in Europe 2010', that each year 38.2% of adults in the European population (18-65 years old) suffer from a mental health disorder [4]. This places a cost of approximately €798 billion on the European economy [5]. Identification of factors which modulate an individual's disease risk provides a platform from which to better understand individual predisposition, and design specialised and preventative treatments, extending as far as targeting individuals prior to the manifestation of mental health disorders.

Bipolar Affective Disorder (BD) is fundamentally a disorder of mood, with a lifetime prevalence of approximately 1% in the general population [6]. The term BD encompasses the central concept of the disorder, in which there is a fluctuation between euthymic states, mania and depression. Diagnosis of BD depends on the occurrence of two or more episodes of mania or hypomania [7]. The presence of depressive episodes is not a pre-requisite for BD diagnoses. Nevertheless, the majority of people with BD do experience depressive episodes during the course of their illness. The peak age of onset for the disorder is 25, with onset after the age of 18 being considered typical. However, cases of early onset BD (13-18 years of age) and very early onset of below 13 years of age have been reported [8].

Bipolar disorder encompasses a number of different cognitive, behavioural and affective symptoms (Figure 1). Mania is defined as an elevated or irritable mood that is prominent and sustained for a period of one week or more [7]. Symptoms of mania are not restricted to mood swings but encompass cognitive, physiological and behavioural changes. Manic symptoms can include an increase in energy and activity, rapid speech, racing thoughts, impaired judgement, insomnia and feelings of elation and grandiosity. Psychotic symptoms such as grandiose delusions and paranoid ideation are often part of manic episodes. The features of mania can severely affect an individual's ability to function on a daily basis. Hypomania is a less severe form of mania, which lasts for a period of four days or more resulting in some interference to an individual's daily life [7]. Depending on the classification system used, BD can be diagnosed by the occurrence of two or more episodes of mania/hypomania as mentioned previously. However, the usual course of BD is one of recurrent swings of mood between mania and depressive states, broken by periods of euthymia [6]. While mania may be the defining symptom of BD disorder, depression is the concept that is more widely known and understood. Depressive episodes are defined as persistent low moods lasting for a period of more than two weeks. During these episodes individuals may experience a loss of interest or pleasure in normal activities. Symptoms can include inability to sleep, a sense of hopelessness, suicidal ideation, fatigue, reduced self-confidence and social impairment [7]. While this list is not exhaustive, it incorporates the main features of a depressive episode (Figure 1). The full list of the International Classification Diseases-10 (ICD-10) criteria used to diagnose the different types of BD is given in Appendix I.

The presence of psychotic symptoms represents a severe manifestation of BD and these are classified as being either mood congruent or mood incongruent. Mood congruency occurs when the symptoms are consistent with the individual's current mood. Mania with mood congruent psychosis may present itself in symptoms such as grandiose delusions or hearing voices telling the subject that they have special powers. Mania with mood incongruent psychosis can be defined as the delusions not reflecting the manic state, such as hearing voices describing

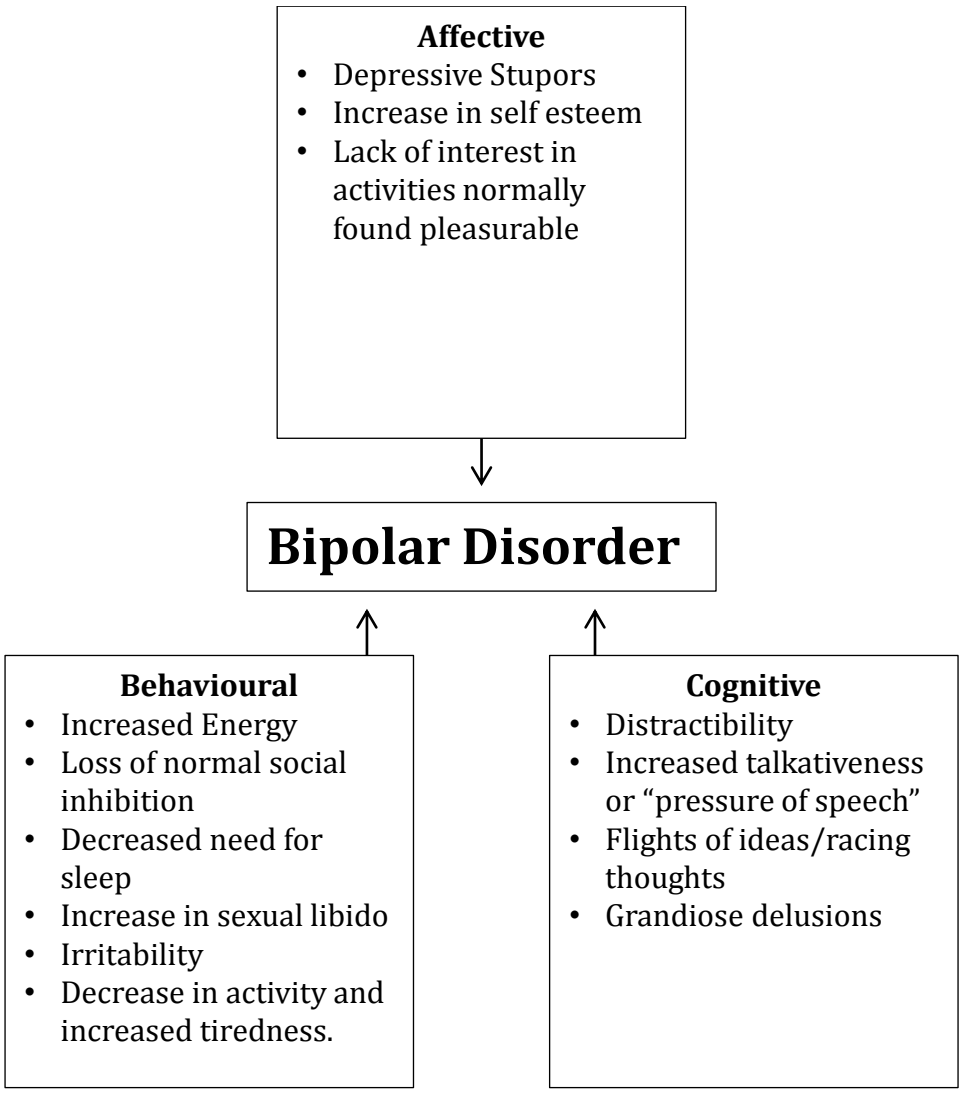


Figure 1. Common symptoms of Bipolar Disorder.

There are three categories of symptoms which occur during the course of BD, these are affective, behavioural and cognitive.

neutral topics or delusions of reference [7]. Examples of mood congruent depressive psychosis are delusions of guilt, nihilistic delusions and delusions of bodily decay, also known as Cotard's delusion [9]. Depressive episodes with mood incongruent psychosis manifest when the hallucinations do not exhibit affective content [7].

Individuals suffering from BD often have comorbid alcohol or substance misuse disorders [10, 11]. It is not known if this association is secondary to the mood state

(ie. a coping mechanism in depression and/or decreased self-control during manic episodes), or due to shared etiological factors (including shared genetic risk).

1.1.2 HISTORY AND CLASSIFICATION

Descriptions of people suffering from psychiatric illnesses, and the familial nature of these disorders date back as far as Hippocrates (460-337 BCE) and the Hippocrates-Galen Humoral Theory [12]. This theory proposed that imbalance in bodily fluids, resulted in specific bodily disease. Imbalance of Hippocrates four 'humours', yellow bile, black bile, phlegm and blood, was proposed as the cause of mood disorders and other illnesses. It can be argued that Hippocrates was the first individual to classify mental health disorders by differentiating the disorders melancholia and mania.

The link between mania and melancholia was first made by the Greek physician Artaeus of Cappadocia (c. 30-90CE). Arteaus was a strong follower of Hippocrates' teachings, and a member of a group called the 'Eclectics' [12]. Eclectics were defined by the multiple approaches they took to defining a disorder from both the observation of the disorder and application of philosophical theories. Artaeus produced very careful descriptions of mental health disorders in his books 'On the Aetiology and Symptomatology of Chronic Diseases' and 'The Treatment of Chronic Diseases' [13]. He described mania and melancholy as two distinct entities of the same disorder, as below.

'' … I think that melancholia is the beginning and a part of mania…The development of a mania is really a worsening of the disease (melancholia) rather than a change into another disease…In most of them (melancholics) the sadness became better after various lengths of time and changed into happiness; the patients then developed a mania''. Aretaeus of Cappadocia (30-90 CE)

Artaeus, in agreement with Hipprocrates, stated that the disorders were biological in nature and were a result of disturbances in the mind and other organs. His classifications, albeit broader than today's classification, were the first model for bipolarity. Artaeus was also the first to make the distinction between depression which was a result of innate biological imbalance, 'melancholia', and depression

which resulted from negative life experiences. In the second century, Alexander of Tralles (525-605 CE) first described the transition that can occur from melancholia to mania and stated that mania is a worse form of melancholia. He also linked melancholia to suicidality and homicidality [14].

Other descriptions of mood disorders can be found in both Chinese and Ayurvedic medical books [15]. One of the most influential books in Ayurvedic medicine; the C*haraka Samhita*, devotes a chapter to the discussion of insanity. In the 12th century BCE a description of *Kuang*, a disease which is characterised by irritable raving can be found. This is later linked to *Dian*, the disease of worry, withdrawal and depression in the medical text *Classics of Difficulties* and *Spiritual Axis* [16].

The delineation of BD as an independent disorder, separate from other depressive disorders and schizophrenia (SCZ), has its origins in the 19th century. In 1854 Jules Baillarger (1809-1890) and Jean Pierre Falret (1794-1870), independently introduced the idea of a "*la folie á double forme"* (madness dual form) or *"folie circular"* (circular madness) to the French Imperial Academy of Medicine in Paris [17]. These terms arose from the observation of patients in asylums who presented with alternating episodes of melancholy and mania. This pattern of variation in patients' moods differentiated their illness quite distinctly from individuals suffering from depression alone. The idea, that mania and depression were two distinct states within one unitary disorder, echoed the ideas of Alexander of Tralles and Artaeus of Cappadoccia described centuries previously.

The differentiation of BD from schizophrenia stems from Emil Kraepelin's (1856-1926) study and classification of psychiatric illnesses. During his lifetime Kraepelin published four editions of the *Compendium der Psychiatrie* in which he proposed that psychiatry was a branch of medical science and the manner by which psychiatric disorders were studied should be akin to the approach taken by the natural sciences. Studying the clinical course and manifestation of BD, Kraepelin noted that, unlike SCZ, the disorder was punctuated by periods of relative normality and symptom free intervals [18]. This led to the formation of what is known as the Kraepelin Dichotomy which denotes BD and SCZ as two separate disorders, each with their own disease progression and symptoms [19]. In 1899, Kraepelin coined the term manic depressive illness. The description of the

different disorders and the manner by which Kraepelin approached the study of such disorders paved the way for modern psychiatry.

In 1953, Karl Kleist, a German neurologist and psychiatrist, coined the terms bipolar and unipolar. Kleist had similar views to Kraepelin and considered both bipolar and unipolar to be part of the same disorder. In 1957 Karl Lenord, Kleist's student, proposed that unipolar and bipolar disorder were in fact separate entities. The definition of BD has continued to evolve and subtypes of BD have been proposed.

Hypomania was originally described in 1881 by Erich Mendel but in the latter part of the 20th century it came to be seen as the defining feature of a milder form of BD. In 1976 Dunner proposed the subdivision of BD and its milder form, Bipolar type II (BD-II) [12].

1.1.3 Diagnostic Criteria

Two main diagnostic classification systems are currently used in psychiatry. These are the ICD-10 from the WHO [7] and the Diagnostic and Statistical Manual of Mental Disorders, 5th edition, (DSM-V) from the American Psychiatric Association (APA) [20]. A third system is currently under development, the Research Domain Criteria from the National Institute of Mental Health (NIMH) [21].

Jacques Bertillion (1851-1922), of the International Statistical Institute, devised a classification system based on causes of death, called the Bertillon Classification of Causes of Death [22]. Bertillion's classification system contained three lists of 44, 99 and 161 conditions upon which the categorisation of causes of deaths and other maladies could be made. This system was adopted by numerous countries for disease and death classification. It became known as the ICD in later years and was revised every ten years. Several changes occurred over these years including the expansion of disease classification. The sixth revision of the ICD document (ICD-6) was undertaken by the WHO and was published in 1948. The ICD-6 was expanded into two books and for the first time contained separate sections for the classifications of mental health disorders. The current edition of the ICD, is the ICD-10 [7].

The second system which is primarily used in North America is the DSM. The DSM was first published in 1952 by the APA and incorporated a number of different classifications from the ICD-6. The DSM-I expanded on the categories established in the ICD-6 and aimed to provide a classification system of clinical utility (http://www.psychiatry.org/practice/dsm/dsm-history-of-the-manual). The following example shows how the classification system has evolved according to the contemporary aetiological theory. The term 'reaction' was used in the DSM-I, as mental health disorders were often thought to arise in response to life events. This term 'reaction' was omitted from the DSM-II, when more biological theories of aetiology held sway. There have been many revisions of the DSM, and the current edition is DSM-V, which was released by the APA in May of 2013. The next publication of the DSM is expected in 2017, in which revision of the classification system based on genetic and neuroscience findings are planned.

In relation to BD, there are differences in the diagnostic criteria between the ICD-10 and the DSM-V. The DSM is a United States based classification system, whereas the ICD is a core function of the WHO which has been ratified by all 193 WHO member countries. DSM-V makes the distinction between BD-I and BD-II whereas no such distinction is seen in the ICD-10. In the DSM-V classification, at least one manic episode is required for the diagnosis of BD-I, whereas at least one episode of hypomania and a major depressive episode is required for a BD-II diagnosis [20]. In ICD-10 a distinction between mania and hypomania is made, but BD is not divided into BD-I and BD-II. ICD-10 tends to be employed in more varied environments than the DSM, which is utilised primarily by professional groups and highly trained individuals [21].

As mentioned previously, the presence of a manic episode is an important criterion for BD diagnosis. Approximately 40% of individuals with BD are initially diagnosed with unipolar depression, as it is quite common for an individual to suffer many depressive episodes prior to their first manic episode [23].

The third classification system is currently under development. Since 2009, the NIMH has been creating a new system to "Develop, for research purposes, new ways of classifying mental disorders based on behavioural dimensions and

neurobiological measures." [21]. This has been termed the Research Domain Criteria (RDoc).

The RDoc will focus on incorporating findings from scientific research in a manner that will allow for a higher degree of sub classification, and less reliance on clinically derived symptom clusters. The RDoc aims to facilitate the elucidation of pathological mechanisms underlying mental disorders. Specifically, the RDoc aims to map disorders to neural circuits for specific behavioural functions [21]. Cuthbert *et al.* set out the four main challenges that need to be addressed to achieve the RDoc.

1. The bringing together of experts from various specialised fields in order to accurately identify "fundamental behavioural components" that underlie mental health disorders.
2. Create methods to determine how far the disruption to these fundamental behavioural components extends, and what can be considered as atypical and what will be considered as pathological.
3. Develop methods to accurately measure these disruptions for clinical diagnoses.
4. Integrate research from a variety of fields including genetics and neuroscience which contribute to these genetic disorders.

The RDoc is not intended to be used, at least initially, for routine clinical diagnoses but rather as a classification system for research. It is the first system to attempt to fully incorporate biological research into classification. The DSM-V has made a move towards doing this, but only to a limited extent [24]. Currently the ICD-10 and DSM-V do not map well onto genetic, system biology or neuroscience findings [21].

Revisions to disease classification systems can have a large impact on patient diagnosis. Given the breadth and heterogeneity of symptoms associated with BD, until the formalisation of the bipolar-unipolar distinction in the DSM-III, many individuals were misdiagnosed and not identified until later in life as suffering from BD [7]. Stoll *et al.*, conducted a study from 1972 until late 1988 which looked at discharge diagnoses for individuals in six psychiatric treatment centres in the

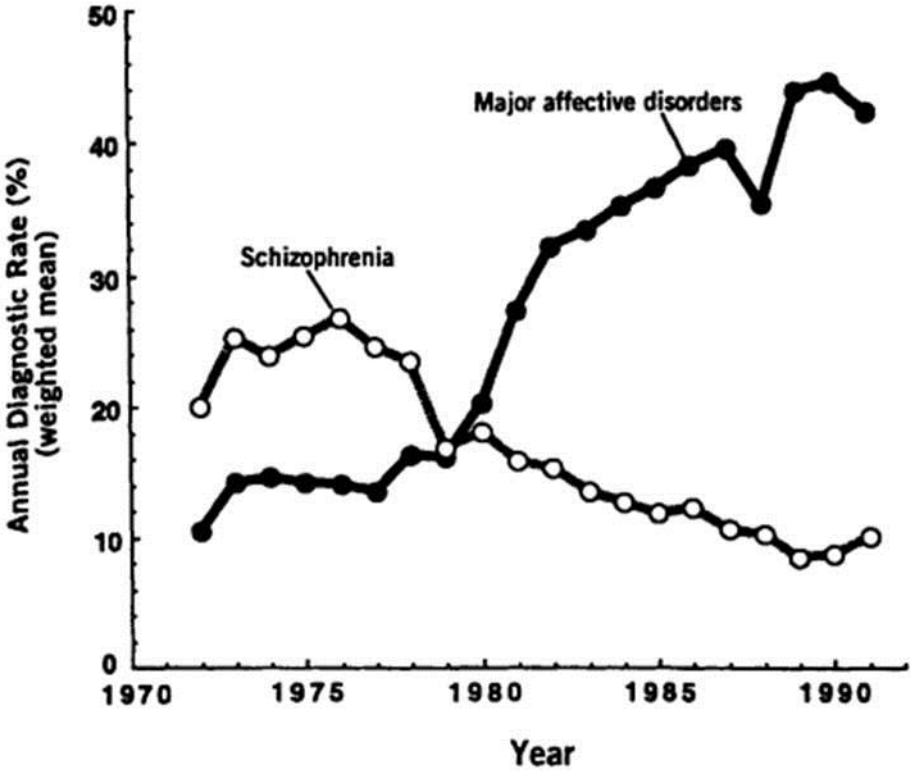


Figure 2. Weighted mean for discharge diagnoses in 6 psychiatric treatment centres in the United States between 1972-1988.

Image taken from [25].

United States [25]. During this period, the DSM-III was released, which contained changes to disorder classifications. In addition to the formalised distinction between unipolar disorder and bipolar disorder, the broad category of schizophrenia was further refined. The results from this study illustrate the impact of the changes to the diagnostic criteria on the incidence of the two disorders (Figure 2). The refining of the SCZ disorder classification led to a drop in the number of individuals diagnosed with SCZ. Conversely, there was a large increase in the incidence of major affective disorders.

The pathology of BD and other psychiatric illnesses have proved elusive. The study of genetic effects in psychiatry is helping to shed light on the underlying mechanisms of the disease. The shift from basing diagnoses on clinical criteria to diagnoses based on underlying biological pathology has the potential to provide more accurate diagnoses for patients.

1.1.4 Bipolar Disorder Subtypes

As discussed in the section above, there are two main subtypes of Bipolar Disorder. These are Bipolar Disorder type 1 (BD-I) and Bipolar Disorder type 2 (BD-II). For an individual to be diagnosed with BD-I under the DSM-V system, they must meet the criteria of presenting with a prominent manic episode sustained for a period of seven days of more. BD-II is characterised by depressive episodes and less extreme periods of mania, referred to as hypomania. These can occur more frequently than manic episodes but are not as severe (Figure 3). Less common subtypes of BD include rapid cycling BD, more than four depressive episodes per year, and mixed affective disorder [26].

Risk of suicide in BD is greatly elevated during depressive periods and has a higher prevalence during disease course in BD-II (24%) than in BD-I (17%) [27]. In 2007, it was reported that whilst the international population average for suicide was 0.017%, the annual number of individuals with BD who attempted suicide was 23-fold higher than this (around 0.4%) [28].

The Standardised Mortality Ratio (SMR) has been used in estimates for BD. SMR is used in epidemiology to measure the increase or decrease in the mortality of a study cohort. The SMR, for suicide alone, measures the ratio of expected numbers of suicides versus observed suicides in the study cohort. SMR for BD is 15.0 (males) and 22.4 (females) for individuals who had been previously hospitalised [29, 30]. A recent meta-analysis, conducted by the International Society for Bipolar Disorders Task Force on Suicide [31], looked at a number of the variables in people with BD who had attempted suicide and died by suicide. Some of the key risk factors identified were depressive polarity of first episode and depressive polarity of last or more recent episode, earlier age of illness onset, and any co-morbid substance abuse. Given the criteria for BD, as defined in the ICD-10, it is likely that there will be a delay in diagnosis if the first mood disturbance is depressive. This lag time between diagnosis of BD and prescription of mood stabilising treatment could be responsible for the increased risk of suicide associated with the initial depressive episode. A study conducted by Angst *et al.*, followed an initial cohort of 406 individuals suffering from BD over a period of 34-38 years in order to assess the SMR of BD [32]. They were followed prospectively for 22 years after which

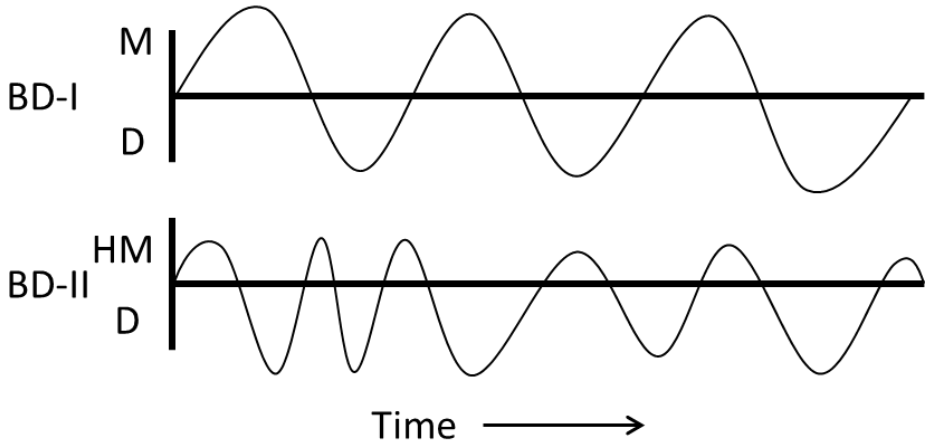


Figure 3. Mood changes in BD-I and BD-II over time.

Fluctuations between moods occurs more frequently in BD-II but has a less severe impact on an individual's daily functioning than the manic episodes in BD-I. M, mania, HM, Hypomania, D, Depression. Image adapted from [33].

mortality was assessed for 99% of the cohort, at this stage 76% of participants had died. The cohort consisted of individuals whom had been hospitalised, representing a group of individuals suffering from a more severe form of BD. These therefore may not be indicative of the general population of individuals with BD. Angst *et al.* showed that patients who received long term treatment were less likely to commit suicide. In addition to this they showed that suicide mortality (SMR=18.04) was similar to the SMR described previously [29, 30].

Depressive episodes and manic episodes both have a marked effect on the quality of life of an individual suffering from BD. Manic episodes often result in the forced hospitalisation of individuals [34]. Depressive episodes have been linked to rates of suicide in individuals with BD which are increased compared to the general population [35]. These results strongly reinforce the idea that delineation of disorder mechanisms and early diagnoses is of paramount importance in the field of psychiatric illness.

1.1.4.1 Disease Aetiology

BD is a complex disorder the aetiology of which is far from being fully delineated. The nature of this complexity holds true for the underpinnings of the disorder. The

aim of this section is to provide a brief summary of our aetiological understanding of BD. I will discuss risk factors and pathological associations from the demographic down to the molecular. Three main topics will be covered in this section, environmental risk factors, neuropathological features and biochemical changes.

There has long been the argument between geneticists and other fields of study about whether the cause of the disorder results more from environmental stresses than genetics. Genetic effects are not fully penetrant as 100% concordance rates are not seen between monozygotic twins [36]. Current evidence points towards a joint effect where some environmental stress factors, combined with genetic variation result in disease manifestation.

1.1.4.2 Environment Risk factors for Bipolar Disorder

Environmental impacts and stresses that occur during an individual's lifetime can have resounding effects on their life. Reported risk factors for BD include, prenatal stress (infection), perinatal stress (hypoxia), childhood maltreatment, traumatic life events and environmental toxins such as drug use.

Bipolar Disorder and SCZ share a number of similarities in terms of cognitive dysfunction [37], anatomical changes [38], and genetic risk factors [39]. Increasingly SCZ is coming to be conceptualised as a neurodevelopmental disorder. Prenatal and perinatal factors have been shown to be associated with an increased risk of developing SCZ [40]. The similarities between BD and SCZ have led to the suggestion that BD may also be considered as a neurodevelopmental disorder [41]. While evidence for the association between BD and prenatal and perinatal stress is less consistent than with SCZ, a number of epidemiological studies proposed that pre- and peri-natal stresses may contribute an increased risk of developing BD. Some of these early life environmental risk factors include season of birth [42], specific obstetric complications [43] and inadequate prenatal nutrition [44]. Consistent evidence from multiple studies has implicated preterm birth as a risk factor for BD [45]. A study conducted by Nosarti *et al.* demonstrated that individuals with a gestational age of between 32 to 36 weeks were 2.7 (95% CI, 1.6-4.5) times more likely to be diagnosed with BD in adulthood and individuals

born prior to 32 weeks were 7.4 (95% CI, 2.7-20.6) times more likely to develop BD compared with term births (37-41 weeks) [46]. Adverse experiences during childhood also comprise an environmental risk factor that is associated with the risk of BD, its severity and the age of illness onset. The most consistent forms of traumas experienced within a cohort of early onset BD are that of physical and sexual abuse. A correlation has been observed between severity of abuse and the complexity of psychopathological manifestation such as lifetime suicide attempt (OR 1.466, 95% 1.095-1.964) and rapid cycling (OR 1.476, 95% CI 1.073-2.030) [47]. Verbal abuse, including that from parents and social peers in the form of bullying has also been implicated in the development and progression of BD [48].

Substance abuse is common in individuals with psychiatric disorders. Alcohol dependence syndrome (ADS) and BD are often co-morbid. While the directionally remains uncertain, study estimates show co-morbid rates of BD and ADS to be between 6-69%, with an average rate of 30% per study [49]. Cassidy *et al.* who studied substance abuse histories of 392 patients hospitalised following manic or mixed episodes of BD, reported comorbidity rates of 14-60% for drug abuse, with a higher rate of substance abuse observed in males with BD versus females. The rate of alcohol abuse was 59.7% in males and 37.8% in females with BD. In a similar manner the rate of drug abuse for males with BD was 54.5% versus 33.8% in females with BD [49]. Substance abuse prior to onset of disease is associated with a higher risk of BD. A bi-directional association between cannabis use and the presence of BD has recently been studied by Feingold *et al.* [50]. They report an association between frequent use of cannabis (from once weekly to daily/almost daily) and incidence of BD, but not a correlation between increased incidence of episodes in BD.

Both environmental risk factors and genetic risks will be mediated through biological substrates which will be discussed in the following sections.

1.1.4.3 NEUROPATHOLOGY

Neuropsychological impairments and alterations to brain structure have been long associated with BD. The onset of BD, peaking after adolescence, closely correlates with brain development during this period [51]. During adolescence brain

development slows and the brain undergoes goes high rates of synaptic pruning. There have been a number of cognitive studies undertaken on BD to determine the extent of neuropsychological impairment on executive function. Cognitive function tests cover seven domains of executive functioning in the brain including processing speed, verbal comprehension, working memory, verbal learning and memory, reasoning and problem solving and verbal comprehension.

Deficits have been reported in individuals with BD in executive functioning, processing speed and memory [52]. Pre-morbid studies on BD cognitive impairment have reported significantly lower effect sizes than seen in SCZ. Effect sizes between -1.0 and 1.5 are typically observed in SCZ, while lower effect sizes in BD are estimated to be between -0.5 and -0.7 [52, 53]. Premorbid testing of childhood cognition demonstrates subtle differences between children who later develop BD and healthy controls. These subtle differences show that approximately 23% of children who later develop BD have neuropsychological impairment, defined as achieving scores below the 10^{th} percentile [54].

In BD, a more severe level of cognitive dysfunction has been found in patients who have a history of psychosis [55]. This level of impairment is closer to the levels exhibited by those individuals with SCZ than BD patients who exhibit no psychotic features. Cognitive impairment in individuals with SCZ is virtually indistinguishable to those found in people with psychotic BD [37, 56]. The similar cognitive signatures in SCZ and BD indicate that in both groups of patients the same underlying cognitive systems are perturbed.

The natural history of cognitive impairment following first manic or depressive episode of BD has also been studied. Emerging evidence suggests that unlike SCZ, where impairment is more severe and pervasive throughout the course of an individual's lifetime, cognitive impairment in BD is less severe. The deterioration is more pronounced in individuals who have a history of psychotic symptoms [57]. Dickerson *et al.* report deficits in overall cognitive abilities and executive functioning in 60 BD patients in the two years following disease onset [58]. Additionally, Zanelli *et al.* [57] describe an observable difference in verbal fluency and verbal memory between BD patients and healthy volunteers with no history of psychiatric illness. Following illness onset there is a marked difference between

the course of cognitive impairment in individuals who have a history of psychosis versus those where psychosis is absent [37].

1.1.4.4 NEUROANATOMICAL

The first volumetric association between brain regions and bipolar disorder came from a report by Rieder *et al.* in 1983 [59]. This computed topographic study showed a significant correlation between age and ventricular to brain ratios in a group of BD patients when compared to individuals with schizophrenia and schizo-affective disorder. MRI studies have shown differences in the size of the lateral ventricle volumes in BD patients compared to healthy controls [60]. The alterations in size also differ in and between age matched BD patients. Those who have had multiple episodes present with increased ventricular volume compared to individuals who have had one episode [61]. Functional magnetic resonance imaging studies have identified structural and functional variation in regions associated with emotional processing in bipolar patients compared to healthy volunteer subjects. Evidence points towards limbic hyperactivation and frontal hypoactivation during emotion processing tasks, principally in manic individuals, but these activation states were present to a lesser degree in euthymic and depressed subjects [60]. Reduction in the volume of the prefrontal and cingulate brain regions have been noted in studies involving youths with BD [62]. Hyperintensities in the deep white matter and subcortical grey matter have been shown to be associated with BD [63]. Hyperintense lesions in magnetic resonance imaging (MRI) studies denote areas of damage in the brain such as axon/nerve loss, focal demyelination and ischaemic damage [63].

A number of meta-analyses have identified regional alterations in brain regions. These include: prefrontal lobe reduction and globus pallidus enlargement. Hallahan *et al.* [64] performed a mega analysis between 321 individuals with BD-I and 442 healthy controls, reporting an inverse correlation between illness duration and total cerebral volume. They further report differences in hippocampal and amygdala volume between patients who were being treated with lithium and unmedicated BD patients, who hadn't taken psychotropic medication for a period of two weeks prior to, and lithium for a period of a month prior to the MRI [65, 66]. An increase in volume in both brain regions in lithium treated

patients was reported. Decreased activity in specific brain regions has also been detected by multiple studies. Reduced activation detected in the rostral and lateral orbiotofrontal cortext subregions has been reported during manic episodes with a corresponding change seen during depressive episodes [67].

Although some of the findings above are well replicated, there is a lack of longitudinal studies to explore the natural history of neurocognitive and neuroanatomical changes in BD. A further drawback of many of the studies is that they only include euthymic or medicated patients for practical reasons. While the study of euthymic BD subjects can elucidate stable trait changes it is difficult to study functional changes associated with manic episodes. The influence of antipsychotics and the drugs used to ameliorate BD episodes impacts cognitive processes and this must be taken into account during studies trying to delineate the underlying mechanisms of the disorder. There is some evidence to suggest that lithium treatment increase brain volumes in individuals with BD [64, 68]. No impact has been seen with other treatments for BD on brain structure.

Neuroimaging studies looking at neural circuits and brain structure can help define clinical outcomes and potentially redefine bipolar disorder diagnostic criteria if implicated appropriately. Cognitive testing is an important tool for examining executive functions in determining the cognitive deficits present during periods of mood disturbances and in euthymic periods [69].

1.1.5 Biochemical pathways implicated in Bipolar Disorder

1.1.5.1 Neurotransmitters

A number of biochemical pathways have been implicated in BD. Given the episodic nature of BD it seems reasonable to look for biochemical changes which correlate with affective changes. Neurochemical and neuronal receptor changes have been reported in individuals with BD. A number of biological processes and pathways that are likely to be involved in BD have been discovered as a result of patient's responses to psychoactive agents. Neurotransmitters such as epinephrine, norepinephrine, serotonin, and dopamine and their pathways have been implicated in the pathology of BD.

In 1972 the hypothesis of a "balanced adrenergic-cholinergic system" underlying different affective states was proposed by Janowsky *et al.* [70]. The central tenet of this hypothesis is that relatively high adrenergic activity results in mania, whereas relatively high cholinergic activity results in depression. Evidence for the involvement of the central nervous system (CNS) acetylcholine neurotransmitter came from observations that a number of cholinergic agonist and acetylcholinesterase inhibitors induced severe depression [71].

At the time of the proposal for a balanced adrenergic-cholinergic system, little was known about dopamine function, composition, or its contribution to the development of mania. There are four major neurocircuits where dopamine neurotransmission predominates. These are the tuberoinfundibularm, the nigrostriatal, the mesocortical, and the mesolimbic pathways [72]. Dopamine signalling is largely mediated through G coupled protein receptors, modulating fast synaptic transmission in glutaminergic and GABAergic neurons [73].

A comprehensive review conducted by Van Enkhuizen in 2014 conclude that there is more evidence for a catecholaminergic–cholinergic balanced hypothesis in BD [74]. Evidence for the involvement of catecholamines (dopamine and norepinephrine) in the etiology of BD has been growing over past decades. Rare increases in both dopamine and norepinephrine, caused by drugs used to treat depression, have been shown to cause a switch in individuals from a state of depression to mania [75]. Generally, antidepressants impact multiple neurotransmitter systems, for example many increase extracellular levels of serotonin in the brain [76]. The method of action for drugs used to treat depression and mania will be discussed in section 1.1.6 of this thesis.

The mood stabilising drugs, lithium and sodium valproate, have an impact on the dopamine signalling pathway. Valproate has been shown to increase dopamine reuptake transporter (DAT) gene expression in a time and dose dependent manner [77] through interaction with the Sp family of transcription factors. Agents such as L-dopa, used in the treatment of Parkinsons disease and a precursor to dopamine, can result in hallucinations. Other dopamine agonists can exacerbate mania [72], further implicating the dopamine pathway in BD. Haloperidol, an antipsychotic

drug used as treatment for BD, decreases dopamergic transmission and is an effective treatment for mania [78].

Lower dopamine type 1 (D_1-type) receptor densities could reflect a higher synaptic dopamine concentration. These findings are consistent with the hyperdopaminergic state hypothesis. Lower concentrations of D_1-type receptor levels has been observed in the frontal cortex of patients with BD in comparison to healthy volunteers as measured by positron emission tomography [79].

Imbalances in both dopaminergic and glutamatergic pathways have been demonstrated post-mortem in the brains of BD patients [80]. Increased glutamate/glutamine ratio and reduced levels of N-methyl-D-aspartate (NMDA) receptor subunits are reported in post-mortem brains from BD patients [81]. Observations of changes in glutamate levels have also been found in the serum, plasma and the CNS of individuals with BD [82]. There is some doubt as to whether these changes reported by multiple groups could be influenced by medication exposure or post-mortem effects on the brain. Studies which have controlled for some of these variables have found increased levels of glutamate in the frontal cortex of patients with BD [83]. Glutamate is a major neurotransmitter in the cortex and mediates fast synaptic excitatory signalling. This evidence suggests a potential role for dysfunction in the glutamatergic neurotransmission in the pathophysiology of BD.

1.1.5.2 Cellular Changes

At a cellular level disruption of calcium (Ca^{2+}) homeostasis and oxidative stress is implicated in the pathophysiology of BD I. Dysregulation of intracellular Ca^{2+} has been shown in lymphocytes and platelets of patients with BD [84]. Increased levels of glutamate may result in excitoxicity following calcium influx [85]. The glutamate pathway is implicated in the regulation of calcium influx and an imbalance in the glutamatergic pathway may lead to calcium dysregulation.

Serotonin or 5-hydroxytryptamine (5-HT) is another neurotransmitter that has been robustly associated with mood disorders. In the brain 5-HT production is found localised in the raphe nuclei. These are moderately sized nuclei that are

located in the brain stem; their axonal projections extend throughout the brain [86].

Studies reportnig reductions in the levels of serotonin have led to a monoamine deficiency hypothesis for the cause of BD [87, 88]. Selective serotonin reuptake inhibitor (SSRI) drugs target the serotonin transporter (SERT) and serotonergic neurotransmission that are widely recognised as playing an important role in depression. The monoamine hypothesis is based on the observation that reduced monoamines neurotransmitters in the CNS result in depression [89]. This hypothesis is based largely on the observation that drugs that increase monoamine levels are effective antidepressants such as the SSRI fluoxetine, or the monoamine oxidase inhibitor tranylcypromine [89]. While more recent discoveries of potent antidepressants have indicated that the hypothesis of monoamine deficiency is not sufficient to explain depression, there is nonetheless a strong body of evidence that suggesting it plays a leading role in regulating mood and depression.

The Ca^{2+} signalling pathway has been implicated in BD as well as SCZ [90-92]. The Ca^{2+} pathway, is a key modulator in many of the processes that regulate neuronal excitability, information processing and cognition. The Ca^{2+} pathway is involved in the regulation of neurotransmitter release from the presynaptic terminals and in a variety of postsynaptic processes. Alterations to these calcium signalling pathways may result in a functional switch in either excitatory or inhibitory neurons. Excitation or inhibition of neurons commences a downstream activation of signalling pathways that modify the level of neuronal activity responsible for controlling brain rhythms [91]. Membrane depolarisation that occurs in response to the activation of these excitatory and inhibitory neurons is also known as the tonic excitatory drive [93]. Alterations to brain rhythms may be connected to physiological processes, such as control of sleep/wake patterns, that are experienced in both depressive and manic episodes of bipolar disorder [94]. As membrane depolarisation occurs sodium (Na^{+}) flows into the cells while potassium (K^{+}) current activity is depressed [95]. An example of the tonic excitatory drive is the hydrolysis of the membrane phospholipid phosphatidylinositol 4,5-bisphosphate (PtdIns4,5P_2), leading to inactivation of K channels responsible for M_1 current [96], which depolarises the membrane to increase neuronal activity. When PtdIns4,5P_2 hydrolysis occurs inositol 1,4,5-triphosphate ($InsP_{3+}$) is formed,

stimulating the opening of Ca^{2+}-activated non-selective cation (CAN) channels. These CAN channels can also be activated by Ca^{2+} flux into the cell through voltage dependent calcium channel (VDCC) genes [93]. Mutations in genes that are members of the VDCC family have been implicated in multiple genetic studies of BD and will be discussed in depth later in this thesis.

A number of other processes and pathways have been implicated in BD. These include inflammation, and the hypothalamic-pituitary-adrenal (HPA) axis. Inflammation has been reported to play a role in both the periphery of the brain and the CNS for BD [97]. Inflammation has also been reported to have a role in the pathophysiology of BD as well as a bidirectional link being postulated for mood symptoms and inflammation [98, 99]. The HPA axis has been shown to be dysfunctional in patients suffering from BD [100]. Molecular abnormalities of the two main brain corticosteroids receptors, glucocorticoid and mineralocorticoid, essential for the HPA function have been found in post-mortem studies of BD brains [101]. Immune responses in the body result in the release of cytokines by macrophages [98]. Cytokine release can have an effect on the HPA-axis as well as microglial activation and brain structure [102].

This brief overview shows that a wide variety of neurochemicals have been implicated in perturbing mood state. While the effect of exogenous pharmacological agents on neurotransmitter systems is often clear, identifying endogenous susceptibility to BD in neurotransmitter pathways has proved to be a challenge.

1.1.6 Treatment of Bipolar Disorder

Currently the main type of treatment for BD is pharmacological. Pharmacological treatments for BD can be divided into two categories. The first is treatments for acute mood disturbances (depression and mania) and the second is prophylactic treatment against both manic and depressive episodes. Often combinations of treatments are used for the treatment of acute phases of illness and for prophylaxis. Several medications are licenced for the treatment of BD and other medicines are prescribed off label.

One major issue in the pharmacological treatment of BD is the significant number of side effects, some of which are both medically and psychologically severe. These range from weight gain, diarrhoea and fatigue, to extrapyramidal symptoms such as parkinsonism, akathisia and tremors [103]. While medications prescribed for BD are effective (eg. Lithium use for BD prophylaxis), these are not completely effective for everyone. A meta-analysis of six studies, focusing on the use of lithium in the treatment of acute mania has reported a response rate of 47% in BD patients versus 32% in controls [104]. Some individuals have a heightened response to the drug than others, while some individuals are treatment resistant for lithium [103]. There are also a number of psychological therapies used in the treatment for BD which will be touched on briefly in this section.

The obstacle to the development of disorder specific pharmaceutical treatments results is our current limited understanding of the pathophysiology of BD. As has been shown in the previous section there are multiple facets to the BD phenotype including mood disturbance, cognitive impairment, psychotic features, and disturbances in various physiological functions.

1.1.6.1 MOOD STABILISERS

Lithium is the first line medication used to treat BD. Lithium has been used for the past 60 years but its mechanism of action is not well understood [105]. There are currently two main hypotheses of lithium action. The first is the "inositol depletion theory" [93] and the second is the "neurogenesis" hypothesis [68]. The inositol depletion theory focuses on the deficits seen in the excitatory and inhibitory neurons, whereas the neurogenesis theory incorporates cellular decline in progenitor cell populations [106].

Acetylcholine acting through through the *muscarine 1* (M1) channel catalyses the hydrolysis of PtdIns4,5P_2 into InsP_3. This hydrolysis leads to the activation of InsP_3 receptors located in the endoplasmic reticulum (ER) releasing Ca^{2+} ions which stimulate the tonic excitatory drive as described in 1.1.5.2 and as shown in Figure 4. InsP_3 is then recycled through the inositol pathway and converted into inositol by IMPase providing a basis for the formation of PtdIns. Lithium impacts on the

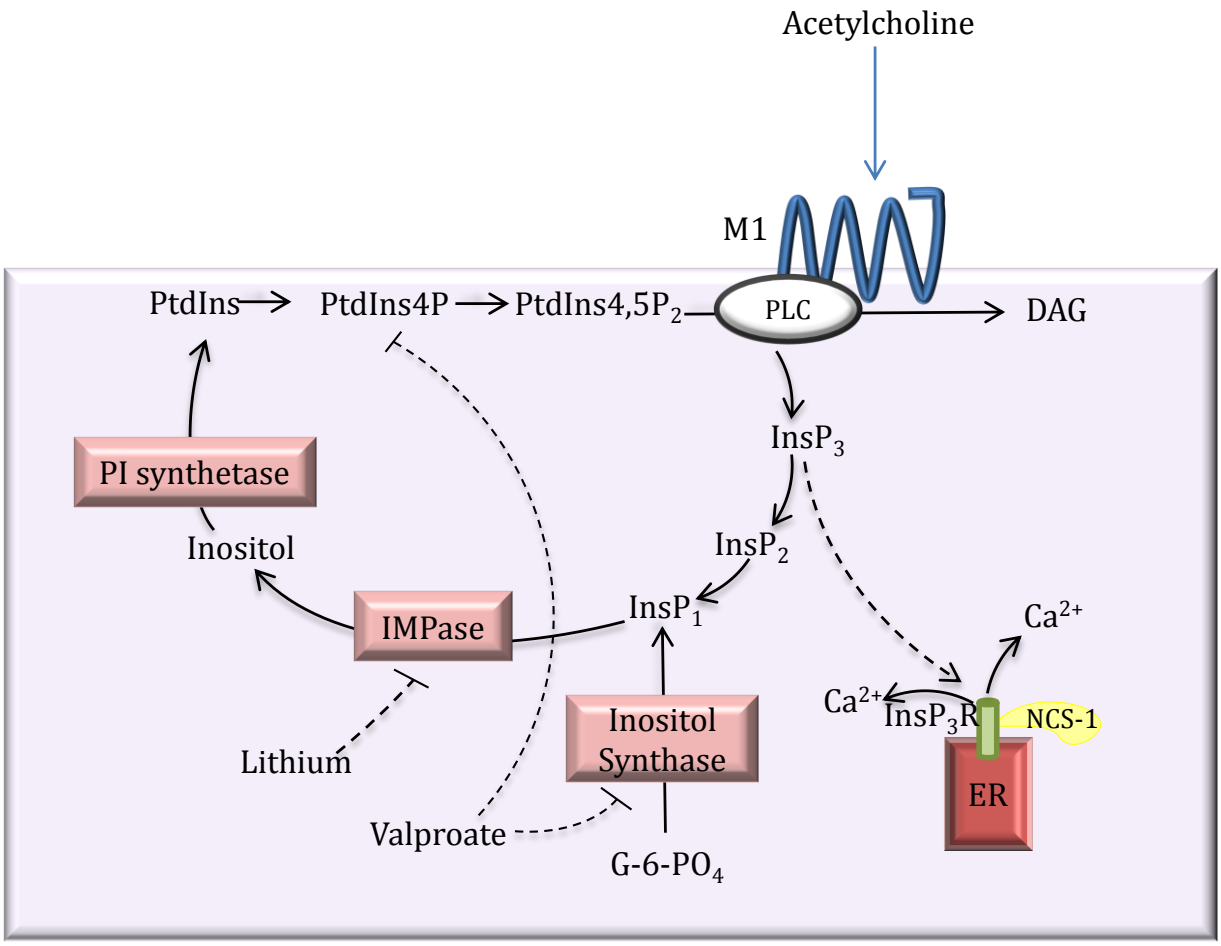


Figure 4. Effect of Lithium and Valproate on the inositol synthesis pathway.

The presence of lithium reduces the supply of inositol and in turn levels of PtdIns4,5P_2. Lithium decreases the amount of PtdIns4,5P_2 present in the system by inhibiting the lithium-sensitive inositol monophosphate (IMPase) which recycles InsP_1 into inositol. This lowers the intracellular levels of Ca^{2+}. Valproate may improve inositol reduction by 1. Lowering the amount of PtdIns4P available to be converted to PtdINS4,5P_2 and then hydrolysed to InsP_3. 2. Inhibiting Glucose-6-Phosphate synthesis into novel InsP_1. Image based on [93].

inositol pathway by blocking IMPase function and reducing the amount of free inositol present for the conversion to PtdIns.

Sodium valproate, originally used as an anticonvulsant, is often used in the course of treatment for BD. This drug is also thought to deplete inositol. Valproate may prevent the production of *de novo* inositol from glucose-6-phosphate (6-G-PO_4) [107, 108]. Valproate use also reduces the amount of PtdIns4P and PtdIns4P,5P_2

[109, 110] which is depicted in Figure 4. This action occurs by blocking of inositol synthase and PtdIns4P.

Lithium and valproate are both thought to increase synaptic plasticity stability, through inositol depletion. Increased synaptic formation, secondary to the reduction of membrane phophoinositides, has been demonstrated in cultured rat hippocampal neuron after four hours of lithium treatment [111]. Lithium's therapeutic benefits are normally seen after one to two weeks when administered for psychosis; however synaptic formation has been previously shown to occur at a faster rate *in vitro* than *in vivo* [112].

Another mechanism of action for lithium has shown to occur through the Glycogen synthase kinase-3β (GSK-3β)/wnt/β-catenin pathway. In 1997 Hedgepeth *et al.* showed that lithium activates the wnt signalling pathway through inhibition of GSK-3 [113]. The activation of the wnt signalling pathway leads to an accumulation of β-catenin protein. This protein is important for activating expression of neuroprotective genes such as *BCL-2* in addition to promoting nuclear transcription of *IGF-I,IGF-II*, *BDNF* and *VEGF* [114]. In the adult and developing brain *IGF-I* acts to promotes neuroplasticity, neurogenesis and angiogenesis [115].

Valproate has also been shown to increase mRNA levels of the molecular chaperone, glucose related protein (GRP78), in C6 glioma cells in culture [116]. GRP78 is an endoplasmic reticulum chaperone protein that regulates intracellular levels of Ca^{2+}. This sequestering of excess Ca^{2+} ions prevents excitoxicity in neurons.

The neurogenesis theory of lithium action has been mainly born out of imaging studies. A meta-analysis conducted by Hafeman *et al.* in 2012 describes the effects of lithium on the brain as measured by structural magnetic resonance imaging (sMRI), function MRI (fMRI) and diffusion tensor imaging (DTI) [68]. Structural MRI studies show that lithium increases or normalises grey matter volumes in several brain regions where deficits were seen prior to treatment, including the amygdala, hippocampus, anterior and subgenal cingulate [68]. Robust correlations have also been observed between the increases in grey matter volume and treatment response [117, 118]. Moore *et al.* reported a significant increase in

prefrontal cortex grey matter in lithium responsive patients compared to non-responders (p=0.003). Long term potentiation (LTP) of CA1 pyramidal cells has also been shown to occur following four weeks of treatment with lithium. This suggests lithium is promoting synaptic plasticity [119]. Lithium has additionally been shown to have a robust effect on neuronal circuits, decreasing neuronal death [120, 121] and facilitating neurogenesis [122, 123]. This mode of action is important as it may facilitate the repair of impaired neuronal communication that have been proposed as an underlying mechanism for BD [124].

Studies have demonstrated that taking lithium reduces suicide risk in BD patients [125]. This is clearly an important outcome of lithium action, given that the annual prevalence of suicide attempt in individuals with BD is 23-fold higher than in the general population (0.017%).

There has been interest in the use of memantine for the treatment of BD. Memantine is one of the four drugs licenced for the treatment of Alzheimer's dementia [126]. Memantine acts primarily on NMDA receptors, antagonising the action of glutamate. Memantine blocks the sensitisation of dopamine D2 receptors in mice, where a cyclic condition mirroring BD had been induced via imipramine administration. Following the discontinuation of imipramine administration many of the animals demonstrated depressive like symptoms. Depressive symptoms were ameliorated by the administration of memantine. Serra *et al.* have reported enhanced mood stabilisation for at least three years in patients who were previously treatment resistant to the standard antipsychotic agents used for BD. Studies suggest that memantine increases hippocampal volume euthymic BD patients [126].

While mania may be the defining characteristic of BD, bipolar depression occurs three times more frequently [127] and causes greater impairment both socially and cognitively [14]. A number of different antidepressants are prescribed, many of which target the biochemical systems which are altered in BD, including SSRIs such as sertraline, monoamine oxidase inhibitors (MAOI), such as moclobemide, noradrenaline reuptake inhibitors including reboxetine and serotonin-noradrenaline reuptake inhibitors such as venlafaxine and duloxetine. SSRI's are some of the most commonly prescribed drugs used to target depression in both BD

and major depressive disorder. SSRIs function by targeting the serotoninergic pathway, activating serotonin transmission and blocking SERT transport of serotonin to receptors [128]. Other common mood stabilising drugs used for the treatment of BD are carbamazepine, gabapentine, and lamotrigine.

However antidepressant treatment brings with it the risk of manic susceptibility therefore antidepressants should be prescribed in combination with an anti-manic agent. Many treatments have anti-manic properties; these include mood-stabilisation treatments discussed above, anti-psychotic drugs, and calcium channel blockers [129]. The three main antipsychotics used in the treatment of BD are olanzapine, haloperidol and risperidone. These have rapid onsets of action and all have been found to be effective in the treatment of mania [130]. Shim *et al.*, investigated the role of olanzapine in synaptic plasticity following a four week course of treatment. Olanzapine was not associated with an increase in LTP in hippocampal CA1 pyramidal cells, but did increase synaptic transmissions.

The UK National Institute for Health and Care Excellence (NICE) publishes guidelines on the assessment and treatment of BD [127]. Lithium is recommended as the first time prophylactic treatment for individuals with BD. Additionally lamotrigine or olanzapine monotherapy is suggested. The antipsychotics, haloperidol, olanzapine, quetiapine or risperidone, are recommended for first line agents in the treatment of mania. For bipolar depression NICE recommends the use of one of the following treatment regimes in combination with lithium: fluoxetine, fluoxetine combined with olanzapine or quetiapine.

Psychological treatments are also recommended for the treatment of BD. These include cognitive behavioural therapy (CBT) [131] and interpersonal therapy (IPT). In addition to cognitive therapy, psychoeducation is recommended for overall management of BD. These psychological sessions are aimed at the education and personalisation of treatment, focusing on identifying early warning symptoms that may be indicators of a relapse into depression. It has been shown that the introduction of coping mechanisms, and plans to follow post treatment in addition to the afore mentioned CBT and IPT strategies confer prophylactic effects on BD relapse, and result in improvements in residual or subsyndromal symptoms [127].

There has been little recent progress in the pharmacological treatment of BD. It is hoped that the elucidation of the specific underlying pathological defect in individuals with BD will eventually enable personalised interventions. Elucidation of the underlying genetic risk factors, alongside work being conducted on neural pathway disruptions is essential to these ends.

1.2 Genetics of Bipolar Disorder

In this section I will focus on the evidence for a major genetic contribution to the aetiology of BD. I will give an overview of the different molecular genetic methodologies that have been utilised to delineate the genetic underpinnings of BD. These will be discussed in five sections covering linkage studies, candidate gene studies, genome-wide association studies (GWAS), pathway analysis and whole genome sequencing studies.

1.2.1 Heritability of Bipolar Disorder

The tendency of both depression and psychosis to run in families has long been observed. Heritability, as a statistical measure, is the proportion of phenotypic variance attributable to genetic variance. It estimates the contribution genetic impact has to a specific phenotype, but provides no information pertaining to the genes involved in a specific phenotype [132]. Lifetime risk in BD, based on heritability estimates, is shown in Figure 5 with respect of the relatedness to a proband with BD. The general population risk of BD ranges from 0.5-1%; the lifetime risk for a first degree relative of an individual with BD is 5-10% higher compared to an unrelated person in the general population [36]. This increase in risk has been observed in numerous studies despite the use of different diagnostic criteria. Concordance rates between monozygotic (MZ) and dizygotic (DZ) twins have been estimated to be 40-70% and 15-20% respectively.

1.2.1.1 Family Studies

Family studies of mood disorders have not always distinguished between unipolar depression (UPD) and BD. The differentiation of the two disorders into separate

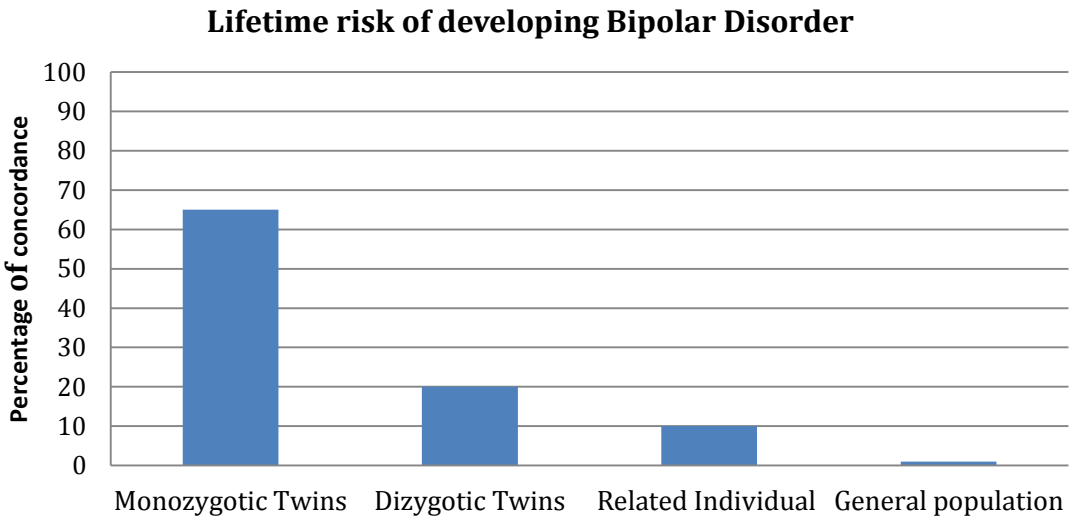


Figure 5. Estimates of the lifetime risk of developing bipolar disorder.

These estimates show the risk of developing BD based on an individual's relatedness to a BD proband. Image based on [133].

entities has only been established in the past 60 years. While studies including both BD and UPD individuals are not specific to disease type, they did show disease aggregation in families [134]. This provides evidence for shared underlying genetic risk factors.

Family studies have reinforced the role of a genetic contribution to BD and other psychiatric illnesses. It is apparent from heritability estimates that the risk of developing a mental illness increases with an increasing relatedness to affected probands with a mental illness diagnoses. Increasing risk of UPD and schizoaffective disorder, has been reported in siblings and family members with BD [135]. The rate at which UPD and schizoaffective disorder segregate in relatives of people affected by a psychiatric illness is twice the rate of the general population [136]. This provides evidence for shared underlying risk factors.

The lifetime risk of developing UPD is estimated to be between 15-20% higher in first-degree relatives of individuals with BD than in the general public. UPD has been proposed to be a more prevalent disease diagnoses in first degree relatives of individuals with BD than any other psychiatric illness, including BD [137].

A recent Swedish study that investigated the relationship of BD and other psychiatric illnesses estimate co-occurrence rates of relative risk between 9.7-22.9 in a total of 54,723 BD patients and their extended families [138]. Heritability for this population in 4,149,748 nuclear families and their offspring is calculated as 58%. In addition to this a study conducted by Lichtenstein *et al.* in 2009 reported on common genetic variants in SCZ and BD from the Swedish multigenerational records. This study included 35,985 first degree relatives of SCZ probands and 40,487 first degree relatives of BD probands from over 2 million nuclear families identified on the Swedish register [139]. The heritability for BD in this study was estimated to be 59%. The co-occurrence between BD and SCZ was 63% adding to the evidence for additive genetic variation which is common between the two disorders. Another study, conducted by Wray and Gottesman, in 2012 showed heritability estimates of 62% for BD in data from >2.6 million individuals from Denmark [140]. This estimate is lower than previously reported from other family studies and could arise through diverse environmental impacts, such as diagnostic interpretation, between national registry data and family and twin studies.

Studies investigating the genetic relationship between BD-I and BD-II have shown no apparent difference in the risk of developing either BD-I (3.6%) or BD-II (3.5%) to the relatives of an individual with BD-I [141]. However, evidence suggests that relatives of BD-II probands are at a greater risk of developing BD-II (6.1%) rather than BD-I (1.8%) [141-143]. The difference between the disorders may suggest a diluted effect of specific genes affecting pathways that lead to full manic episodes. Early onset BD has been reported to have a strong familial link, suggesting that early-onset cases are more genetically homogeneous in nature [144].

In the past shared underlying genetic predisposition in different psychiatric illnesses has often been overlooked but with the development of the Psychiatric Genomics Consortium (PGC) cross disorder group it has become possible to study the shared underlying genetic variation between psychiatric disorders [145, 146]. While each disorder can be classified individually and separately, a common genetic background may result in some of the abnormal phenotypic symptoms which are similar between disorders.

1.2.1.2 TWIN STUDIES

The contribution of genetic and environmental factors on the heritability of a disease or disorder is often measured using concordance rates from twin studies that have used MZ and DZ twin populations. Concordance is usually measured between MZ and DZ twins to determine the level of underlying genetic component as twins often share a common environment and either all of their genes (MZ) or half of their genes (DZ). Measurements of concordance rates are similar to that of lifetime risk estimates. The measurement reflects the presence or absence of the same characteristics or in this case presence of bipolar disorder between individuals. Concordance estimates in MZ twins are higher than those seen in DZ twins [36]. If one assumes that the shared environmental factors are the same between the MZ and DZ sets of twins then these higher levels of concordance rates reflect the actions of genes.

Twin studies in BD have estimated that inherited genetic variation could account for 60-85% of variance in disease risk [4, 5]. Additionally, heritability estimates from twin studies based in the UK, including 67 twin pairs and in a Finnish population based analysis of 19,124 same-sex twin pairs, born from 1940-1957, report heritability estimates of 89% and 93% respectively [6, 7]. BD is estimated to be one of the most heritable psychiatric conditions (Figure 6) second only to autism whose heritability has been estimated as greater than 90% in MZ twins [147].

In relation to the above studies the estimations of concordance rates for MZ and DZ twins in the UK study were 67% and 19% [148] and in the Finnish population was 43% (MZ) and 6% (DZ) [149].

Another form of twin study reporting on concordance rates has focused on the differences observed between the spectrums of the different BD diagnoses. Edvardsen *et al.* (2008) looked at MZ and DZ twin concordance rates between BD-I, BD-II and cyclothymia in 303 individuals from the Danish population [150]. Cyclothymia is viewed as a less severe form of BD where individuals experience low and high mood elevations, but not severe enough to result in hospitalisation. It sometimes can progress into a diagnosis of BD later in life. Higher concordance

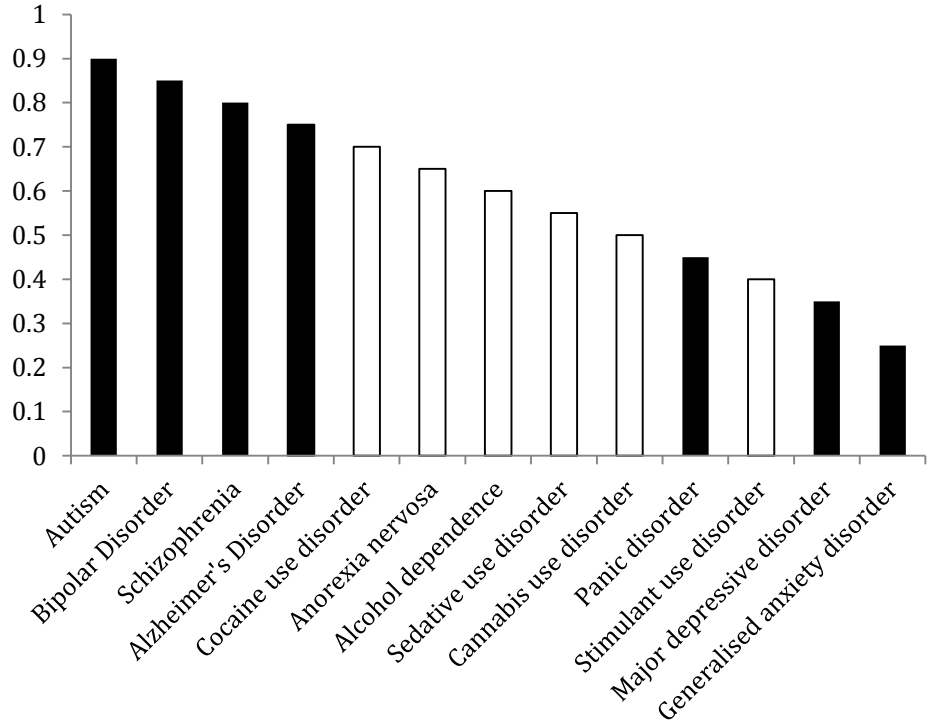


Figure 6. Heritability estimates for a number of Psychiatric Illnesses.

BD has the second highest estimate of heritability amongst psychiatric (0.85) and psychological disorders after autism whose heritability estimates are >0.9. Figure adapted from [151].

rates observed between MZ twins compared to DZ twins followed the trends previously mentioned between BD and other psychiatric illnesses. Half of the MZ twins had either BD-I, BD-II or cyclothymia. This is significantly less than the 10% of DZ twins presenting with one of the disorders (Fishers Exact p=0.047). The difference in MZ and DZ concordance rates and the diminishing levels of relative risk between individuals genetically further away from the proband reflect the true action of genetic influence on BD.

1.2.2 Molecular Genetic Studies of Bipolar Disorder

1.2.2.1 Linkage Studies

Linkage analysis is a statistical method which aims to detect regions of the chromosome which harbour variants that effect the risk of human disease. For Mendelian disorders, linkage analysis has been quite successful in discerning

which genes segregate with disease phenotypes. One of the earliest and most successful examples of linkage analysis of Mendelian disorders was the identification of a *CF* locus on chromosome 7.q31 where a specific deletion of 3 base pairs accounts for 70% of mutations seen in cystic fibrosis patients [152]. During this period the markers utilised for linkage analysis were restriction fragment length polymorphisms (RFLP) and this progressed on to the use of microsatellite markers, repeating sequences of 2-5 base pairs which are found throughout the genome.

Linkage analysis can locate a susceptibility locus to a region of the genome with a resolution of approximately 20-30 centimorgans (cM) [153]. In physical distance this is in the region of 20-30 million base pairs. Fine mapping is then usually employed to investigate the region of interest. Thus linkage analysis was often the first step to gene mapping, and followed by linkage disequilibrium analysis. Linkage provides a powerful tool to detecting variants with an odds ratio (OR) of three or greater [154]. Some of the earliest linkage studies in BD conducted by Reich *et al.* reported linkage of the X chromosome with the disorder. Linkage studies pointed towards the Xq28 region, in individuals with BD [155] using colour blindness, blood group Xg and G-6-PO_4 deficiency as phenotypic markers [156]. Subsequent studies have found evidence for association of the X chromosome with BD from pedigrees in different populations including eight Belgian pedigrees [157], an Iranian pedigree [158], two Sardinian pedigrees [159] and a north American pedigree [160]. However other studies failed to find linkage between BD and the X chromosome [161-163]. Many of the studies published on X linkage and BD were scrutinised over the decades following these publications. There was a serious doubt as to the actual evidence and biases that may have arisen as a result of publishing only pedigrees where X-linked transmission appeared to be observed. Both markers are at a significant difference from each other on the X chromosome. The Xg blood group marker is located at Xq22.3 and the marker for colour blindness is located at Xq28 making the chances of co-linkage unlikely [156]. As is evident from above no one marker appear to be directly linked to BD with conflicting results reported between linkage of markers and BD in different pedigrees.

One of the first linkage analysis studies to use DNA markers was conducted on an old order Amish family, by Egeland *et al.* in 1987. The old order Amish population can be regarded as a defined population in isolate. This linkage study found linkage of BD to chromosome 11p in one pedigree [164]. However the evidence of linkage to 11p disappeared as the family was extended laterally.

Although replicated linkage was found for a few regions, linkage findings in BD have generally been inconsistent. Comparative meta-analyses have been conducted on linkage study results using two different techniques. The meta-analysis approach is a method for combining results from a number of different studies which have similar research methods in an attempt to increase power. Strict inclusion criteria must be used to prevent study bias.

Two meta-analytic methods have been applied to linkage studies in BD. The first method is a multiple scan probability (MSP) method and the second is a genome scan meta-analysis (GSMA). The MSP method modifies a meta-analysis of Fisher to allow for regional *P* values instead of single locus *P* values and was proposed by Gershon and Badner in 2001 as a novel approach for investigating results from linkage studies [165, 166]. Using whole genome linkage data, from eleven published studies, they reported evidence for susceptibility loci at 13q ($p<6x10^{-6}$) and 22q ($p<1x10^{-6}$) for BD [167]. MSP detection focuses on the effect size of a variant in a smaller number of individuals.

The GSMA method was described first in 3 separate papers in 2003. The first publication was used to determine whether GSMA could suitably detect linkage across multiple genome scans [168], the second study focused on meta-analysis of genome-wide linkage studies for SCZ [169], and the third paper focused on BD [170]. Eighteen BD data sets, with greater than 20 affected individuals, were incorporated for this study in an effort to identify linkage association. GSMA was proposed as a method to identify regions which consistently harbour one or more disease related locus.

No significant region of linkage was observed in BD, although some chromosomal regions showed suggestive evidence of linkage (9p22, 10q11.21-22.1 and 14q24.1-32.12) [170]. The use of the GSMA method on data from SCZ linkage studies

identified a significant association on chromosomal region 2q (p<0.000417) [169]. Compared to the SCZ cohort size (1200 pedigrees) the BD cohort was still underpowered (347 pedigrees). This might account for the resulting lack of linkages seen in the BD GSMA study.

The lack of linkage findings is not exclusive to the study of psychiatric genetics but to the field of complex disorders in general. The inability to detect regions of the genome linked with complex disorders may have been arisen from false negative results due to inadequate power, but even linkage studies incorporating large BD datasets proved unable to detect regions conferring relative risk scores of less than 1.5 [154, 171]. Inability to replicate true BD linkage findings may stem from locus heterogeneity. Alternatively, false-positive associations may be reported from studies due to multiple model testing [170]. For complex disorders many of the original linkage studies focused on family pedigrees from populations that were considered to be isolates. This prevents the dilution of linkage evidence that may be experienced if some families have a different gene linked to a common disease phenotype. BD has complex underlying mechanisms and interplay between genetic variants and other risk factors. Linkage analysis has not proved to be extremely effective in delineating the involvement of any one particular gene.

1.2.2.2 Candidate Gene Studies

In the wake of gene linkage studies and the findings of putatively associated chromosomal regions using DNA markers, the focus turned towards identifying genes located in these regions which may harbour variants which influence disease susceptibility. These progressions coincided with advances in technology and development of software which enabled higher throughput sample analyses. Hundreds of samples could be genotyped for one or multiple variants simultaneously, providing an increased production in data available to researchers. Many of the early candidate gene studies focused on genes involved in the pathways for which there was some evidence of disturbance, for example the aforementioned serotonin and dopamine pathways. The original proposed hypotheses of the balanced adrenergic-cholinergic system and a balanced catecholaminergic-cholinergic system, led to a focus on genes which play roles in these systems and other related neurological pathways. In this section I will

provide a summary of some of the main candidate genes which were investigated following linkage studies.

One gene that was investigated in the early days of candidate gene studies was brain derived neurotrophic factor gene (*BDNF*) [172]. *BDNF* is oriented in the reverse direction on chromosome 11p13 and encodes a precursor peptide which, when cleaved, forms the mature protein. It is a member of the neurotrophin superfamily. *BDNF* plays a key role in axonal development, including the promotion and modification of growth alongside the development and survival of neuronal populations [173]. The gene is strongly conserved and only one frequent non-synonymous variant is reported in the gene at codon 66, resulting in a valine to methionine change. Three family based association studies reported overtransmission of the valine amino acid in individuals with BD [174-176]. Four case-control association studies were also conducted which showed no evidence for allelic or genotypic association between BD and this variant [177-180]. A UK case-control association study found no overall association between the *BDNF* variant and BD. However, they reported a significant association between rapid cycling in individuals with BD and the valine amino acid. More recently, reports have stated that individuals with BD have significantly higher circulating levels of BDNF in their plasma compared to controls (p=0.001) [181].

Another gene focused on following the era of linkage studies was *DAOA*. Five independent linkage studies reported evidence for variant association at the *DAOA/G30* locus on chromosome 13q [173]. The gene in this region was named *DAOA* as its protein product was believed to initiate the enzyme D-amino acid oxidase. This region was originally associated with SCZ [182]. Linkage disequilibrium was found in two US families with BD and was replicated in three additional studies [183-186]. No single pathogenic variant was identified in these studies, instead evidence for linkage disequilibrium came from multiple single nucleotide variants and multilocus haplotypes [173]. Genetic variation at two loci in the *DAOA* gene, M23 and M24, have been linked to decreased grey matter volumes in the temporal pole and amygdala areas of the brain [187].

Early linkage studies led to the suggestion that some susceptibility genes are common to both BD and SCZ [188]. A number of linkage studies identified a region

associated with BD on chromosome 22q11 [189-191]. This region contains a microdeletion which incorporates a number of genes [192]. One of the genes in this region was the catechol-o-methyltransderase (*COMT*) gene, which encodes an enzyme involved in the degradation of catecholamines such as dopamine, epinephrine and norepinephrine [193]. Several studies testing the association of a methionine/valine amino acid change with BD reported negative findings [194-196] while others report an association between the methionine and rapid cycling within BD [197, 198]. Studies focusing on associations with variants in *COMT* and SCZ reported three variant allele associations and two haplotype associations [199]. These variants were additionally studied in a BD case-control cohort. Two of the allelic variants and the three haplotypes were associated with BD [200].

The monoamine oxidase A gene (*MAOA*) is located on chromosome Xp11 [201] and encodes a protein which has similar functions to that of *COMT*. Its protein product is involved in the degradation of neurotransmitters ephinpherine, norephinephrine, serotonin and dopamine [202]. The first report of *MAOA* variants associated with psychiatric illness was in 1993. Brunner *et al.* demonstrated that individuals with disturbed monoamine metabolism and X-linked mental retardation had a behavioural phenotype which bore a resemblance to a manic syndrome [203]. Studies of association between the *MAOA* gene and BD have identified three variants in this gene. The first is a dinucleotide CA repeat [204], the second was a variable number tandem repeat [205, 206], and the third variant was a single nucleotide polymorphism [206]. Associations for the three variants have been reported in BD case control cohorts from the UK [207, 208] a Japanese study [209] and an Iranian study [210]. However other candidate gene studies focusing on the *MAOA* gene did not report associations [211-213]

Members of the serotonin receptors (5-HT) have been investigated as a result of findings implicating abnormal serotonergic neurotransmission in BD [214-216]. Allelic variation in the serotonin transporter gene (*5-HTT*) was hypothesised to be involved in the prophylactic response to lithium in BD patients. Currently researchers propose conflicting results as to whether the short form (S) or the long form (L) of the gene show a higher benefit from lithium prophylaxis [217]. The S form of 5-HTTLPR, located in the promoter region of the Solute Carrier Family 6 Member 4 gene (*SLC6A4*), is associated with depression in response to childhood

maltreatment [218]. Individuals who experience childhood maltreatment are likely to develop persistent depressive episodes later in life. The *5-HTTLPR* polymorphism is also associated with cognitive impairment [219].

Other genetic variations which have been investigated in the study of BD are copy number variants (CNV). CNVs are structural genetic variation that are more than 1kb in size and consist of deletions, duplications, insertions or inversion of part of the chromosomes [220]. Large, rare CNVs have been associated with an increased risk of developing SCZ. Given the evidence for a shared underlying genetic architecture between BD and SCZ there have been a number of studies conducted in individuals with BD in an attempt to investigate CNVs. Studies report that the frequency of singleton deletions more than 100kb in length are more common in BD [221, 222]. Additionally, certain subtypes of BD, such as early-onset BD, have a higher frequency of CNVs [223] and of *de novo* CNVs [224]. Most studies conducted on CNVs however suggest that CNVs have a more prominent role in SCZ than in BD, with many researchers reporting no significant frequency difference observed between CNVs in BD and healthy controls [225-227].

As further evidence was produced to support heterogeneity at the level of genetic variation for BD, different methods of statistical approaches were improved upon to account for disease modelling. As was discussed in relation to the meta-analyses conducted on genome-wide linkage data, disease modelling and modes of inheritance have an important role in experimental design. Evidently, if common variation with low effect size played a role in disease manifestation in individuals, then focusing solely on familial pedigrees would not identify these regions. Instead larger samples of unrelated individuals would provide a platform for investigating variation that would have been previously missed due to the limitations of linkage analysis. Candidate gene studies were plagued by an inability to replicate associations. Variant associations in one cohort did not map well on to other population cohorts. Candidate genes studies did not really prove effective until the facilitation of GWAS by high throughput genotyping. Results from GWAS studies will be further discussed in the following section of this thesis, but suffice to say currently numerous candidate genes of interest are emerging.

1.2.2.3 GENOME-WIDE ASSOCIATION STUDIES

The completion of the Human Genome Project (HGP) in 2001 was a turning point for genetic studies. It has opened up the genome for more comprehensive investigations during the past 15 years. Genome-wide association studies (GWAS) genotype hundreds of thousand single nucleotide polymorphisms (SNPs) across the genome in multiple people. GWASs prove a highly attractive approach to the investigation of variance observed in complex disorders. Compared to linkage studies they are capable of detecting multiple loci that have expected smaller effect sizes than would be possible in linkage studies with the same sample size. Additionally, there is a higher probability of defining potential candidate genes located in regions of association than basing candidate gene choice on the potential genes involved in pathogenesis. Similar to the approach taken in whole genome linkage studies, many meta-analyses have been conducted between genotyping data which is present from previously conducted BD GWAS to infer greater evidence for regions of association. GWAS studies are defined as an unbiased, hypothesis free approach to determining regions of the genome which may be involved in genetic risk to disease as there is no predilection for specific genes or candidate regions [228, 229]. To date 18 GWASs in BD have been completed [230-247] and numerous meta-analyses undertaken. The GWASs are outlined in Table 1. Genes involved in ion channels, lithium response and neuronal networks were among the first sets of associated regions reported in BD GWASs.

The first BD GWAS was conducted by Baum *et al.* in 2007. The study implicated a SNP located in the first intron of the diacylglycerol kinase eta (*DGKH)* gene (p=1.5x10^{-8}, OR=1.59). *DGKH* is a gene involved in the lithium sensitive phosphatidyl inositol pathway. This study is an example of a pooled GWAS. Baum *et al.*, genotyped 550,000 SNPs in two separate cohorts and then followed up SNPs of interest with individual genotyping. This proved to be a cost effective manner to identify associated genes.

The Wellcome Trust Case Control Consortium (WTCCC) conducted a GWAS for seven common disorders, of which BD was one [231]. For each of the disorders 2,000 individuals were collected for analysis and a control sample of 3,000 healthy volunteers was used for the study.

Study	Year	Journal	Population	Cases	Controls	Genome-wide significant genes	Implicated genes
Baum	2007	Mol Psychiatry [230]	NIMH and German	461 & 772	563 & 876	_____	*DGKH*
WTCCC	2007	Nature [231]	WTCCC	1,868	3,000	_____	*KCNC2, GABRB1, GRM7, SYN3*
Ferreira	2008	Nature Genetics [232]	STEP-BD & UK & Ireland	4,387	6,209	*CACNA1C, ANK3*	
Sklar	2008	Mol Psychiatry [233]	STEP-BD	1,461	2,008	_____	*MYO5B, TSPAN8, EGFR*
Schulze	2009	Mol Psychiatry [248]	NIMH wave 5 & Meta-Analysis	466	212	*ANK3*	
Smith	2009	Mol Psychiatry [235]	GAIN	1,001 & 345	1,033 & 670	_____	*NAP5, DPY19L3, NTRK2*
Scott	2009	PNAS [236]	NIMH/Pritzker & GSK GWAS	1,177 & 899	722 &904	_____	*CPS1, NEK7, ATP6V1G3*
Hattori	2009	Am J Med Genet B [237]	Japanese	107 & 395	107 & 409	_____	
Djurovic	2010	Jour of Affective Disorders [238]	Norwegian, Icelandic	194 & 435	336 & 10,258	_____	*DLEU2, GUCY1B2, PKIA, CCL2, DPP10, CNTNAP5, FBN1*
PGC-BD	2011	Nature Genetics [239]	PGC-BD	7,481 & 4,496	9,250 & 42,542	*CACNA1C, ODZ4*	
Smith	2011	PLoS Genetics [240]	BiGS	2,191	1,434	_____	
Yosifova	2011	Gene Brain Behav [241]	Bulgarian	188 & 122	376 & 328	_____	*GRIK5, PARD6B, CTSH*
Cichon	2011	AJHG [242]	MooDS (German)	6,030	31,749	*NCAN*	
Lee	2011	Mol Psychiatry [244]	Han Chinese	1,000 & 409	1,000 & 1,000	_____	*SP8, ST8SIA2, KCTD12, CACNB2*
Chen	2013	Mol Psychiatry [245]	meta-analysis	6,658 & 484	8,187 & 1,823	*TRANK1*	
Kuo	2014	Prog. Neuro-Psychopharm [249]	Taiwanese, multistage GWAS	200 & 351	200 & 341	_____	*KCNH7, MYST4, NRXN, SEMA3D*
Mühleisen	2014	Nature Communications [246]	PGC-BD & MooDS	9,747	14,278	*ANK3, ODZ4, TRANK1, ADCY2*	
Xu	2014	BMC Med Genet [247]	Canadian and UK	950	950		*SYNE1 , ZNF659, CNTNAP5, CDH*

Table 1. Summary of all the GWAS studies of BD to date

Summary includes the number of samples involved in each study and lists of genes which reached genome-wide significance or came close to the threshold cut off for whole genome-wide significance. Some of the studies conducted meta-analyses in addition to the original GWAS; if no associations were observed prior to meta-analysis then the results from the meta-analysis are included. NIMH (National Institute of Mental Health), WTCCC (Welcome Trust Case Control Consortium), STEP-BD (Systematic Treatment Enhancement Program for Bipolar Disorder), GAIN (Genetic Association Information Network), GSK (GlaxoSmithKline), BiGS (Bipolar Genome Study), PGC-BD (Psychiatric Genomics Consortium Bipolar Disorder), MooDS (Systematic Investigation of the Molecular Causes of Major Mood Disorders and Schizophrenia)

The highest association signal for BD in the WTCCC GWAS was detected on chromosome 16p12 (p=6.3x10^{-8}). This association disappeared in the expanded reference group analysis (p=5.41x10^{-6}, Supplementary Table 8 [231]) which included individuals from the 1958 birth control cohort, and UK blood services controls. Other genes of interested implicated in the Wellcome Trust GWAS were *KCNC2,* encoding a potassium voltage-gated channel (p=2.2x10^{-7}), *GABRB1,* GABA A receptor β1 subunit responsible for encoding a ligand gated ion channel (p=6.2x10^{-5}), glutamate neurotransmitter receptor *GRM7* (p=9.7x10^{-5}) and finally *SYN3*, encoding synpasin-3 a neuronal phosphoprotein (p=7.2x10^{-5}).

Genome-wide tests look at the correlation between a SNP and a particular disease or trait. A standard GWAS association threshold of p=5x10^{-8} is used to adjust for the multiple tests conducted on the data [250] explaining why the variants from the WTCCC do not classify as genome-wide significant loci in this study. This adjustment is used to control for false positive results. GWASs have the ability to accurately identify variants with an MAF of >5% through improved imputation methodologies, provided the reference panel is matched for population structure [251, 252]. Rare variants are more difficult to identify as they are often not included on the reference panel. In order to address the issue of false negatives arising from GWASs two strategies are used. The first involves the generation of a weight sum of contribution, or polygenic risk score, of variants above nominal significance in a discovery sample and then determining whether they are predictive of phenotype in a second cohort [253]. The second method involves the use of multiple regression to evaluate the genetic variance explained by all SNPs [254].

Sklar *et al.* (2008) conducted a GWAS using the STEP-BD (Edinburgh-Dublin-Systematic Treatment Enhancement Programme 2) and UCL samples and investigated whether previously implicated SNPs from both the WTCCC and the Baum study replicated [233]. No SNP reach genome-wide significance in the study with the highest P-value seen in the *MYO5B* gene, located on chromosome 18, (p= 1.66x10^{-7}, OR=1.51). *MYO5B* encodes a Ca^{2+} sensitive motor which partners with recycling endosomes (REs) upon calcium influx into the cell following the insertion of the α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor (AMPA-R) [255]. It is involved in spine recruitment in neurons during long term potentiation.

Sklar *et al.* did not find an association with the previously reported *CACNA1C* variant rs1006737 (p=3.15x10^{-6}) [231]. However, a number of SNPs in a region of linkage disequilibrium (LD) located in the third intron of the *CACNA1C*, also including the rs1006737 mutation were associated, but not at a genome-wide significant level. Another gene which showed association with BD in the Sklar *et al.* GWAS was tetraspanin eight *TSPAN8* (p=6.11x10^{-7}). Haplotype analysis further strengthened associations with these genes, *MYO5B* p=2.4x10^{-8}, and *TSPAN8* p=7.57x10^{-7}, and also implicated the epidermal growth factor receptor (*EGFR*) p=8.36x10^{-8}.

Ferreira *et al.* conducted a meta-analysis of the WTCCC GWAS, STEP-UCL and ED-DUB-STEP2 in 2008 and reported an association with rs1006737 (p=7.2x10^{-8}) located in the third intron of the *CACNA1C* gene, encoding the L-type α1C subunit of the voltage gated calcium channel gene family. They further reported an association with rs10994336 located in Ankyrin-3 (*ANK3)*, p=9.1×10^{-9}. A meta-analysis conducted by Schluze *et al.* [248] shows association evidence for the previously implicated variant in *ANK3* rs10994336 (p=1.7×10^{-5}, OR=1.54). This meta-analysis contained more data, primarily from the National Institute of Mental Health (NIMH) BD cohort than that included in the Ferreira *et al.* study.

Scott *et al.* (2009) conducted a GWAS on approximately 2,100 bipolar samples and 1,700 control samples all of which had European ancestry from the NIMH repository, Pritzker repository and samples from the UK and Canada which were part of a GlaxoSmithKline sample [236]. Each of the sample cohorts were genotyped using the Illumina HumanHap 550K chip and a fixed effects meta-analysis was conducted. Neither of the individual GWAS studies found associations reaching the genome-wide significant threshold, although three genes showed nominal significance. The first, *CPS1,* encodes a mitrochondrial enzyme involved in synthesis of carbamoyl phosphate (p=7.6x10^{-7}, OR 1.34), the variant rs2813164 (p=8.3x10^{-7}, OR 1.31) is located between *NEK7* (never in mitosis-related gene 7) and *ATP6V1G3* encoding ATPase, H+ Transporting, Lysosomal V1 subunit. To increase the power of this study Scott *et al.* added in genotype data from the WTCCC study mentioned previously. Following quality control they report further evidence for the implicated genes (*CPS1, NEK7* and *ATP6V1G3*) but their analysis of the combined sample data did not reach genome-wide significance. The strongest

association from the meta-analyses was found in an intron of the *MCTP1* gene (p=1.8x10^{-7}, OR=1.21). A number of non-synonymous SNPs in *ITIH1,* encoding the heavy chain 1 of the Inter-α trypsin inhibitor (33 variants p<10$^{=6}$) also shows association in the meta-analysis. Regions on chromosome 1p32.1 and chromosome 3p21 were implicated in this study. The membrane protein encoded by *MCTP1* has a high affinity for binding Ca^{2+} ions in the absence of phospholipids and is highly expressed in the brain [256].

Genes implicated in the first series of GWAS reports and meta-analyses include a number of genes which are involved in neuronal excitability through the formation of ion channels such as *CACNA1C*, *GABRB1*, *KCNC2* and *MCTP1*. Other regions of association contained genes that are in the pathways where lithium may have some effect such as *DGKH*.

It is clear from such studies that combining samples for meta-analysis further benefited the number of results seen and strengthened the evidence for certain genes, while other associations disappeared and did not replicate across sample cohorts. These studies proved more sensitive in detecting variants with small effect sizes than linkage analysis, and some of the early studies showed replicated associations in a number of the same genes. However, given the evidence for the modes of genetic transmissions not enough candidate genes were seen if the hypothesis of polygenic inheritance was to be supported.

In 2011, the Psychiatric Genomics Consortium for Bipolar Disorder (PGC-BD) published results from one of the largest BD GWAS studies to date. They conducted the GWAS on 7,481 BD samples, and 9,250 healthy volunteer samples. SNPs associated with BD at a genome-wide significance level were then tested in an independent sample of 4,496 BD cases and 42,422 controls to determine whether replicated associations with BD could be demonstrated. Further analysis was conducted on the dataset, and a combined GWAS of both BD and SCZ was conducted to determine common underlying associations. One standard quality control method that is used for GWAS studies is genomic control (GC). This method follows the concept that only a small number of variants should show true associations with a phenotype [257-259]. It corrects for the number of false positives which may occur due to population heterogeneity. The genomic control

inflation for the PGC-BD GWAS was calculated to be 1.148 providing evidence that BD will have a large number of underlying variants each with a small effect size. For the primary GWAS the cut off threshold for association was $p \leq 5x10^{-8}$. Imputation was performed to address missing genotypes using the programme BEAGLE [260]. Association of the previously reported *ANK3* variant rs10994397 ($p=7.1\times10^{-9}$, OR= 1.35) was confirmed. A variant in the *SYNE1* gene, located on chromosome 6q25 showed association with BD ($p=4.3x10^{-8}$, OR=1.15). A variant in the *TENM4* gene (encoding a human homolog of the *Drosophila* pair-rule gene *ten-m*), rs12576775 was nominally associated at $p=2.1x10^{-7}$, OR=1.18 as it did not pass the cut off threshold for genome-wide association at $p=5x10^{-8}$. Thirty four SNPs which had a P value of less than 10^{-5} were tested in the second sample to see if their associations followed the same direction of effect. Associations were found with more SNPs than was to be expected by chance. Two SNPs remained genome-wide significant following correction for multiple testing in the replication sample, rs10896135 located in *C11orf80,* and another variant present in the *CACNA1C* gene, rs4765913 ($p=1.60\times10^{-4}$, OR=1.13) which has been shown by Green *et al.* to be in moderate LD with the top GWAS hit rs1006737 [261]. These variants did not reach genome-wide significance in the primary GWAS. Additionally a combined analysis conducted with the PGC-BD cohort and the PGC-SCZ samples reports a stronger association with the *CACNA1C* variant rs4765913 ($p_{raw}=7.7 \times 10^{-8}$) compared to BD alone ($p_{raw}=1.35 \times 10^{-6}$).

In the combined GWAS and replication analysis other members of the L-type family of calcium channel genes showed nominal association. These include *CACNB3* ($p=2.51x10^{-6}$, OR=0.86) encoding the β3 subunit of the VDCC gene family and *CACNA1D,* ($p=3.73x10^{-5}$) which encodes for the 1D α pore of the VDCC family. The study showed the effect that larger sample sizes can have on the number of loci reaching the genome-wide threshold of significance ($5x10^{-8}$) [250]. This suggests that greater GWAS sample numbers are required to identify all the genetic variants contributing to the risk of developing BD.

Since the publication of the 2011 PGC GWAS BD study, there have been three further GWAS publications [242, 245, 246]. Cichon *et al.* report an associated variant, rs1064395, located in the Neurocan *NCAN* gene ($p= 3.02x10^{-8}$, OR=1.31). This finding was replicated in an independent sample. In a combined analysis of

both the replication and original GWAS rs1064395 was found to be associated with BD (p=2.14x10^{-9}, OR=1.17). During early embryonic development *NCAN* is reported to be solely localised to the central nervous system. However evidence from RNA in situ hybridisation studies show expression in the cortical plate and subventricular zone of the neocortex at embryonic days 14.5-16.5 in rats. In the subventricular region transcripts were localised to the caudate region where the neurocan protein may be involved with axon guidance [262]. Expression was also reported in the hippocampus during development. Following birth *NCAN*'s general expression was found to be decreased. Analysis of *NCAN* expression in the human brain reports expression in the hippocampus. These brain regions are known to be involved in cognition and emotional processing which have previously been implicated in BD [41, 101, 111]. Mice who are *NCAN* deficient do not exhibit structural brain changes. However the loss of gene function or variation in the *NCAN* gene may have a more subtle impact such as an effect on LTP which in turn has been reported to be associated with impaired memory [263]. Verbal memory impairments have been reported in euthymic BD individuals [264, 265]

Chen *et al.* conducted a meta-analysis on a combined sample of 14,000 individuals with European and Asian ancestry in 2013 [245]. An association between rs9834970 in *TRANK1* (p=2.4x10^{-11}) and BD was reported [245]. Two additional SNPs on chromosome 2q11 reached the genome-wide significance threshold of p<4x10^{-8}, located near the *LMAN2L* gene. T*RANK1* expression in a number of cell lines (HeLa, SH-SY5Y and HEK293) was also measured in response to commonly used mood stabilizers lithium and valproate. While no expression difference was detected in response to lithium, valproate increased expression of *TRANK1* mRNA levels in all cell lines in both a -time and -dose dependent manner. No alteration in expression was noted when the cells used were exposed to corticosteroids which are reported to provoke mania in BD patients. The variant rs9834970 in *TRANK1* had also previously been reported in the WTCCC (p=4.50x10^{-7}) [231] and a Brazilian family study (p=0.0025) [266].

Mühleisen conducted a GWAS combining unpublished data from the Systematic Investigation of the Molecular Causes of Major Mood Disorders and Schizophrenia (MooDs) PGC data cohort and the previously published PGC-BD samples [246]. The MooDs cohort consisted of 2,266 individuals with BD and 5,028 matched healthy

controls. Associations were found between *ANK3*, *TRANK1, ODZ* and BD, and two new susceptibility loci were also implicated. These were in the *ADCY2* gene (p=9.89 × 10^{-9}, OR=1.14) and a region on chromosome 6q21 between two genes *MIR2113* and *POU3F2*. Only one variant in the previously reported BD susceptibility gene, *CACNA1C*, was reported as being nominally significance (rs4765913 p=9.69 × 10^{-6}, OR=1.12). This variant did not pass the threshold for genome-wide significance in this cohort. The remaining SNPs in the *CACNA1C* were not associated with BD in the MooDs PGC cohort.

Another GWAS study, conducted by Lee *et al.* in a Han Chinese population of BD-I samples, reports a number of suggestive associations. These were in *SP8* (p=4.48x10^{-7}), *ST8SIA2* (p=6.05x10^{-6}), *KCTD12* (p=9.4x10^{-6}). Additionally they reported an association between rs11013860 in *CACNB2,* the β2 subunit of the calcium channel family of genes (rs11013860, p=5.15×10^{-5}) [244]. Nominal association is additionally reported near *ANK3* (rs1938526, p=5.15x10^{-5} and rs11720452, p=1.48x10^{-5}). These results have yet to be replicated in any of the larger GWAS studies. Other studies conducted on different population cohorts have not reported differences between associations seen in individuals with Asian ancestry or European ancestry [245].

A multi stage GWAS was also published in 2013 by Kuo *et al.* [249] starting with a primary study sample of 200 BD-I cases and 200 controls. From this 15 individual variants and eight genes were fine mapped in an extended cohort of 240 samples and 240 controls, followed by replication analysis in 351 cases and 341 controls. No gene reached genome-wide significance and the highest ranking SNP rs7619173 (p=2x10^{-5}) was not in a gene coding region. SNPs located in four other genes showed nominal significance in this data *KCNH7* (p=0.0047), *NRXN* (p=0.0095), *SEMA3D* (p=0.037) and *MYST4* (p=0.0047).

In 2014 Xu *et al.* conducted a GWAS on case and control samples from the UK and Canada. This was supplemented by analysis from 229 Canadian families, each with a BD proband [247]. Certain samples from the UK and Canadian cohorts have been previously included in GWAS meta-analysis reports from the PGC-BD GWAS [239] and the Scott *et al.* GWAS and meta-analysis [236]. In addition to this a subset of individuals from the UK sample were included in the WTCCC GWAS [231]. No

individual genes exceeded the genome-wide significant cut off threshold of 10^{-8}, but a number of genes showed nominal significance in the preliminary data. Fourteen SNPs in *SYNE1* were observed to have a P value of $\leq 10^{-6}$. This however could be due to the fact that the signal originated in the UK cohort that had previously been involved in the WTCCC and the Liu *et al.* meta-analysis between BD and depression so it cannot be described as in an independent signal in this study. A variant located in *ZNF659* was among the top 68 SNPs $p=3.25\text{x}10^{-4}$ and this gene had already been reported as being associated with BD [233].

The design of GWAS in psychiatric illness has evolved in recent years. Instead of conducting GWAS of single disease case-control samples, broader case definitions have been used in an attempt to investigate shared genetic susceptibility. Additionally, information from previous GWAS has been used by research group to target specific candidate genes for sequencing of genes of interest in an effort to determine variants which may lead to the signal seen in the GWAS.

Pereira *et al.* used the UCL data from the GWAS conducted by Sklar *et al.* in 2008 to investigate the *IGF1* gene [243]. The sample consisted of 506 BD-I samples and 510 matched healthy controls. They identified ten SNPs from the 37 studied in the *IGF1* gene as having nominally significant P values. This gene was then sequenced in individuals who had a predefined haplotype, *IGF1* susceptibility HAP, to determine whether any additional variation, potentially missed by the GWAS study, was present. Six additional single base changes were identified in these subjects alongside a CA tandem repeat. These seven variants were screened in an extended cohort sample of 937 individuals with BD and 990 healthy control individuals. These samples included the ones which had been previously undergone GWAS analysis. Of the seven variants found, three were reported to be associated with BD in the UK cohort, rs5742620 ($p=0.0454$), rs78352613 ($p=0.006$) and the CA microsatellite ($p=0.0132$). These variants were not previously covered by the GWAS study. This study provides an example of how focused sequencing of genes implicated through GWAS can identify other relevant variants. *IGF1* mutations may play a role in the neurological deficits that are present in individuals with BD. An *IGF1* knock out mouse was found to have neurodevelopment deficits indicating a role for IGF1 in myelination and axonal growth [267]. These results have not been replicated in any other GWAS to date.

The PGC cross disorder group has conducted a meta-analysis of five psychiatric disorders in an attempt to identify loci with shared effect between Autism Spectrum Disorder (ASD), Attention Deficiency-Hyperactivity Disorder (ADHD), BD, Major Depressive Disorder (MDD) and SCZ. The meta-analysis included a total of 33,332 cases and 27,888 controls [268]. In the primary analysis four loci passed the genome-wide significant threshold. These included two L-type calcium channel gene subunits *CACAN1C* (p=1.87x10^{-8}) and *CACNB2* (p=4·29x10^{-8}, OR=1.08). Using the Bayesian information criterion (BIC) they determined the best-fit model for a variants effect on the five diseases. While *CACNA1C* had previously been associated in MDD, SCZ [269] and BD [270], the BIC model in this meta-analysis favoured a role for *CACNA1C* in BD and SCZ only. *CACNB2* was shown by the BIC model to have a cross disorder effect. The strongest association signal in the meta-analysis came from rs2535629 in an intron of *ITIH3* (p=2.54x10^{-12}, OR 1.07). *ITIH3* has been previously implicated in the PGC Bipolar GWAS in a combined analysis of SCZ and BD samples [236, 270]. The *ITIH3* gene encodes the inter-α-inhibitor heavy chain 3. The second highest association signal was found at rs11191454 in an intron of *AS3MT* (p=1.39x10^{-8}, OR=1.13). The BIC model analyses for *ITIH3* and *AS3MT* also found a shared effect between all five disorders.

The PGC cross disorder group also applied polygenic risk profiling to the five disorders. One set of data was used as the discovery set and the other samples were used as target sets. Polygenic risk score analysis calculates the proportion of variation in the targets set using risk scores from the discovery set [39]. An overlap between the adult onset disorders, MDD, BD and SCZ was identified (P<10^{-5}). The strongest overlap was seen between SCZ and BD.

Curtis *et al.* [271] performed a meta-analysis on a cohort of 506 BD-I, 523 SCZ and 505 controls from the UK described in chapter 3.1.1 of this thesis. An association signal from rs17645023 (p=10^{-6}) was reported. This variant is located between two members of the γ family of subunits which interact with L type family of calcium channel genes, *CACNG4* and *CACNG5*. This variant was not associated in either of the samples when they were analysed separately against healthy control populations. Mutations in the related *CACNG2* gene are associated with epileptic seizures and ataxic gait in mice [272]. These genes encode transmembrane AMPA

receptor regulators which traffic AMPA receptors from the endoplasmic reticulum to the post-synaptic density.

Many of the SNPs associated with BD through GWAS are localised in intronic or intragenic regions making it difficult to determine their functional effect. Studies focusing on identifying and annotating functional and regulatory regions in the genome have found that only 4.6% of variants associated in GWAS studies are located in coding regions [273]. For example one of the variant most strongly associated with BD is located in the third intron of *CACNA1C* gene. It may be that the most strongly associated variants are not the causative ones but rather they are in strong LD with causative variants.

In addition to conducting a meta-analysis Chen *et al.* also investigated the potential risk loci that may be still be discoverable through GWAS using a method devised by Park *et al.* [274]. This method utilises previously published GWAS to determine the power and risk prediction of future studies. They estimate that there is a minimum of 105 additional susceptibility loci that can be determined for BD through GWAS studies but for this to happen the number of samples involved would need to be greater than 63,000 [245].

The most recent SCZ GWAS, conducted by the PGC, has identified 108 regions of association across the genome harbouring 128 independent associations in a large sample of 37,000 cases and 113,000 controls [275]. This is compared to the results from the original PGC mega analysis GWAS which identified 22 independent loci, 13 of which were new, when the sample sizes were approximately 14,000 and 18,500 cases and controls respectively [276].

This increase in sample numbers correlating to the number of resulting associated loci is depicted in Figure 7 from a review by Crowley and Sakamoto [277], looking into the advances from psychiatric genetics and the predicted outcomes for the next 5 years. It is hoped that an increasing jump in the number of associated loci for BD will follow the trend observed in SCZ. GWAS studies have proved successful in the identification of risk variants of small effect sizes which are associated with BD. A number of new technologies and methods have been developed to investigate the contribution rare variants may have on disease risk.

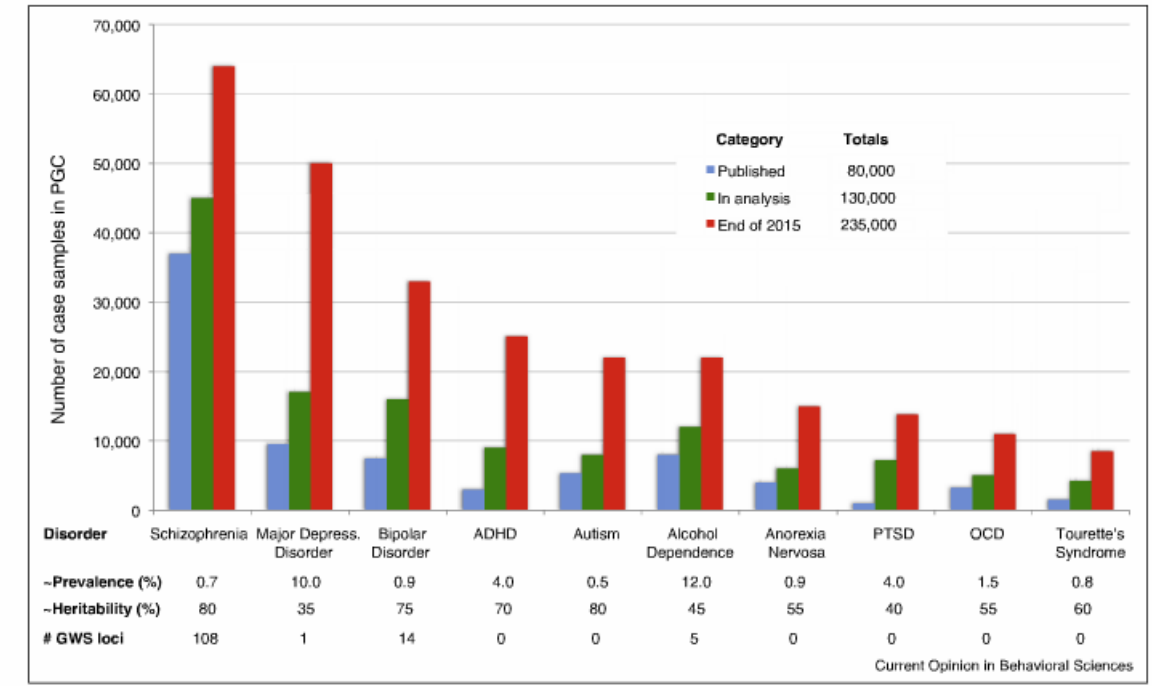


Figure 7. Relationship of associated loci in psychiatric illness and sample size.

Progression of outcomes in genetic hits for a range of psychiatric illnesses compared to increases in sample sizes. Image taken from [277]

The many constraints on large sample collection has meant that alternative pragmatic approaches to analysing GWAS data are beginning to be employed by researchers in the field of psychiatric genetics. Additionally, the data may harbour further associations which are not obvious following the primary GWAS analysis. The most consistent manner in approaching these large sets of data has been to begin to look at the data in a new manner and to apply different statistical tests to the data. Variants of small effect sizes may contribute to disease risk by the accumulation of multiple variants in a particular pathway or gene family. Investigation of the function of these variants through pathway analysis can provide further insight into the underlying genetic mechanisms of BD.

1.2.2.4 Pathway Analysis

As the number of BD and control samples for which GWAS data is available has increase it has become possible to explore genetic associations at a pathway level in a meaningful way.

Pathways analysis was conducted on the data from the PGC-BD GWAS of 7,481 BD cases and 9,250 healthy controls [239]. Using the top 34 independent SNPs from the GWAS study (p>5x10^{-5}) the PGC-BD group performed an analysis to determine whether there was an enrichment of any gene ontology (GO) pathways in the genes where these variants were located. The used an interval-based enrichment analysis (INRICH) which tests for an enrichment or burden of associated variants amongst predefined gene sets. They employed a permutation based approach to control for differences amongst genes. Controlling for gene size, SNP density and gene density is essential to avoid introducing bias to the analysis. The analysis was conducted by binning genes based on size and SNP density. An enrichment of nominally associated SNPs was found in GO:0015270 which contains eight genes, three of which belong to the L-type calcium channel family, *CACNA1D*, *CACNA1C*, and *CACNB2* (p=2x10^{-5}) [239].

INRICH pathway analysis has also been conducted on the data from the PGC five disorder meta-analysis and reported further evidence for pathways involving calcium channel genes. Associations in the calcium channel genes originated from the BD GWAS, however in the cross disorder group the removal of the BD cohort did not result in the calcium channel gene enrichment disappearing. This suggests that this family of voltage gated ion channels has an underlying role in a number of psychiatric illnesses [268].

INRICH is not the only pathway analysis software available to conduct gene enrichment analysis. Other methods include FORGE, ALIGATOR, SET-SCREEN and MAGENTA. These were compared in a recent study investigating pathways in psychiatric illness [278]. It is to be assumed that if one pathway is robustly associated using different methods it is more likely to signify a true result. A combined q value was devised to collate the p values from the different pathway analysis software. The q value is taken from the false discovery rate, where the q-value of a test measures the proportion of false positives incurred when the test is called significant. Each of the gene sets were ranked by this value. The pathways were defined using five different databases GO, Kyoto Encyclopedia of Genes and Genomes (KEGG), Panther, Reactome, TargetScan. This study implicated a number of different pathways including histone methylation marks, neuronal signalling pathways and with genes involved at the postsynaptic density and the immune

system. The pathway most associated across adult psychiatric disorders (SCZ, MDD, and BD) was GO:51568, histone H3-H4 methylation (combined $p=5.75\times10^{-8}$, $q=0.0003$). The second strongest association was seen a pathway for histone methylation, GO=16571 ($p=1.46\text{x}10^{-5}$, $q=0.0362$). It is not strange that pathways involved in gene regulation and translation may be involved in BD or other psychiatric illnesses. Recent assessment of the results from a number of GWAS studies has shown an abundance of enrichment with SNPs in the promoter, 5'UTR and the 3'UTR [279]. This may help to explain the GWAS association signals that occur outside protein coding regions. A nominal association for the calcium channel activity pathway GO:005262 was also seen amongst all five disorders ($p=3.13\text{x}10^{-3}$) [278].

Pathway analysis was conducted on the GWAS data used in the Xu *et al.* publication [247]. The pathway analysis in the primary data set was performed based on a method devised by Beyene *et al.* [280]. SNPs that had a nominal association value of $p<0.01$ were chosen for pathway enrichment analysis; this included 5,111 SNPs from the discovery cohort in the study which mapped to 2,155 genes. These were the closest genes to each variant as defined using the Illumina SNP annotation file. The number of variants in each gene was used to calculate an aggregate summary measure [280]. The authors calculated the statistical significance of predefined pathway with Ingenuity Pathway Analysis software (IPA, version 11904312) using the aggregated summary measure. This used the Fisher's exact test to calculate the level of association reported based on the number of genes annotated, the number of genes studied and representatives from the input data and the entire set of genes being assessed in the analysis. The threshold for association was set at $p\leq0.05$ for pathway enrichment. Thirty pathways were found to be significantly enriched in the data presented in this study, a number of which show support for roles of genes already implicated through GWAS such as calcium signalling ($P_{adj}=0.003$), glutamate receptor signalling ($P_{adj}=0.003$), axonal guidance signalling ($P_{adj}=0.003$), cyclic-AMP- response element (CRE)-binding protein (CREB) signalling in Neurons ($P_{adj}=0.0161$). Other pathways implicated may reflect alternative disorders present in the cohorts, or other underlying mechanisms which may play small roles in BD but have not yet been implicated through GWAS studies such as the "molecular mechanisms of cancer" pathway ($p=0.003$).

Current results from pathway analyses have implicated pathways that would be expected to be involved in BD given previous evidence from GWAS studies. These include the calcium signalling pathway. Members of this pathway have been implicated through numerous GWAS. Other pathways of interest which could further our understanding of BD, which have not previously been implicated through GWAS or linkage studies, are the histone methylation pathways. These pathways are highly involved in the regulation of gene expression and transcription. In addition to these results, these analyses have highlighted pathways previously not implicated in BD such as cancer pathways which appear to have no direct impact on the phenotypes observed in individuals with BD but genes in this category may have an impact on signal cascade pathways.

1.2.2.5 Next Generation sequencing in Bipolar Disorder

Next generation sequencing (NGS) is essential for all fields of biological research. In the past decade it has become increasingly more affordable for researchers and collaborations to invest in both whole exome sequencing and whole genome sequencing (WGS). Sanger sequencing has been utilised to investigate genetic variation in the genome using capillary electrophoresis, however, this method is not suited to high through put screening of the genome. NGS works in a highly parallel fashion across the genome concurrently. Genomic DNA is cut into small uniform DNA fragments which allow parallel sequencing reactions to occur (Figure 8). Each of the small fragments of DNA is screened and bases identified from the signal emitted as each fragment is resynthesized. These are then aligned with either a reference genome or they undergo novel reconstruction to produce *de novo* sequence alignments. All of the aligned reads allows for the genomic sequence to be read accurately. Most genomic bases are read more than once, and this is described as sequencing read depth, or coverage. High coverage can be used to determine whether the base that is observed at a specific position in the genome has been accurately called and to determine whether the call is likely to be accurate.

Figure 9 shows a timeline, taken from a review by Lehner, Senthil and Addington [281], marking the progress that has occurred in the field of psychiatric genetics during the past 26 years. The number of biosamples stored by the NIMH

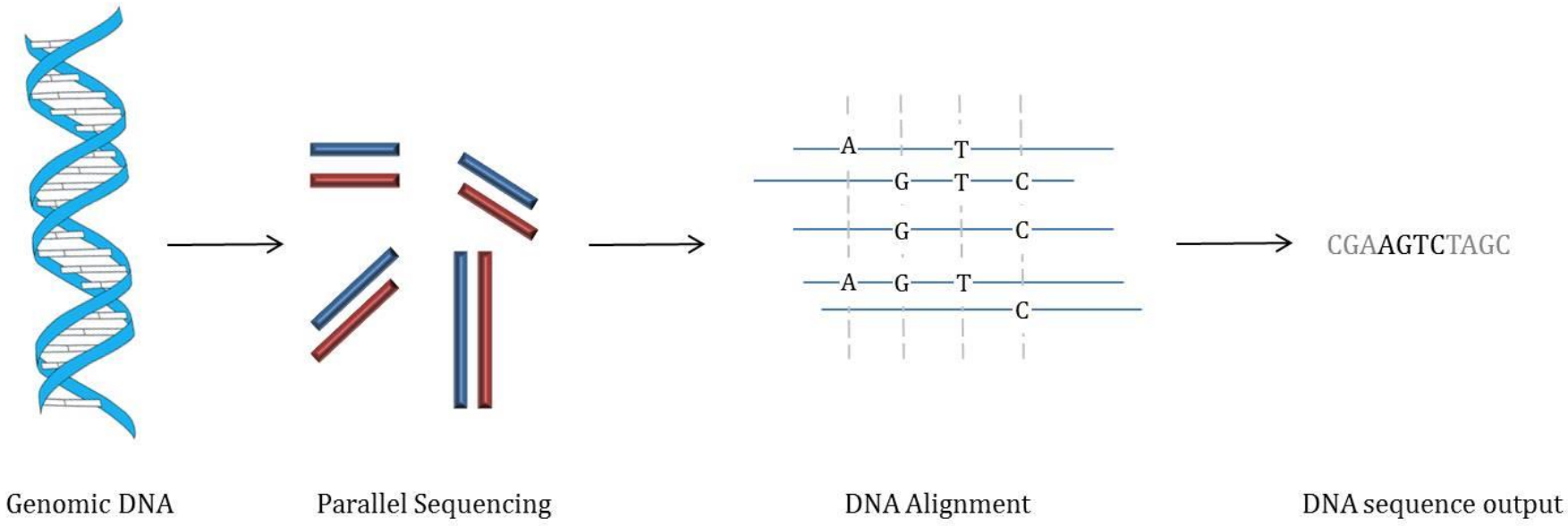


Figure 8. Concept of Next Generation Sequencing through parallel sequencing of genomic DNA.

Genomic DNA is fragmented into short reads which facilitates the concurrent parallel sequencing reactions. DNA alignment shows how sections are read multiple times and the DNA sequence is read from the base which occurs in the majority of reads. In some reads where a common SNP may be present this will show in the vcf output file from the sequencing reaction. These sequences are then aligned to produce an entire whole genome sequencing output file.

Repository and Genomics Resources (NRGR) shows an accompanying increase in the technology available to analyse the data and to handle the samples (green circles). There are also an increasing number of policies which have been created to promote the sharing of data and depositing both human and non-human genomic data in repositories (Orange). There are two different sample repositories currently available for depositing such data, the NRGR and the Database of Genotypes and Phenotypes (dbGAP) (Purple Circles). Additionally in Europe there is a repository called the European Genome-phenome archive (EGA) that is a repository for the WTCCC data, which is not shown in this figure (https://www.ebi.ac.uk/ega/home). There are a growing number of large-scale collaborations that have greatly facilitated genetic research into psychiatric illnesses, providing additional samples and resources for a combined approach to understanding genetics. This originally started with the HGP, aimed at producing a reference human genome and was completed in 2001. The establishment of the PGC has been previously discussed in this thesis and has had a previously unseen impact on delineating genetic underpinnings in psychiatric illness. The Autism Sequencing Collaboration (ASC) formed in early 2010 and involves over 20 research groups working together to exome sequence 20,000 Autism patients (Light Blue circles).

To date there have seven next generation sequencing studies conducted in cohorts of patients with BD (Table 2). Four of these were exome sequencing studies conducted on BD case control samples and family pedigrees [282-285]. In addition to this two whole genome sequencing studies have been published [286, 287]. Two of these studies have been conducted on the large Amish pedigree that was originally used to investigate linkage in BD [164] (see section 1.2.2.1).

There are advantages and disadvantages associated with both whole genome and exome sequencing. The deciding factor for many researchers arises from cost benefit analysis; exome sequencing is cheaper so a greater number of samples can be sequenced with a higher depth of coverage achieved for the same price as a fewer number of samples and lower coverage at a whole genome level. With GWAS studies pushing the mind-sets of researchers to the “greater number of samples coincides with increased number of loci” resolve, it is understandable why this appears an attractive option.

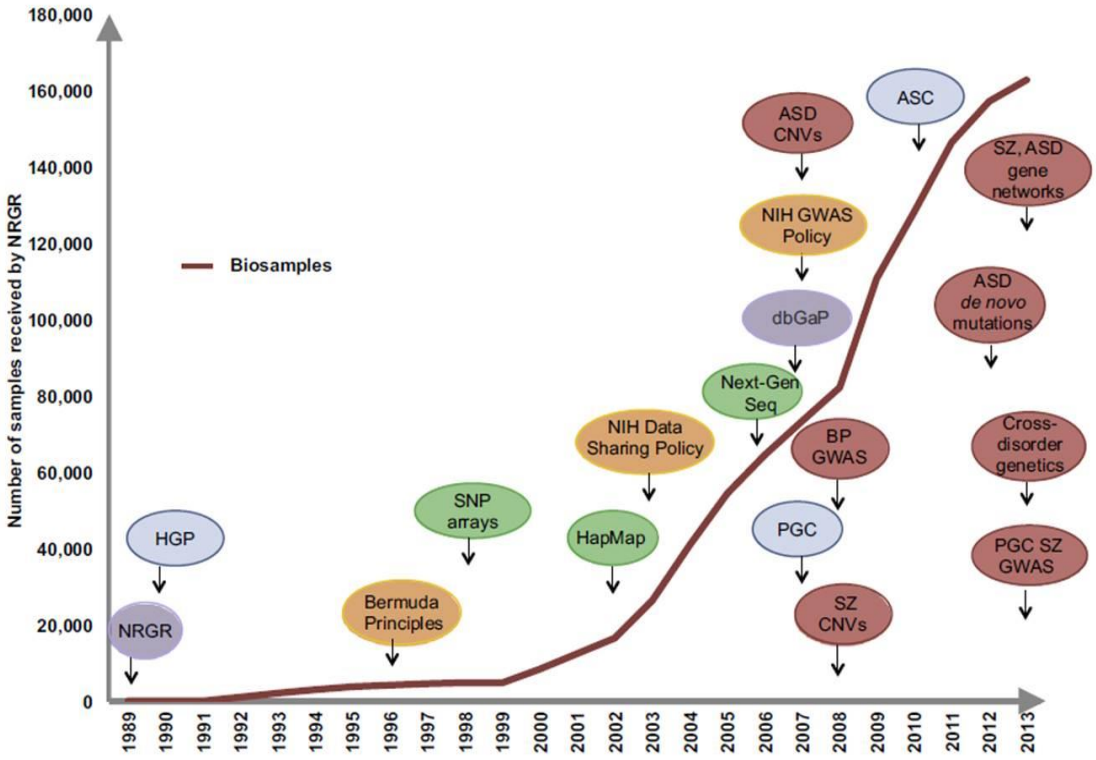


Figure 9 Timeline of progress in psychiatric genetics from 1989-2013.

This graph depicts advances that have been made in the field of psychiatric genetic studies since 1989. New technologies, such as SNP arrays, are depicted by green circles. Sample repositories such as dbGAP and the NIMH Repository and Genomics Resource (NRGR) are depicted in purple. Data sharing policies which have been implemented over the past 26 years are denoted by orange circles. Large scale collaborations such as the Psychiatric Genomics Consortium (PGC) and the Human Genome Project (HGP) are denoted by light blue circles. Finally pivotal findings such as *de novo* variation in autism spectrum disorder and the first BD GWAS are denoted by red circles. Image [281]

Study	Year	Journal	Population	Study Design	NGS	N		Families	Genes Implicated
						Cases	Controls		
Chen	2013	PLoS Genetics [282]	Unknown	Case-Control	Exome	191	107	_____	Seven genes sets nominally associated
Cruceanu	2013	Genome [283]	Canadian	Large Pedigree	Exome			250 samples from 25 families	*ZNF259*,segregation of a rare allele among affected individuals
Kerner	2013	Front. Psychiatry [284]	NIMH*	Single Family	Exome	3	1	One family consisting 3 affected daughters and 1 unaffected son	*IQUB*, *JMJD1C*, *GADD45A*, *GOLGB1*, *PLSCR5*, *VRK2*, *MESDC2*, *FGGY*
Georgi	2014	PLoS Genetics [286]	ASMAD**	Large Pedigree	Whole Genome Sequencing combined with Linkage analysis	50		Originating from one pedigree with 497 individuals	Some rare deleterious mutations found but these may be results of founder effects
Strauss	2014	Hum. Mol. Genetics [285]	ASMAD	Large Pedigree	Exome	7		Originating from four families	*KCNH7*
Fiorentino	2014	Bipolar Disorders [287]	UK UCL	Case Only	Whole Genome	99		_____	*CACNA1C*, *ANK3*
Ament	2015	PNAS	European	Large Pedigree	Whole Genome combined with resequencing and genotyping	200		Originating from 41 families	*GABRA6*, *ANK3*, *CACNA1B*, *CACNA1C*, *CACNA1D*, *CACNG2*, *CAMK2A*, *NGF*

Table 2. Next generation sequencing studies conducted on Bipolar Disorder.

*National Institute of Mental Health sample set. **Amish Study of Major Affective Disorders. For some of the samples additional analysis was conducted following whole genome or whole exome sequencing.

On the flip side of the coin, an ability to investigate genetic variation present in all regions of the genome is potentially a bigger advantage given that GWAS points to numerous variants located in different regions outside of the classically defined protein coding regions of genes.

In 2013 Chen *et al.* published a case control study focusing on exome sequencing from 191 BD cases and 207 controls [282] . In part, the main goal of their study was to investigate different methods of statistical analysis that can be employed to find variation present in exome sequence data for case control analyses. Using the Burden or Mutation point test (BOMP) method they searched for functional variants in a specific genomic region. This specific defined region can be in either a single gene, in a set of genes, in a defined region of a chromosome, or in a pathway of genes. The method is designed to allow for the inclusion of weightings for variants calculated by their bioinformatically determined functional impact. The authors identifyed pathways that were enriched for BD related variants. Seven pathways were nominally significantly associated with BD. The *MAPK* pathway was the most associated and after correction for multiple testing showed association with BD (p=0.0065). Of the 263 genes in the MAPK pathway, a number have been previously implicated in BD such as *CACNA1C*. The MAPK pathway is highly conserved and is thought to be involved in synaptic plasticity and cellular proliferation, migration and differentiation [288]. Additionally both lithium and sodium valproate activate the *MAPK* signalling cascade pathway, which may be involved in the anti-manic effects of these compounds [289].

Cruceanu *et al.* published an exome sequencing study that highlighted a rare missense mutation in *ZNF259* which appears to be segregating in affected individuals in one pedigree [283]. The sample population included individuals where there was a high incidence of BD in the family pedigree and where probands showed good response to lithium monotherapy. Aside from the *ZNF259* finding there were no other striking results or association seen in the sample

Conflicting results were seen in the exome data published by Kerner *et al.* [284]. They investigated and sequenced a nuclear family in which three daughters were affected with BD and the fourth sibling was unaffected. There were no previous reports of psychiatric illness in this family on either the paternal or maternal side. Rare heterozygous mutations were found in eight brain expressed genes, which

occurred in protein coding regions of the gene, in the three individuals with BD but were absent in the remaining family members and 200 additional control individuals screened for the mutations in the general population. Of the eight mutations found, five were novel mutations and three had been previously described in some international studies but still were rare in healthy controls. All were rare variants and the genes in which they arose play a role in the both the ERK/MAPK and CREB cycle pathway which regulates gene expression and has been show to play a role in neuronal plasticity [290, 291]. While this study had only had three probands it was able to determine shared rare genetic variation. Additionally these variants were absent in unrelated individuals with the same ancestral background and in a large control population from an additional study.

Two groups have focused on the Amish Study of Major Affective Disorders (ASMAD) cohort. Strauss *et al.* conducted exome sequencing in seven out of fourteen patients from four different Amish family pedigrees. Ten of the variants in this study were chosen for association analysis following sequencing. One variant, present in a potassium channel gene, *KCNH7*, was found in all fourteen of the individuals with exome sequencing data and was also associated in the wider pedigree. As previously discussed, mutations in potassium channels and ion channels have been implicated in the aetiology of BD.

Georgi *et al.* conducted WGS on 50 individuals, including 23 with BD, from a single large Amish pedigree. They used linkage analysis to prioritise variants to genotype in 488 members of the extended pedigree. Several genomic regions showed a trend towards significance but none of these findings survived multiple correction.

Fiorentino *et al.* conducted WGS on 99 BD individuals. This study was conducted in the Molecular Psychiatry lab in UCL and the results will be discussed in further detail in Chapter 6.2 [287].

Most recently, Ament *et al.* [292] conducted a next generation sequencing study on 200 individuals with BD from 41 multiple affected BD pedigrees of European ancestry. The focus of their study was elucidating rare variants that may confer an increased susceptibility to individuals in the population and their risk of developing BD. From the families, two to seventeen individuals were sequenced per pedigree and each of the individuals were separated into one of three clusters.

The first consisted of closely related affected individuals. The second included affected individuals separated by multiple meiosis and the final group consisted of pairs of subjects who were discordant for BD. The reason for the multiple clusters was to test different genetic disease models such as rare variants with highly penetrant effects which may follow a monogenic pattern of inheritance or an oligogenic pattern were rare or uncommon variants may cluster in individuals and additively cause disorder symptoms. Three types of hypotheses were then employed to choose genes to further investigate. Genes located within 100kb from loci reaching significance ($P<10^{-4}$) in the PGC BD GWAS [270]. Genes which are intrinsic for neuronal functions were also investigated to see whether they harbour an increased number of rare variants in individuals with BD. The last cohort of genes to be investigated originated from gene ontology pathway analysis. Variants with MAF < 0.05 present in coding or regulatory regions of the genome were also considered. Genotyping of rare variants was then conducted in a larger cohort of 3,014 BD individuals (GAIN) [235] versus 1,717 healthy controls recruited from the Translational Genomics Research Institute [240]. Variant investigation highlighted a rare SNP, rs3811993 $P=3x10^{-5}$, in the *GABRA6* gene, which encodes the α6 subunit of one the main inhibitory neurotransmitter families, GABA receptor family, in the CNS. Resequencing of this variant in the additional case control sample further identified this variant in a further 7 BD cases and 1 single healthy control.

Pathway analysis in the 200 WGS BD individuals and family members highlights an increased burden of rare and uncommon variants in their genome, specifically in genes involved in neuronal pathways (see Figure 10). These pathways include GABA pathway genes ($P=1.5x10^{-5}$) and GTPase genes ($P=6.8x10^{-3}$). Additionally voltage gated calcium channel genes ($P=3.6x10^{-3}$) and calmodulin-dependent protein kinase genes ($P=1.3x10^{-2}$) showed trends towards significance.

Combined analysis of both the modelling approach and the pathway analysis suggested that rare variation in individuals affected with BD and their relatives is found predominantly in non-coding regions of the genome with 164 variants found outside exonic regions. Of these 88% were located in promoter regions in addition to the 5' and 3'UTR.

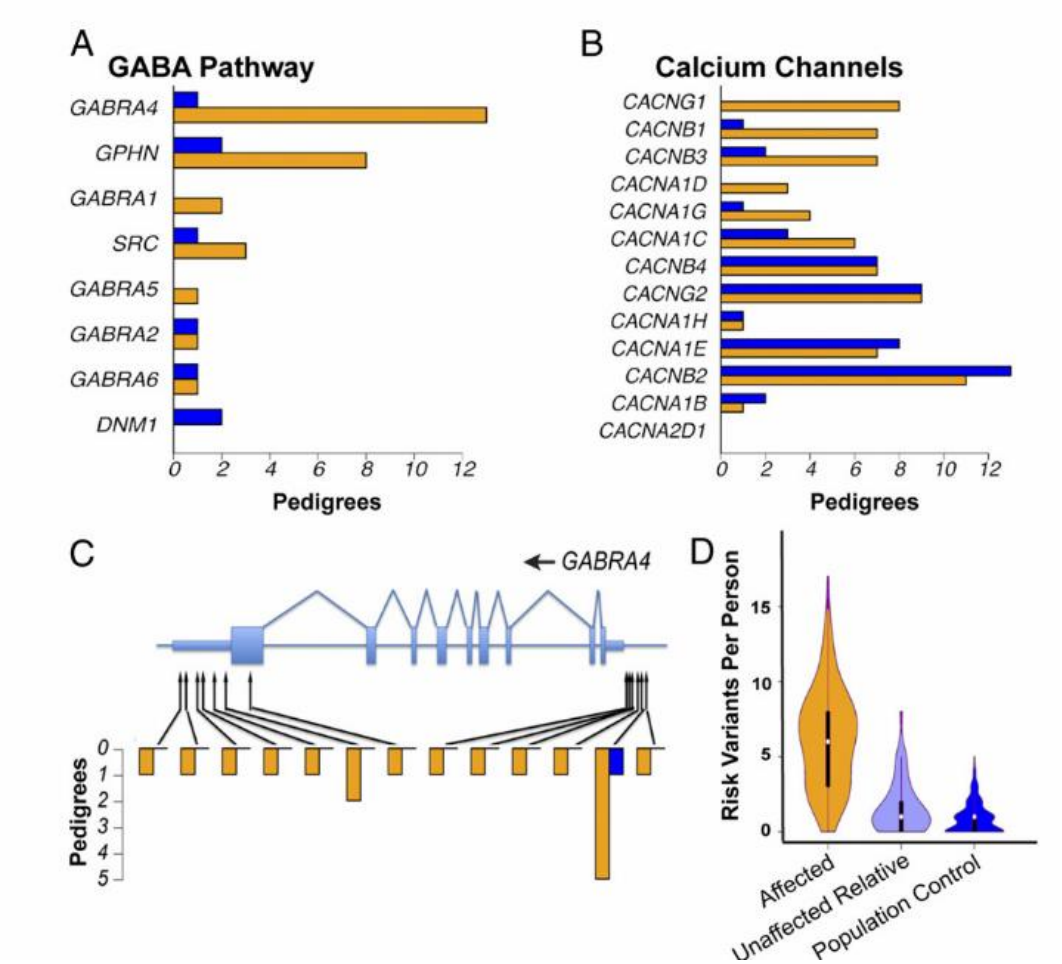


Figure 10. Families of genes highlighted from pathway analysis in 200 WGS BD patients.

A. The number of pedigrees which contain a fully or almost fully segregating variant from a coding or regulatory region in a gene which is defined as being a member of the GABA pathway in BD (Orange) and healthy controls (blue). B. The number of pedigrees which contain a fully or almost fully segregating variant from a coding or regulatory region in a gene which is defined as being a member of the Calcium channel pathway in BD (Orange) and healthy controls (blue). C. Segregation of rare coding or regulatory SNPs or indels in the *GABRA4* gene from the GABA family. Each of the thick bars denote an exon, and the 5' and 3' UTR regions are denoted by the extended thick lines before and after the first and last exon respectively. D. The number of rare risk variants in each of the different cohorts, BD (orange), unaffected relations of a BD proband (light blue), healthy controls (dark blue) in the sample used for the Ament *et al.* study. Risk variants were defined as those with a $P<0.001$, SNPs from the mixed model disease type with $P<0.05$ and that had an odds ratio of great than 1. Image taken from [292].

Ament *et al.* conducted further variant analysis by targeted sequencing of 26 genes, 20 of which were implicated in their study (10 voltage gated calcium channel genes, 5 GABA receptor subunits and five camodulin-kinase signalling genes), and six others genes from the PGC GWAS association candidate loci (*ANK3, TENM4, NTRK2, NTRK1, NGF, BDNF)*. This target resequencing focused primarily on the coding and regulatory regions of the genes, and found little evidence for coding variants in these genes with only *ANK3* reaching significance in a model which supports the presence of only rare risk variants and the absence of protective variants. In contrast, significant associations were reported between rare variants located in enhancer and promoter regions, which have an impact on transcription factor binding sites. Analysis of the data using the sequence kernel association test (SKAT) [293] identified six genes, *CACNA1C, CACNA1B, CACNA1D, CACNG2, CAMK2A* and *NGF* that were significantly associated with BD.

The dawn of the new era for genomic studies come with a need for more powerful and further developed statistical and informatics approaches. NGS data often produces thousands of novel variants which may impart a functional effect on disease susceptibility. Identifying the variants of interest and their association between phenotypes can be difficult given that some of these variants have small to moderate effect sizes. Increasingly powerful analytical methods are being devised in an attempt to address this issue. One of the common approaches to sorting through NGS data is to conduct burden analysis of variants in different genes and between gene sets [282]. Other approaches used to identify variants which may result in phenotypic changes are regression models [294] and over dispersion tests [295]. Both regression and over dispersion tests can analyse variants regardless of direction of effect and therefore these tests do not suffer from the loss of power that protective or neutral variants have on burden tests. Other approaches are starting to incorporate mixed models as well as using weighting of different types of genetic variants.

NGS offers researchers the ability to select a sample cohort in which a more thorough investigation of genetic variation may be undertaken. The steps that follow after the initial data collection are where the integral approaches to analysis change. The investigators above have included resequencing of genes of interest in larger sample populations, using linkage analysis to investigate variation in the

NGS data, genotyping of variants in larger pedigrees, and using publically available data to determine variant frequency.

1.3 CALCIUM CHANNEL GENES

1.3.1 INTRODUCTION

Bipolar disorder genome-wide association studies have identified a number of disease related genes and pathways. The L-type calcium channel gene family has been repeatedly implicated in BD. Calcium channels are formed from a combination of auxiliary subunits, based around a calcium selective α pore, first identified by their association with purified dihydropyridine [296, 297] (Figure 11).

Researchers have identified 26 genes with five different subunits, encoding ten α, four β, four $\alpha_2\delta$ and eight γ calcium channel subunits in the voltage dependent family of calcium channel genes (Table 3) [298]. The main alpha subunit forms a calcium selective pore in the membrane of the cell allowing the influx of Ca^{2+} ions into the cell while other auxiliary subunits interact directly with the α subunit to regulate channel gating in response to extracellular signals. The Ca^{2+} signalling pathway is a key component for initiating a multitude of different physiological changes in response to the influx of Ca^{2+} ions. This signalling pathway acts as a secondary messenger and launches a variety of signalling cascade pathways. Calcium channel pores are found in nearly all cells of the body and neurologically they participate in the mechanisms which are responsible for regulating neuronal excitability, information processing and cognition [93].

The role that calcium plays as a secondary messenger is an integral part of the biochemical transduction of synaptic signals and this in turn impacts gene regulation and expression [299]. The allosteric nature of calcium as a divalent cation makes it suitable for its vital role in biological systems. Its large radius and electron shell configuration make it suitable for interacting with large biological molecules, and the size and radius offers Ca^{2+} ions a flexibility that other smaller

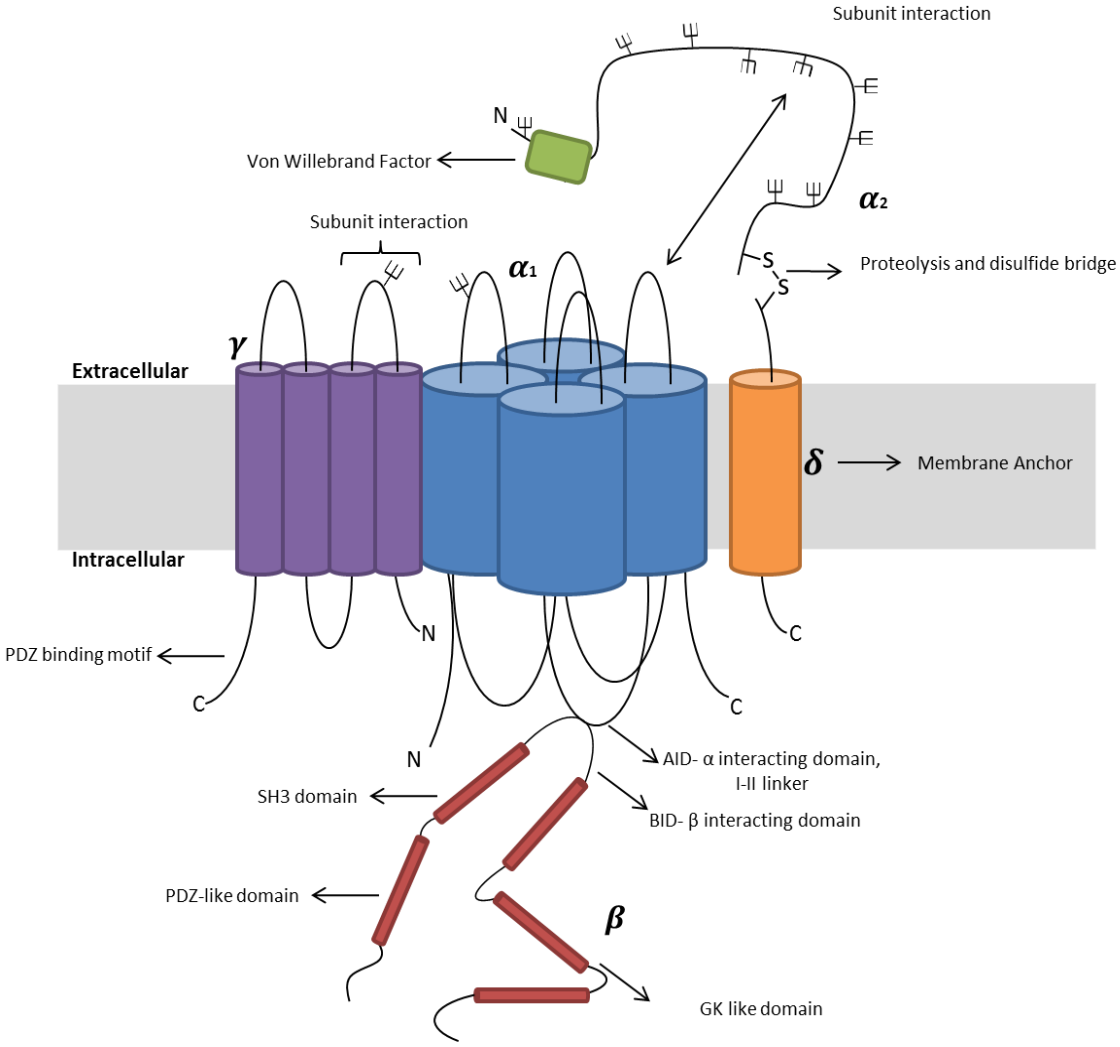


Figure 11. Formation of the Voltage-Gated Calcium Channel in the cell membrane.

The main α1 subunit forms a pore in the cell membrane allowing Ca2+ influx into the cell following membrane depolarization. The other auxiliary subunits control the voltage gating and influx of Ca2+ into the cell. Each of the protein subunits interact via various domains. The α1 subunit interacts with β subunits via an α interacting domain (AID) and β interacting domain (BID). The α2 and δ subunits bind via proteolysis disulphide bridge formation. Other binding motifs such as the, post synaptic density protein (PSD95), Drosophila disc large tumour suppressor (Dlg1), and zonula occludens-1 protein (zo-1) (PDZ) domain allow the auxiliary subunits to scaffold to the post synaptic density. The SRC homology 3 (SH3) domain controls the biophysical properties of the calcium channel activity and inactivation . GK, Guanylate kinase-like domain is essential for plasma membrane expression of β subunits of the calcium channel family of genes [300]. Image based on [301].

Gene Name	Type	Chromosomal Location
CACNA1A	P/Q type, alpha 1A subunit	19p13
CACNA1B	N type, alpha 1B subunit	9q34
CACNA1C	L type, alpha 1C subunit	12p13.3
CACNA1D	L type, alpha 1D subunit	3p14.3
CACNA1E	R type, alpha 1E subunit	1q25.3
CACNA1F	L type, alpha 1F subunit	Xp11.23
CACNA1G	T type, alpha 1G subunit	17q22
CACNA1H	T type, alpha 1H subunit	16p13.3
CACNA1I	T type, alpha 1I subunit	22q13.1
CACNA1S	L type, alpha 1S subunit	1q32
CACNA2D1	alpha 2/delta subunit 1	7q21-q22
CACNA2D2	alpha 2/delta subunit 2	3p21.3
CACNA2D3	alpha 2/delta subunit 3	3p21.1
CACNA2D4	alpha 2/delta subunit 4	12p13.33
CACNB1	beta 1 subunit	17q21-q22
CACNB2	beta 2 subunit	10p12
CACNB3	beta 3 subunit	12q13
CACNB4	beta 4 subunit	2q22-q23
CACNG1	gamma subunit 1	17q24
CACNG2	gamma subunit 2	22q13.1
CACNG3	gamma subunit 3	16p12.1
CACNG4	gamma subunit 4	17q24
CACNG5	gamma subunit 5	17q24
CACNG6	gamma subunit 6	19q13.4
CACNG7	gamma subunit 7	19q13.4
CACNG8	gamma subunit 8	19q13.4

Table 3. Calcium channel genes.

Calcium channel alpha pores and their chromosomal locations in the human genome

divalent cations such as magnesium do not have. Intercellular calcium concentration is regulated through a number of specific mechanisms to maintain a low cytosolic level. A calcium gradient is formed across the cell membrane by a sarco/endoplasmic reticulum Ca^{2+}-ATPase (SERCA) pump that removes calcium from the cytosol, and places it into internal stores in the endoplasmic reticulum (ER) membrane. This gradient allows calcium entry into the neurons through the voltage dependent calcium channel located on the plasma membranes. Additionally calcium can be released from the ER stores through $InsP_3$ receptors [302] which will be further explained below (Figure 12). Influx of calcium through ion channels occurs in response to excitatory neurotransmitters signals. Calcium can enter the cell through either VDCCs, or ligand gated channels such as (N-methyl- D-aspartate)-type glutamate receptors (NMDARs) and α-amino-3-hydroxy-5-methyl-4-isoxazolepropinonic acid receptors (AMPA-R). In addition to this, calcium entry may be instigated via neuronal inputs stimulating G protein coupled receptors (GPCRs). Both of these mechanisms create intracellular and nuclear calcium channel transients for further downstream signalling (Figure 13) [303].

The mechanism of membrane depolarisation, that lead to the opening of VDCCs and the influx of Ca^{2+} ions, occurs through the stimulationof excitatory postsynaptic potentials (EPSP). EPSPs induce action potentials that propagate actively through the axon and then back through the dendrite [304], leading to plasma membrane depolarisation and opening of L-type VDCCs. L-type calcium channels are ubiquitously expressed in the somatic and perisomatic area [305] and allow for a high influx of Ca^{2+} ions. This occurs through synaptic input stimulating NMDA and AMPA-R on the postsynaptic density.

The second mechanism that leads to increased nuclear and intracellular calcium transients occurs through the coupling of GPCRs on the postsynaptic density. The coupling of the GPCRs leads to the activation of phospholipase C (PLC), that results in cleavage of phosphatidylinositol 4,5, biphosphate which in turn leads to the formation of $InsP_3$ [306]. $InsP_3$ generation causes the opening of $InsP_3$ receptor pumps which are located on the ER and Golgi apparatus allowing the efflux of Ca^{2+} into the cytosol and then into the nuclear pore complex.

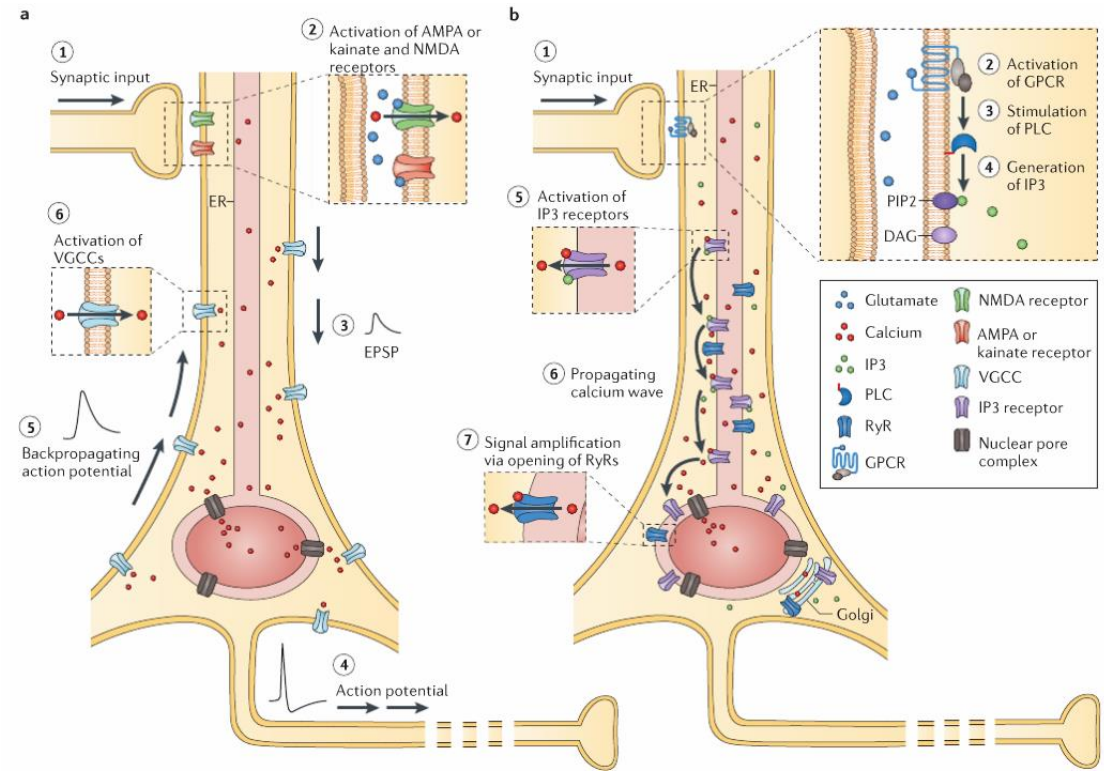


Figure 12. Mechanisms by which neuronal activity generates increased levels of intercellular and nuclear calcium transients.

A. (1) Excitatory neuronal signals activate both (2) NMDA and AMPA receptors in the post-synaptic density. (3) Upon activation of these ligand gated channels excitatory postsynaptic potentials (EPSP) are invoked (4) firing action potentials in the neurons. (5) These action potentials disseminate back through the dendrites causing membrane depolarisation. (6) Depolarisation of the membrane results in the opening of VDCC, allowing for calcium influx into the neuron. **B.** (1) Input from excitatory neuronal signals activates (2) G protein coupled receptors (GPCRs) which lead to the (3) activation of phospholipase C (PLC). (4) The activation of PLC results in the production of inositol 3 phosphate (IP3). (5) When IP3 is produced IP3 receptors are activate and lead to the release of Ca^{2+} from intracellular stores in the neuron. (6) Release of intracellular Ca^{2+} results in an intercellular propagating calcium wave. (7) Further amplification of the calcium signal is achieved through release of calcium from the endoplasmic reticulum or golgi through ryanodine receptors. Image taken from [303].

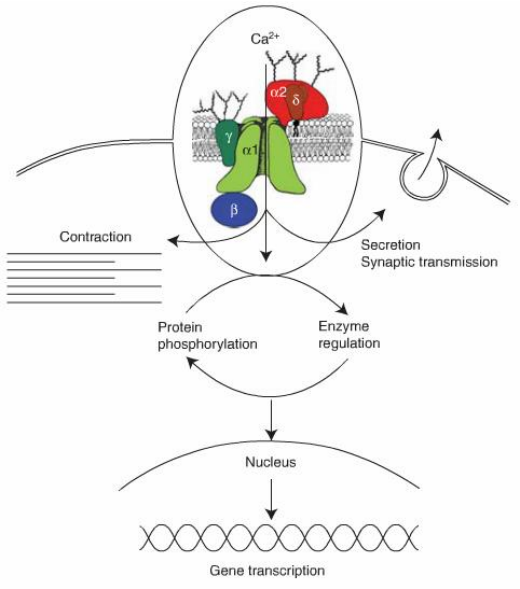


Figure 13. Processes activated following the influx of Ca^{2+} ions.

Calcium can enter the cell through a number of different mechanisms and activates an assortment of signal transduction pathways. These processes act at different cellular levels, stimulating tissue contraction in muscles, allowing communication between cells through the transport and movement of chemicals such as hormones and cellular signalling both by secretion and through synaptic signalling. At a protein level, enzyme regulation is controlled and protein phosphorylation can be instigated. At the genetic level the control of gene transcription is influenced by the influx of Ca^{2+} ions. Image taken from [307].

Calcium's control of gene expression can occur through a number of mechanisms. Upon influx of calcium through the calcium pore the Ras-MAPK cascade signalling pathway is initiated, leading to the phosphorylation of serine residue 133 in CREB. Cell depolarisation, resulting from the influx of calcium into the cell, can lead to the alternative splicing of pre-mRNAs impacting on gene expression [308]. Crucially, Ca^{2+} influx plays a vital role in gene regulation through transcriptional initiation. The role of nuclear calcium in transcriptional regulation was first reported by Hardingham *et al.* in 1997 following studies conducted on a mouse pituitary cell line [309]. Transcriptional activation initiated by the CREB transcription factor was completely blocked following injection of the cell nucleus with BAPTA, a calcium chelator. BAPTA was fused to a large 70-kDA carbohydrate, dextran, which prevented the diffusion of the chelator through the nuclear membrane, preventing the chelator interaction with cytosolic calcium. In addition, the presence of dextran

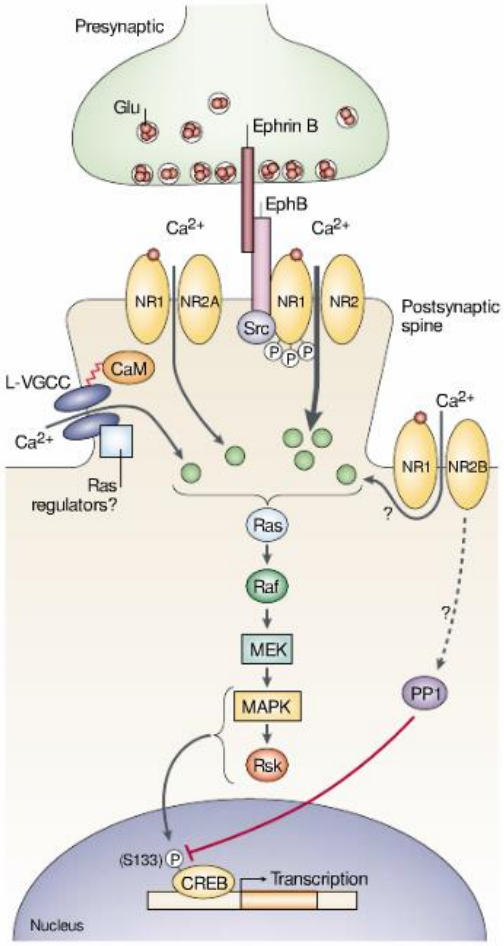


Figure 14. Activation of gene transcription by the influx of calcium through plasma membrane calcium channels.

As calcium enters cells through α pores on the plasma membrane it initiates the Ras-MAPK signalling pathway. The binding of calmodulin to the IQ domain in the carboxy terminus of the α1C subunit allows for the activation of the Ras-MAPK signal transduction pathway. Activation of this pathway leads to the phosphorylation of Ser133 in CREB. Image from [310].

did not interfere with membrane depolarisation which results in the influx of Ca^{2+} ions through VDCC, on the cell membrane, as described previously. Transcriptional activation initiated through other pathways such as the serum response element (SRE) [309], were not blocked by BAPTA. The ERK/MAP kinase pathway targets SRE following an increase in concentration of calcium in the submebranous space which is located close to the calcium channels in the membrane [311-313].

Ca^{2+} influx through different channels mediate different responses. L-type VDCCs are highly associated with activation of the CREB cycle through phosphorylation of CREB at Ser133 [312, 314]. This phosphorylation at Ser133, and other serine residues in the CREB protein enhance initiation of transcriptional activation and activity [315]. This occurs through the action of calmodulin-dependent protein kinase IV (CaMKIV), a tyrosine/serine kinase which is located in the cell nucleus [316]. The effect of phosphorylation is thought to recruit the transcriptional coactivator CREB-binding protein (CREBBP) to the phosphorylated Ser133 residue [317, 318]. Phosphorylation of other serine residues in the CREB protein occurs in response to specific neuronal activity [319], stimulated solely through the phosphorylation of Ser133 in CREB. Activation of CPB, by CaMKIV and nuclear calcium, is a requirement for the initiation of gene transcription in the nucleus [320]. It is proposed that CBP acts to initiate transcription by increasing the acetylation of histones bound to target genes in the nucleus [318].

In addition to commencing the downstream cascades of signalling pathways, the influx of calcium ions into cells is directly monitored by a family of proteins called neuronal calcium sensors (NCS) [321]. One member of the NCS family is the downstream response element (DRE)-antagonist modulator (DREAM) which binds directly to calcium through its EF-hand calcium binding motif [322, 323]. The EF-hand is a helix-loop-helix motif that has high affinity for binding calcium, and with a lower affinity, magnesium [324]. DREAM can block transcription by binding to the downstream response element (DRE) (TACCGACAT) located in many genes [325]. As calcium concentration increases in the nucleus it interacts with DREAM, binding to the EF hands and allowing transcription to occur.

Changes in cytoplasmic calcium levels mediated by entry through L-type VDCCs regulate transcription through the initiation of the Ras-MAPK (mitogen-activated protein kinase) pathway (Figure 14) [310]. This initiation of transcription is highly coupled to a calmodulin-binding isoleucine-glutamine (IQ) domain in the carboxy terminus of the CACNA1C α pore [326]. As calmodulin binds to L-type VDCCs in the plasma membrane, it stimulates prolonged Ras-MAPK dependent phosphorylation of Ser133 in the CREB transcription factor. Calmodulin binds to L-type VDCCs in a calcium dependent manner and activates the Ras-MAPK pathway through localisation of the signalling pathway.

1.3.1.1 ALPHA SUBUNITS

The α1 subunit of voltage gated calcium channels is the main pore which allows the influx of Ca^{2+} in and out of the cell [298]. Ten genes encode the calcium channel alpha pores (Table 4). The general structure of the α pore consists of four repeat domains, each of which contains six transmembrane segments. This subunit can function independently to allow calcium influx in the cell. Under normal physiological situations it is associated with auxiliary subunits which control the opening and closing of the transmembrane pore, and trafficking of the pore from the ER.

There are five different types of voltage dependent calcium channel α pores, N-type, P/Q-type, T-type, R-type and L-type. These are grouped into three different categories, Ca_v1, Ca_v2 and Ca_v3, depending on their function. The Ca_v1 subgroup is involved in excitation and contraction in cardiac, skeletal and smooth muscle cells, transcriptional regulation, enzyme activity, and neuronal Ca^{2+} transients in dendrites and cell bodies. Ca^{2+} transients occur when calcium enters the cell through membrane depolarisation-activated calcium channels as inward calcium currents. Ca_v2 types are involved in neurotransmitter release and the Ca_v3 types are involved in cardiac pace making and repetitive firing [307].

The classification system for the different types of calcium channel pores originates from their different physiological and pharmacological channel properties. L-type VDCC are characterised by high voltage activation with large single channel conductance, slowed voltage dependent inactivation, up-regulation marked by cAMP-dependent phosphorylation pathways and inhibition by Ca^{2+} blockers [327, 328]. T-type VDCCs are so named as a result of their transient nature, are activated at a much lower negative membrane potential than the L-type family of VDCCs. Additionally they have small single channel currents, are inactivated rapidly and deactivate slowly [329]. N-type calcium channels were originally distinguished by their intermediate voltage dependence activation rate which is between that of the L- and T- types. They have a more negative and faster rate of inactivation than the L-type of VDCC and but in contrast have a more positive and slower rate of inactivation than that of the T-channels.

Channel Name	Channel Subtype	Gene	Subunit	Location	Tissue	Associated Disorders
L-Type	Cav1.1	*CACNA1S*	α_1 pore	1q32	Skeleton	Hypokalemic periodic paralysis; Malignant Hyperthermia; Thyrotoxic periodic paralysis, susceptibility
	Cav1.2	*CACNA1C*	α_1 pore	12p13.3	Cardiac cells, Endocrine System, Neurons	Timothy Syndrome; Brugada syndrome 3
	Cav1.3	*CACNA1D*	α_1 pore	3p14.3	Endocrine System, Cochlea, Cardiac cells, Neurons	Primary aldosteronism, seizures, and neurologic abnormalities; Sinoatrial node dysfunction and deafness
	Cav1.4	*CACNA1F*	α_1 pore	Xp11.23	Retina	X-linked cone-rod dystrophy type 1; X-linked congenital stationary night blindness type 2
N-Type	Cav2.2	*CACNA1B*	α_1 pore	9q34	Neurons	Neuropathic pain; Myoclonus-Dystonia syndrome
P/Q type	Cav2.1	*CACNA1A*	α_1 pore	19p13	Neurons	Episodic ataxia; Familial Hemiplegic Migraine; Spinocerebellar Ataxia type 6; Sporadic Hemiplegic migraine
R-Type	Cav2.3	*CACNA1E*	α_1 pore	1q25.3	Neurons	-
T-Type	Cav3.1	*CACNA1G*	α_1 pore	17q22	Neurons and Heart	Abberantly methylated can result in Primary Tumour Growth
	Cav3.2	*CACNA1H*	α_1 pore	16p13.3	Neurons, Heart, Kidney, Liver	Epilepsy, Childhood Absence, Idiopathic generalisation
	Cav3.3	*CACNA1I*	α_1 pore	22q13.1	Neurons	-

Table 4. Classification of the alpha pores of the calcium channel family

There are ten genes which encode α pore subunits of the VDCC. Here the localisation of gene expression is shown and disorders which are associated with dysfunction in the α pore are listed.

The N-type calcium channel genes are found predominantly in the presynaptic termini and participate in the transmission of neurotransmitters upon depolarization [330].

The current is extremely variable in different neurons and N-type channels are largely distinguished through their insensitivity to organic L-type Ca^{2+} blockers but sensitivity to the ω-conotoxin GVIA from the cone snail [331, 332]. Several other types of calcium channels were identified through the effects of peptide toxins.

The P/Q type channels are separate entities but there is little to distinguish between them. The P-type channel is highly sensitive to the spider toxin ω-agatoxin IVA [333], while the Q type channels are blocked with a lower affinity than the P type [334]. The R-type VDCC are resistant to most typical calcium channel blockers [334]. While the N, P/Q and R type VDCCs appear to be predominantly expressed in neurons, T- and L-type Ca^{2+} currents are found in a wide number of cells.

1.3.1.2 BETA SUBUNITS

For L-type VDCC and other types of high-voltage-activated calcium channels, the β subunit is specifically required for plasma membrane expression and correct gating [335]. The four different members of the β subunit family share a conserved 31 amino acid sequence known as the β interacting domain (BID) [336]. The -N and -C termini of all four members are variable, and there are a number of different splicing isoforms of each variant. The β subunit binds with high affinity to the α pore of the calcium channel via a specific 18 amino acid anchoring domain in the cytoplasmic linker found between the first two homologous repeats in the α subunit (I-II linker), also known as the α interaction domain (AID) [337-339]. Studies have shown that the β subunits are important for surface expression, degradation and promoting the gating of VDCC. Binding of the β subunit prevents the degradation of the α subunit, and exposes the ER exporting signals on the α pore to facilitate trafficking to the cell membrane. Additionally the β subunits play a protective role in trafficking of the voltage gated calcium channels to the cell membrane by preventing degradation by proteases and promotes the correct

folding of these proteins [340] [341]. They also play a role in regulation of the VDCC by lipids, G coupled proteins, kinases and other signalling proteins [342].

1.3.1.3 ALPHA-2-DELTA SUBUNITS

Currently four members of the α2δ family of subunits have been cloned, *CACNA2D1*, *CACNA2D2*, *CACNA2D3,* and *CACNA2D4*. These subunits are expressed in a variety of tissues including CNS expression of *CACNA2D1*, *CACNA2D2*, and *CACNA2D3* [343]. *CACNA2D4* is expressed in the retina and endocrine tissue [344]. Molecular studies investigating the composure of the α2δ subunit have shown that, following disulphide bridge reduction, the α2 subunit is 150kDa, and the δ subunit is between 17-25kDa [345]. Post translational proteolytic cleavage occurs on the C terminus of a pre-protein which forms the δ subunit [345]. Sulphide bond formation normally occurs co-translationally in the lumen of the ER and it is not yet known which protease causes the proteolytic cleavage that result in disulphide bond formation. However there is the suggestion that this occurs during trafficking of the α2δ protein [346]. The native α2δ protein is anchored to the extracellular membrane, with the α2 in the extracellular space, and the δ subunit anchored to the membrane as can be seen in Figure 11. Some of the functions of the α2δ subunits depend on expression of the β subunit of the calcium channel gene family [347].

The α2δ subunit contains a Von Willebrand factor-A domain (VWA) which binds to a variety of cell adhesion and extracellular proteins [348] through its metal ion adhesion domain, this interacts with a number of proteins to promote trafficking and to increase the density and expression of the α pore from L, N, P/Q and R-type calcium channels at the plasma membrane [349].

1.3.1.4 GAMMA SUBUNITS

The *γ* family of calcium channel subunits are comprised of eight different members. The *γ*-subunits are expressed in different VDCCs and can be clustered into three separate groups, (Figure 15), *γ1* and *γ*6 subunits are mainly expressed in skeletal muscle tissue and lack the PSD-95/DLG/ZO-1 (PDZ)-binding motif characteristic of the rest of the *γ* family [350]. Additionally *γ*6 is expressed in

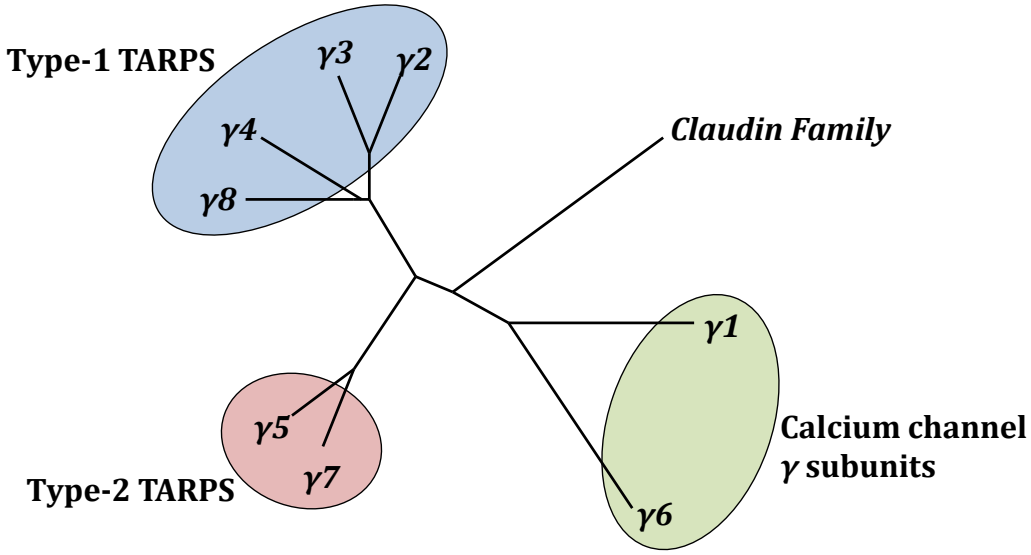


Figure 15. TARP Phylogenetic tree

Phylogenetic tree showing the relationship between calcium channel γ subunits and TARPs and the closely related claudin tight junction proteins. Image based on [351].

cardiac myocytes, where it is found in two alternative isoforms, the only *γ*-subunit to have an alternative isoform.

Both *γ1* and γ6 stabilise L-type VDCC in an inactive state. Studies have shown that *γ*2 (stargazin) has a similar inhibitive effect on the VDCC by stabilizing the interaction between another auxiliary subunit of the calcium channel, α2δ [352], while *γ5* regulates Ca^{2+} permeability via AMPA-R assembly. Additionally, *γ7* expression has an effect on calcium channel activity, but rather than promoting this effect via subunit interaction [353], it does so through alterations on mRNA stability [354]. While originally classified as a VDCC subunit studies have shown stargazin's main function is in synaptic signalling in the central nervous system where it, and the other members of the *γ* family of subunits, are tightly associated with AMPA-R regulation and localisation [355]. *γ2, γ3, γ4* and *γ8* are classified as type-I transmembrane AMPA receptor regulators (TARPs) and are intrinsically involved in the control of both the trafficking of AMPA-R to the cell surface as auxiliary subunits. They are also involved in channel gating and mediating fast synaptic transmission in the CNS in specific brain regions [351], [356].

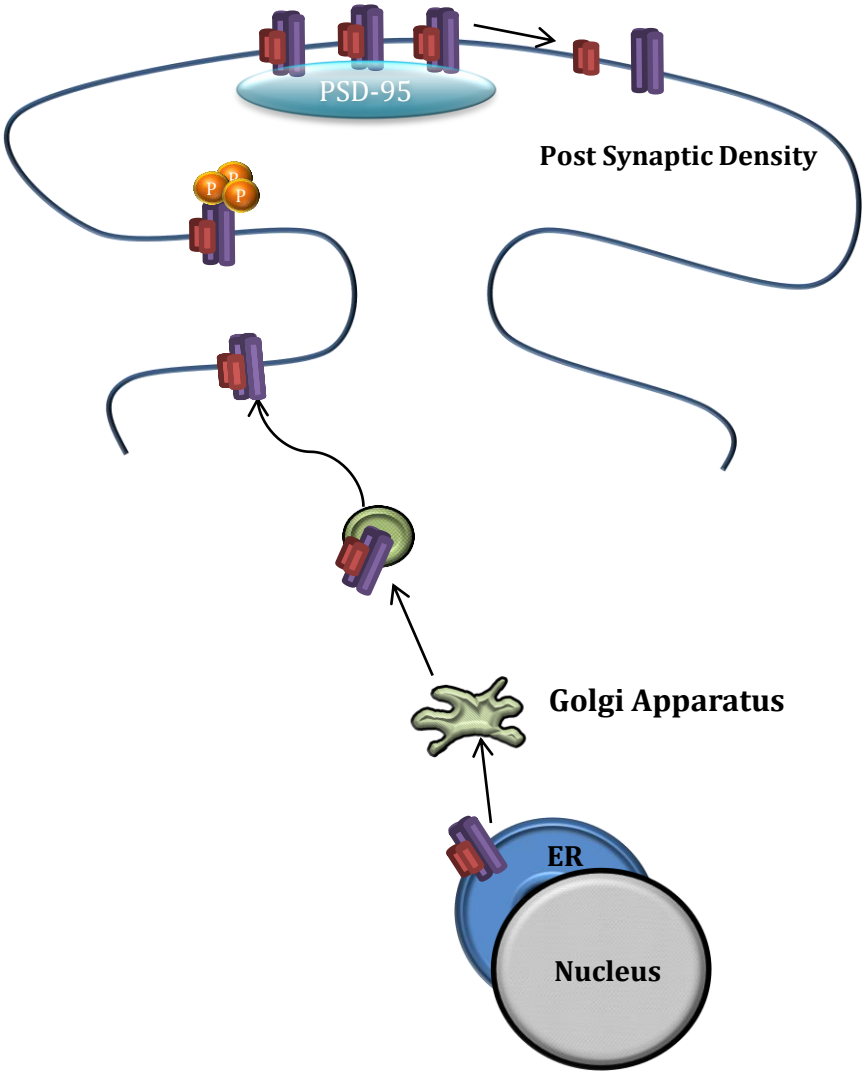


Figure 16. Trafficking of AMPA-R to the postsynaptic density through TARP interaction.

ER, Endoplasmic Reticulum, PSD-95, Post-synaptic Density Protein 95

They are expressed neuronally and play a role in the localisation and regulation of gating of AMPA-R at the post-synaptic density (PSD) (Figure 16) [357, 358]. *γ5* and *γ7* (type-II TARPS) are more closely related to the type-I TARPs, than γ6 and γ1 and the claudin family of proteins. The type-II TARPs are specifically enriched in specific neuronal areas, with the highest expression of γ7 occurring in the Purkinje cells and *γ5* being found associated with AMPA-R in the cerebellum, and expression of the protein found to a lesser extent in the kidneys [359].

Research into the main functions of TARPs have shown them to play an integral role in mediating fast excitatory transmission in the CNS, where post translational modifications occur to the specific glutamate receptors resulting in long-term

potential or long-term depression [360]. Alterations to the transport of these receptors could alter CNS plasticity resulting in abnormal synaptic connections forming. The role of TARPs in establishing connections during development is another potential hypothesis for how variation in these subunits may lead to an increased susceptibility to psychiatric illnesses.

Association of TARPs and AMPA-R in the endoplasmic recticulum facilitates trafficking from the ER to the plasma membrane [361]. Following this binding of the TARP, through the C terminus, to PSD-95, it localises and secures the AMPA-R to the postsynaptic membrane. Trafficking of AMPA-R to the postsynaptic membrane is differentially regulated depending on which TARP it is bound too. Type-I TARPs binding to AMPA-R type I (GluR1) slows the desensitization and deactivation of the glutamate channels and prolongs their activation following the removal of glutamate [362]. This binding forms a specific complex between the TARPs and AMPA-R. Binding of TARPs to AMPA-R allows the AMPA-R to leave the ER and be trafficked to the postsynaptic density. Here they are anchored by scaffolding proteins through PDZ-95 binding domain.

1.3.2 Genetic Variation in Calcium Channel Genes

There are a number of studies which have shown that variation in calcium channel gene subunits plays a prominent role in the manifestation of severe psychiatric disorders and phenotypes.

As previously discussed, clear evidence pointing to abnormalities in calcium levels has been seen in patients suffering from BD. Discerning whether mutations in these genes are having an effect on the assembly of the calcium channel subunit or controlling the activation of gating and calcium permeability into the cell would help researchers further understand the basis of the calcium imbalance found in BD.

1.3.2.1 Alpha Subunits

Genetic variation in *CACNA1C* leads to the manifestation of Timothy syndrome and Brugada syndrome. Missense mutations (p.G406R and p.G402S) in exon 8A of

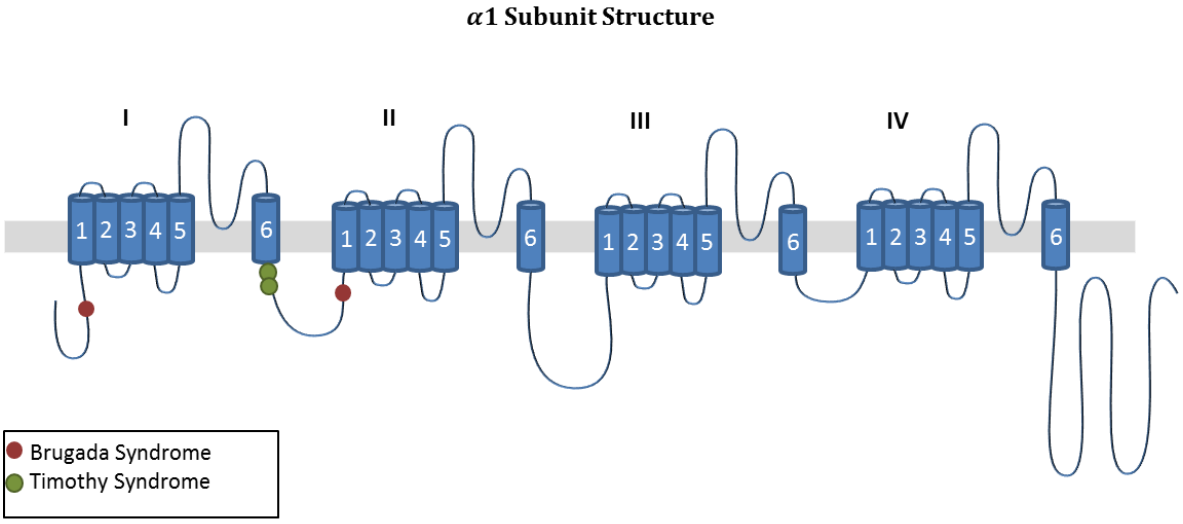


Figure 17. Structure of the Alpha Pore of VDCC.

The *CACNA1C* α pore subunit is comprised of four repeat domain which have 6 transmembrane domains each. Variants which are associated with Brugada syndrome are represented by red circles. The variants associated with Timothy syndrome are represented by the green circles. All of the variants are located in intracellular loops of the α pore.

CACNA1C, an alternatively spliced exon, are responsible for the rare disorder, Timothy syndrome (TS) [363] (Figure 17). This disorder is characterised by developmental delays and autism, in addition to severe cardiac manifestations. The two mutations were originally reported first in 2004 by Splawski and colleagues showing links with autism and cardiac arrhythmia [363].

In vitro studies have shown that iPS stem cells, derived from patients with TS, which were used to develop cortical neuronal precursor cells, displayed aberrant development and differentiation. Defects in Ca^{2+} signalling and decreased expression of genes in brain regions, such as the lower cortical layers, are additionally observed in these neurons [364]. In the same publication Pasca *et al.*, showed that tyrosine hydroxylase is abnormally expressed and there is increased production of both dopamine and norepinephrine. This can be corrected through the use of roscovitine, which in addition to acting as a cyclin-dependent kinase, functions as a L-type calcium channel blocker [365]. Calcium channel blockers are also used in the treatment of bipolar disorder and may function in a similar

manner to the method by which the aberrant gene expressions in TS are ameliorated.

Additionally, variants known to cause Brugada Syndrome (BrS) a severe cardiovascular disease are found in *CACNA1C* (Figure 17). Individuals suffering from BD have been consistently associated with an elevated risk for cardiovascular disease [366]. A Danish study following 5.5 million individuals, from 1978 to 2001, reports that of the 11,648 people diagnosed BD, 3,669 had passed away when the study was complete. Estimates of the standard mortality rate (SMR) of cardiovascular disorder on the cohort show an elevated risk of 1.59 in men and 1.47 in women compared to that of the general population [367]. Estimates of cardiovascular SMR in a Swedish study report similar values of 1.9 for men and 2.6 for women [29]. BrS is a serious cardiovascular disease which can cause sudden cardiac death [368]. Variants in the *CACNA1C* and *CACNB2B* genes lead to a reduced peak calcium current in Brugada syndrome which results in the loss of gene function [369, 370].

Variants in the P/Q type alpha subunit, *CACNA1A,* have been shown to result in a number of autosomal dominant neurological defects such as familial hemiplegic migraines, episodic ataxia type-2 and spinocerebellar ataxia [371].

One of the top association signals in BD comes from rs1006737 (P= 7.0 × 10^{-8}) [232, 233], lying in the third intron of *CACNA1C*. Another two SNPs located in this intronic region, rs4765913 (p= 1.52 × 10^{-8}) [232] and rs1024582 (p= 1.7 × 10^{-7}) [270], are also strongly associated with BD. GWAS results across five different psychiatric illnesses further implicates rs1024582 in susceptibility to both BD and SCZ [268].

Research shows that that the presence of rs1006737 *CACNA1C* BD risk variant leads to alterations in certain brain activities. One study showed that the rs1006737 risk allele in healthy males is associated with lower extraversion, trait anxiety, paranoid ideation and higher harm avoidance [372]. The rs1006737 risk variant has been shown by many MRI studies to modulate amygdala functioning during emotional processing, the enhancement of such activation leads to impaired facial emotion recognition in BD patients [268, 373-376]. There has been

conflicting evidence as to whether the presence of the *CACNA1C* variant results in brain volumetric alteration, with reports stating this SNP has been associated with brainstem alterations, increased grey matter density as well as a cortical volume increase [377-379], while a recent study looking at multiple variants did not find any association between this SNP and brain volumetric alterations [380].

CACNA1B has been reported to show evidence for association with SCZ (p=0.002) in genome-wide analyses of genome-wide association datasets [381]. A recent study by Ament *et al.* [292] has shown that there are rare variation in four genes that are members of the calcium channel family of genes, *CACNA1B, CACNA1D,* and *CACNG2*, associated with BD.

1.3.2.2 BETA SUBUNITS

Gene ontology enrichment analysis of 34 SNPs reaching a genome-wide significant association ($P \leq 5 \times 10^{-8}$) in a large-scale genome BD association study have found enrichment in the GO: 0015270, dihydropyridine-sensitive calcium channel activity. Of the eight genes in this category, 3 are members of the VDCC family, *CACNA1C, CACNA1D* and *CACNB3* [239]. Other genes encoding calcium channel subunits have shown association in BD and other neurological disorders. In addition to the associations seen in the PGC GWAS for *CACNB3,* null β3 mice show impaired working memory, coinciding with increased aggression and a decrease in anxiety [382]. A case-control study focused on patients with epilepsy has identified 3 mutations in *CACNB4* found only in people suffering from epilepsy and none in healthy volunteers [383]. *CACNB2* has been found to be associated with BD in the Han Chinese population (rs11013860, $P=5.15 \times 10^{-5}$) [244]. Jan *et al.* have shown support for an association between the same variant, rs11013860 in *CACNB2,* and BD-I in a homogenous Taiwanese cohort [384].

1.3.2.3 ALPHA-2-DELTA-4 SUBUNITS

A rare deletion in *CACNA2D4* has been found in two unrelated late onset bipolar I patients and one control, all of the break points of these deletions caused the removal of CACHE domain, which is involved in small molecule recognition in a wide range of proteins [385]. Linkage disequilibrium contrast analysis between

SNPs in cases and controls has shown association between SNPs in *CACNA2D4* in BD (LD-contrast test, p=4.6x10^{-14}) [386].

1.3.2.4 GAMMA SUBUNITS

Mutations in stargazin (*CACNG2*) were first identified in stargazer mice [387]. The genetic variation in *CACNG2* leads to the loss of AMPA-Rs at excitatory synpases [388]. These mice suffer from recurrent epileptic seizures, cerebellar ataxia and ataxic gait [389].

A combined burden analysis of non-synonymous variation in *CACNG5* has shown an association with SCZ (p=1.46x10^{-8}). This data also reports that four variants in the *CACNG5* gene result in decreased cell surface expression of the AMPA-R2 [390]. A case-case study between BD and SCZ has identified an association between rs17645023 and both disorders (p=1x10$^{-6.1}$) which lies between *CACNG4* and *CACNG5* [271]. Studies focusing on SCZ have shown TARP dysregulation at a both transcript and protein level, with γ-3 and γ-5 protein levels increasing and *γ-4, γ-7* and γ-8 protein levels decreasing in the anterior cingulate cortex [391].

1.4 INTRONIC GENETIC VARIATION

1.4.1 INTRODUCTION

It is clear and evident from the above descriptions of the genetic models proposed for BD that while variants located in exonic regions may have a role to play in increasing disease risk, it is also likely that many other risk variants in regulatory regions have an important role to play in disease susceptibility. These regulatory domains will include regions such as the promoter, the 5'UTR and the 3'UTR and specific regions of the introns. These intronic regions may have a functional impact on transcription factor binding sites and gene expression. A large proportion of GWAS candidate hits occur in introns (such as with the *CACNA1C* gene) or in regions of the genome where there is no gene present. In the past these results have been dismissed because it has been difficult to functionally implicate these variants in disease risk. Fortunately, in the past five years, there has been a

significant move in improving the resources and databases available for such investigation, primarily starting with the ENCyclopaedia of DNA Elements project (ENCODE). In addition to this large scale project additional online resources and software are becoming increasing available, facilitating stronger research and experiment design.

1.4.2 ENCODE DATA

The aim of the ENCODE project is to characterise all functional variants in the human genome [392]. In the first phase of publications the consortium targeted 1% of the genome to determine techniques and methods which would accurately identify functional elements [393]. The second phase, resulting in a co-ordinated set of 34 publications in 2012, targeted technologies which had been used in phase one, to improve techniques used to identify genomic elements. The second phase covered 147 different cell lines and tissues, resulting in 1640 data sets. Phase two cell lines where divided into three different tiers (http://genome.ucsc.edu/ENCODE/cellTypes.html). Tier 1 included three cell lines; K652, a leukaemia cell line, GM12878, an immortalized B lymphocyte Epstein-Barr line and H1-hESC, a human embryonic cell line. Tier 2 was comprised of 15 different tissues and cell types. There are two neuronal cell lines in tier two, H1-neurons derived from human embryonic stem cells and SK-N-SH cells from a neuroblastoma. Tier 3 included another 338 cell lines and tissue samples. Of these 18 were related to the central nervous system, 14 from brain tissue, one from the cerebellum, one from the frontal cerebrum tissue, one from the frontal cortex and one from the pons.

Functional elements include RNA transcribed sites, protein-coding regions, transcription factor binding sites (TFBS), open chromatin and histone modifications, and DNA methylation sites [392]. Evidence strongly points to over 80.4% of the genome containing functional genomic elements in at least one cell type.

Regions of accessible chromatin provide researchers with a baseline map of areas of possible transcription in the genome. Regions of open chromatin are marked by the presence of DNAse-hypersensitivity sites (DHSs), indicating regulatory active

regions [394, 395]. Of the 7 million peaks found, approximately 40% were located within 2.5kb of previously categorised transcriptional start site (TSS). Chromatin immunoprecipitation followed by sequencing, (ChIP-seq) was generated for 119 different TF over the main cell lines [396]. ChIP-seq is used to identify TFBS and histone modifications [397]. Histone marks combined with DHSs can be used to identify putative functional elements involved in transcriptional regulation, such as promoters or enhancers (Table 5) [398]. Transcription factor binding sites play an important role in activation of transcription. Mutations in TFBS, located in ChIP-seq peaks, were introduced into cells to determine whether the proposed regions were functionally active or just putative elements. Over 70% of the reported TFBS showed altered function in functional reporter assays in the ENCODE data [396]. It was noted that affinity for TF binding at specific genome regions was allele specific, with reduced functionality seen in the presence of novel SNPs. Use of these data provides a novel method for determining functionality of novel SNPs found in non-coding regions of the genome which have no previous genotyping data.

The National Human Genome Resources Institute has a catalogue of all SNPs found by GWAS. Approximately 93% of SNP-phenotype associations lie in non-coding regions [399], making it difficult for researchers to infer biological impact for these findings. SNPs with novel functional data found using the ENCODE methods show strong enrichment for GWAS disease associations. Maurano *et al.* looked at the localisation of disease associated variants and their results show that 76.6% of these non-coding GWAS SNPs are present in DHSs (2,931 SNPs) or are in complete LD (r^2=1) with SNPs in neighbouring DHS [399]. Individual analysis showed enrichment of GWAS variants in TFBS modulated chromatin structure, as previously mentioned.

Characterising the potential etiological function of previously annotated SNPs associated with disease could implicate new genes underlying certain disorders highlighting novel therapeutic and treatment pathways. The ENCODE data set encompasses a large range of variants which have a vast range of non-coding functions. Currently it has the potential to add novel functional data to genomic regions which were previously uncharacterised.

Histone modification	Putative Functions
H2A.Z	Variant of H2A Histone protein. Enriched at insulators (with CTCF), and upstream and downstream of TSS [400]
H3K4me1	Enriched at enhancers and is a mark of regulatory elements associated with enhancers. Enriched downstream of TSS [401]
H3K4me2	Mark of regulatory elements associated with promoters and enhancers [402]
H3K4me3	Mark of regulatory elements at active promoter sites and transcription starts [402]
H3K9ac	Mark of active regulatory regions with a preference for promoters [403]
H3K9me1	Unknown, preference for 5' end of gene [404]
H3K9me3	Mark of transcriptional repression associated with constitutive heterochromatin and repetitive elements
H3K27me3	Mark of active regulatory elements which may distinguish active promoters and enhancers from their inactive counterparts
H3K36me3	Mark of actively transcribed regions found predominantly near exons [402]
H3K79me2	Transcription-associated mark with a preference for the 5' of genes [404]
H4K20me1	Unknown, preference for 5' end of gene [404]

Table 5. Histone Markers in the Genome

1.4.3 UTILISING DATA SETS SUCH AS ENCODE TO INVESTIGATE NGS

ENCODE data provide a rich resource for researchers investigating variation that is located outside protein coding regions. Prior to its publication researchers depended on bioinformatic extrapolation from TFBS prediction software and conducting extensive research to determine whether these findings were real or not. Even then the evidence may not have conclusively defined the regulatory effect that variants may have been having. The accessible nature of the ENCODE data provides a platform for researchers to tailor their experiments to specific regulatory functions. The range of cell types investigated is extensive but a number of cell types, as described previously, are over represented in the data release. Projects such as ENCODE have led to a paradigm shift in the idea that intragenic and intergenic regions once classified as containing "junk" DNA, now represent a treasure trove of biological structures which may be extremely important for disease studies. As such the ENCODE data provide an exemplary model for the effects a collaboration can have on instigating alternative approaches to research.

Chapter 2 Aims of Thesis

The main aim of this thesis was to investigate members of the L-type group of calcium channel genes, and genes encoding auxiliary subunits of the calcium channel family to elucidate genetic variation that may increase the susceptibility of an individual to develop bipolar disorder.

The specific aims of the thesis were:

- To screen calcium channel genes, previously implicated in GWAS studies, for de novo or rare mutations using a method called High-resolution Melting Curve Analysis to identify variation conferring disease susceptibility to individuals with bipolar disorder.
- To investigate variation detected using High-resolution Melting Curve Analysis that is present in promoter regions, 3' and 5'UTR regions because these are commonly missing from exome capture sequence data.
- To genotype variants of interest in the UCL extended case control cohort.
- To investigate variation present in any of the calcium channel genes from whole genome sequencing data of 99 individuals with bipolar disorder. To follow up any potentially interesting mutations using genotyping in a larger expanded case control cohort.
- To screen the third intron of *CACNA1C* for mutations which may be giving rise to the strongest association signal which has been implicated through multiple GWAS studies and determine if these mutations are located in a functional region as denoted by ENCODE data.
- To determine the potential role of any base pair changes found to be associated with bipolar disorder in the calcium channel family of genes in this study utilising functional assays.

Chapter 3 Material and Methods

3.1 General Methods

In this section all of the materials and methods used for the different experiments will be outlined and discussed. Methods which were specific to both the high-resolution melting curve analysis and the next generation sequencing analysis will be discussed separately.

3.1.1 UCL research Subjects

3.1.1.1 Bipolar Sample

This study included 1,890 affected BD research subjects. These were sampled in three cohorts. The first cohort, UCL1, includes 506 BD-I research subjects defined by the presence of mania and hospitalisation according to research diagnostic criteria (RDC) [232, 233]. The second and the third cohorts, UCL2 and UCL3, consist of 593 and 791 subjects respectively with classifications of BD-I or BD-II, described in the introduction. Ancestry screening was used as a selection criterion for the inclusion of cases. Samples were included if at least three out of four grandparents were English, Irish, Scottish or Welsh and if the fourth grandparent was non-Jewish European, before the EU enlargement in 2004. All participants provided signed consent and National Health Service (NHS) multicentre research ethics was obtained.

BD research subjects had been given a clinical diagnosis by the ınternational Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) of BD and then needed to fulfil the criteria for the lifetime version of the Schizophrenia and Affective Disorder Schedule (SADS-L) [405]. They were additionally rated with the Operational Criteria Checklist (OPCRIT) [406], which provides DSM–III-R (1987) 3rd ed [407].

3.1.1.2 Schizophrenia Sample

The schizophrenia sample cohort consists of 605 SCZ samples which had previously fulfilled the ICD10 criteria for SCZ made by NHS psychiatrists and needed to fulfil the criteria for SCZ (SADS-L) [405]. Ancestral requirements for samples were the same as those described previously for the BD cohort. Samples were further rated with the OPCRIT [406] provided by the DSM–III-R (1987) 3rd ed [407]. NHS multicentre research ethics approval was obtained. All participants provided signed consent.

3.1.1.3 Alcohol Dependence Syndrome Sample

The UCL Alcohol dependence syndrome sample consisted of 1225 case samples, recruited as part of the UK-COGA (United Kingdom Collaborative Study on the Genetics of Alcoholism) study [408]. Diagnosis was conducted using a modified version of the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA-I) questionnaire for the United Kingdom. These samples met the diagnostic criteria according to the DSM-IV and ICD-10 and were of UK or Irish ancestry as defined previously 3.1.1.1 .These samples were additionally rated using the OPCRIT [406] provided by the DSM–III (1980) 3rd ed [407]. NHS multicentre research ethics approval was obtained. All participants provided signed consent.

3.1.1.4 Control Samples

The sample of 1095 controls comprised 614 screened subjects who had no first degree family or personal history of psychiatric illness and an additional 480 unscreened normal British subjects obtained from the European Collection of Animal Cell Cultures (ECACC). National Health Service multicentre research ethics approval was obtained. All participants provided signed consent. Controls were screened using the SADS-L to ensure they did not meet the criteria for any psychiatric illness were included if they had no family or personal history of mental illness and additionally if they met the ancestry criteria as defined above (3.1.1.1) [409].

3.1.1.5 OTHER DATASETS

A number of data sets were used in this thesis to determine the allelic frequency of variants in healthy controls in the general population and two other SCZ datasets. Three control datasets were used; the exome variant server (EVS) (http://evs.gs.washington.edu/EVS/), the 1000 Genome dataset (http://www.1000genomes.org/) and data from the Exome Aggregation Consortium (ExAC) (http://exac.broadinstitute.org)

The EVS is part of the National Heart, Lung, and Blood Institute GO Exome Sequencing Project (ESP). It contains exome sequencing data on 6503 unrelated individuals with heart, lung and blood disorders from a number of different populations. This dataset contains information on 2203 African-Americans and 4300 European-Americans. The 1000 Genomes project dataset contains whole genome sequencing information on 2577 individuals, including 503 unrelated individuals with European ancestry. The ExAC comprises exome sequencing data of 60,706 unrelated individuals. The European cohort in the ExAC consisted of 33,370 individuals [410].

Two SCZ datasets were used in this study. These were the UK10K SCZ cohort that contains exome sequencing data for 1390 individuals and the Swedish schizophrenia cohort comprising of 2,536 individuals [411]. The ExAC dataset contains data from Swedish SCZ and BD studies (12119); these include the 2453 individuals from the previously mentioned Swedish exome cohort. Where possible the 2453 individuals were removed from the overall control sample to ensure no overlap occurred as the sequence calls for each individual was previously known.

3.1.2 GENOMIC DNA EXTRACTIONS

DNA samples were collected from blood samples from the UCL1 cohort, saliva samples for the UCL2 cohort and a mixture of both blood and saliva for the UCL3 samples. DNA was obtained from samples in UCL1 and UCL3 from blood samples using a modified Puregene protocol in a two part process. DNA from UCL2 and UCL3 was extracted from saliva samples following the Oragene protocol for DNA extraction (DNA Genotek, Ottowa, Canada).

3.1.2.1 GENOMIC DNA EXTRACTIONS FROM BLOOD

The blood samples, stored at -20°C, were defrosted in a water bath at 37°C over a period of 30 minutes to one hour. This is done over a short period of time to prevent cellular damage and the release of enzymes that may interfere with the quality of DNA. The defrosted blood was transferred to a 50ml Falcon tube, to which 30ml of 1x red blood cell lysis buffer solution [100mM NaCl, 100mM EDTA] (Appendix II, II.I) was added. The blood lysis mixture was incubated at room temperature (RT) for five minutes, inverting three times during this period to ensure mixing occurred. The lysis solution breaks down the erythrocytes (red blood cells) and leaves the white blood cells, containing the DNA, intact. The blood lysis mixture was then centrifuged at 3,000rpm for 5 minutes at 4°C. The supernatant was discarded and 20mls of 1x red cell lysis buffer was used to resuspend the pellet. The incubation and centrifugation steps were repeated a final time with 10mls of 1x red cell lysis solution. The addition of three separate volumes of red blood cell lysis buffer aids the removal of excess proteins and other contaminants that may be found in the blood. When a clear white blood cell pellet was left at the bottom of the tube, 15ul of Proteinase K was added to the pellet and the sample was vortexed until it was homogeneous. Proteinase K is a broad spectrum serine protease that digests any proteins that may be remaining in the solution (Appendix II, II.I). Additionally proteinase k limits the activity of intracellular DNAse that could degrade the DNA. Following the addition of Proteinase K, 10mls of cell lysis buffer [10 mM Tris-HCl pH 8.0, 25 mM EDTA, 0.5% SDS] (Appendix II, II.I) was added and the samples were incubated in a water bath at 55°C for one hour (the optimum temperature for the protease activity of Proteinase K is 65°C). After one hour the samples were removed and placed directly on ice. At RT 3.33mls of protein precipitation solution [5M Ammonium Acetate] was then added. The samples were left to cool for 10-15 minutes on ice, vortexing intermittently. The main component of the protein precipitation solution is Ammonium Acetate (NH_4OAc) (Appendix II, II.I). Addition of NH_4OAc increases the protein-protein interactions that occur in solution. The increase of salt concentration causes proteins to aggregate and precipitate out of solution. Samples were then spun at 3000rpm for ten minutes at 4°C. Following the spin the supernatant was removed using a sterile pipette and transferred to 10mls of isopropanol. Each tube was gently inverted a number of times until the DNA in

solution precipitated out. The samples were then centrifuged for five minutes at 3000rpm. This ensures that any DNA that has precipitated out is tightly pelleted on the bottom of the tube. The isopropanol is then removed from the tubes and the DNA washed using 10mls of 70% ethanol. The samples were centrifuged again for five minutes at 3000rpm. The ethanol was carefully poured off and the Falcon tubes were inverted on tissue paper to remove any remaining ethanol. Samples were left to air dry for 15-20 minutes until all of the ethanol was removed. Low tris-EDTA [Tris-Cl (1 M, pH7.5-8), EDTA (0.5 M, pH 8.0)] (TE) (Appendix II, II.I) buffer solution was added to rehydrate the DNA and samples were left shaking overnight at RT. Once the DNA was dissolved samples were quantified using the procedure described in 3.1.3. DNA samples were normalised to form stock solutions of between 50ng/µl 500ng/µl and working aliquots were created by diluting samples to 25ng/µl.

3.1.2.2 Genomic DNA Extraction from Saliva

DNA was collected from some participants using the Oragene DNA Self-Collection Kit (DNA Genotek). The Oragene Self-Collection Kit is a non-invasive manner of sample collection, which increases the ease at which samples may be collected from participants that are not able to comply with blood collection. Upon collection of the sample and closure of the tube a DNA stabiliser is released and mixed with the sample. The samples are stable at RT for up to 5 years from collection. DNA from the saliva samples was extracted as per the directions detailed in the manufacturer's instructions. The saliva samples containers were incubated at 50˚C for a minimum of one hour. The purpose of the incubation is to ensure that nucleases present in the sample are permanently inactivated to prevent any DNA degradation. Samples were transferred to a 15ml centrifuge tube and the volume of each sample was noted. Impurities were removed from the sample by adding 1/25th volume of prepIT-L2P (DNA Genotek), vortexing the samples for a few seconds, and incubating each tube on ice for ten minutes. Samples were centrifuged at 3000rpm at RT for ten minutess. The supernatant was then transferred to a clean 15ml tube and the pellet was discarded. This step ensures that the majority of the contaminants are removed from solution. DNA was precipitated from the supernatant by adding 1.2x volume of 100% ethanol to the

solution and the tubes were mixed by inversion. The samples were left to stand for ten minutes at RT. Lower temperatures are not recommended as impurities may co-precipitate with the DNA. The tubes were centrifuged at 3000rpm for ten minutes to pellet the DNA. Following centrifugation the supernatant was removed from the tubes and the DNA was washed with 1ml of 70% ethanol. Samples were left for one minute at RT and the ethanol was completely removed after this time. The DNA was rehydrated using between 200-500ul of TE buffer solution depending on the amount of DNA that precipitated in the tube. Samples were left shaking for a minimum of one hour at 50°C and overnight at RT to ensure the DNA dissolved in the TE buffer. Once the DNA was dissolved samples were quantified using the procedure described in 3.1.3. DNA samples were normalised to form stock solutions of between 50ng/µl 500ng/µl and working aliquots were created by diluting samples to 25ng/µl.

3.1.3 DNA QUANTIFICATIONS

All DNA samples were quantified by the Qubit® 2.0 Fluorimeter (Invitrogen, UK) using the dsDNA BR (broad range) kit, which is highly selective for double stranded DNA over RNA. The Fluorimeter was calibrated using two standards provided by the manufacturer: 0ng/µl and 100ng/µl. 1ul of each sample was added to a master mix containing 199ul of Qubit® dsDNA BR buffer and 1ul of Qubit® dsDNA BR reagent. Samples were then incubated in the master mix for two minutes to allow the assay to reach maximum fluorescence and read using the fluorimeter. Each sample was read twice and samples were quantified in duplicate to ensure the average readings did not cross a standard error threshold of 5. Samples which passed the standard error cut off threshold were requantified following further shaking of the sample tubes. The reason for aberrant concentration readings is often due to undissolved DNA.

In addition to the use of the Qubit, the NanoDrop 2000 spectrophotometer (Thermo Scientific, UK) was used for quantifying DNA when there were lower sample numbers and the purity of the DNA sample needed to be checked. This works on low volumes of DNA and outputs DNA concentration per µl using 260nm absorbance rate. DNA purity is measured using the 260/280nm, and 260/230nm

outputs. For DNA the 260/280 ratio should be around 1.8 and for pure DNA the 260/230 ratio should be between 2.0-2.2. Deviation in these ratios generally reflects contamination in the DNA sample or degradation of the DNA.

3.1.4 Polymerase Chain Reaction

Polymerase chain reaction (PCR) is an *in vitro* technique that enables researchers to clone a specific region of DNA from small amounts of starting product. It is inexpensive and efficient and allows specific targeting of regions of interest. PCR consists of three steps that are vital for the appropriate region to be amplified: DNA denaturation, primer annealing and polymerase strand elongation. Denaturation of the sample causes the two strands of DNA to separate; this is done by heating the DNA to 95°C for two minutes. PCR buffers provide an optimal pH for the reaction and contain cations such as K^+, which aids stabilization of primer annealing by binding to the phosphate groups on the backbone of DNA. Another important cation present, NH_4^+, can destabilise the hydrogen bonds between mismatch DNA bases by interactions during thermocycling, which increases the ratio of specific: unspecific final product.

Longer denaturation time can be used if the region of interest has a high GC content. Primer hybridisation occurs when assay of DNA, primers, dNTPs and polymerase are incubated at an optimal annealing temperature, generally 5°C lower than the primer template melting temperature. Polymerase elongation time is calculated based on one minute per kilobase (kb) of DNA to be amplified. The ratio of specific product to unspecific product increases exponentially as the PCR reaction progresses.

3.1.4.1 Primer Design for PCR

Primers were designed for PCR and real-time PCR using a number of guidelines. Primers should be between 18-22bp long, each having a melting temperature (T_m) between 42-65°C. The T_m of the primer pairs should not differ by more than 5°C and the GC content of each primer should be in the range of 40- 60%. Primer design should be avoided in genomic regions where repeats, or a run of 3 or more

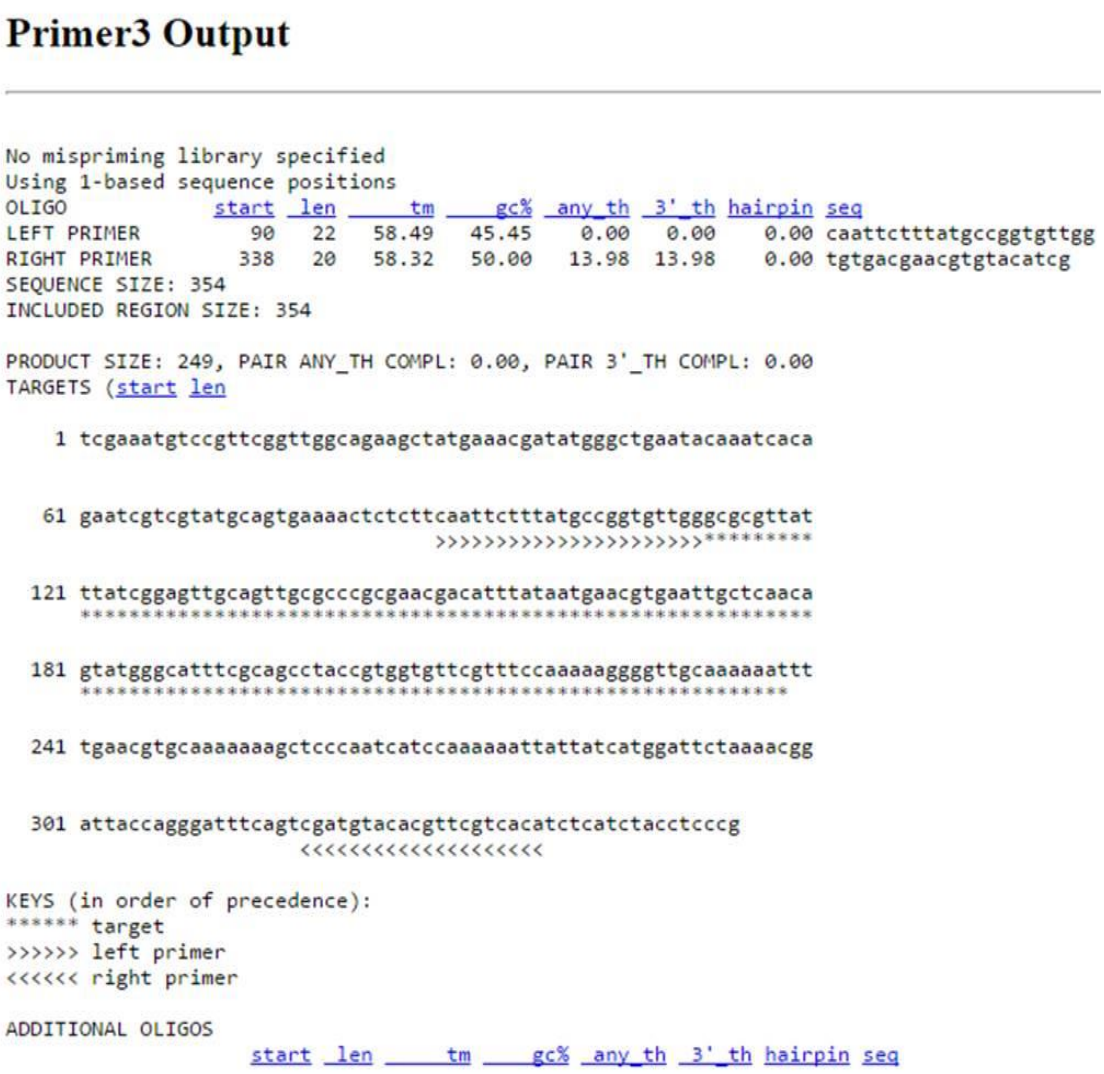
bases are present. Primers should contain a C or G base at the 5' or 3' end: this aids specific binding due to the stronger bonding of G and C bases.

The specificity and concentration of the primers designed are essential for maximum efficiency and targeting the specific region of interest. If the concentration of the primers is too high it raises the probability of the primers misaligning or annealing to regions other than the sequence of interest resulting in a number of spurious PCR products. Similarly, primer concentrations below a certain threshold will also have adverse effects on the PCR amplified as it restricts the amount of primer template available to anneal to the single stranded DNA. The final concentration for PCR primers should be between 0.05-1μM for the final reaction, with typically 0.1-0.5μM of each primer.

Additionally some primers were designed using Primer3 online software (http://primer3.ut.ee/) [412, 413]. The region of DNA to be amplified is used as an input file and the primers are designed in the adjacent regions, defined using square brackets [ ]. An example of the output from Primer3 is shown in Figure 18. Following their design, primers were analysed using NetPrimer (Premier Biosoft, http://www.premierbiosoft.com/NetPrimer/AnalyzePrimer.jsp). This analysis covers a number of primer properties such as determining the primers propensity for the formation of secondary structures such as hairpin loops or dimers, -3' and -5' end stability and repeats and runs. Secondary structures often result in poor or no yield of the targeted amplicon as they reduce the amount of primer that is available to anneal to the target DNA sequence

3.1.4.2 PCR Optimisation

Optimisation for the PCR assay was conducted using four previously established conditions. These were the primary conditions used when starting an optimisation. If a genomic region appeared difficult to amplify using the standard optimisation conditions other conditions were tried and tested that specifically targeted the problem encountered in this region, such as high GC content. GC-rich genomic regions are refractory to amplification as they have a higher probability of forming secondary structures due to the stronger nature of the bonds present between the cytosine and guanine base. DMSO lowers the Tm of the DNA by binding to the



Primer3 Output

```
No mispriming library specified
Using 1-based sequence positions
OLIGO            start  len      tm      gc%  any_th  3'_th hairpin seq
LEFT PRIMER         90   22   58.49   45.45    0.00   0.00    0.00 caattctttatgccggtgttgg
RIGHT PRIMER       338   20   58.32   50.00   13.98  13.98    0.00 tgtgacgaacgtgtacatcg
SEQUENCE SIZE: 354
INCLUDED REGION SIZE: 354

PRODUCT SIZE: 249, PAIR ANY_TH COMPL: 0.00, PAIR 3'_TH COMPL: 0.00
TARGETS (start len

    1 tcgaaatgtccgttcggttggcagaagctatgaaacgatatgggctgaatacaaatcaca


   61 gaatcgtcgtatgcagtgaaaactctcttcaattctttatgccggtgttgggcgcgttat
                                >>>>>>>>>>>>>>>>>>>>>>*********

  121 ttatcggagttgcagttgcgcccgcgaacgacatttataatgaacgtgaattgctcaaca
      ************************************************************

  181 gtatgggcatttcgcagcctaccgtggtgttcgtttccaaaaaggggttgcaaaaaattt
      ************************************************************

  241 tgaacgtgcaaaaaaagctcccaatcatccaaaaaattattatcatggattctaaaacgg


  301 attaccagggatttcagtcgatgtacacgttcgtcacatctcatctacctcccg
                       <<<<<<<<<<<<<<<<<<<

KEYS (in order of precedence):
****** target
>>>>>> left primer
<<<<<< right primer

ADDITIONAL OLIGOS
                   start  len      tm      gc%  any_th  3'_th hairpin seq
```

Figure 18. Primer 3 results for primers in the intronic region of CACNA1C.

Information pertaining to hairpin loops, secondary structures, GC% is also given by the software.

cytosine residue changing DNA conformation; this change in conformation facilitates the annealing of primers to the amplicon region and enhances the amplification of the region of interest.

The addition of 7-deaza-dGTP has been reported to weaken the bonds in base-base interactions, lowering the melting temperature required for the PCR [414]. Aside from the above mentioned exceptions, the conditions used for standard PCR optimisation varied the concentration of $MgCl_2$ present in the reaction, and the addition or absence of betaine. The addition of Mg^{2+} ions is essential as they form soluble complexes with the dNTP mixtures removing the beta and gamma

Volume (µl)	2.0 mM $MgCl_2$		2.5 mM $MgCl_2$	
	5 M Betaine	No Betaine	5 M Betaine	No Betaine
10x Buffer	2.5	2.5	2.5	2.5
Betaine (5M)	5	0	5	0
$MgCl_2$ (50mM)	1	1	1.25	1.25
dNTPs (25mM)	0.2	0.2	0.2	0.2
F oligo (10µM)	1	1	1	1
R oligo (10µM)	1	1	1	1
Water	11.8	16.8	11.55	16.55
Taq	0.5	0.5	0.5	0.5
DNA (12.5ng/µl)	2	2	2	2
Total	25	25	25	25

Table 6. Standardised PCR reactions mixes for optimisations.

Step	Temperature (°C)	Time	Cycles
Denaturation	94	5 mins	1
Denaturation	94	30 sec	
Annealing	60	30 sec	35
Extension	72	30 sec	
Hold	72	10 mins	1

Table 7. PCR conditions used for primer optimisation

phosphates from the dNTPs allowing for the reaction to continue. The four primary conditions used for primer optimisation are listed in Table 6 and the PCR thermocycling conditions used are presented in Table 7.

3.1.5 Agarose Gel Electrophoresis

For simple and efficient detection of PCR products following PCR amplification agarose gels are used. Agarose is a natural polymer, extracted from seaweed, which upon heating and cooling in a buffer forms a gel through hydrogen bonds. Agarose gels are porous in nature and allow for the DNA to run through the gel when an electric current is applied. We used a continuous buffer to prepare the gel and allow for the migration of DNA along the gel. Continuous buffer systems are ones where the concentration of the buffer is the same in both the gel and the tank. PH of the buffer is essential for the electrophoretic mobility; a 1x Tris-Borate-

EDTA [TBE, 90mM Tris, 90mM Boric, 2mM EDTA, pH 8.0] (Appendix II, II.II) was used for these experiements. Optimal agarose concentration is dependent solely on the size of the band that is being run. The lower the concentration of agarose, the faster the DNA will travel through the gel. For PCR products >1kb it is recommended to use a 0.8% gel, and for products <1kb it is essential to use a 1% gel. This ensures the adequate separation of the bands, allowing for accurate detection of the specific PCR product. If there are a number of small bands presents and the agarose concentration is too low, these will cluster together making it difficult to discern whether the correct band is present.

For the every gel 0.9g of agarose was added to a conical flask (1% w/v) and sufficient 90ml of 1xTBE buffer was added to the flask (Appendix II, II.II). The amount of agarose and TBE buffer depends solely on the size of the gel needed. The agarose was left to hydrate for a minute and then boiled for over one minute to ensure all of the agarose powder had dissolved in the buffer and cooled until the temperature was 50-60°C. Ethidium Bromide (EtBr) was prepared at 10mg/ml and added to the gel prior to casting, at a final concentration of 5ug/ml (Appendix II, II.II). EtBr is an intercalating agent that fluoresces under UV lights and is used to visualise the DNA at the end of the gel run. The gel was then poured into a casting tray which had combs placed to form wells and left to set for 30-45 minutes. After the gel had fully solidified it was placed in the electrophoresis gel tank and the buffer was added until it is 3-5mm above the gel.

Five µl of PCR products were mixed with 1µl of 5x loading buffer (Bioline UK) prior to the addition of the DNA to the wells in the gel. The loading buffer contains glycogen or sucrose and adds density to the DNA allowing it to remain in the gel well rather than diffuse in the buffer. The buffer additionally adds a visible colour allowing for the progress of DNA migration to be seen on the gel. Specific molecular markers, Hyperladder markers I and IV (Bioline UK) were added to determine the size of the bands seen as they ran in tandem. 5µl of the PCR product and the molecular marker were pipetted into separate wells. The electrophoresis tank was then assembled fully and connected to the power supply. Gels were primarily run for 30 minutes at 120V; this differed depending on the size of the bands of DNA (the bigger the molecular weight of the product the slower it moves through the gel). Following sufficient separation of the bands the gel was

visualised using a UV transilluminator (UVP Gel-doc-it Imaging systems, UK) to determine the correct amount of PCR product was present. This image was captured using an internal camera.

3.1.6 MAMMALIAN CELL CULTURE

To investigate the effect variation in specific genes may have on a cellular system, or gene expression we used the human embryonic kidney 293 transformed cell line (HEK293) that were a kind gift from Josef Kittler (UCL) to transfect in DNA. These cells line were first characterised by Graham, Smiley, Russell, and Nairn [415] following transformation of a cell line using human adenovirus type 5. HEK293 cells have been widely used for recombination protein experiments since the cell line's first description. A versatile cell line, its cellular machinery is capable of generating mature functioning proteins, allowing for biological evaluation of cellular trafficking and expression can be studies in tandem. Some speculation around the origin of these cells has persisted, while they have been obtained from the kidney, researchers such as Van der Eb predict that they may be neuronally derived given the expression of some mRNA and other expressed proteins that are normally found in neuronal lineages [415].

3.1.6.1 SUBCULTURING OF HEK293 CELLS

HEK293 cells were cultured as mono layers in a plastic tissue culture flask at 37°C in 10ml of DMEM (Dulbecco's Modified Eagle Media), 10% FBS (Fetal Bovine Serum), 1% Penicillin (50U/ml)/Streptomycin (50µg/ml) (Appendix II, II.III). The media is buffered by biocarbonate and must be kept in an atmosphere of 95% O_2 and 5% CO_2 and is pH sensitive. Cells were grown in Nunc T-75 sterile filter cap flasks (Thermo Scientific, UK) until they were between 80-90% confluent. When they reached an appropriately confluent stage they were split into the required number of flasks. Cells were split in a class II microbiological safety hoods to prevent contamination. All surfaces and anything placed inside the hood were sprayed with 70% ethanol prior to the introduction of the cells to the hood to further prevent contamination. Cells were removed from the incubator and the DMEM culture media was removed. The cells were washed using 5ml of PBS

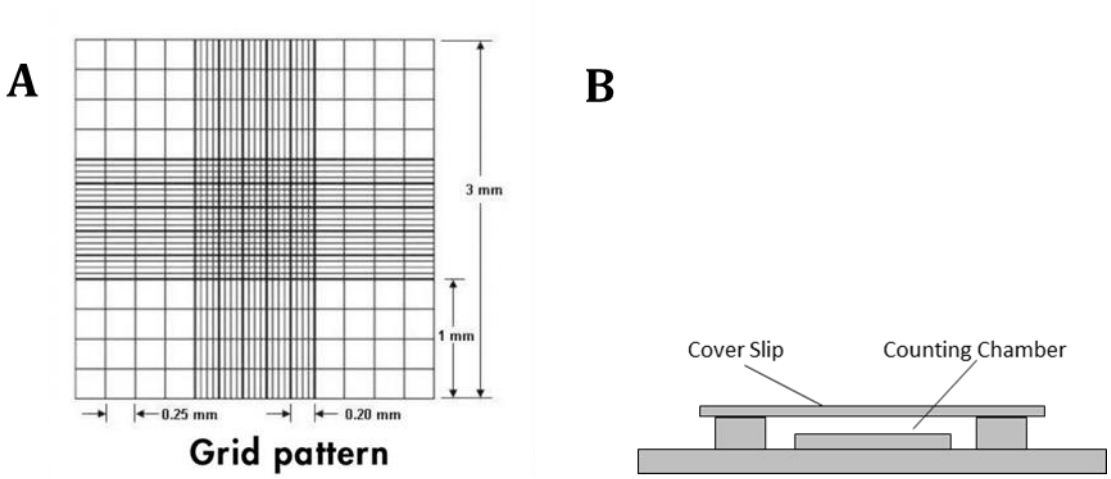


Figure 19. Neubauer-Improved Haemocytometer.

A. The grid pattern of the counting chamber of the haemocytometer is divided into specific sections of known dimensions. Each of the small squares are 1mm in height, and 0.2mm in width. This allows for the accurate measurement of the number of cells in solution by the counting the cells in the specific squares [source: www.peqlab.de, Neubauer grid] **B**. The construction of the Neubauer Haemocytometer showing the counting chamber and the cover slip.

(Phosphate Buffered Saline) to remove any remaining culture media as foetal calf serum contains protease inhibitors (Appendix II, II.III). Trypsin is an active protease and the cells were treated with trypsin to remove the monolayer of cell growth from the bottom of the T-75 flask. Five ml of Trypsin was then added and cells were left in the incubator for 5 minutes.

Following the incubation period the flask was removed from the incubator, the flask was gently tapped to ensure the full detachment of cells. The flask was checked under the microscope to ensure all cells had detached fully and trypsinization was stopped by the addition of 5ml of DMEM, 10% FBS, 1% Pen/Strep. The mixture was then removed from the flask and placed in a 15ml Falcon tube and spun for 5 minutes at 1000rpm to pellet the cells. The supernatant was discarded and the cells were resuspended in 10ml of DMEM, 10% FBS, 1% Pen/Strep. Various volumes were then added to newly prepared T75 flasks containing an appropriate volume of DMEM, for example a 1:10 dilution would involve 9ml of DMEM in the T75 flask and the addition of 1ml from the

resuspended cells. The remaining unused cells were disposed of in a 4% Virkon solution (RMS, UK).

3.1.6.2 Counting of Cells

Prior to preparation of cells for transfection cell counting was undertaken using a Neubauer-Improved haemocytometer (Marienfeld, Germany) (Figure 19). The haemocytometer is a modified glass slide with indentations and line markings around a specific region of known area and depth that facilitates the counting of cells. As the specific measurements are known, the amount of cells in a volume of liquid can be deduced by taking an average count of cells found in specific areas. The gridded area of the haemocytometer consists of nine $1mm^2$ squares. These are further divided up so that the central square results in $0.0025mm^2$ squares.

The distance between the cover slip and the counting chamber is $0.1mm^2$. The haemocytometer and the cover slip were washed prior to use with 70% ethanol. The cover slip was moistened and pressed down gently on top of the chamber until it was aligned and fixed. Cells were prepared following the protocol for subculturing, as outlined in 3.1.6.1, but following the centrifugation step, cells were resuspended in 5ml DMEM, 10% FBS and 1% Pen/Strep. The cellular suspension was mixed gently to ensure the suspension was homogeneous and 10µl of cell suspension was applied to the edge of the cover slip and enters each of the counting chambers through capillary action. Cells were counted four times using the centre square (Figure 19) to estimate the average number of cells in the 10ul added. A hand counter was used to count the number of cells in the selected area. The average number of cells counted was determined by dividing the total number of cells counted by 4. The number of cells in the area is equivalent to the number of the cells per square $x10^4$/ml. The volume needed to achieve the appropriate number of confluent cells was then determined for seeding of the cells into 6-well plates.

3.1.6.3 Seeding of cells

Twenty four hours before transfection, HEK293 cells were seeded into 6 well plates at approximately 5×10^5 cells per well in 3ml of DMEM, 10% FBS and 1%

Penicillin/Streptomycin. Cells were left for 24 hours in the incubator at 37°C and 5% CO_2.

3.1.6.4 CATIONIC LIPID MEDIATED CELLULAR TRANSFECTION

Lipofectamine was used to transfect the created plasmid DNA vectors into the HEK293 cells (Appendix II, II.IV). Lipofectamine is a cationic liposome that mediates the introduction of DNA into the cell with higher efficiency then previously used methods such as calcium phosphate coprecipitation, which can result in low transfection yields and cell toxicity. The cationic lipids consist of a positively charged head group with a number of hydrocarbon chains. This attribute of the lipid is responsible for mediating interaction between the cationic lipid and the phosphate backbone of the nucleic acid. It also instigates DNA condensation for complex formation. The negatively charged DNA binds spontaneously to the positively charged head of the cationic lipid. Not much is known about the exact entry of the DNA-lipid complex into the cells but it is thought that the positively charged head interacts with the negatively charged cell membrane and the transfection complex enter the cells through endocytosis. Cells were plated so that they reached 70-90% confluency at time of transfection. To prepare the cells for transfection the growth media was removed and the cells were washed with 1ml of PBS to ensure the removal of any antibiotic, as this can interfere with transfection efficiency. Growth media was then replaced with 1.2ml of Opti-MEM® reduced media serum (Life technologies, UK), which is a modified Eagle's Minimum Essential Media suitable for cationic lipid transfections. Due to its higher concentration of HEPES (2-[4-(2-hydroxyethyl) piperazin-1-yl] ethanesulfonic acid) and sodium bicarbonate in the buffer it is more suitable for transfecting cells when creating DNA complexes. Media such as DMEM, when supplemented with antibiotics and FBS, can negatively affect transfection efficiency in wide variety of cell lines. Opti-MEM reduced media serum is also supplemented with L-glutamine, hypoxanthine, thymidine, sodium pyruvate, trace elements and growth factors and was used during the transfection period to facilitate greater transfection efficiency.

Transfections were performed using Lipofectamine 2000® (Life Technologies). In an eppendorf tube 150µl of Opti-MEM media was mixed with 4µl of Lipofectamine 2000® and the mixture was left to incubate for 10 minutes. In a second eppendorf

tube 150µl of Opti-MEM was combined with 500ng of the DNA of interest. Following the lipofectamine and Opti-MEM incubation period the solution was added to the DNA/Opti-MEM mixture and Lipofectamine/Opti-MEM mixture were incubated for 20 minutes at RT. The transfection complex was added into each well, one drop at a time, and the plate was left in the incubator. After four hours the Opti-MEM media was removed and replaced with DMEM, 10% FBS and 1% Pencillin/Streptomycin.

3.2 Methods for results Chapter 4 and 5: High-resolution melting curve analysis in Calcium Channel genes

3.2.1 Research Samples

High-resolution melting (HRM) curve analysis was conducted on the first 1000 individuals collected for the UCL BD sample as described in section 3.1.1.1. Follow up KASPar genotyping of variants was conducted using the BD sample and the initial control sample as described in section 3.1.1.4 Additional BD samples, following the ancestry classifications previously established, and control DNA samples were collected following the primary HRM analyses and were included in the KASPar genotyping analysis. The SCZ and ADS samples, described in sections 3.1.1.2 and 3.1.1.3 respectively, were used for follow-up genotyping. Additional databases were used to determine variant frequency. These include the 1000 Genomes project, the Exome Aggregation Consortium cohort, the Exome variant server database, the UK10K SCZ sample and the Swedish SCZ Exome sample.

3.2.2 High-resolution Melting

HRM is a fluorescent PCR technique that allows for the rapid high-throughput analysis of genomic regions for the presence of variants. HRM is used a priori to sequencing to allow for the rapid screening of multiple cases and control samples, thus reducing cost and time and effort in identifying novel and previously identified variation in the genome. HRM is a real time PCR technique that amplifies the region of interest using standardised PCR techniques, as mentioned in (3.1.4), in the presence of an intercalating fluorescent dye. There are a number of dyes that

can be used in for the HRM analyses, we utilised a saturating dye. These dyes do not interfere with the melting temperature (T_m) of the amplicon, or inhibit polymerase activity. In addition to this saturated dyes can be used at higher concentrations than unsaturated dyes. This ensures a higher rate of incorporation of the dye into dsDNA. The fluorescence measure is indicative of the DNA state, with high fluorescence seen when the dye is incorporated in dsDNA, and low levels of fluorescence when it is unbound. In conjunction with the fluorescent dyes the PCR was carried out using a LightCycler 480 (Roche, UK) thermal cycle. Optimal PCR conditions were determined using the conditions in Table 8. Each fluorophore has a specific wavelength which is emitted when it is excited and the incorporation of the dye into the DNA and whether it is bound can be measured using quantitative PCR.

Following the amplification step, the product was slowly denatured in incremental temperatures producing a characteristic melting curve profile. For HRM, the temperature increments increase each cycle by between 0.008 - 0.2 °C increments. This slight decrease in temperature increases the resolution of the melting profile, making it easier to discern differences in the DNA sequence. Unlike low resolution melting curve analysis, which is viewed as the derivative of fluorescence vs temperature ($-\frac{dF}{dT}$) against T, the data derived from HRM must be analysed using the data in the raw form. A comparison is done between the pre-melt temperature and the post-melt temperatures, which is normalised to identify mutations (Figure 20) [416].

3.2.3 PURIFICATION OF SAMPLES

Purification of PCR products ensures the elimination of PCR fragments and probes prior to downstream experiments. Purification of the DNA was completed using the ethanol/EDTA precipitation method. Following PCR it is likely that an excess of salts, the presence of PCR primers, primer-dimers, reaction buffers or dNTPs, could interfere with sequencing or other experiments. MicroCLEAN (Microzone Limited, UK) is a DNA clean up reagent that concentrates the PCR product and removes the previously mentioned reagents (Appendix II, II.V).

Reagents	Sensi $MgCl_2$ 3mM	Sensi $MgCl_2$ 4mM	Sensi $MgCl_2$ 4.5mM	Sensi $MgCl_2$ 5mM	300nM Quanta	400nM Quanta
SensiMix HRM	2	2	2	2	0	0
50mM $MgCl_2$	0.3	0.4	0.45	0.5	0	0
EvaGreen Dye	0.16	0.16	0.16	0.16	0	0
F Oligo(10mM)	0.08	0.08	0.08	0.08	0.15	0.2
R Oligo(10mM)	0.08	0.08	0.08	0.08	0.15	0.2
Accumelt	0	0	0	0	2	2
DNA (12.5ng/µl)	1	1	1	1	1	1
Water	1.38	1.28	1.23	1.18	1.7	1.6
Total	5	5	5	5	5	5

Table 8. Optimisation condition for HRM assays.

Standard conditions used to optimise HRM primers for the specific amplicon of choice. The main changes between the conditions lie in the increased in the $MgCl_2$

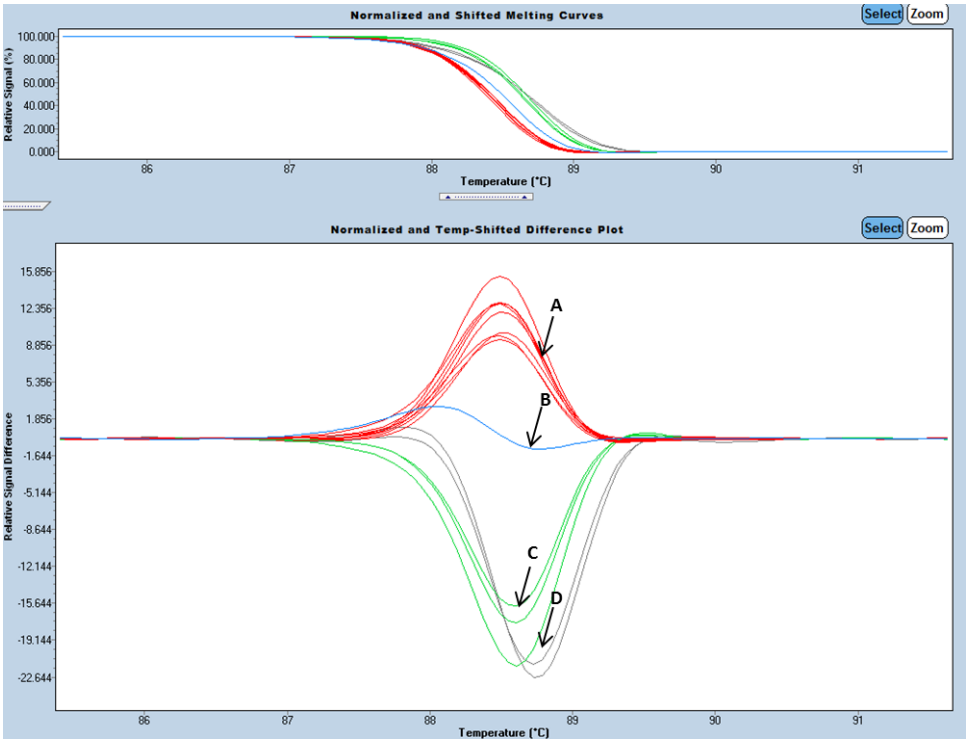


Figure 20. HRM melt curve analysis.

HRM melt curve analysis of *CACNA1D* Exon 1 after normalization. Curve A and C (Red and Green) represent the presence of the wildtype allele. Curve B (Blue) is a different shape and Curve group D show a shift to the right. These are samples which were choosen to follow up by Sanger sequencing.

3.2.3.1 MICROCLEAN OF PCR PRODUCTS FOR PURIFICATION IN 96-WELL PLATES

An equal volume of microCLEAN e.g. 10µl of Microclean for 10µl of PCR reaction, was added to the DNA sample and the plate was shaken using a Mixmate (Eppendorf, Germany) to mix the solution. The plate was placed in a centrifuge at 4000rpm for one hour. To remove the supernatant, plates were overturned on a tissue and the plate was spun in the centrifuge at 1000rpm for one minute. The pellet was then resuspended in 5µl of pre-PCR water (Sigma-Aldrich, UK) and allowed to rehydrate at RT.

3.2.3.2 MICROCLEAN OF PCR PRODUCTS PURIFICATION IN TUBES

An equal amount of microCLEAN reagent was added to the DNA sample and mixed by pipetting. Following five minutes incubation at RT, the tubes were spun in a bench top centrifuge at 13,000rpm for seven minutes. The supernatant was aspirated and spun again for one minute to remove any dregs that may be still present in the tube. The pellet was then resuspended in 5µl of pre-PCR water (Sigma-Aldrich, UK) and allowed to rehydrate at RT.

3.2.4 SANGER SEQUENCING

DNA sequencing was used to validate and confirm the presence of mutations that may have been identified using HRM analysis and genotyping, or created via site directed mutagenesis. DNA samples to be sequenced were tagged with fluorescent dye using BigDye1.3 (Life Technologies) reagents. Each of these fluorescent markers was incorporated into the DNA sequence via thermo cycling (Table 10). Thermo cycling allows for the dyes to be incorporated and creates and amplifies the sequencing extension products, each of which are terminated by one of the four genetic bases (Figure 21) Each of the fluorescent dyes emits a specific wavelength when excited by light making it possible to discern the presence of a specific base at the 3' terminal dideoxynucleotide as A, C, T or G.

The forward and reverse primers used to sequence the region of interest were designed as previously described in section 3.1.4.1 and the region of amplicon to be sequenced was amplified under optimum conditions.

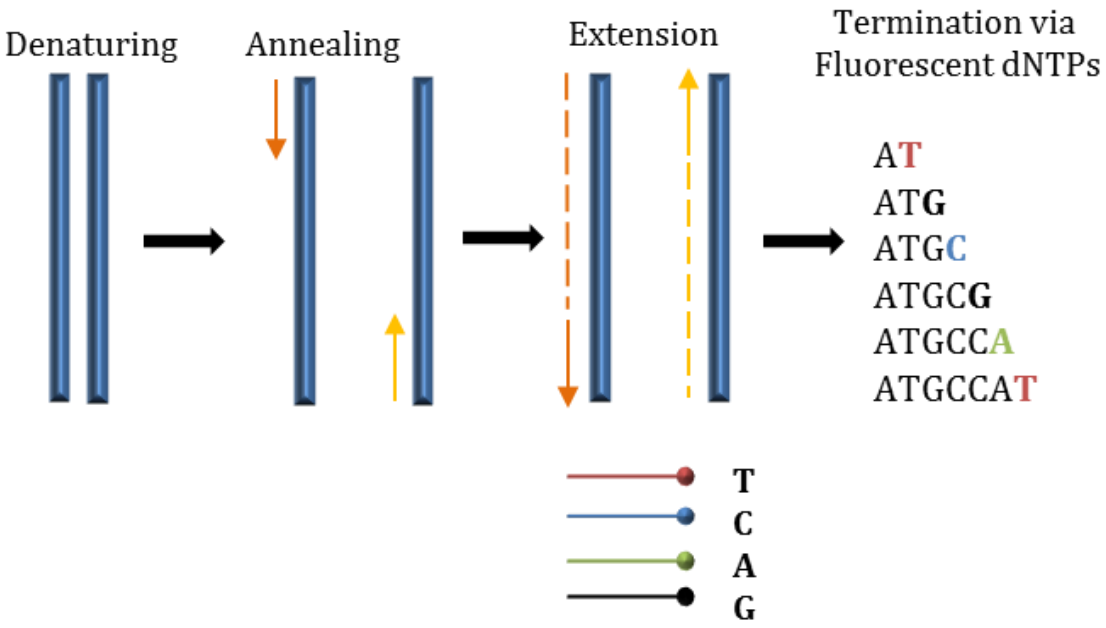


Figure 21. BigDye reaction for Terminator Cycle Sequencing.

Reagents	Volume (µl)
Big Dye v3.1	1
Buffer x5	1.5
Primer F/R	0.32
Water	6.18
PCR Product	1
Total	10

Table 9. Big Dye v3.2 Reaction Mix for sequencing

Step	Temperature (°C)	Time	Cycles
Denaturation	94	5 mins	1
Denaturation	96°C	10 sec	
Annealing	50°C	5 sec	25
Extension	60°C	4 min	

Table 10. Big Dye v3.1 Thermocycling conditions

3.2.4.1 BIG DYE THERMOCYCLING

PCR for the sequencing reaction was conducted using amplified product according to the manufacturer's protocol (Table 9) using the resuspended PCR product which had been cleaned using Microclean (3.2.3). One forward and one reverse condition was done for each sample to ensure adequate coverage of the genetic region was achieved as fidelity to the true sequence somewhat deteriorates as the amplicon gets longer. Having a forward and reverse sample alleviates this issue. Purification and precipitation of the DNA for sequencing was done using EDTA/Ethanol precipitation. 2.5µl of EDTA [125mM] and 30µl of 100% ethanol were added to the 10µl of extension reaction from the Big Dye v3.1 PCR reaction. The precipitation reaction was left at RT for ten mins and was then spun in a centrifuge for one hour at 4000 rpm. To remove the supernatant the samples were spun upside down at 1000g for 10 mins. Samples were then washed with 30µl of 70% ethanol and spun for ten minutes at 4000g. The supernatant was removed again using the procedure previously described. Samples were left at RT for 15 minutes to ensure that all of the ethanol was completely removed. Following this, 15µl of Hi-Di™ formamide was added to each of the wells on the plate to be sequenced or to the individual tubes. Sequencing was conducted using the ABI 3730xl (Applied Biosystems, UK) at a number of locations including North Thomas Regional Genetics Lab, Queen Square, Wolfson Institute, UCL and Source BioScience, Cambridge.

3.2.4.2 SEQUENCING ANALYSIS

DNA sequencing analysis was conducted using the Staden Package [417]. The package assembles the DNA, covering the most essential requisites for sequencing analysis. The Staden package contains two programmes *gap4* and *pregap4*. *Gap4* performs the assembly of the sequence files, joins contigs, allows for pattern recognition, read pair analysis and allows for the visualisation of all contigs present in the sequencing of samples. *Pregap4* provides a graphical interface where the user can conduct quality control on each individual sample. It is then possible, following the input of the sequencing files, to investigate whether variation is present in the sequence which would deviate from the reference sequence obtained from the UCSC genome browser (Figure 22).

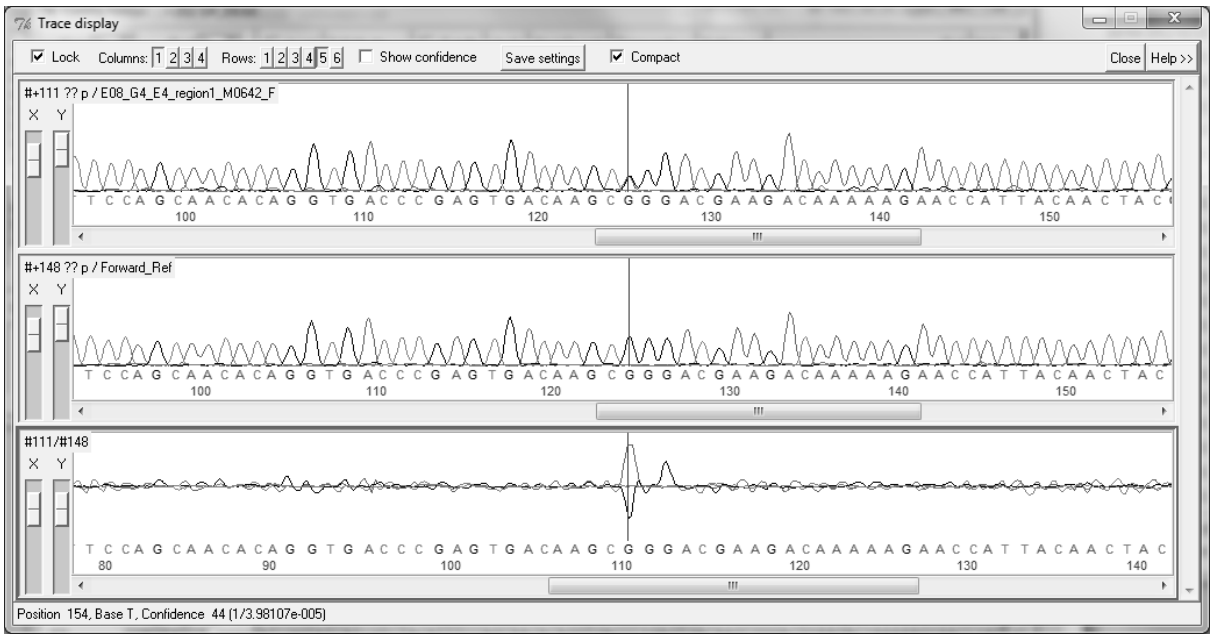


Figure 22. Staden Trace display of a *CACNG4* variant.

The staden package assembles all of the sequencing files inputs and creates contigs through which the user can search for variation in the sequence. Figure 22 shows the presence of a heterozygous mutation G/A in an individual in the fourth exon of the *CACNG4* gene

3.2.5 VARIANT SELECTION FOR KASPAR GENOTYPING

Following HRM screening of variants and sequencing analysis each sample was investigated to determine whether there were mutations present in the genomic regions of interest. Only a minority of samples were chosen to send for sequencing to determine whether variation was present. To investigate allele frequencies in the entire BD and control cohorts the variants need to be genotyped. To qualify for the following genotyping, variants needed to satisfy a number of criteria which are outlined below.

3.2.5.1 BIOINFORMATIC FUNCTIONAL PREDICTION

Bioinformatic prediction software tools are used to infer functionality of variants present in the genome. Different predictive software can produce conflicting results, as the characteristic of the variant analysed are not the same.

For variants present in exons, PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) and Sort Intolerant from Tolerant (SIFT) (http://sift.jcvi.org/) were used to predict whether these variants could be deleterious, or benign. PolyPhen-2 calculates a variant's impact on the structure of the resulting protein by contrasting conservation of the wild-type base with the frequency with which the substitution is seen in the family of related homologous proteins. It also measures the number of structural impacts a non-synonymous variant could have on the protein structure and secondary structure. This is achieved by determining the nature of the amino acid change, polar to non-polar, and whether the inserted amino acid is spatially different in the structure, altering necessary regions in the protein structure such as hydrophobic bonds.

SIFT on the other hand purely focuses on the conservation of a specific base pair, utilising sequence alignment data closely related sequence data using Position-Specific Iterated Basic Local Alignment Search Tool (PSI-BLAST).

The promoter, 5'UTR, and 3'UTR variants were investigated using Transcription Element Search System (TESS) (http://www.cbil.upenn.edu/) [418] and Alibaba2.1 (http://www.gene-regulation.com/pub/programs/alibaba2/index.html). These programmes examine whether variants may have an impact on transcription factor binding sites (TFBS) which are predicted from the DNA sequence. Often a base change may not result in a change of TFBS, but sometimes it may result in a loss of the TFBS and/or the creation of a new one.

3.2.6 SNV GENOTYPING WITH KASPAR

Following variant validation, using sequencing in the genes of interest, and preliminary investigation using the above mentioned software, the single nucleotide variants (SNVs) where further investigated in the complete cohorts of interest. HRM is a useful tool for determining whether variation is present but does not offer the researcher complete information on the frequency of the SNV in that population. The main reason for this is often the variant does not have a significant enough effect on the change of the melting curve to distinctly set apart all of the individuals that carry the variant. As the control cohorts are also not screened

using HRM analysis, genotyping is the perfect way to get a complete overview of the frequency of the variant in both case and control populations.

KASPar (KBiosciences Competitive Allele-Specific Polymerase Chain reaction) genotyping is a simple, rapid, cost effective way to determine allele frequencies in a sample set. KASPar is a homogeneous, FRET based, endpoint genotyping technology consisting of two main components. The first component is the KASPar assay mix, which consists of competing, non-fluorescently labelled, forward primers targeting the variant of choice, and one common reverse primer. The second component is the KASPar master mix, which contains the FRET cassette and Taq polymerase in an optimised buffer solution. The two forward competing primers, targeting the wild-type allele and variant allele, contain complementary sequences to the FAM and HEX dye respectively, which are present in the master mix. During PCR one of the allele-specific primers matches the genomic region and the target region is amplified with the common reverse primer. As PCR progresses the level of allele specific tails increase and the fluor labelled part of the FRET cassette releasing the fluor from the 3' end quencher to produce a fluorescent signal (Figure 23).

3.2.6.1 KASPAR PROTOCOL

Primers were designed to target the desired variant using Primer Picker (K Bioscience, UK) software. This primer designing software creates two specific forward primers which target either the wild-type or the variant, and two reverse primers. Only one of the reverse primers is used for the actual genotyping procedure which is determined using optimisation. Optimisation of genotyping conditions was done using both specific forward primers and one of the two common reverse primers. The volumes for the allele mixtures are described in Table 11 and the six different master mixes used for the genotyping optimisation are outlined in Table 12. Optimisation was carried out using a DNA sample that contained the variant of interest and a wild-type control DNA sample. The optimised conditions were used to carry out case-control genotyping on previously prepared DNA samples. The samples were previously aliquoted onto 384-well PCR plates and dried down. Prior to the addition of the optimised master mix for the PCR procedure the positive control samples (12.5ng/µl) were added to the plate

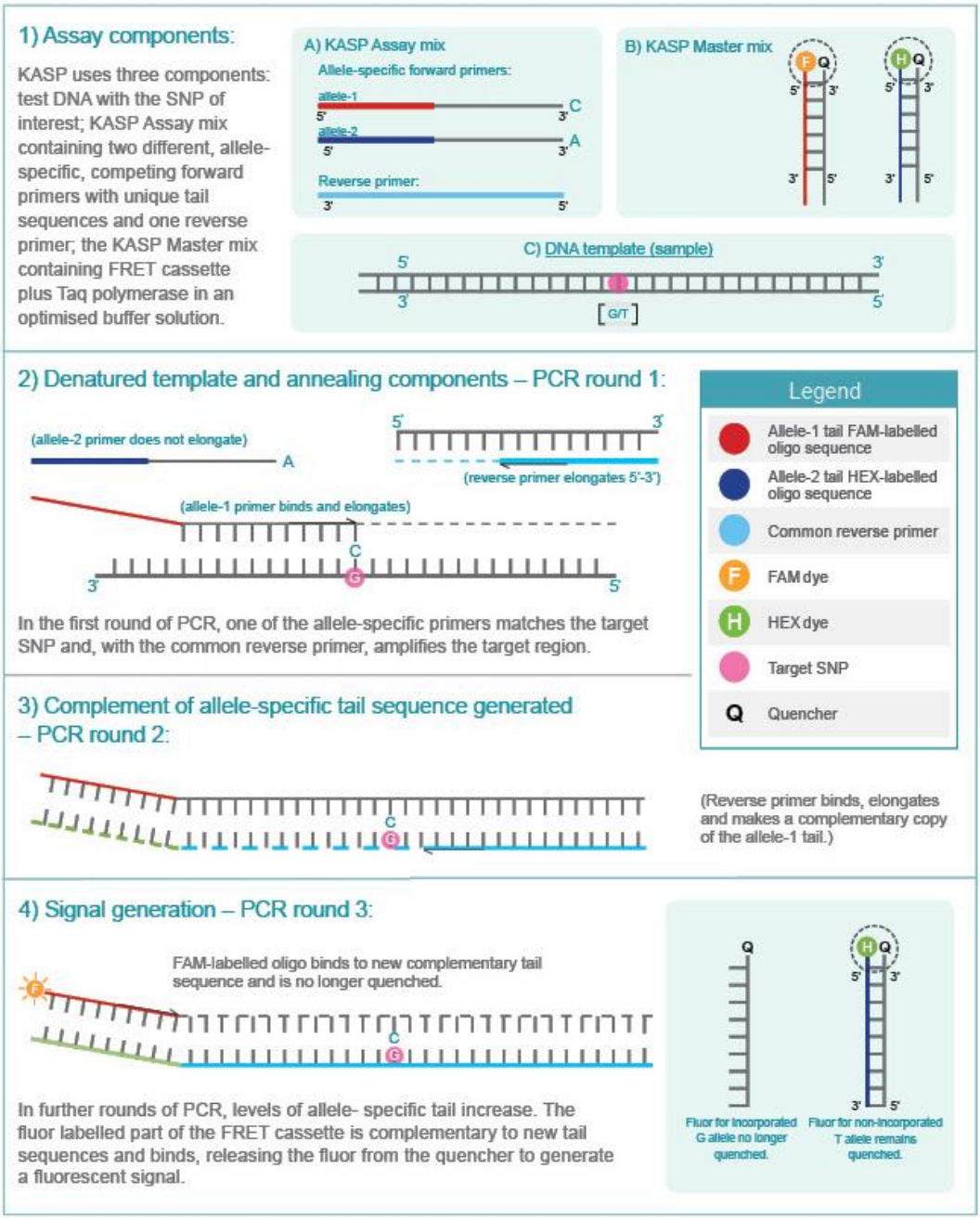


Figure 23. Diagrammatic representation of the mechanisms underlying KASPar Chemistry.

Image taken from LGC KASP genotyping chemistry user guide

Conditions	Concentration in Assay Mix	Volume in 1X Assay Mix (µl)
Allele Specific Primer 1 (100µM)	12	12
Allele Specific Primer 2 (100µM)	12	12
Common (reverse) Primer (100µM)	30	30
Water/TrisHCl (10mM,pH8.3)		46
Total volume		100

Table 11. Allele Mix for KASPar Assay

Reagents	A 1.8mM $MgCl_2$	B 2.2mM $MgCl_2$	C 2.5mM $MgCl_2$	D 2.8mM $MgCl_2$	E 5% DMSO	F 10% DMSO
DNA (12.5ng/µl)	1	1	1	1	1	1
2X Rxn mix (+KTAQ)	2	2	2	2	2	2
Assay mix 1/2	0.11	0.11	0.11	0.11	0.11	0.11
$MgCl_2$ 1:10	0	0.32	0.56	0.8	0	0
Water	0.89	0.57	0.33	0.09	0.69	0.5
Total	4	4	4	4	4	4
DMSO	0	0	0	0	0.2	0.4

Table 12. Optimisation Conditions for KASPar Genotyping

and dried at 50°C. The master mix for the PCR procedure was dispensed into each well of the plate using an Epmotion 5075 (Eppendorf, UK). The plate was then mixed using a Mixmate and spun in a centrifuge at 1500rpm for 1 minute before loading into the LightCycler 480 (Roche Diagnostics, UK). The thermal cycling conditions for the endpoint genotyping are outlined in Table 13. During the PCR the readings for the PCR were read three different times, as some readings provided better separation than others. Only one optimal reading was used for each of the analyses. For all SNPs genotyped, 17% of samples were duplicated to detect error and confirm reproducibility of genotypes on a cross check plate consisting of case and control duplicates. All the data were analysed to confirm Hardy-Weinberg equilibrium (HWE). Additionally rare variants present in a low number of samples were sequenced from stock DNA in all samples to confirm genotyping calls.

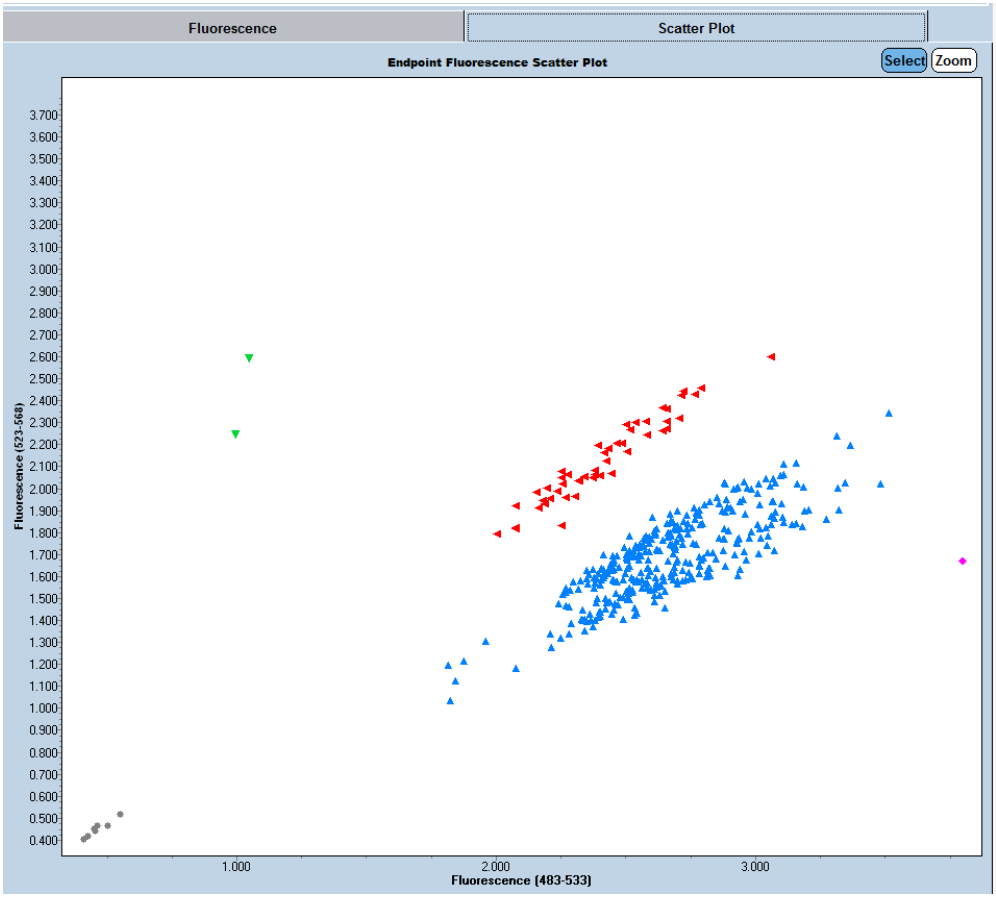


Figure 24. KASPar genotyping for *CACNA1C.*

KASPar endpoint genotyping for one of the case-control plates for variant rs79398153, located in the third intron of CACNA1C. Wildtype alleles are represented by the blue triangles, heterozygous individuals are represented by the red triangles and individuals homozygous for the variant are represented by green triangles. Both the grey and pink circles are samples which have failed or are undefined and will be repeated on the following plate.

3.2.6.1.1 ENDPOINT GENOTYPING

Endpoint genotyping was conducted using LC480 software (see Figure 24). Genotyping data was taken from the whole amplification process and not the post-PCR melting step. Each fluorescent output was measured and samples were differentiation depending on the signal recorded, any variants which did not cluster closely were manually called as unknowns and were re genotyped on the cross check plate. Additionally samples which failed to amplify in the primary genotyping run were plated onto the cross check plate from a stock sample of 25ng/µl and diluted to the working concentration of 3.33ng/µl. Genotype data was

Step	Temperature (°C)	Time	Ramp Rate °C/s	Cycles
Hot Start Activation	94	15 min	4.8	1
1st Amplification	94	20 sec	2.5	10
	65	1 min	2.5	
2nd Amplification	94	20 sec	2.5	26
	57	1 min	2.5	
Reading	37	1 sec	2.5	1
	38	1 sec	0.06	
3rd Amplification	94	20 sec	2.5	3
	57	1 min	2.5	
Reading 1	37	1 sec	2.5	1
	38	1 sec	0.06	
4th Amplification	94	20 sec	4.8	3
	57	1 min	2.5	
Reading 2	37	1 sec	2.5	1
	38	1 sec	0.06	
Cooling	40	1 sec	2.5	1

Table 13. Thermocycling conditions for KASPar Genotyping

accumulated and compiled to double check no differences were present between the original calls and the cross check plate. If such deviations were present samples were sequenced to verify calls. As to be expected for each SNP genotyped there was a number of samples which failed both the original and cross check genotyping. To ensure reliable assay mixes were used, runs which had less than 98% of the samples called were re optimised.

3.2.6.2 Statistical Analysis

Allelic associations for SNPs were performed using the Chi-square test and the Fisher's exact test. A cut-off significance value of $p < 0.05$ was used

3.2.7 Plasmid Construction

Vectors are constructs of DNA that allow the insertion of genomic regions of interest by a number of different cloning techniques. There are several important benefits that arise from this procedure of construction vectors in molecular biology. Facilitation of *in vitro* experiments is the primary function of the construction of vector sequences. The introduction of genetic material into cells outside their normal biological location allows for easy manipulation of the newly created recombinant DNA. Diverse experimental conditions allow investigation of the effect that variants may have on gene expression, cell trafficking, localisation of specific proteins and much more. Construction of the vectors allows for the specific labelling of the recombinant DNA with fluorescent proteins such as green fluorescent protein.

3.2.7.1 Preparation of *CACNAG4* tagged with RFP

The cDNA for *CACNG4* was obtained from Origene, USA. The *CACNG4* open reading frame (ORF) was present in the pCMV6-XL4 vector (Figure 25). The pCMV6-XL4 vector contains a CMV promoter for *in vivo* expression in mammalian cells and a T7 promoter for *in vitro* experiments. There are multiple restriction sites that can be used for the removal of the insert from the vector for cloning into the pTAG-RFP vector (Figure 26).

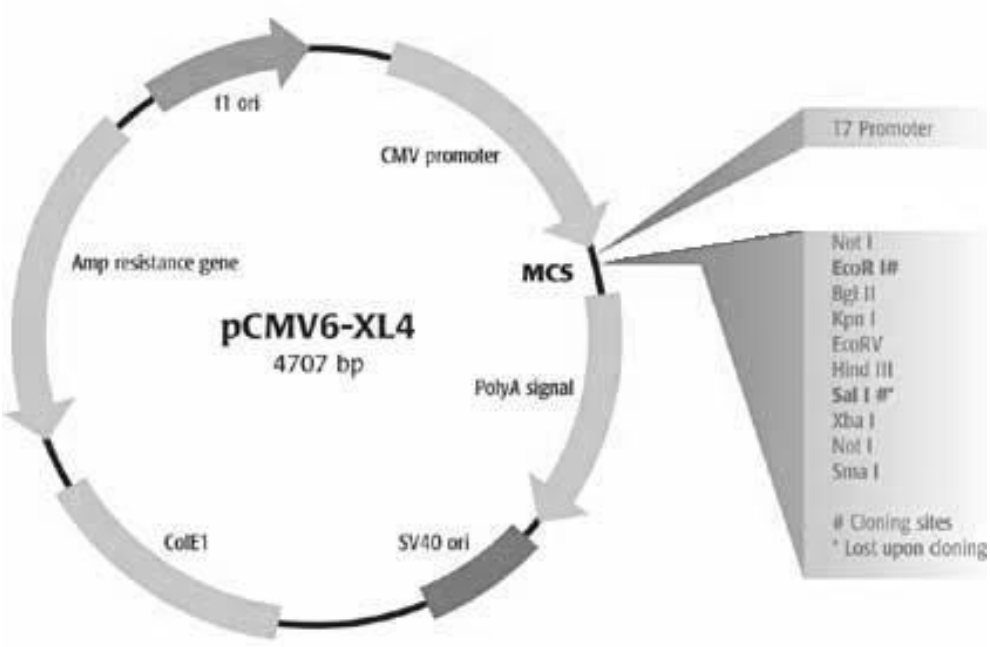


Figure 25. Vector sequence for pCMV6-XL4.

CACNG4 was present in the multiple cloning site and was removed using restriction digestion.

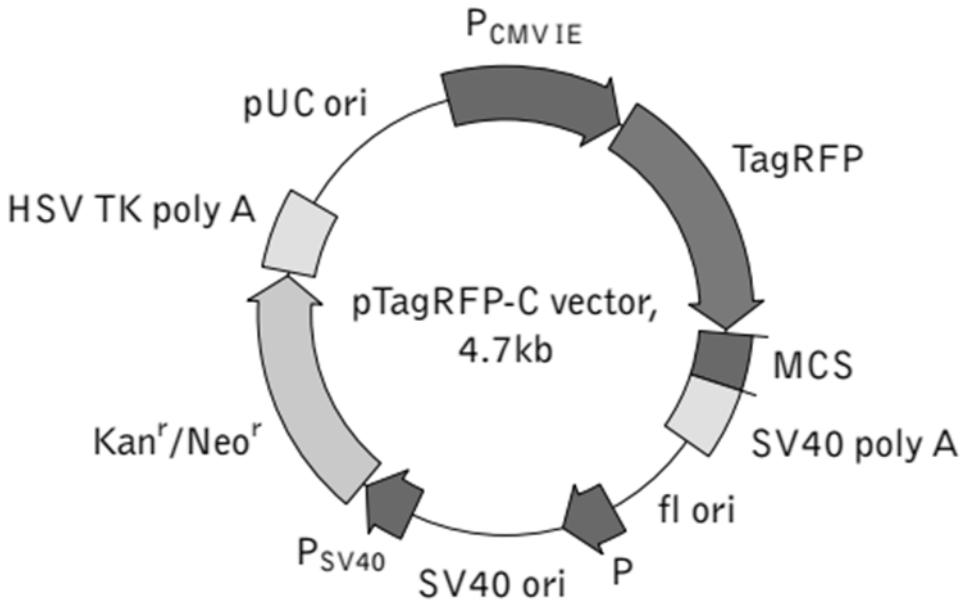


Figure 26. Vector Sequence for pTAG-RFP.

CACNG4 was inserted into the multiple cloning site using restriction digestion.

3.2.7.2 GluR1

Two GluR1 constructs were used for this project. The first is GluR1 in the pGEM-T vector (Figure 27), which is cloned in using TA cloning and the second is pCI-SEP-GluR1 plasmid which will be described later.

The pGEM-T vector is 3kb in size, contains multiple cloning sites and is a high efficiency TA cloning vector. The plasmid was sent on a Whatman FTA elution card, containing 5-10µg of DNA, obtained from Sino Biological Inc Japan. The vector was isolated from the elution card by first cutting around the filter paper containing the vector and cutting the resulting circle into small pieces. The filter paper was transferred to a microcentrifuge tube and 200µl of sterile H_2O was added. The mixture was briefly centrifuged to ensure all of the paper was in the water. The microcentrifuge tube was left to incubate at 95°C for 30 minutes in the water bath. The tube was vortexed gently every ten minutes. Following the incubation period the mixture was centrifuged for 30 seconds and 100µl of the supernatant was removed from and placed in a new clean tube. 10µl of the solution was run on a 1% agarose gel (EtBr) for two hours against hyperladder 1 (Bioline) to verify the presence of the vector (Figure 28).

3.2.7.2.1 LIC cloning

Ligase independent cloning (LIC) is a fast and efficient way of creating vector insert clones that does not utilise DNA T4 ligase or need the presence of restriction sites to work as is common with cloning procedures. The protocol first described by Aslanidis and Jong in 1990 [419] uses specific complementary base overhangs of approximately 10-15 base pairs to insert the DNA fragments into the vector of choice. The method is based primarily on the 3'→5' exonuclease nature of T4 DNA polymerase that creates the specific complementary base allowing non-covalent bi-molecular associations between the vector and the insert. The vector contains complementary overhangs to the insert but not to itself so that it cannot recirculize upon linearization.

The vector is linearised using a single restriction site digestion and then is treated with dCTP (deoxycytidine triphosphate). This allows the 3'→5' exonuclease nature

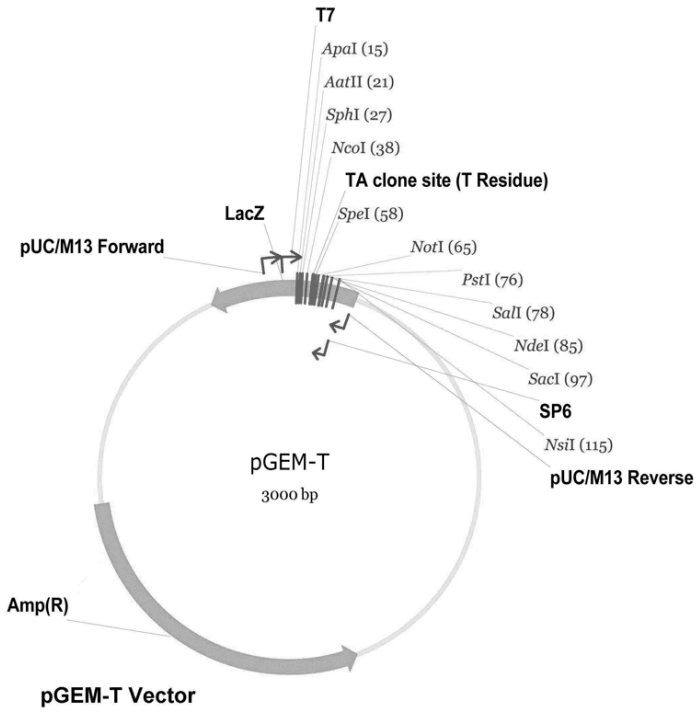


Figure 27. pGEM-T Vector

The GluR1 insert was present in the TA cloning site and was removed using LIC cloning. Image taken from http://www.sinobiological.com/Vector-pGEM-T-a-1636.html

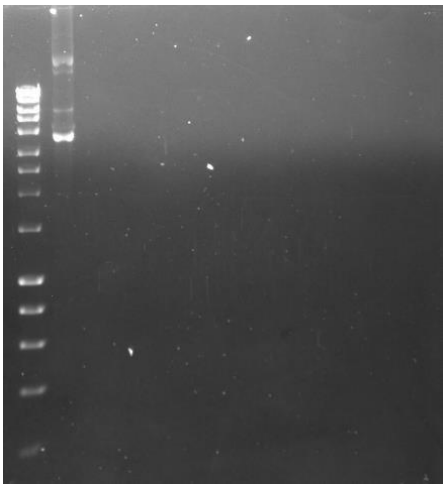


Figure 28. GluR1Elution

The 10µl of eluted GluR1 cDNA run against Hyperladder 1 (Bioline). The observed pGEM-T vector and GluR1 should be approximately 5.7kb long. However, due to the supercoiled nature of the vector it is able to run faster on the gel giving the observed band of approximately 3.8kb.

of T4 DNA polymerase to remove all of the nucleotide bases until the first cytosine residue is reached. The result of this linearization followed by dCTP treatment results in the 10-12 base pair overhang allowing for the specific ligation independent annealing reaction to proceed. Similarly the insert is amplified using primers with complementary overhangs to the extra bases in the vector and is treated with dGTP to compliment the exonuclease activity seen in the vector. The vector and insert are then combined in a tube and the overhangs are long enough to provide very specific enzyme annealing between both the vector and insert.

3.2.7.2.1.1 LINEARIZATION OF THE LIC VECTOR

The pcDNA3 green fluorescent protein (GFP) Lic cloning vector (6D) was first linearised using *SspI* (Promega, UK). In a microfuge tube 5µl of Buffer E (Promega, UK) 2.5µl of *BSAI* (10 units/µl), and 5µg of plasmid DNA were mixed together. Following a short vortex, 1µl of *SspI* was added and pre-PCR water added to culminate in a final volume of 50µl. The digestion mixture was incubated at 50˚C for one hour. Following digestions the sample was run on a 0.8% agarose gel to ensure complete digestion and that linearization had occurred. As a positive control an undigested vector was run in the well beside the digested pcDNA3 GFP. The sample was added to 10µl of 5x loading buffer and hyperladder 1 was used as a marker to determine band size. The gel was run for one hour and 100V. The gel was analysed using the UV transilluminator (UVP Gel-doc-it, Imaging Systems, UK) as previously mentioned in section 3.1.5 and the linearised band of LIC vector was extracted. The plasmid DNA was purified from the gel using a QIAquick Gel Extraction Kit (Qiagen, UK) based on the manufacturers protocol. Quantification of the extraction sample was determined using a Nanodrop 2000 (Thermo, UK).

3.2.7.2.1.2 PCR OF THE INSERT FOR LIC CLONING

To prime the insert for inserting into the LIC cloning vector PCR was performed with specific primers. The primers used were *F: 5'*-TACTTCCAATCCAATGCCACCATGCAGCACATTTTTGCCTTC-3' and *R*: 5'-CTCCCACTACCAATGCCCAATCCCGTGGCTCCCAAG-3'. The forward primer contains the complementary overhang, denoted by the underline, and 21 bases of the

beginning of *GLUR1*, including the start codon. The reverse primer contains the specific complementary overhangs, also denoted by the underline, but skips the stop codon at the end of *GLUR1* to ensure correct read through of GFP is achieved. It is essential that both primers have a melting temperature of 60˚C. The insert was amplified using the high fidelity polymerase *Pfx*. An optimisation was run to determine the best optimizing conditions for the PCR reaction. The conditions are listed in Table 14 . The master mix for each optimisation was created and the PCR reaction undertaken Table 15. Following PCR, the amplified sample was run on an 0.8% agarose gel at 100V for one hour. It is essential for the protocol to remove any free dNTPs in the reaction mixture. The sample was analysed as previously mentioned in section 3.1.5 and extracted from the gel using the QIAquick gel extraction kit according to the manufacturer's instructions.

3.2.7.2.1.3 T4 DNA TREATMENT OF BOTH LIC VECTOR AND PCR PRODUCT

The LIC vector and PCR product were incubated in the presence of LIC-qualified T4-polyermase (EMD Millipore, USA) and dCTP and dGTP respectively. 0.2pmol of the insert was treated with dGTP to provide enough material for 10 annealing reactions. The formula used to calculate the amount of DNA to give 0.2pmol can be found below:

$$A = \frac{B \times 10^6}{N \times 660}$$

Equation 1. Calculation of pmol of DNA

Where A is the [pmol of DNA], N is the length of the DNA region in base pairs and B is the starting µg concentration of DNA length N. 660 is the approximate molecular weight of a DNA nucleotide pair.

For the PCR product 2µl of T4-DNA polymerase buffer, 0.2pmol of PCR product, 0.5µl dCTP, 1µl DTT (Dithiothreitol), 2µl 10x BSA and 0.4µl T4 DNA polymerase (3units/µl, Promega, UK) were mixed in a 1.5ml microfuge tube. Sterile water was added to a final volume of 20µl. The solution was vortexed briefly and then spun at 13,000rpm for 1 minute. The reaction mix was incubated for 30 minutes at RT.

Step	Time	Temperature	Cycles
Denaturation	2 min	95˚C	1
Denaturation	30 sec	95˚C	
Annealing	30 sec	56˚C	30
Extension	3min*	72˚C	
Extension	10min	72˚C	1
Hold		4˚C	1

Table 14. PCR conditions for LIC cloning.

*1 minute per kb is used for *Pfx* polymerase

Volume (µl)	**2.0 mM $MgCl_2$**			**2.5 mM $MgCl_2$**		
	A 5M Betaine	**B** No Betaine + No Enhancer	**C** No Betaine + Enhancer	**D** 5 M Betaine	**E** No Betaine + No Enhancer	**F** No Betaine + Enhancer
10x Buffer	2.5	2.5	2.5	2.5	2.5	2.5
10 x Enhancer	2.5	0	2.5	2.5	0	2.5
5 M Betaine	5	0	0	5	0	0
50 mM $MgSO_4$	1	1	1	1.25	1.25	1.25
25 mM dNTPs	0.2	0.2	0.2	0.2	0.2	0.2
10 µM F oligo	1	1	1	1	1	1
10 µM R oligo	1	1	1	1	1	1
Taq	0.2	0.2	0.2	0.2	0.2	0.2
DNA (25ng/ul)	0.5	0.5	0.5	0.5	0.5	0.5
Water	11.1	18.6	16.1	10.85	18.35	15.85
Total volume	25	25	25	25	25	25

Table 15. *pfx* primer optimisation conditions.

Following the incubation period the enzyme was inactivated by incubation at 75˚C for 20 minutes and then spun at 13,000rpm for 1 minute.

For the LIC vector 600ng of *SspI* digested vector was used, which is enough for 20 annealing reactions. 2µl of T4-DNA polymerase buffer, 600ng of LIC vector, 0.5µl dGTP, 1µl DTT (dithiothreitol), 2µl 10x BSA and 0.4µl T4 DNA polymerase (3units/µl) were mixed in a 1.5ml microfuge tube. Sterile water was added to a final volume of 20µl. The solution was vortexed briefly and then spun at 13,000rpm for 1 minute. The reaction mix was incubated for 30 minutes at RT. Following the incubation period the enzyme was inactivated by incubation for at 75˚C for 20 minutes and then spun at 13,000rpm for 1 minute.

3.2.7.2.1.4 ANNEALING OF THE LIC VECTOR AND PCR INSERT

The insert and vector were annealed using the complementary base pair overhangs that had been generated using the procedures described previously. For this reaction, 0.02pmol of insert DNA, and 25-50ng of LIC prepared vector was used. The concentration of the vector varies depending on the vector: insert ratio, which is suggested to be 1:2.

In a microfuge tube 3µl of LIC prepared vector was mixed with 3µl of LIC prepared insert and sterile water was added to a final volume of 20µl. The mixture was incubated at RT for 5-30 minutes. Annealing of the complexes generally occurs in first five minutes but the insert and vector can be incubated for up to one hour with equivalent results.

3.2.7.2.1.5 *E.COLI* TRANSFORMATIONS

Following the incubation, 6µl of annealing mixture was transformed into 900µl of Library Efficiency® DH5α™ Competent Cells (Life Technologies, UK). Cells were chilled on ice for 30 minutes. They were then placed in a water bath at 42°C for 45 seconds before placing them on ice again for 2 minutes. The cells were then placed in super optimal broth (SOC, Life Technologies, UK) medium at 37°C for one hour and shaken at 225rpm (Appendix II, II.VI). To check transformation efficiency, a parallel transformation was done using the 0.1ng of pUC18 DNA.

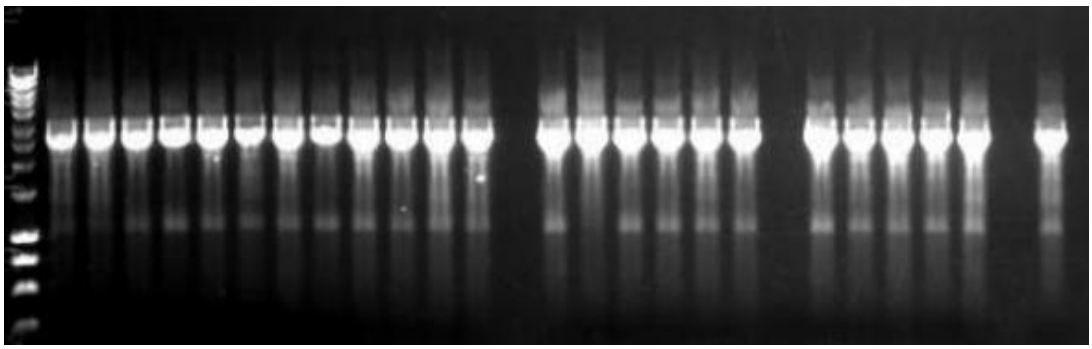


Figure 29 PCR of *E.coli* single colonies for LIC insert.

3.2.7.2.1.6 SELECTION OF *E.COLI* WITH INSERT

Following overnight incubation a number of single colonies were removed A sterile single pipette tip was used to pick up each individual colony. The tip was then dipped into 5µl of pre-PCR sterile water. The tip was then placed in 500µl of Luria-Bertani broth (LB broth) [10g NaCl, 10g Trypton, 5g Yeast] with a vector specific antibiotic (Ampicillin (100µg/ml) or Kanamycin (50µg/ml)). The cells were shaken in LB Broth for 8 hours. DNA from the *E.coli* was extracted using a mini prep procedure (Qiagen) which is described below. In parallel a restriction digest was run on the 5µl of pre-PCR water, in which the tip used to pick single colonies was dipped. The digest was conducted using *Xbal* and *HindIII*. These two restriction sites are unique to the vector and flank the area in which the insert should anneal. In a microfuge tube 5µl of pre-PCR water containing the *E.coli* DNA, 2µl of Promega Buffer E (Promega, UK), 2µl of BSA 10X, 1µl of *Xbal* and 1µl of *HindIII* were mixed together. Sterile water was added to make a final volume of 20µl. The samples were incubated at 37°C for 4 hours. Following incubation 10µl of digested DNA was mixed with 2µl of 5x Loading Buffer (Bioline, UK) and run for 2 hours at 100V on a 0.8% agarose gel. Following gel analysis, samples that contained the insert could be selected to grow up (Figure 29).

3.2.7.2.2 MINI PREP

DNA was isolated using the QIAprep® miniprep Kit (Qiagen, UK) from *E.coli* following the restriction digest in order to verify by sequencing that the insert was present. Samples which showed the presence of a correct band size were selected for mini-prep. Following the 8 hour incubation, 250µl of insert containing *E.coli*

was transferred into a 50ml falcon tube containing 5ml of LB Broth with appropriate antibiotic (Appendix II, II.VI). The cell solution was left shaking at 37^0C for ≥16 hour. Following overnight incubation, the *E.coli* was pelleted by centrifugation at 8000rpm for 5mins. The waste LB broth was discarded and the pellet was resuspended in 250µl of Buffer P1 and transferred to a microcentrifuge tube. 250µl of Buffer P2 was then added to the samples and each tube was invert 4-6 times to mix the solutions. The samples were left to incubate at RT for five minutes for lysis of the bacterial cells to occur. 350µl of Buffer N3 was then added to terminate lysis and the samples were spun at 13,000rpm for 10mins to remove the cellular debris. The supernatant from this step is then removed by pipetting and added to a Qiaprep spin column. The samples were spun for 60seconds and the flow through was discarded. 500µl of Buffer PB was added to the spin column and the samples were spun for 60 seconds and the flow through was discarded. Finally 750µl of Buffer PB was added and the samples were spun for 60 seconds. The flow through was then discarded. 50µl of sterile water was added to the column and the samples were left for one minute at RT. The flow through was then collected and the samples were quantified using the Nanodrop as described in 3.1.3.

3.2.7.2.3 MIDI PREP

To ensure an adequate volume of plasmid DNA was obtained prior to the onset of *in vitro* experiments a midi prep was done for each plasmid containing the insert. Samples which contained the correct insert were transformed into Library Efficiency® DH5α™ Competent Cells (Life Technologies, UK) (3.2.7.2.1.4) The plates were left overnight at 37^0C and a single colony was taken and placed in 500µl of Luria Bertani (LB) Broth with a vector specific antibiotic (Ampicillin (100µg/ml) or Kanamycin (50µg/ml)) (Appendix II, IV). The cells were shaken in LB Broth for 8 hours. Following this they were transferred into a 50ml Falcon tube containing 5ml of LB Broth with appropriate antibiotic. The cell solution was left shaking at 37^0C for ≥16 hours. The following morning the sample was spun down in a centrifuge at 6000rpm, at 4^0C for 15 minutes. The supernatant was disposed of and a 200µl pipette tip was dipped in the pellet of *E.coli* to make a preparation for the midi prep without having to first mini prep the DNA.

Following the incubation period the vector DNA was extracted according to the manufacturer's protocol (Qiagen, UK). The 100ml of bacterial cell culture was split into two 50ml Falcon tubes. The cells were harvested by spinning at 6000rpm for fifteen minutes at 4^0C. The bacterial pellet was then resuspended in 4ml of buffer P1, and the sample mixed vigorously to ensure it was completely homogenous. Four ml of buffer P2 and the sample was mixed by inverting the tube a number of times. The sample was incubated at RT for up to five minutes. Following the incubation period, four ml of pre-chilled buffer 3 was added and the sample was mixed to ensure the suspension was completely homogenous. Upon addition of buffer 3 a white fluffy lysate should form in the Falcon tube. The addition of chilled buffer 3 ensures the precipitation of proteins, cellular debris and even genomic DNA. The suspension is then added to a QIAfilter cartridge and the sample is left for 10 minutes for the lysate to rise to the top of the cartridge.

During the ten minute incubation period the QIAGEN-tip was equilibrated by adding 4ml of buffer QBT and the column was allowed to empty via gravity. The clear liquid from the QIAfilter cartridge was filtered into the QIAGEN-tip using the syringe plunger. Once the sample was loaded onto the QIAGEN-tip, the tip was washed twice with 10ml of buffer QC twice. The first wash removes the majority of contaminants still residing in the sample. The DNA was then eluted from the QIAGEN-tip by adding 5ml of buffer QF. The elution was collected in a clean 50ml Falcon tube. DNA was precipitated by adding 3.5ml of room-temperature isopropanol and spun at 6000rpm at 4^0C for one hour. The supernatant was discarded and the pellet was washed using 2ml of 70% ethanol. The sample was then spun at 6000rpm for twenty minutes. Following the spin the supernatant was carefully decanted and the tube was left inverted to ensure the removal of the majority of ethanol. The samples were then left open under a heat source for 5-10 minutes to guarantee the removal of all of the ethanol. The DNA pellet was redissolved using pre-PCR water.

Samples extracted, using the Qiagen midi prep kit, were quantified using the Nanodrop 2000 (Thermo Scientific, UK). A260/280 and A260/230 ratios were noted to ensure sample purity. A ratio of 1.8 for the A260/280 reading and between 2.0-2.2 for A260/230 reading is considered pure.

3.2.7.2.4 pCI-SEP-GluR1/GluR2

The pCI-SEP-GluR1 (Figure 30) and GluR2 (Figure 31) vectors were provided from the Addgene repository from Robert Malinow (Plasmid 24002) [420]. The pCI mammalian vector contains the super-ecliptic pHluorin (SEP) coding sequence on the N-terminus of the GluR1/GluR2. The pCI-SEP vector contains a signalling peptide, at the extreme end of the N-terminus, which is essential for direction of the AMPA-Rs to the endoplasmic reticulum following translation [421, 422].

GFP proteins can be utilised in a number of different molecular biological techniques. Flow cytometry analysis was used to determine the expression of both GluR1 and GluR2 at the cell surface in the presence of the CACNG4 wild-type and mutant proteins. The ecliptic pHluorin is an enhanced pH sensitive variant of the GFP protein. Under normal conditions GFP has a major peak at 395nm and a minor peak at 475nm when measured with an emission maximum of 509nm. Ecliptic pHluorin is non-fluorescent in an acidic environment (pH>6.0) but becomes fluorescent under neutral conditions, enhancing both emission peaks [423]. SEP tagging has been used to measure AMPA-Rs surface expression previously [424]. The presence of the AMPA-R/SEP complex at the cell surface increases the fluorescence compared to that of normal GFP and ecliptic pHluorin. Additionally there is a 100-fold increase in the fluorescent signal when the SEP is present at the neutral cell surface, compared to the relatively acidic internal compartment [425].

3.2.7.3 Site directed Mutagenesis

Site directed mutagenesis was used to introduce the variants of interest into the wildtype vectors previously created using PCR and ligation. Site directed mutagenesis is a molecular biology technique involving a high fidelity DNA taq. This technique allows researchers to investigate the implications of a variant present in DNA sequence and create functional assays *in vitro* to determine whether it causes a change when present versus what would normally occur. Variants such as point variants, amino acid replacements and insertion of amino acids can be accomplished using this technique. Primers are designed to be complementary to the bases surrounding the region of interest, and contain the desired variant. During thermal cycling the primer directed replication of both

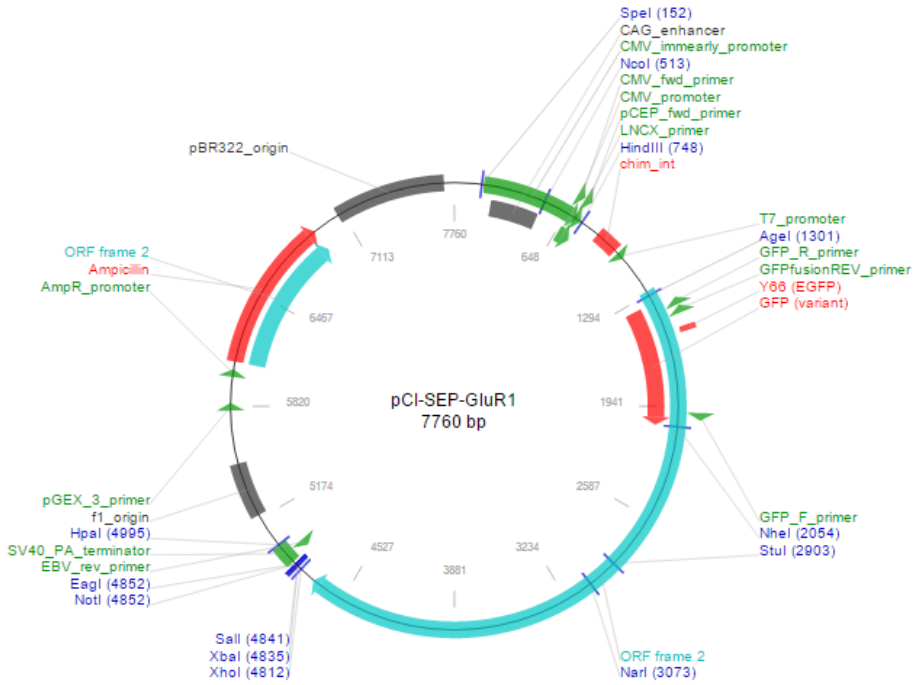


Figure 30. Circular map of the pCI-SEP-GluR1 vector.

Image taken from https://www.addgene.org/24000/

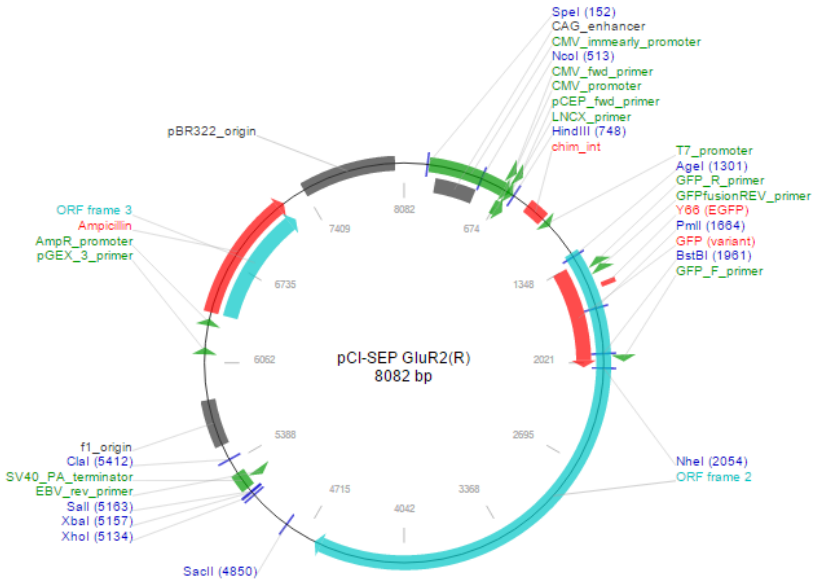


Figure 31. Circular vector map of the pCI-SEP-GluR2 vector.

Image taken from https://www.addgene.org/24001/

strands of the plasmid vector occurs. The replicated plasmid contains the mutated base and staggered nicks. As the parental plasmid DNA has been isolated from dam$^+$ *E.coli* strains it is methylated and is susceptible to *DpnI* digestion. This technique is used to remove the parental plasmid DNA following temperature cycling. This ensures that only the plasmid containing the variant is transformed into the competent cells.

Forward and reverse primers, complementary to both strands of the plasmid, were designed using the Agilent QuikChange Primer Design programme (www.agilent.com/genomics/qcpd). To ensure high fidelity annealing the primers were between 25-45 base pairs long with a melting temperature $\geq$78°C. HPLC purified primers were obtained from Sigma Aldrich, guaranteeing highly purified primers. Site directed mutagenesis was conducted using the QuikChange XL kit (Agilent). The reaction for the thermal cycling conditions was set up in a PCR plate. A control reaction was run in parallel using the pWhitescript 4.5kb control plasmid. The PCR reaction mixture is listed in Table 16 and the thermocycling conditions are listed in Table 17.

These samples were then transformed into QuikChange XL Ultra competent cells by adding 2µl of *DpnI* digest DNA to 50µl of competent cells, to which 2µl β-mercaptoethanol had been added. Cells were incubated on ice for 30 mins and then placed in a water bath at 43˚C for 30 seconds. The samples were then placed back on ice for two minutes. 500ml of SOC media was added to the cells and they were left to shake for one hour before being 250µl of each sample was plated out onto an LB Agar plate containing appropriate antibiotic. These samples were left at 37˚C overnight. The following morning single *E.coli* colonies were picked from each plate and added to 500µl of LB Broth. If no colonies were observed the site directed PCR reaction was repeated with increasing concentrations of DNA. These concentrations ranged from 50ng/µl to 200ng/µl. Colonies which were selected for growth were shaken at 37˚C for 8 hours. After 8 hours 250µl was added to 5ml of LB broth containing antibiotic and the samples were left overnight. DNA from these samples was isolated using the mini prep protocol described in 3.2.7.2.2. Once the DNA was extracted it was sent for sequencing to confirm the presence of the variants of interest. The samples which contained the variants were then prepared for midi prep 3.2.7.2.3.

Reagent	µl
10x Reaction Buffer	5
dsDNA template (10ng/µl)	1
Primer 1 (125 ng)	1.25
Primer 2(125ng)	1.25
dNTP mix	1
Quik Solution	3
ddH_2O	37.5
PfuUltra	1
Total	50

Table 16. PCR reaction mixture for site directed mutagenesis

Step	Temperature	Cycles	Time
1	95	1	1 min
	95		50 sec
2	60	18	50 sec
	68		1 min/kb of plasmid length
3	68	1	7 min

Table 17. Thermocycling conditions for site directed mutagenesis

3.2.8 CELLULAR TRANSFECTIONS

The DNA transfection was done as previously described 3.1.6.3 and 3.1.6.4 . 500ng of GluR1 and *CACNG4* were transfected into the cells 24 hours after seeding in 6 well plates. Each time six samples were transfected and the pGL3-basic vector was used as a control for transfection in wells where co-transfection of GluR1 and GluR2 and CACNG4-wt/CACNG4-mutant were not added. Cells were seeded at a density of 0.3 x 10^6 so that they would reach 75-80% confluence in 24 hours.

3.2.9 FLOW CYTOMETRY

Hek293 cells were grown as described previously at 37˚C, 5% CO2 in DMEM supplemented with 10% FBS and 1% Penicillin/Streptomycin. Cells were transfected when they were 75-80% confluent. The constructs used for transfection were wild-type CACNG4-RFP, mutated CACNG4-RFP, GluR-1-SEP and GluR2-SEP. Co-transfections of wild-type CACNG4-RFP with either GluR-1 or GluR2

and mutated CACNG4-RFP with GluR1 or GluR2 was conducted. The pgl3-basic vector was used as a control for the co-transfections when only one of the main plasmids of interest was being transfected in. Twenty four hours after transfection all cells were trypsinised and collected. The cells were spun at 1500rp for five minutes to pellet the cells. The cells were washed using 1ml of PBS and spun again to remove any remaining DMEM and PBS. Cells were fixed in 4% paraformaldehyde for 15 minutes. The paraformaldehyde was removed and the cells were washed twice with PBS. They were resuspended in an appropriate volume of PBS and stored at 4˚C.

Flow cytometry was conducted on fixed HEK293 cells using the BD LSRFortessa (Becton & Dickinson, New Jersey, USA) and collected using BD FACSDiva software (Becton & Dickinson, New Jersey, USA). Voltage settings for forward and side scatter were optimised and the same conditions were used across all samples. Where possible, 10,000 events were captured for each sample, and experiment run to generate enough data for further analysis. Further sample analysis was conducted using Flowjo software (Tree Star, USA). Cell populations were disseminated based on the negative control for each experiment using the forward scatter and side scatter parameters. These were then gated to control for background fluorescence which may arise from negative controls. The mean fluorescence intensity values were generated from the population positive for SEP fluorescence. The difference between wild-type CACNG4-GluR1/GluR2 and mutated CACNG4-GluR1/GluR2 was calculated using a two-tailed Student T-Test. The data is representative from a total of three independent replicate experiments.

3.3 Methods for results chapter 5: Investigation of variants using Next Generation Sequencing Analysis

3.3.1 Research Samples for Next Generation Sequencing

Whole-genome sequencing (WGS) was performed on 99 of the BD-I subjects who had a positive family history of bipolar disorder or bipolar spectrum disorder and a relatively early age of onset. A further four criteria were applied when selecting individuals for WGS; whether the individuals were lithium responsive or unresponsive and whether the cases were also alcohol dependent or not. Selection of the samples for whole genome sequencing, and variant calls were conducted prior to my commencement on this project. A list of the processes undertaken for the WGS project, from sample selection to genotyping of variants in the UCL cohort, is shown in Figure 32 These BD-I samples sent for WGS were selected by colleagues in the Molecular Psychiatry Laboratory, UCL. The WGS was conducted Illumina (Illumina, Inc. San Diego, CA, USA) and annoted by KNOME (Knome Inc, Boston, MA). Allele frequencies and comparisons to the 1000G allele frequencies, collated by genomic position, were conducted by a statistician at UCL (Dr. Curtis, UCL). The frequency cut off for variant selection, a minus log p (MLP) >1.3 for the Fisher's exact test between the 99 BD subjects and 1000G, was decided by all individuals involved in variant selection. I was involved in selecting the remaining criteria including which region of interest in the *CACNA1C* intron 3 to investigate, the use of ENCODE marks, GERP scores and conservation marks and criteria for SNP selection in coding regions of the calcium channel family of genes. Once these variants were selected they KASPar genotyping was conducted in the BD case and healthy volunteer sample as described in sections 3.1.1.1 and 3.1.1.4 respectively. Additionally I decided to use externally available databases to investigate variant frequencies for alleles which were rare.

3.3.2 Next Generation Sequencing and Analysis

The genomic DNA was sequenced using one hundred base pair paired-end reads on a Hi-Seq 1000 (Illumina, Inc. San Diego, CA, USA). Sequence data alignment to the NCBI human reference genome 37.1 (hg19) and variant calling was performed

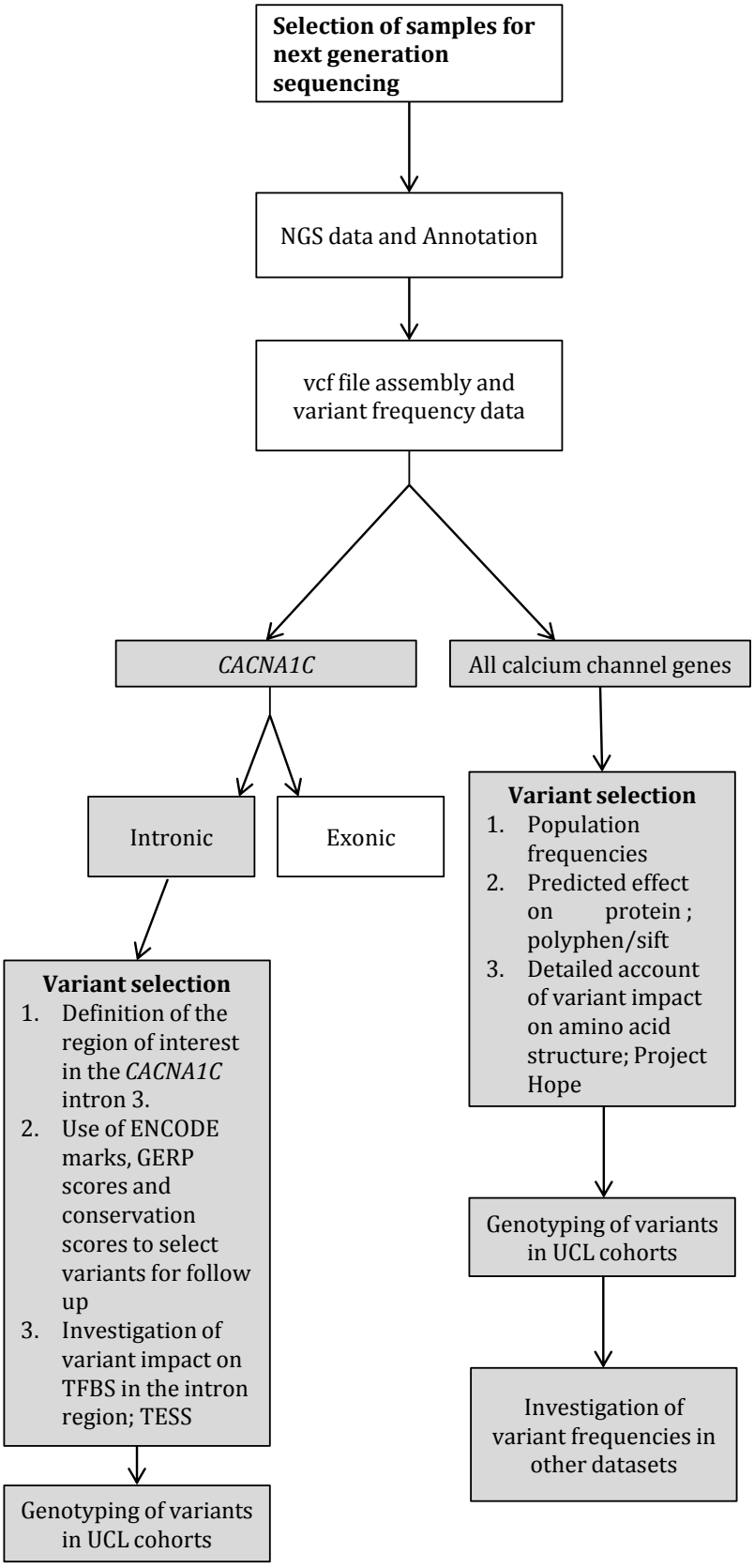


Figure 32. From Samples to variant genotyping. A stepwise depiction of the experiments involved in whole genome sequencing

The white boxes denote steps which were performed either by colleagues at UCL or KNOME and the grey boxes denote processes which I conducted as part of the whole genome sequencing project in the lab.

using the CASSAVA pipeline at Illumina. The sequence data from these individuals was further analysed and annotated using kGAP (Knome Inc., Boston, MA).

3.3.3 VARIANT SELECTION

3.3.3.1 CACNA1C

All variants present in the third intron of *CACNA1C* were pooled from the Knome VARIANTS software (Knome). Allele counts of each SNP in the BD samples were compared to those in the European cohort of the 1000 genomes (1000G) project. SNPs with a minus log p (MLP) >1.3 for the Fisher's exact test between the 99 BD subjects and 1000G were chosen to be followed up by genotyping in the case-control samples. Variants present in poly-base regions or insertions in repeat regions were excluded from genotyping.

The variants were selected for genotyping in the UCL case-control cohort if they met the following four criteria:

- Located in the third intron of CACNA1C between flanking higher recombination peaks (chr12:2,230,353-2,559,413 hg19);
- Located in a putative functional site defined by being an ENCODE marked element [392];
- Located in a conserved non-coding sequence, determined using the mammalian conservation track on UCSC or by their GERP scores;
- Not in a repeat region.

3.3.3.2 OTHER CALCIUM CHANNEL GENES

As evidence points towards calcium deregulation in individuals with BD, and pathway ontology analysis has shown that the calcium channel gene set is implicated with BD, variants present in the remaining calcium channel genes were investigated using KNOME software. The variants were selected for genotyping in the UCL case-control cohort if they met the following four criteria; Variants were non-synonymous; Genes have been previously implicated in Bipolar Disorders,

either through gene ontology analysis, pathway analysis, or have an association signal in the PGC1 BD data set; An MLP > 0.7. This indicates that there was one individual from the BD sample with the variant and no individuals in the 1000Genome sample; PolyPhen-2 prediction of Damaging, or if the prediction was benign the variant was not present in the 1000G.

3.3.4 KASPAR GENOTYPING

3.3.4.1 CACNA1C

Alignment software (KNOME) annotated one hundred and thirty four SNPs before filtering in the 99 BD 1 cases. After filtering, seven polymorphisms (two previously unreported SNPs and five that were on dbSNP) were selected to be followed up by genotyping analysis in the UCL sample. Each variant was validated by confirming the base pair call confidence using individual BAM files before genotyping. Bioinformatic analysis to determine potential function was carried out using the ENCODE track on the UCSC genome browser (http://genome.ucsc.edu/).

3.3.4.2 OTHER CALCIUM CHANNEL GENES

KNOME software predicted 26 non-synonymous variants in *CACNA1S, CACNA1D, CACNA2D4, CACNA1C, CACNB1*, and *CACNB2.* After filtering on the above mentioned criteria, 8 SNPs were selected to be followed up using KASPar genotyping. Each variant was validated using base pair call confidence using individual BAM files before genotyping.

3.3.5 STATISTICAL ANALYSIS

Allelic associations for SNPs were performed using the Chi-square test and the Fisher's exact test. A cut-off significance value of $P < 0.05$ was used.

3.3.6 PLASMID CONSTRUCTION

3.3.6.1 PJET1.2

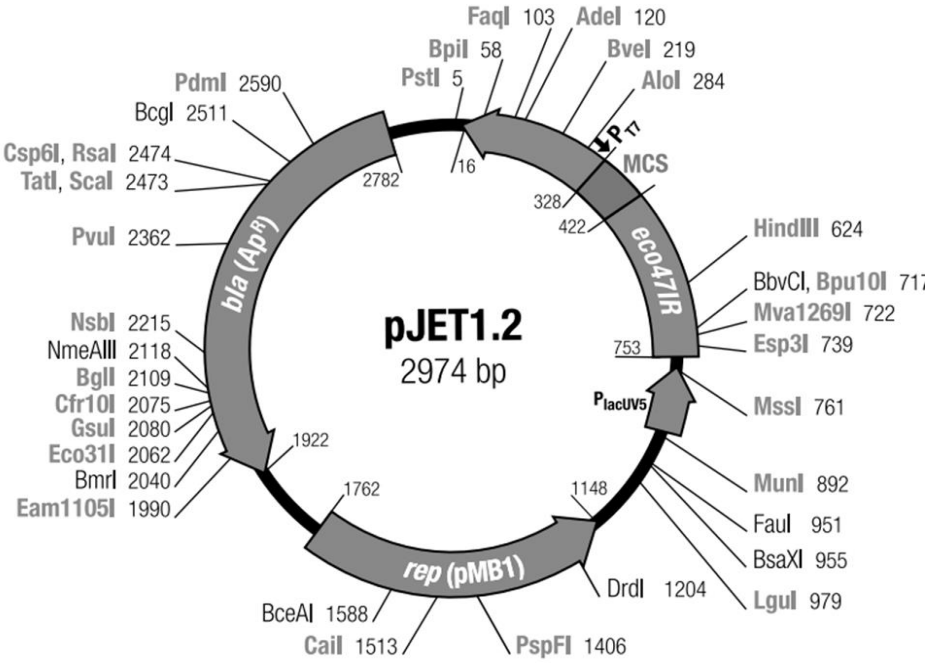


Figure 33. pJET1.2 Blunt End Cloning Vector

(http://www.thermoscientificbio.com/uploadedFiles/Resources/pJET1.2-blunt-cloning-vector-map.pdf)

The pJET1.2 linearised cloning vector (Figure 33) is part of the CloneJET PCR cloning kit that allows blunt end cloning of PCR generated DNA fragments. The vector contains a lethal enzyme gene that is only removed upon the ligation of the DNA insert allowing for positive selection of the bacteria that contain the recombinant vector. E.coli containing the recircularised vector without the insert will express the lethal enzyme upon transformation preventing further colony formation. Blunt end PCR products were produced using high fidelity *Pfu* DNA polymerase and could be directly ligated into the pJET-1.2/blunt vector within 5 minutes.

3.3.6.1.1 BLUNT END LIGATION

To determine whether the variants located in the third intron of *CACNA1C* have a functional effect on the regulation of gene expression we focused on a 729bp

Step	Three-Step Cycling	Time
Denaturing	94^0C	15 sec
Annealing	55^0C	30 sec
Extension	68^0C	1 min per kb

Table 18. *Pfx* cycling conditions

region that incorporates both SNPs and the regulatory regions as denoted by the ENCODE track on the UCSC genome browser. Both variants were identified during the analysis of the next generation sequencing data and are in complete LD in European populations as seen in the 1000 genomes (1000G), British in England and Scotland (GBR), Iberian population in Spain (IBS), and Toscani in Italy (TSI), but the T allele appears to be absent from the Finnish population (FIN).

3.3.6.1.2 AMPLIFICATION OF THE THIRD INTRON OF *CACNA1C*

The variants rs79398153 and rs116974827 located in the third intron of CACNA1C were amplified using *F:* 5'-cgctagcctgcaccacgtgtctttgcc-3' and *R: 5'*-gaagcttagtttgtctcacagaccctgacc-3' (Sigma Aldrich, UK). The primers were originally created with *NheI* and *HindIII* restriction sites and a buffer of 6 bases before the restriction site. This allows the insert to be cloned directly into the pGL3-Basic vector using the method of double restriction digest. However, following the first transformation and sequence verification we found that the 6 bases found prior to the restriction site and the first base of the restriction site had not been cloned into the vector following transformation. We decided to use the blunt end cloning procedure to clone the insert into the pJET1.2/blunt vector and created the second set of primers noted above. Optimisation of the PCR conditions for the amplification of the insert was undertaken using the conditions described in Table 15. The action of blunt end cloning relies primarily on the action of *Pfx* DNA polymerase (Lifetechnologies, USA). *Pfx* polymerase is a high fidelity thermostable enzyme which does not contain 5'-3' exonuclease activity. It contains 3'-5' proofreading activity, which proof reads the template produced ensuring high

fidelity output was used. To 0.5µl of DNA (25ng/µl), 2.5µl of *pfx* buffer, 2.5µl of Enhancer, 1µl of $MgSO_4$ (50mM), 0.2µl of dNTPs (25mM) 1µl of both the forward and reverse primer (10uM), 0.2µl of *Pfx* DNA polymerase were added. Pre-PCR water (Sigma Aldrich, UK) was added to give a final volume of 25µl. The PCR was run under the conditions listed in Table 18.

The samples were run on a 1% agarose gel to determine which condition was optimal (Appendix II, II.II). The insert was cloned into the pJET1.2/blunt vector as per the manufacturer's protocol (Thermo Scientific, UK). The suggested concentration of DNA to use for the ligation is 500ng for a PCR product of length 1000bp. The ligation reaction for the cloning of the 729bp insert into the blunt end vector was set up on ice. In a 1.5ml microcentrifuge tube 10µl of reaction buffer (2x), 1µl of DNA (500ng), 1µl of pJET1.2/blunt cloning vector (50ng/ul), pre-pcr water (Sigma Aldrich, UK) up to a volume of 19µl and 1µl of T4 DNA ligase. The sample was vortexed gently for 5 seconds. The ligation mixture was incubated at RT for 5 minutes.

3.3.6.1.3 TRANSFORMATION OF VECTORS USING BLUNT END LIGATION

The ligation mixture was transformed into the cells using heat shock transformation. Prior treatment of the competent cells by chilling in the presence of Ca^{2+} prepares the cell wall for the change in permeability Heat shock transformation is based on the principle that a sudden increase in temperature creates a pressure difference between the inside and the outside of the *E.coli* cells. This induces the formation of pores in the cell membrane, and returning the cells to a normal temperature returns the cell membrane to normal and allows for the cell to heal. Following the ligation 2.5µl of the mixture was added to 900µl of Library Efficiency® DH5α™ Competent Cells (Life Technologies, UK). Cells were chilled on ice for 30 minutes. They were then placed in a water bath at 42°C for 45 seconds before placing them in ice again for 2 minutes. The cells were then placed in SOC medium at 37°C for one hour and shaken at 225rpm. To check transformation efficiency, a parallel transformation was done using the 0.1ng of pUC18 DNA.

Cells were plated on to agar plates containing ampicillin (10µg/ml) at different volumes. The volumes used were 50µl, 100µl (x2), 200µl and 500µl. This ensures that if there if the *E.coli* growth is slow, the plate with the highest volume of cells should contain some for selection and vice versa for a faster rate of growth. The cells were left in an incubator overnight at 37°C. Following overnight incubation a number of single colonies were removed and grown up for mini prep (Qiagen, UK). Each of the tips that picked the colony was dipped into 5µl of pre-PCR sterile water and PCR was conducted using the optimised primer conditions. Following PCR 10µl of PCR DNA was mixed with 2µl of 5x Loading Buffer (Bioline, UK) and the samples were run for 2 hours at 100V on a 0.8% agarose gel. Following gel analysis the samples containing the correct samples were chosen to midi prep to create a working stock of the plasmid for the transfection.

3.3.6.2 PGL3-BASIC VECTOR

The pGL3 Luciferase Reporter Vectors are a series of vectors that can be used to determine whether putative regulatory regions in the genome have an effect on gene expression. As discussed above, studies such as the ENCODE study are identifying a plethora of genomic regions that could be functionally active at the level of gene expression regulation. The pGL3-Basic vector was obtained from promega (Figure 34) (Promega Corporation, USA). Due to the lack of eukaryotic promoter and enhancer sequences, it provides maximum flexibility for cloning numerous different putative functional elements. The luciferase expression is solely dependent on the insertion of a functional promoter upstream of the luciferase gene, although some basal level of luminescence may be seen.

Following the insertion of the *CACNA1C* region into the pgl3-basic vector site directed mutagenesis (3.2.7.3) was used to create three different plasmid vectors. The first contained the rs79398153 variant, the second contained the rs116947827 variant and the third contained both variants. Stock solutions of all four plasmids were created using the mini and midi prep methods outlined previously (3.2.7.2.2 and 3.2.7.2.3).

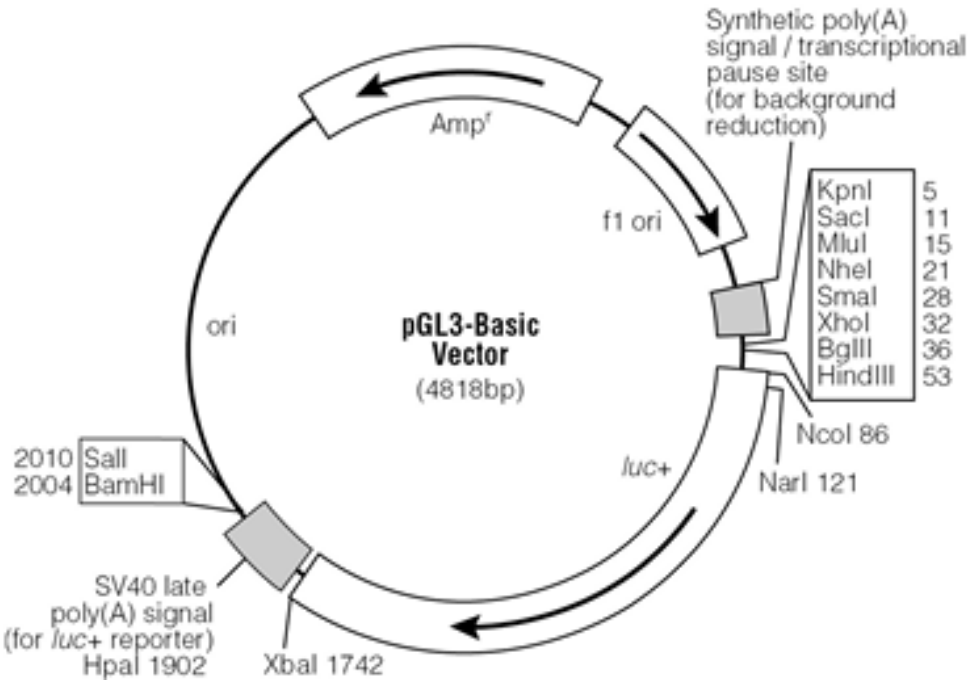


Figure 34. pGL3-Basic Vector

Image taken from the Promega pGL3 Luciferase Reporter Vectors technical manual

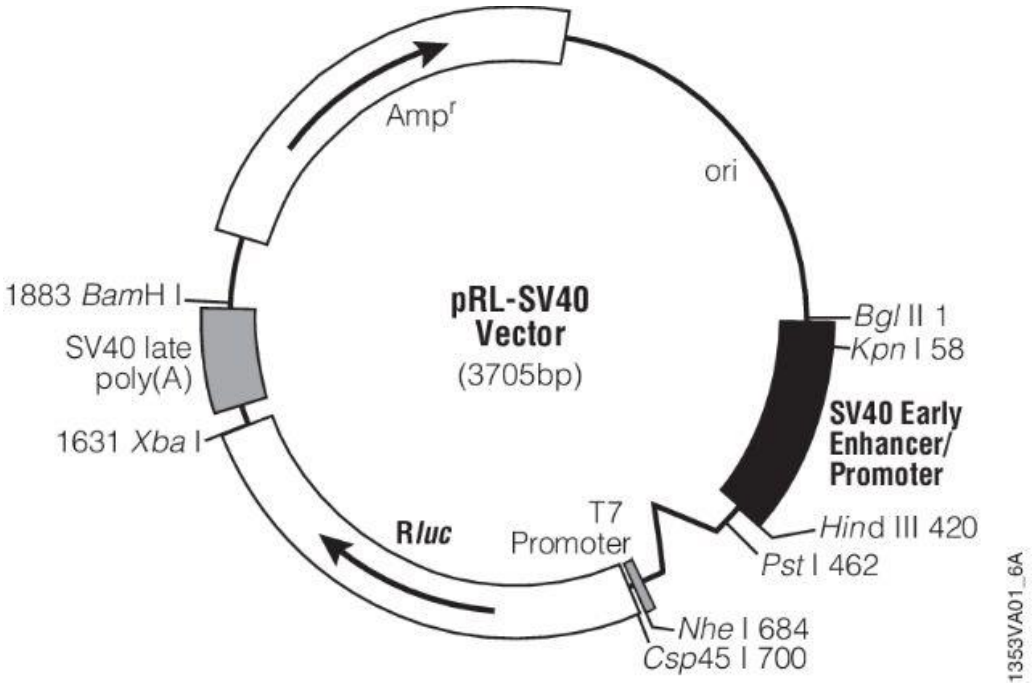


Figure 35. pRL-SV40 *Renilla* Vector

Image taken from the Promega pRL-SV40 technical manual

3.3.7 LUCIFERASE ASSAY

The Dual-Luciferase® Reporter (DLR™) Assay System (Promega Corporation, USA) was used to measure luciferase expression following the transfection of the *CACNA1C* intronic vectors into HEK293 cells. As a dual reporter assay system the DLR™ assay from promega quantifies firefly and *Renilla* luminescence sequentially thus providing an internal control that measures for baseline response. The vector containing the gene responsible for *Renilla* luminescence is the pRL-SV40 (Figure 35). This gene contains the SV40 enhancer and early promoter elements that ensure high levels of *Renilla* expression in co transfected cells. This is an essential attribute of the kit as experimental variabilitys such as pipetting errors, cell lysis efficiency and assay efficiency can have an impact on number of cells producing luminance regardless of the amount originally seeded. The ratio of firefly to *Renilla* luminescence can be calculated giving a more accurate indication of whether the insert is effecting the expression of luciferase. The components of the DLR™ kit are listed in Table 19.

3.3.7.1 PREPARATION OF CELLS FOR LUCIFERASE EXPRESSION MEASUREMENT

Measurement of luciferase luminescence was done following the manufacturer's protocol. Cells were co transfected with the pgl3-basic and prl-SV40 vectors and were left to grow for 48 hours. Prior to the addition of passive lysis buffer (PLB), DMEM was removed from each well and the cells were washed using 1ml of PBS. Following the addition of the PBS the plate was gently swirled to ensure all of the growth media was removed. To each well 500µl of PLB was added, which is sufficient to cover the cell monolayer in a 6 well plate. The plate was put on a shaker and left at RT for 15 minutes. The cell lysates are then removed and placed in a clean 1.5ml microcentrifuge tube and spun at 13,000rpm in a table top centrifuge for 30 seconds to remove the cellular debris. Following the spin the supernatant was placed in a new microcentrifuge tube.

3.3.7.2 MEASUREMENT OF LUMINESCENCE

Luminescence was measured on a Turbo 20/20 luminometer (Turbo Biosystems, UK). To the predispensed 100µl of the LAR-II solution 20µl of cell lysate was

Reagent	Function
Passive Lysis Buffer (PLB)	Passive Lysis Buffer promotes rapid lysis of cultured mammalian cells without the need for scraping of adherent cells from the wells. In addition it is designed to stabilise the firefly and *Renilla* enzymes and maintains minimal autoluminescence Prior to use, the 5x PLB was diluted in 4 volumes of distilled water to create a 1x working stock solution.
Luciferase Assay Reagent II (LAR II)	The addition of the LAR II generates a stabilised luminescent signal from the firefly luciferase gene. The LAR II solution is prepared by resuspending the lyophilised Luciferase assay substrate in 10ml of Luciferase Assay Buffer. Aliquots of 1ml were made and the mixture was stored at -80°C.
Stop & Glo®	Stop & Glo® reagent simultaneously quenches the firefly luminescence while activating the *Renilla* luciferase expression and stabilises the *Renilla* luminescence signal. This reagent was freshly prepared prior to each experiment. One volume of Stop & Glo substrate was added to 50 times Stop & Glo® buffer for the appropriate number of reactions.

Table 19. Components of the Dual Luciferase Reporter Assay Kit

added. The cell lysate-LAR II solution was mixed by pipetting once or twice. The sample is not vortexed as this can coat the sides of the tube preventing it mixing with Stop & Glo® once it is added in. The programme used to measure the reporter activity on the TURBO 20/20 was the Promega DLR II assay measurement programme. This programme incorporates a two second delay prior to reading and is followed by a ten second measurement period for each reporter assay sample. Following the reading of firefly luminescence, 100µl of Stop & Glo® reagent was added and the sample was vortexed briefly to ensure complete mixing of the LAR-II reagent and the Stop & Glo® reagent. The tube was replaced in the Turbo 20/20 machine and the *Renilla* luciferase activity measurement was taken over a period of ten seconds, and the final output and ratio was noted. To ensure reproducibility of the results the luciferase expression reporter assay was performed in triplicate each time and it was repeated three times.

Following the completion of all of the replications the results were analysed using an independent, two tailed T test using SPSS v.17. The two tail test was employed to discern whether there was an increase or decrease in the production of luciferase luminescence compared to the cells where there was no *CACNA1C* insert.

3.3.8 Electromobility Shift Assay

Electromobility shift (EMSA) assays are designed to detect the formation of protein complexes with nucleic acids [426]. EMSA assays are based on the principle that the formation of protein-nucleic acid complexes will move more slowly through the gel that free nucleic acid. It is used both as a qualitative and quantitative method for determining whether DNA binds to complexes using fluorescently labelled oligos. These are then used to perform a competition assay against unlabelled oligos. If the binding of the labelled DNA probe to the protein complex is specific, the visible band should completely disappear when an unlabelled oligo, with the same sequence as the labelled oligo, is added in excess.

3.3.8.1 Preparation of nucelar extracts from HEK293 Cells

HEK293 cells were grown in Nunc T-75 sterile filter cap flasks (Thermo Scientific, UK) to approximately 75-85% confluency in DMEM, 10% FBS, 1% Pen/Strep. The growth medium was removed from the flasks and the cells were washed with ice cold 1xPBS. Cytoplasmic lysis buffer was then added to the flasks and the cells were harvested into pre-chilled 1.5ml eppendorf tubes using cell scrapers (Thermo Scientific, UK) (Appendix II, II.VII). One millilitre of cells and cytoplasmic lysis buffer was placed in each tube. The cells were lysed by addition of 90µl of 10% NP-40 solultion to each tube and cells. The homogenates were centrifuged in a table top centrifuge at 2000g for 10mins at 4˚C. Following centrifugation the supernatant containing the cytoplasmic proteins was removed and discarded. The nuclear pellets were suspended in 32µl of nuclear lysis buffer. Rapid freeze-thaw cycles in dry ice and ethanol were used to disrupt the nuclei. Cellular debris was removed by centrifuging the samples from 10 minutes at 13000g. Following centrifugation the supernatant from each tube was removed and mixed together.

The sample was gently pipetted to mix the nuclear lysate and then samples were stored in aliquots of 20µl at-80˚C.

3.3.8.2 BRADFORD ASSAY

The Bradford assay quantification assay (Sigma-Aldrich, UK) was used to quantify protein concentration in the nuclear lysates according to the manufacturer's instruction. The reagent contains Brilliant Blue G, a dye which forms a complex with the proteins in the nuclear lysate solution. The formation of these complex results in a shift to the absorption maximum that is proportional to the amount of protein bound. Using a protein standard, such as bovine albumin (BSA), allows for the quantifications of proteins of unknown concentration.

The Bradford assay solution was brought to room temperature prior to performing the assay. BSA (10mg/ml) was diluted in nuclear to make protein standards of 1 µg/ml, 2 µg/ml, 5 µg/ml, 10 µg/ml, 20 µg/ml and 50µg/ml. One part of the protein sample was mixed with 30 parts of the Bradford reagent in a disposable cuvette and incubated at room temperature for five minutes. Following the incubation period, sample absorbance was measured using a UV spectrophotometer (Pye Unicam PU8600 UV/VIS spectrophotometer, Philips) at 595nm. Each sample was performed in duplicate and a standard curve was created to determine the concentration of the nuclear lysate by plotting the absorbance at 595nm against the protein standard concentrations.

3.3.8.3 PREPARATION OF THE 4% ACRYLAMIDE GEL

In a 50ml falcon tube, 42.12ml of MilliQ H_2O, 5ml of 40% polyacrylamide stock solution, 2.5ml of 10X TBE were added together (Appendix II, II.VII). The solution was mixed and then 300µl of 10% Ammonium Sulphate (APS) and 80µl of Tetramethylethylenediamine (TEMED) were added in. The addition of APS and TEMED initiates the polymerization of the acrylamide gel. The solution was thoroughly mixed and then poured between two 40cm gass plates with 0.75m^2 spacers and combs and allowed to set for one hour.

3.3.8.4 Oligo Preparation

Primers were designed for the *CACNA1C* intronic variants, rs79398153 (C/T) and rs116947827 (C/T). These oligonucleotides were 5' end-labelled with infrared dye DY-682 (Eurofins MWG, Germany). Complementary primers and unlabelled primers targeting the wildtype and mutated alleles of both variants were designed as described previously in section 3.1.4.1. Generation of the dsDNA from the oligos was completed by denaturing aliquots of 5µl 20pmol/µl at 100˚C for three minutes. These complexes were then cooled to room temperature prior to use.

3.3.8.5 Binding Reaction

The binding reaction for the EMSA was set up as described below in Table 20. 10µg of nuclear lysate was used per sample (Appendix II, II.VII). The use of Poly (dI.dC) in the reaction prevents non-specific binding of proteins in the nuclear extract leading to a reduction in background noise. The reaction was left in the dark at room temperature for 20 minutes for the binding reaction to proceed.

3.3.8.6 Electrophoresis

The 4% polyacrylamide gel was pre run for an hour on a LiCor 42000L sequencer (LiCor, Nebraska, USA). The gel was run at 10V/cm at 20˚C in 0.5xTBE buffer (Appendix II, II.VII). This pre run ensures that all traces of APS are removed and ensures the gel is at a constant temperature. Following the binding reaction incubation time, 1µl of Orange loading dye (10X) was added to the reaction mixture and the samples were mixed thoroughly. The samples were then resolved by electrophoresis on the 4% native polyacrylamide gel. The DNA-protein complex was detected using the LiCor 42000L.

Competition assays were performed for both rs79398153 and rs116947827 to determine binding specificity and to determine binding efficiency. The labelled probes were outcompeted using varying amounts of unlabelled oligos (10x, 20x, 50x, 100x and 200x). Ideally if the binding of the labelled probe is specific, the addition of the excess unlabelled for the same variant should result in the observed band disappearing. The principle behind the competition reaction is that excess

Reagents	LP	LP & NCL	LP & NCL & 10x ULP	LP & NCL & 20x ULP	LP & NCL & 50x ULP	LP & NCL & 100x ULP	LP & NCL & 200x ULP
Parker Buffer	2	2	2	2	2	2	2
2.5mM DTT	2	2	2	2	2	2	2
$MgCl_2$ (100mM)	1	1	1	1	1	1	1
10% NP40	1	1	1	1	1	1	1
Poly dI.dC	1	1	1	1	1	1	1
Labelled Oligo (50fmol)	1	1	1	1	1	1	1
Unlabelled Oligo (50fmol)	0	0	1	1	1	1	1
H_2O	12	11	10	10	10	10	10
Nuclear Lysate (10 µg)	0	1	1	1	1	1	1
Total Volume	20	20	20	20	20	20	20

Table 20. EMSA Conditions

Each EMSA shift assays used an increasing excess of unlabelled probe to determine the specificity of the binding seen on the gel. LP, labelled probe, NCL, nuclear cell lysate, 10x, 10-fold excess, ULP, Unlabelled probe.

Reagents	LP	LP & NCL	LP & NCL & YY1
Parker Buffer	2	2	2
2.5mM DTT	2	2	2
$MgCl_2$ (100mM)	1	1	1
10% NP40	1	1	1
Poly dI.dC	1	1	1
Labelled Oligo (50fmol)	1	1	1
H_2O	12	11	9
Nuclear Lysate (10 µg)	0	1	1
YY1 antibody (H-10)	0	0	1
Total Volume	20	20	20

Table 21. EMSA supershift reaction mixture.

The EMSA reaction mixture was prepared for each of the variants labelled probes with nuclear cell lysate (NCL), Labelled probe (LP) and YY1 antibody. The antibody was incubated with NCL for 20 mins prior to the addition of the labelled probe.

unlabelled DNA should compete for binding sites in the protein complex and eliminate the labelled oligo resulting in the removal of the observed band. Each of the competition reactions were conducted three times.

3.3.9 EMSA SUPERSHIFT ASSAY

EMSA supershift assays are slight modifications to the EMSA shift assay. In this assay antibodies specific to the putative transcription factor binding site are incubated with the nuclear cell lysate prior to the addition of the labelled probe. If the antibody recognises the protein in the nuclear cell lysate one of two results will be observed. If the antibody does not inhibit the binding of the labelled probe a super shift in the band will be observed due to the high molecular weight of the antibody. Secondly, the antibody could interfere with DNA and protein interactions and result in an absence of a band. YY1 Antibody (H-10) X (Santa Cruz Biotechnology) is a monoclonal mouse IgG_1 representing the full length of YY1 of human origin and was used for the supershift assays. The EMSA supershift assay is listed in Table 21; the antibody was incubated with the nuclear cell lysate for 20 minutes away from light prior to the addition of the labelled probe. The samples were then added to the gel as described in 3.3.8.6. The DNA-protein complex was detected using the LiCor 42000L.

Chapter 4 High-resolution melting curve analysis of regulatory regions in Calcium Channel genes

4.1 INTRODUCTION

Calcium influx into cells plays a prominent role in our biological systems, from cardiac and skeletal muscles to neurons and the CNS, it acts as a secondary messenger, instigating cascade pathways and regulating transcription. Given calcium's ubiquitous presence throughout the human body, and the physiological properties it contributes to our systems, it is not surprising that variants in calcium channel genes can have detrimental effects *in vivo.* Perhaps the best example of a susceptibility gene for psychiatric illness is *CACNA1C* which encodes the α pore of the L-type voltage dependent family of calcium channel genes (VDCC). As previously discussed in section 1.3.2, the *CACNA1C* variant rs1006737 has been repeatedly associated through genome-wide association studies in bipolar disorder.

L-type calcium channels genes are expressed ubiquitously throughout the brain. They are highly associated with activation of the CREB cycle through phosphorylation leading to a role in transcriptional regulation and activation [315]. Additionally, L-type VDCC's play a role in transcriptional regulation through calcium influx into the cytoplasm which initiates the Ras-MAPK pathway [310]. The Ras-MAPK pathway mediates the transfer of signals received on the cell surface to the nucleus of the cells through a protein cascade signalling pathway.

L-type calcium channels are characterised by high voltage activation with large single channel conductance, slow voltage dependent inactivation, up-regulation marked by cAMP-dependent phosphorylation pathways and inhibition by Ca^{2+}

blockers [327, 328]. The *CACNA1C* alpha pore is expressed in a number of brain regions including the cerebellum and the hippocampus. Variants p.G406R and p.G402S in the *CACNA1C* gene lead to Timothy syndrome, a disorder characterised by developmental delay and autism [363].

One of the top association signals in BD comes from rs1006737 (P= 7.0 × 10^{-8}) [232, 233], lying in the third intron of *CACNA1C*. Another two SNPs located in this intronic region, rs4765913 (p= 1.52 × 10^{-8}) [232] and rs1024582 (p= 1.7 × 10^{-7}) [270], are also strongly associated with BD. GWAS results across five different psychiatric illnesses further implicates rs1024582 in susceptibility to both BD and SCZ [268].

An accumulation of evidence has been building which suggests that some of the main variation associated with an increased susceptibility to BD is to be found in regions that are not currently being covered by exome sequencing. A recent study conducted by Ament *et al.* has identified a burden of variants in promoter, 5' and 3'UTR regions which are associated with BD [292]. GWAS studies, such as the study from the PGC-BD, show regions associated with BD but the top associated SNPs are often not thought to be causative or functional, and are located outside exonic regions [270]. Due to the heterogeneity and complex nature of BD, the ability of GWAS to identify causative genes is limited. In general, GWAS studies have reported robust association between disease or common traits and SNVs. Many of the associations explain only a fraction of the heritability estimated from family studies. It is hypothesised that rare variants contribute to the genetic variance of disease risk and quantitative trait variation. Current GWAS platforms are not designed to capture rare variants making it difficult to infer the association of rare variants using this method. Utilising data and results from GWAS studies allows variant investigation into SNVs located in regions such as the 5 and 3'UTR, and promoter regions. It can aid in the identification of regions where association signals are seen which may be arising as a result of LD.

Gene ontology enrichment analysis of 34 SNPs reaching a genome-wide significant association ($P \leq 5 \times 10^{-8}$) in a large-scale genome BD association study have found

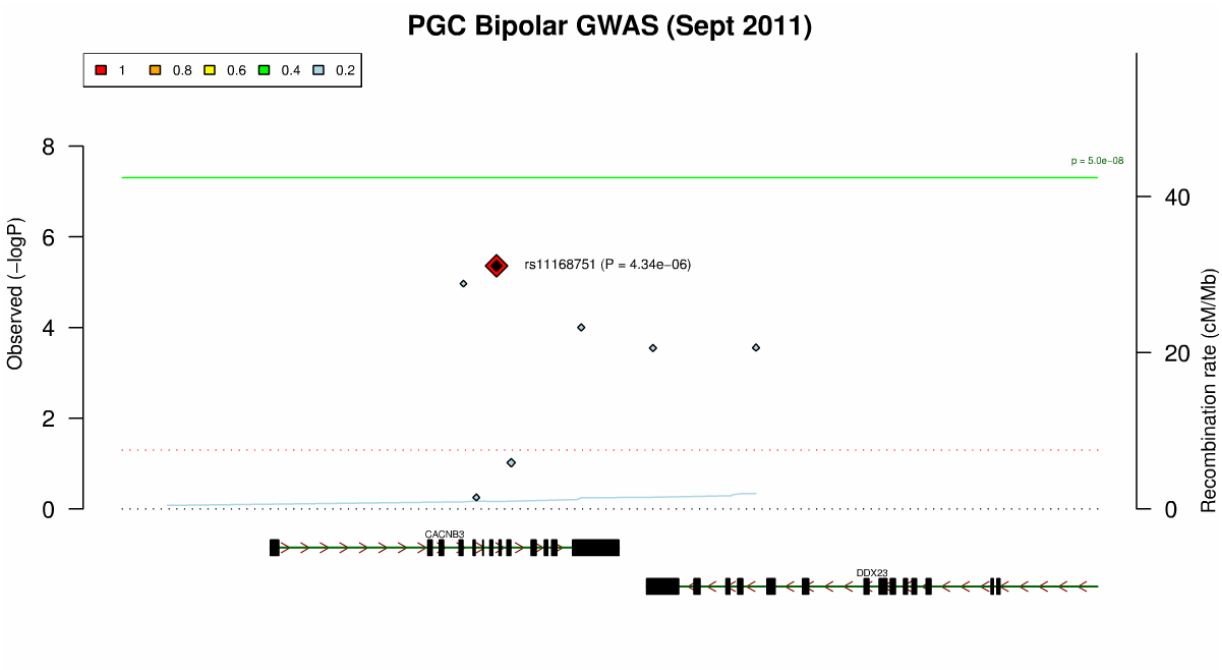


Figure 36. Regional association plot of CACNB3.

The regional association plot was obtained from http://www.broadinstitute.org/mpg/ricopili/. The data is from the PGC-BD GWAS [270]. One of the polymorphisms, rs11168751, shows nominal significance of P=4.43x10^{-6} but this does not reach the threshold level of genome-wide significance(P=5x10^{-8}). The Y axis shows the –log10 of the P value.

enrichment in the GO: 0015270, dihydropyridine-sensitive calcium channel activity. Of the eight genes in this category, 3 are members of the VDCC family, *CACNA1C, CACNA1D* and *CACNB3* [239]. The regional association plot for *CACNB3* is shown in Figure 36. No SNVs in this gene pass the threshold for genome-wide significance which is set at P=5x10^{-8} [250].

Numerous genome-wide association studies have shown consistent association signals originating from a SNP in the *CACNA1C* gene. The PGC-BD GWAS has found a large association peak in the third intron of *CACNA1C* (Figure 37). The group has also reported nominal association signals for the *CACNA1D* region (see Figure 38). The combined evidence of calcium channel gene associations from BD GWAS studies and the involvement of the calcium channels in neuronal networking pathways make this set of genes an appropriate candidate family to study in an attempt to identify variation that may increase an individual's susceptibility to BD.

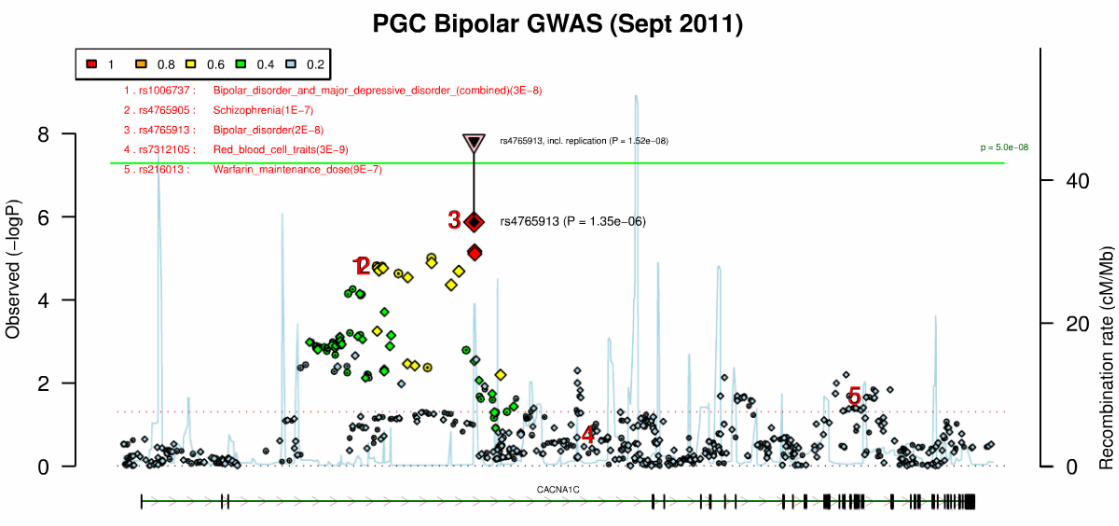


Figure 37. Regional association plot of *CACNA1C*.

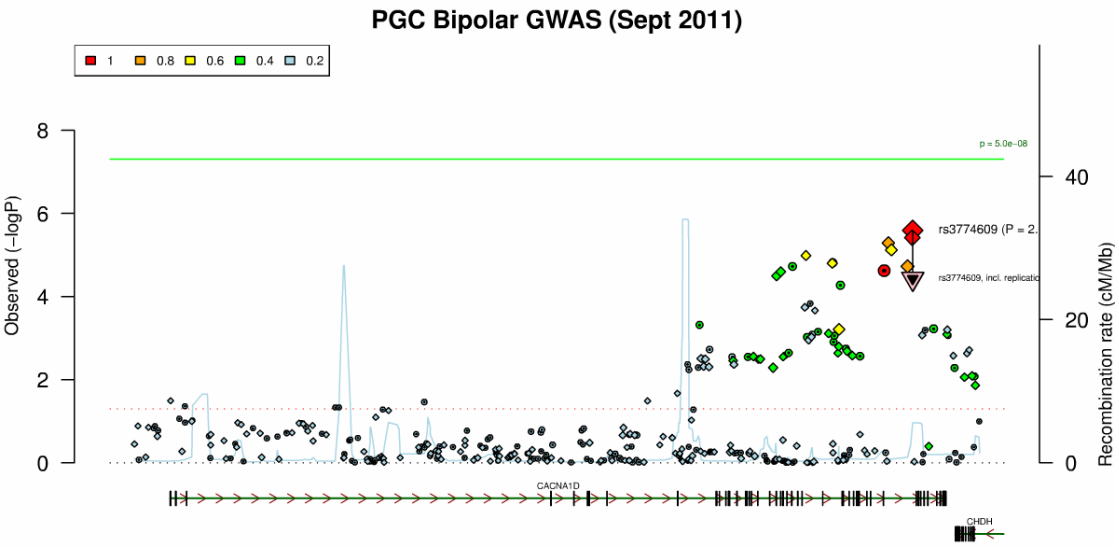


Figure 38. Regional association plot of *CACNA1D*.

The regional association plots for *CACNA1D* and *CACNA1C* were obtained from http://www.broadinstitute.org/mpg/ricopili/. The Data reflects the GWAS conducted by the PGC-BD [270]. The X axis shows the chromosome number and the position of the genetic region in kilobases. The Y axis shows the $-\log_{10}$ of the P value. In *CACNA1C,* one of the polymorphisms, rs1006737, in intron 3 reaches genome-wide significance (p=3x10^{-8}). rs4765905 and rs4765913 have also been associated with SCZ (p=1x10^{-6}) and BD (p=2x10^{-8}) respectively. None of the or polymorphisms reach genome-wide significance in *CACNA1D*, however close to the 3' end of the gene there appear to be some variants with nominally significant levels of association.

4.2 HYPOTHESIS

Members of the calcium channel family of genes have been associated with BD through GWAS, such as *CACNA1C,* or through gene pathway enrichment analysis such as *CACNA1D* and *CACNB3*. There is a certain limitation to the variants discoverable through GWAS as many rarer variants are not present on GWAS array platforms used to conduct these analyses. Current exome capture techniques often missed variants which are located in regulatory regions of the genome Figure 39. I focused on the afore mentioned genes in an attempt to identify polymorphisms in BD cases which were completely absent or present at a lower frequencies in ancestrally matched controls.

Genetic variants located in regulatory regions can have a detrimental impact on the control and initiation of gene expression which would lead to aberrant expression of either the α subunits or the auxiliary subunits in the calcium channel gene family. It is possible that changes to expression of these genes could lead to disruption of intracellular calcium homeostasis and have an impact on downstream processes such as long term potentiation or synaptic plasticity in neurons.

4.3 AIMS

The aim of this study was to screen the promoter, 5'UTR, 3'UTR and first exons of *CACNA1C*, *CACNB3* and *CACNA1D* using HRM for rare novel variants which may confer an increased susceptibility to BD. The data from ricopili as shown in Figure 36, Figure 37 and Figure 38 from the PGC BD GWAS demonstrates that there is association between these genes and BD. Common variation in the regulatory regions of these genes do not appear to be associated with BD from the PGC BD GWAS. We used this data for selection of the calcium channels to follow up, with a specific focus on the regulatory regions and rare variants that may be present in the genes. On detection of a novel variant, genotyping would be undertaken in the full case-control cohort to investigate whether the variant was associated with BD

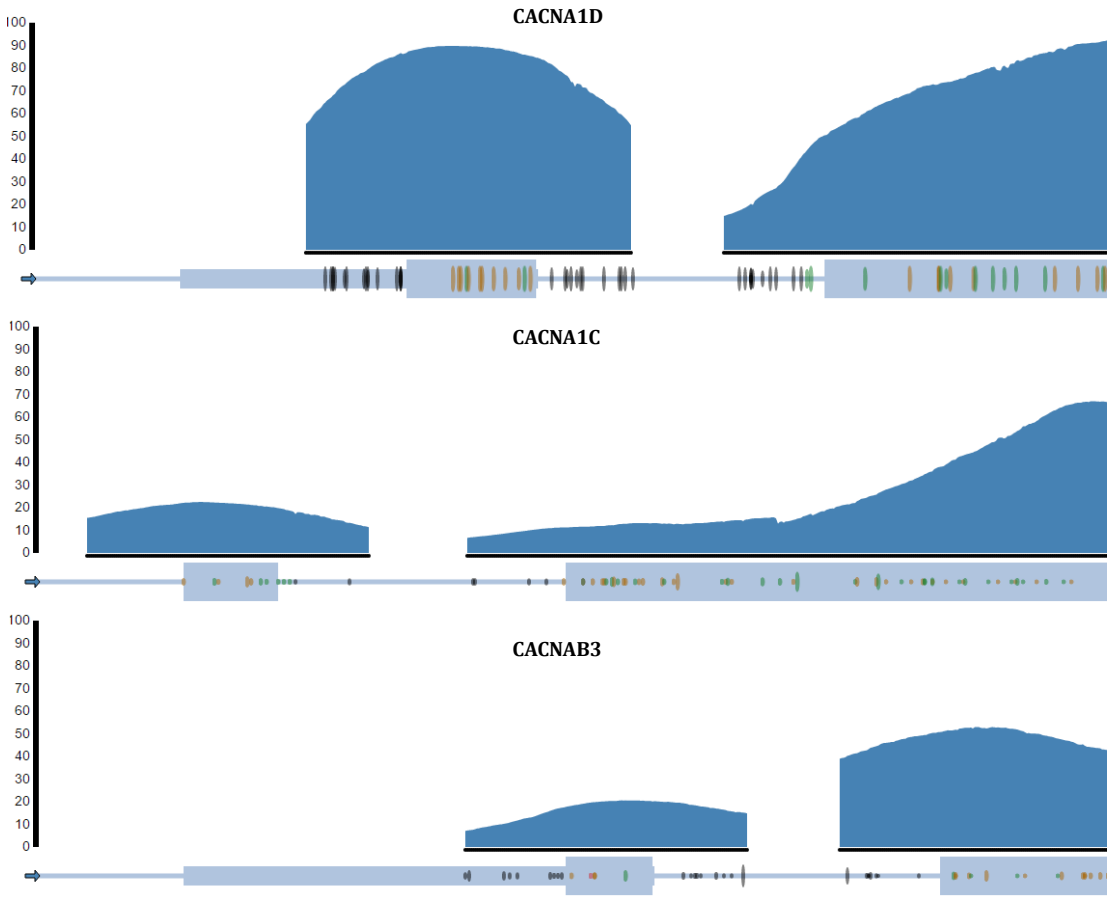


Figure 39. Exome sequencing coverage of regulatory regions in *CACNA1D*, *CACNA1C* and *CACNB3*.

Exome capture over the 5'UTR, promoter region and the first exon of *CACNA1D*, *CACNA1C* and *CACNB3* is relatively low in exome sequencing from the ExAC database (http://exac.broadinstitute.org/)

4.4 HRM METHODS FOR SCREENING *CACNA1C*, *CACNA1D*, AND *CACNB3*

The genetic sequence for each of the genes was retrieved using the UCSC genome browser (https://genome.ucsc.edu/). Primers for each genetic amplicon were designed using Primer 3 software as described previously (3.1.4). The list of primers used for HRM analysis of *CACNA1C, CACNA1D* and *CACNB3* are listed in Appendix II**Error! Reference source not found.**. The regions screened by HRM were the 5'UTR and the promoter region, defined as 1000bp upstream of the transcriptional start site.

To assess the presence of variation in the regulatory regions of *CACNA1C*, *CACNA1D*, *CACNB3* 1,098 BD cases were utilised for HRM analysis. Following sample normalisation individuals who presented with aberrant curves were sent for Sanger sequencing to determine whether they were in fact real variants and not false-positives. Variants deemed to have potential aetiological effects, were followed up KASPar genotyping in 1098 BD research subjects and 1095 ancestrally matched controls.

4.5 RESULTS

4.5.1 HRM SCREENING OF GENES

HRM analysis of the promoter region and the first exon of *CACNA1C, CACNA1D* and *CACNB3* identified three novel variants and three previously annotated variants from dbSNP37 in the UCL1 and UCL2 cohort. These were validated using Sanger sequencing. The validated results are summarised in Table 22. The number of cases found with each variant through Sanger sequencing is also listed in Table 22. These numbers are not reflective of allele frequency as variants were chosen for sequencing based on aberrant curves. Novel variants are named by their chromosomal location and nucleotide change. The promoter region of *CACNA1C* harboured four variants rs538323756, rs11062091, rs112412740 and 12:2162570 (C/G). All original SNV verification was conducted using GRCh37/hg19 build of the genome on the UCSC genome browser using the dbSNP137 release. Variant 12:2162217 (G/A) was not present on dbSNP137 but has been since given an 'rs' number, and so was considered a novel variant for this study. Variant rs312480 was detected in *CACNA1D* and 12:49212493 (C/G) was detected in *CACNB3* in the promoter regions of both genes. Due to the extremely GC rich sequence and the presence of several repeat runs of sequence the second promoter amplicon in *CACNA1D* was not screened. This was because the PCR primers were unable to bind and amplify the regions efficiently due to the refractory nature of regions of high GC content and the presence of secondary structures. Secondary sets of primers were designed in an attempt to rectify this issue and several alternative PCR assay mixes were used that included deaza-7-dGTP and/or Phusion high fidelity DNA polymerase (New England Biolabs) that

Gene	Region	SNV ID	Chromosomal position (hg19)	Variant	1000G MAF	AA/Aa/aa
CACNA1C	Promoter	12:2162217 (G/A)*	chr12:2162217	G/A	-	38/2/0
	Promoter	rs11062091	chr12:2162369	G/A	0.148	19/14/7
	Promoter	rs112412740	chr12:2162548	G/A	0.0024	63/1/0
	Promoter	12:2162570 (C/G)	chr12:2162570	C/G	-	63/1/0
CACNA1D	Promoter	rs312480	chr3:53529140	A/T	0.1406	30/21/2
CACNB3	Promoter	12:49212493 (G/C)	chr12:49212493	G/C	-	22/2/0

Table 22. HRM variants from analysis in *CACNA1C, CACNA1D* and *CACNB3*.

Some variants currently have dbSNP numbers but dbSNP137 was used for SNV validation. SNVs which are present on dbSNP138 or above are as following *rs538323756. MAF was obtained from the 1000G data set.

had previously been shown to improve the amplification of GC-rich DNA sequences [427]. These strategies were not successful and it was not therefore possible to accurately identify variants in these regions using HRM analysis.

4.5.2 Investigation of the Effects of Variants

Bioinformatic analysis was conducted on all variants to discern whether their presence had the potential to alter transcription factor binding sites in the promoter regions of each gene. This was done using Transcription Element Search System (TESS) which has been described in section 3.2.5 for variants located in non-coding regions of the genome. TESS scores the likelihood of transcription factor binding sites using positional weighted matrices against TFBS databases TRANSFAC 4.0, JASPAR, IMD and CBIL-GibbsMat databases. Positional weightings are calculated and the log-odds score is shown for the putative transcription factor binding site. The results for each variant are listed below.

The *CACNB3* variant 12:49212493 (G/C) is located in the promoter region of the first two isoforms of CACNB3, and in the intron of the secondary isoforms, transcript 3 and transcript 4. It is 221bps upstream of the start codon. The 12:49212493 (G/C) variant results in the creation of a CDC5 TFBS. The CDC5 transcription factor has been implicated in transduction of cytoplasmic signals to the nucleus following serum stimulation (Figure 40)[428].

12:49212493 (G/C)

Wild-type G

```
ctcagccgcg ctcggggtgg gaccGgctgg gtttgggggg
                          ====== = ACE2 (9.53)
                          ====== == SWI5 (10.60)
```

Mutation C

```
ctcagccgcg ctcggggtgg gaccCgctgg gtttgggggg
                        ======== === CDC5 (8.25)
                          ====== = ACE2 (9.53)
                          ====== == SWI5 (10.60)
```

Figure 40. TESS predictions for 12:49212493 (G/C) in the *CACNB3* promoter.

The TFBS highlighted in green indicate new TFBS which are gained in the presence of the variant. The 12:49212493 (G/C) variant in the promoter region of *CACNB3* results in the creation of the CDC5 TFBS. The final two TFBS's ACE2 and SWI5 are not affected by the resulting base change suggesting that the G/C allele is not essential for the creation of this binding site.

TESS results for rs538323756 are shown in Figure 41. It was predicted that when the wild type allele of rs538323756 was present an estrogen receptor (EsR) alpha Sp1 TFBS would be created. EsR-alpha SP1 is a member of a family of nuclear reporters and are ligand-activated transcription factors [429]. EsR-alpha activates transcription either directly or by activation of other transcription factors in a ligand dependent manner. EsR-alpha forms a complex with the SP1 transcription factor in a ligand activated manner [430]. The binding of ER-alpha SP1 to nuclear transcription factors is involved in activation of the cell cycle [431]. In the presence of the variant allele this transcription factor binding site appears to be lost, and the change results in the creation of two new TFBS. No information pertaining to the T-AG transcription factor could be found. The E2F family of transcription factors are involved in a number of cell cycle processes such as regulation of cellular differentiation, apoptosis and proliferation [432, 433]. The p107 transcription factor is a member of a tumour suppressor family and regulates E2F DNA binding activity and leads to repression of transcription through recruitment of repressor complexes to promoters [433].

TESS results for 12:2162570 (C/G) in the *CACNA1C* promoter region also predicts changes to the TFBS sequence (Figure 41). When the wildtype C allele is present there are three TFBS which are lost in the presence of the variant G allele. These

12:2162217 (G/A)

Wild-type G

```
agcggcgctc ggcgcggcgc ggcgGgcccg gagcggcggc ggcggctctt
              ======= = GCF (11.1281)
                      ====== ER-Alpha (12.00)
                                    ======= GCF (10.00)
                                    ======== = WT1-KTS (18.00)
```

Variant A

```
agcggcgctc ggcgcggcgc ggcgAgcccg gagcggcggc ggcggctctt
              ======= = GCF (11.1281)
                   == === T-AG (6.4952)
                      ==== E2F+p107 (8.00)
                                    ======= GCF (10.00)
                                    ======== = WT1-KTS (18.00)
```

12:2162570 (C/G)

Wild-type C

```
gggaatcagg taatCgtcgg cggggaag
  ====== GCN4 (12.00)
     ===== == NF-GMb (13.00)
           ====== Prd (9.4321)
            ===== ICP4 (12.00)
                ===== ====== ABF1(10.00)
                 ==== RAF (8.00)
```

Variant G

```
gggaatcagg taatGgtcgg cggggaag
  ====== GCN4 (12.00)
     ===== == NF-GMb (13.00)
         = ===== IPF1 (10.00)
           ===== YYI (10.00)
                 ==== RAF (8.00)
```

Figure 41. TESS predictions for 12:2162217 (G/A) and 12:2162570 (C/G) in the *CACNA1C* promoter.

The TFBS highlighted in red indicate ones which are lost in the presence of the mutated allele and the green TFBS are ones which are gained in the presence of the variant allele. The 12:2162217 (G/A) variant allele results in the creation of two new TFBS, T-AG and E2F+p107 and the loss of the ER-Alpha TFBS. The 12:2162570 (C/G) variant results in the loss of three TFBS Prd, ICP4 and ABF1, and the addition of two new TFBS IPF1 and YY1. The numbers after the transcription factors in the diagrams are the log of the odds score for the conservation of the bases here for each TFBS and represent the strength of the prediction for the TFBS.

are PRD, ICP4 and ABF1. The PRD transcription factor controls transcriptional activation and anti-termination [434]. ICP4 is a regulator of viral gene expression associated with herpes simplex virus type 1 [435], it can act as both a promoter and inhibitor of transcription depending on its promoter structure. ABF1 is a member of the basic helix-loop-helix proteins. It contains a transcriptional repressor domain and it is purported to be involved in regulating cell differentiation [436]. The IPF1 (independent of pancreatic development 1) transcription factor is vital both during development and throughout later life for pancreatic development. It was originally thought that expression of this transcription factor was solely located to cells in the developing foregut [437], however other studies suggest that IPF1 transcription factor is also found in neuronal precursor cells in developing mice [438] and may play a role in the development of the CNS. Yin Yang I (YY1) is a ubiquitous transcription factor which plays a fundamental role in transcription regulation [439]. It has been linked to initiation of transcription in addition to activation and repression of transcription, giving it its name Yin Yang [439, 440].

4.5.3 GENOTYPING OF VARIANTS FROM HRM

Variants selection for follow up genotyping was based on previously described criteria (see section 3.2.5). The variants were required to be novel and result in changes to transcription factor binding sites in the regulatory regions. Three variants from the HRM analysis of *CACNA1C, CACNA1D* and *CACNB3* fitted these criteria. Two promoter region SNVs from *CACNA1C*, 12:2162217 (G/A) and 12:2162570 (C/G), and 12:49212493 (G/C) in the promoter region of *CACNB3.* Optimization of assays for the genotyping of the SNV from the promoter region *CACNB3* failed repeatedly. This variant was validated through Sanger sequencing from the stock DNA but due to the fact that the *CACNB3* variant is located in a region of high GC content it was not possible to optimise an assay to genotype this variant in the case control cohort.

The *CACNA1C* SNV 12:2162217 (G/A) was present as a heterozygote in 42 controls, and as a homozygote in two controls. This high frequency was reflected in the case population with 43 heterozygous cases confirmed through genotyping (p=0.79).

Gene	SNP	Eur MAF	Cohort	Sample number	Genotype calls AA/Aa/aa	MAF	P	Fishers Exact	OR (95% CI)
CACNA1C	12:2162217 (G/A)	-	BD	924	881/43/0	0.023	0.79	0.44	1.06 (0.7-1.6)
			Control	934	890/42/2	0.027			
CACNA1C	12:2162570 (C/G)	-	BD	938	930/1/0	0.001	0.31	1.00	NA
			Control	922	922/0/0	NA			

Table 23. Genotyping results for calcium channel variants detected by HRM.

MAF, minor allele frequency, P, Chi Square p value, OR, Odds Ratio, CI, Confidence Interval, NA, Not Applicable.

12:2162570 (C/G) was present in one BD sample, and not present in the control cohort (p=0.31). Neither of these variants were associated with BD in our sample (Table 23).

4.6 DISCUSSION

HRM failed to identify any genetic variants associated with BD in the regulatory regions and first exons of *CACNB3*, *CACNA1C*, and *CACNA1D*. This is typical in the field of complex genetics. Various disease models have been proposed for BD, including common genetic variants with small effect sizes and rare variants with large effect sizes. This is mirrored in the results that are coming from GWAS and next generation sequencing studies where large sample sizes are needed to detect association between BD and rare SNVs. There are a number of reasons why this may have occurred. Firstly the parameters set for SNV follow up genotyping in the complete UCL cohort limited the follow up variants to novel variants which had not been previously reported. This process led to the inclusion of three variants in the follow up analyses. The remaining variants may contribute susceptibility to BD but this cannot be determined given the original experimental parameters set for variant inclusion.

HRM analysis allowed for the verification of SNPs in regions which were GC rich. However these GC rich regions were refractory to the development of follow up genotyping assays because we were not able to design assays to adequately genotype these variants.

In silico estimates of the predicted effect of the 12:2162570 (C/G) variant were that this change would lead to the loss of three TFBS and the creation of two new TFBS. This information does not provide enough detail to determine whether the variant may result in an increased or decreased expression of *CACNA1C.* All three of the transcription factors whose binding sites were lost are involved in transcriptional regulation, ABF1 was purported to play a role in cell differentiation, IPF1 involved in regulation of gene expression which can act as both an activator or repressor of gene transcription depending on the shape of the promoter structure and finally Prd that controls transcriptional activation and anti-termination. IPF1 has been shown to be present in regions that generate neurons during the time of active neurogenesis, additionally the tethering of YY1 to the histone deacetylase protein, HDAC3 and HDAC2, result in transcription repression through histone deactylase activity [441, 442]. If the creation of the two new transcription factor binding sites were to lead to a repression of transcription of *CACNA1C* it could have knock on effects on transcriptional regulation and control of the CREB cycle.

It is interesting that the 12:2162570 (C/G) variant was present in only one individual with BD in the approximate 1,000 individuals genotyped. With this data alone it is difficult to determine whether this variant may play a role in the aetiology of BD. Evolutionary theory holds that deleterious variants by default are likely to be rare due to purifying selection [443, 444]. The sample size of the original exploratory cohort used to detect variant present in the regulatory regions of *CACNA1C, CACN1D* and *CACNB3* was of adequate size in which to find common and rare variation. The ability to detect an association between rare variants and a disease, is based on a number of factors, one is the relative risk of the variant, as estimated by 1+ λ, where λ is the excess relative risk of disease conferred by the allele [445]. The second is the sample size. Zuk *et al* have determined that to detect allele association with a 2-fold increase to disease risk, approximately 28,000 samples are needed. For variants with larger relative risk scores, such as a 20-fold

increase in relative risk, 260 cases would be needed [445]. Our sample size is appropriate for variant identification and investigation of either common variants, or rare variants with larger effect sizes.

The finding of a rare case-only variant suggests that replication of this variant in a larger sample size would be needed to determine whether it confers a relative risk to the development of BD. Previous reports have suggested that variants located in the promoter, 5' and 3'UTR regions provide a stronger contribution to risk for BD than coding variants [292]. In general, regulatory regions are less conserved than coding regions of the genome which could account for the greater presence of rare variants here.

It is clear to see from the above data that HRM provides a valuable resource for high throughput variant detection in specific regions of genes. Given adequate amplicon and sample sizes both common and rare novel variants were detected using this method. If known variants are present in a region, smaller amplicon regions can be utilised to genotype all variants. This could be used in future for variants, such as the *CACNB3* 12:49212493 (G/C), to screen for variants whose genotyping assay optimisation failed in the lab.

Chapter 5 High resolution melting curve analysis of *CACNG4*

5.1 INTRODUCTION

Previous observations have indicated that several members of the calcium channel family of genes, including the TARP family of subunits, could potentially be implicated in BD susceptibility. TARPs consist of two subgroups, the canonical type I consisting of γ2, γ3, γ4 and γ8, and the atypical type II γ5 and γ7 [446]. Two members, γ1 and γ6, of this family of subunits are involved in interactions with the calcium channel pore and modifications of calcium influx into the cells through the α pore. These proteins play a vital role in trafficking and scaffolding of the AMPA-R to the postsynaptic density in neurons and are involved in promoting AMPA-R maturation [351, 356, 359, 388]. TARPs are expressed throughout the central nervous system (CNS) and display differing functions in their interactions with AMPA-Rs. The association of TARPs with AMPA-R have an important impact on AMPA-R function and induce AMPA-R modifications in function. This includes increasing single channel conductance, slowing their desensitization and deactivation and weakening blockage of calcium permeable AMPA-R by endogenous intracellular polyamines [362]. Each TARP differs in their influence on AMPA-R properties [351, 447]. Members of the TARP family can bind to a number of AMPA-R, but it is thought that only one TARP binds to the AMPA-R subunit in conjunction with their function at that given time [351, 359]. Additionally, TARPs play a role in canonical forms of synaptic plasticity through regulation of AMPA-R numbers in the hippocampus following long term potentiation (LTP) or depression (LTD) [351] and in Purkinje cells following LTD [448].

TARPs play a role in calcium entry into the cell through interaction with AMPA-R complexes that lack a GluR2 subunit. AMPA-Rs consisting of GluR-2 subunits are the most common receptors in the CNS and are calcium impermeable CI-AMPA-R [449, 450]. AMPA-R complexes lacking the GluR2 subunit are permeable to calcium

(CP-AMPA-R) and demonstrate greater channel conductance than GluR2 containing subunits [451, 452]. The interaction of TARP-γ2 with CP-AMPA-R results in increased calcium permeability through the attenuation of polyamine blocking [453, 454]. It has been previously demonstrated that AMPA-R desensitization and deactivation is prolonged through their interactions with TARPs [455]. This property is mediated by the first extracellular loop of the TARP protein [362].

As mentioned in section 1.3.2.4 members of the γ family of subunits have been associated with severe neurological phenotypes. *In vivo* studies into *CACNG2* gene function, encoding TARP-γ2, demonstrate that in KO- γ2 mice demonstrate severe ataxia [272, 456]. This ataxia is a result of the absence of surface and synaptic expression of AMPA-R in cerebellar granule neurons. This is due to the role γ2 plays in trafficking the AMPA-R to the PSD. Motifs in the γ2 protein have recently been found to interact with the Arc protein [457], which is transcriptionally upregulated in models of learning [458]. Arc functions at neural synapses and its transcription has been be linked to neural activity which is involved in information processing and storage [459]. Arc binds to γ2 amino acid Y229 through a unique hydrophobic pocket in the N-lobe domain of the Arc protein [457]. This binding of Arc to γ2 can mediate down regulation of AMPA-R receptors during synaptic scaling, and this regulation may be determined by TARP gene expression [457]. While γ2 does not show a genetic association with bipolar disorder or schizophrenia, the presence of the critical residue for Arc binding is of interest. Association between schizophrenia and genes encoding proteins containing an Arc N-lobe consensus sequence have been reported [411, 460]. The γ2 and γ4 share 58.5% homology as determined by Uniprot (http://www.uniprot.org/) (Figure 42) . The consensus sequence for the Arc N-lobe protein appears to be present in the γ4 protein. This however is not indicative of binding to Arc N-lobe as the TARP γ3, which has 73.7% homology with TARP γ2, has evolved to avoid Arc binding in higher vertebrates [457].

A case-case analysis between BD and SCZ, in the UCL cohort conducted by Curtis *et al,* has shown an association in the 81kb intragenic region between *CACNG4* and *CACNG5* (see Figure 43 [271]). Both *CACNG4* and *CACNG5* are members of the voltage dependent calcium channel genes and encode a type-I TARP and a type-II

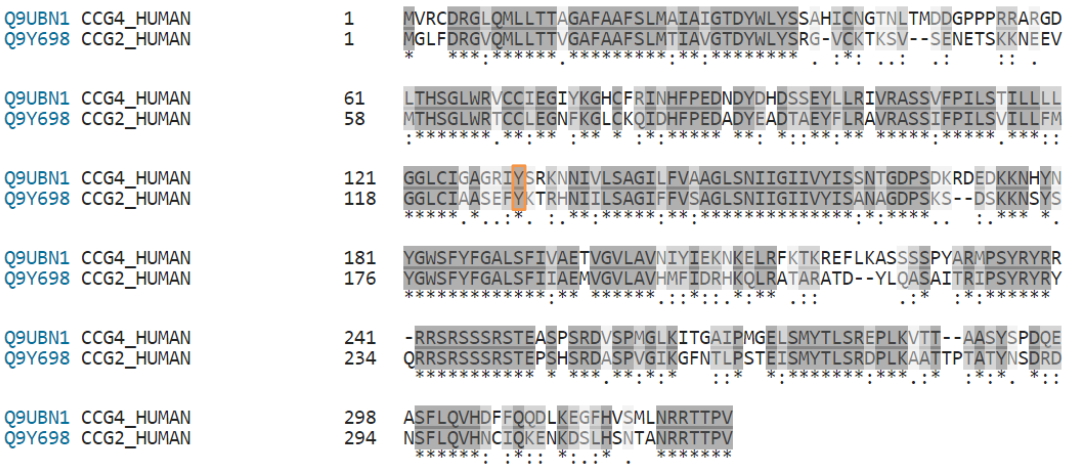


Figure 42. TARP γ2 and γ4 protein homology.

The Arc binding domain at *CACNG2* Y229 (orange square) is also conserved in *CACNG4* at Y236. This is not a pre-requisite for binding at the Arc N-Lobe domain because the *CACNG3* protein also has this domain and it doesn't bind the Arc protein.

TARP respectively. Data from the PGC-BD GWAS does not report any association signal arising in the *CACNG4* gene as shown in Figure 44. The association signal seen in the case-case study between the SCZ cohort and the BD cohort could be driven by variants located in either of the genes. The properties of the TARP family of proteins, and their role in synaptic plasticity, make them an interesting gene family to study in the search for variants underlying genetic contribution to psychiatric disorders such as BD and SCZ.

5.2 HYPOTHESIS

While specific genes have been identified through GWAS or follow-on candidate gene studies, aside from the meta-analysis from Curtis *et al* there has been very little research of disease variant associations in the *CACNG4* gene. This is due, in part to the limited power of GWAS studies with many rare variants not reaching the significant threshold levels due to their lower frequencies in BD cases. Therefore, we focused on the *CACNG4*, a member of the TARP family of genes which regulates AMPA-R trafficking to the PSD in an attempt to identify polymorphisms found in

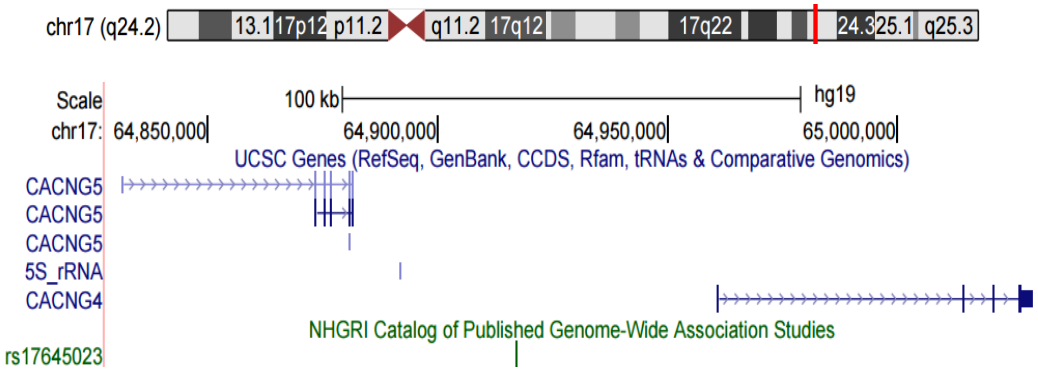


Figure 43. Location of the associated variant from case-case analysis between *CACNG4* and *CACNG5.*

A case-case study between the UCL BD and SCZ cohort found an association between *CACNG4* and *CACNG5*. The variant, rs17645023, is located 45kb *CACNG4* and 35kb from *CACNG5* [271]. Image from https://genome.ucsc.edu/.

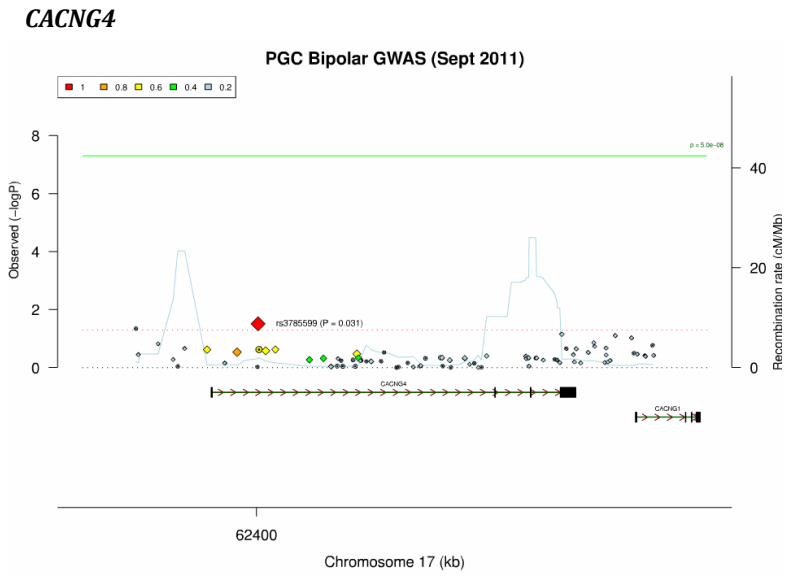


Figure 44. Regional association plot of *CACNG4*.

The plot is taken from the PCG BD GWAS. The plot is taken from the PCG BD GWAS. The x-axis shows the observed –logP value of each variant and the y-axis shows the chromosomal position. Variant rs3785599, in red, is associated with BD at p=0.031. The other SNPs coloured are in linkage disequilibrium with the associated SNP. The level of linkage disequilibrium is denoted by the colour of each variant. This variant does not pass the threshold level for genome wide significance The x-axis shows the observed *P*The ricopili plot was obtained from http://www.broadinstitute.org/mpg/ricopili/. The data reflects the PGC-BD GWAS [270].

BD cases and which were completely absent or present at lower frequencies in the ancestrally matched controls.

Amino acid changing variants in specific regions of the gene could result in aberrant trafficking of the AMPA-R to the PSD. In addition to this, it is still undetermined if γ4 protein interaction with AMPA-R subunits is specific to one of the four, or if it binds to all GluR subunits. The presence of variants which effect trafficking of the AMPA-R to the PSD could have an effect on synaptic plasticity, and calcium entry to the cell, through the CP-AMPA-R GluR1, GluR3, GluR4.

5.3 AIMS

All regulatory regions and exons of *CACNG4* were screened for variants using HRM analysis to establish whether the association signal seen in the BD vs SCZ analysis may be driven by a variant, or variants, harboured in these genetic regions. Further functional analysis was then undertaken to understand the biological effect of these variants.

Previous work has been conducted into novel variation in *CACNG5* using HRM analysis and this gene was not covered in the course of this PhD [390].

5.4 HRM SCREENING OF *CACNG4*

Primers for *CACNG4* are listed in Appendix II and were designed to cover the promoter region, the 5'UTR, all four exons and the 3'UTR. Full coverage of these regions was ensured by creating primers which overlapped with the amplicon before and after it, an example of which is shown in Figure 45.

1,098 BD cases were utilised for HRM analysis. Following sample normalisation individuals who presented with aberrant melt curves were sent for Sanger sequencing to determine whether they were in fact real variants and not false-positives. Variants deemed to have potential aetiological effects, were followed up KASPar genotyping in 1900 affected BD research subjects and 1350 ancestrally matched controls. The extra samples for genotyping came from newly extracted

***CACNG4* Exon 4 & 3'UTR**

taaagaggggaactgggggctgagagtgtgttttattgtatacctcatttgaactttgtactctatt
acccattcaacattaattactgaaaagaaaaattagcaaccagaataaagaaatactcatcaga
gaggggagtgtccacctgttccctacactgccgtcccactgtgggtctaacctctgcctctctccc
tccctggccccgctccccgctccccgcttccctttcagGCCTCAGTAACATCATCGGTAT
CATCGTCTACATTTCCAGCAACACAGGTGACCCGAGTGACAAGCGGGAC
GAAGACAAAAAGAACCATTACAACTACGGCTGGTCTTTTTACTTTGGAG
CTCTGTCTTTCATTGTGGCTGAGACCGTGGGCGTCCTGGCTGTAAACAT
TTACATTGAGAAAAATAAAGAGTTGAGGTTTAAGACCAAACGGGAATT
CCTTAAGGCGTCTTCCTCTTCTCCTTATGCCAGGATGCCGAGCTACAGGT
ACCGGCGACGGCGCTCGAGGTCCAGCTCAAGGTCCACCGAGGCCTCGCC
CTCCAGGGACGTGTCGCCCATGGGCCTGAAGATCACAGGGGCCATCCCC
ATGGGGGAGCTGTCCATGTACACGCTGTCCAGGGAGCCCCTCAAGGTGA
CCACCGCAGCCAGCTACAGCCCCGACCAGGAGGCCAGCTTCCTGCAGGT
GCATGACTTTTTCCAGCAGGACCTGAA<G>GAAGGTTTCCA<C>GTCAG
CATGCTGAACCGAC<G>GACGACCCCTGTGTGAgccgcctgccctttctctccgct
ccagcctctccccagaa<c>ggctcttttgtcacacaggatggcatgtgatcctcaagacgacg
aacaatgaactaaagccaaat<g>cagccctccctggcctccagaggtggcgtgggctggcttt
gcacgaaggttgtgctgggagaccggacccggggctgcagaagaagctgaaggctgactttgtc
ccctccccg<a>aaaggggtgttttgatgcctcagggtctctgaaatctcc<c>gggaagcccc
agagctttcctgaggctgcctggccttgatcaacttgggaagacaaaattgagccattatctcctc
ttggaaacgaatcttgccagaaaaacgggatttcagggccttccctccctgcctgggtgtcgggc
caccagaaggctctgc

Primer Pair 1
Primer Pair 2
Primer Pair 3
Primer Pair 4
Primer Pair 5

Figure 45. An example of primer design for exon 4 and part of the 3'UTR of *CACNG4*.

Each of the primer pairs are labelled in the same colour. Introns and the 3'UTR are denoted by lower case letters, and the exon of *CACNG4* by the capital letters. Any base where common variation is reported is denoted through use of < > brackets. This is to ensure primers are not designed to fall on sequences containing common variation. All of the primer pairs overlap with each other ensuring full coverage of the genetic region of interest.

DNA which had not been extracted prior to HRM screening. Given the sample size of the primary HRM analyses we deemed it large enough to determine variation at the genomic level.

5.5 RESULTS

5.5.1 HRM SCREENING OF GENES

HRM analysis identified 22 variants in *CACNG4* which are listed in Table 24. Thirteen of these variants have been annotated previously on dbSNP, and nine novel variants were identified and verified using HRM analysis and Sanger sequencing. Allele counts from the Sanger sequencing are shown in Table 24 but these are not representative of allele frequencies in the UCL cohort due to the selection criteria for samples. *CACNG4* contains four exons which encode a 37kDa protein, 327 amino acids in length. HRM analysis identified two non-synonymous mutations located in the fourth exon of *CACNAG4.* In addition to these, HRM analysis found one variant in the 5'UTR, six synonymous mutations, four intronic variants and nine 3'UTR variants. Due to the nature of certain genomic regions the promoter region of *CACNG4* were not screened. These regions were extremely GC rich and contained several repeat runs of sequence. This has previously been described in 4.5.1.

5.5.2 INVESTIGATION OF VARIANT EFFECTS

The potential functional impact of SNVs in *CACNG4* was investigated using a number of different programmes depending on the type of variant found. SNVs found in coding regions of the genome were analysed using Polyphen2 and SIFT (section 3.2.5.1). There is a discrepancy between the scores predicted by both Polyphen2 and SIFT. Primarily Polyphen2 predicts that the two variants located in the fourth exon of *CACNG4* are possibly damaging changes. In contrast SIFT predict low damage scores for each of the variants. This result potentially stems from the method by which the programmes investigate variant differences. SIFT uses PSI-Blast to look at base conservation from a number of closely related sequences.

SNP ID	Chromosomal position (hg19)	Variant	1000G MAF	AA/Aa/aa	Class of Variant
rs34743848	chr17:65014360	C/T	0.0114	122/22/0	Synonymous variant
rs186352782	chr17:65014366	C/T	0.0002	143/1/0	Synonymous variant
17:65014414 (C/A)	chr17:65014414	C/A	-	129/15/0	Intronic variant
17:65021046 (C/T)*	chr17:65021046	C/T	0.002	127/2/0	Synonymous variant
rs139952030	chr17:65021094	C/T	0.006	128/1/0	Synonymous variant
17:65021126 (C/T)	chr17:65021126	C/T	-	125/4/0	Intronic variant
rs187070398	chr17:65021132	G/A	0.001	127/2/0	Intronic variant
17:65021173 (G/T)	chr17:65021173	G/T	-	128/1/0	Intronic variant
rs2240648	chr17:65026515	A/G	0.4726	200/2/0	Intronic variant
rs139092398	chr17:65026555	G/A	0.0088	200/2/0	Intronic variant
chr17:65026648 (G/A)**	chr17:65026648	G/A	-	102/1/0	Non-Synonymous variant Arginine to Glutamine
17:65026851 (C/T)	chr17:65026851	C/T	-	101/1/0	Non-Synonymous variant Arginine to tryptophan
rs11649752	chr17:65026886	C/T	0.0597	80/23/0	Synonymous variant
rs138591032	chr17:65027293	G/A	0.0046	100/11/0	3'UTR variant
17:65027367 (G/T)	chr17:65027367	G/T	-	150/2/0	3'UTR variant
17:65027506 (C/T)	chr17:65027506	C/T	-	102/1/0	3'UTR variant
rs201238001	chr17:65027556	A/G	0.001	162/3/0	3'UTR variant
rs740555	chr17:65027740	G/C	0.3498	82/20/18	3'UTR variant
17:65028637 (A/G)	chr17:65028637	A/G	-	102/2/0	3'UTR variant
rs188250082	chr17:65028695	G/C	0.0002	152/6/0	3'UTR variant
rs1051774	chr17:65029052	C/T	0.1703	68/34/0	3'UTR variant
rs3803815	chr17:65029264	G/A	0.0727	134/14/0	3'UTR variant

Table 24. Variants found using HRM analysis in *CACNG4*.

MAF was obtained from the 1000G data set. Some variants currently have dbSNP numbers but dbSNP137 was used for SNV validation. SNPs which are present on dbSNP138 or above are as following *rs555971941, **rs371128228. Genotype counts come from the number of samples sequenced in each region following HRM analysis. These numbers are not representative of allele frequencies.

Name	Location	Amino acid	Polyphen2	SIFT
rs371128228	chr17:65026648	R171Q	0.97 Probably Damaging	0.36
17:65026851 (C/T)	chr17:65026851	R239W	1.00 Probably Damaging	0.07

Table 25. Polyphen2 and SIFT prediction effects for non-synonymous variants in *CACNG4.*

Polyphen2 on the other hand, in addition to calculating base conservation, measures the impact the amino acid change will have on both the secondary structure and tertiary protein structure.

These variants were then analysed using Project Hope (http://www.cmbi.ru.nl/hope/method) [461] to determine the effect of the amino acid changes on protein structure. All predicted amino acid changes are shown in Figure 46. SNV rs371128228 results in an arginine to glutamine change at position 171 in the amino acid chain. Arginine is a larger positively charged amino acid, while glutamine is smaller and is neutrally charged. The wildtype residue is not conserved at this position, however glutamine is not observed in other homologous sequences at this position. The loss of the positive charge and smaller size at position 171 may result in a loss of interactions with other molecules or residues, effecting protein folding or confirmation. SNV 17:65026851 (C/T) results in an arginine to tryptophan change at residue 239. The wildtype arginine residue is positively charged and smaller than tryptophan. Additionally tryptophan is hydrophobic in nature. Arginine is extremely conserved in this position of the sequence. While a few other amino acids do occur in homologous sequences to *CACNG4*, residues with properties similar to tryptophan are not observed. The loss of the positive charge and the change in hydrophobic nature could lead to a loss of hydrogen bonds and disturb correct folding of the resulting protein.

NextProt beta (http://www.nextprot.org) was utilised to determine the specific location of all two variants in the native protein, these are shown in Figure 47. The R171Q variant is located in the second extracellular region. R239W is located directly upstream of the Protein Kinase C (PKC) domain and the nPIST domain.

5.5.3 Genotyping of Variants detected by HRM

The two non-synonymous SNVs from the fourth exon of *CACNG4* were genotyped in the extended BD sample. Additionally, given the reasons for choosing *CACNG4* as a candidate gene to investigate, the variants were genotyped in both the SCZ and ADS cohorts. SNV rs371128228 trended towards significance in the BD population and was present in five individuals p=0.0617, one individual with SCZ p=0.14, and three individuals with ADS p=0.725.

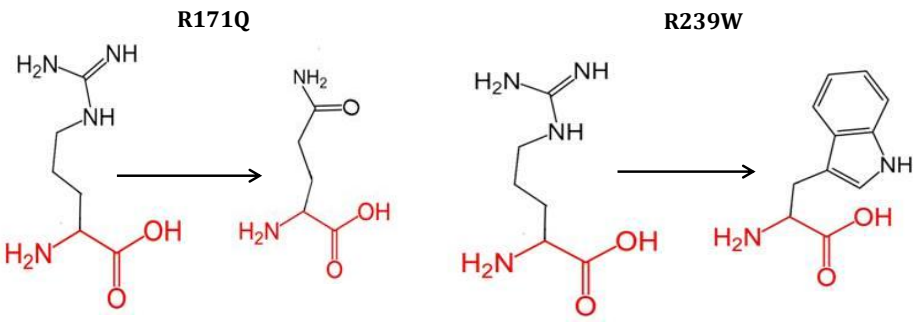


Figure 46. Amino acid changes in *CACNG4.*

R171Q results in an arginine to glutamine change and R239W results in an arginine to tryptophan change.

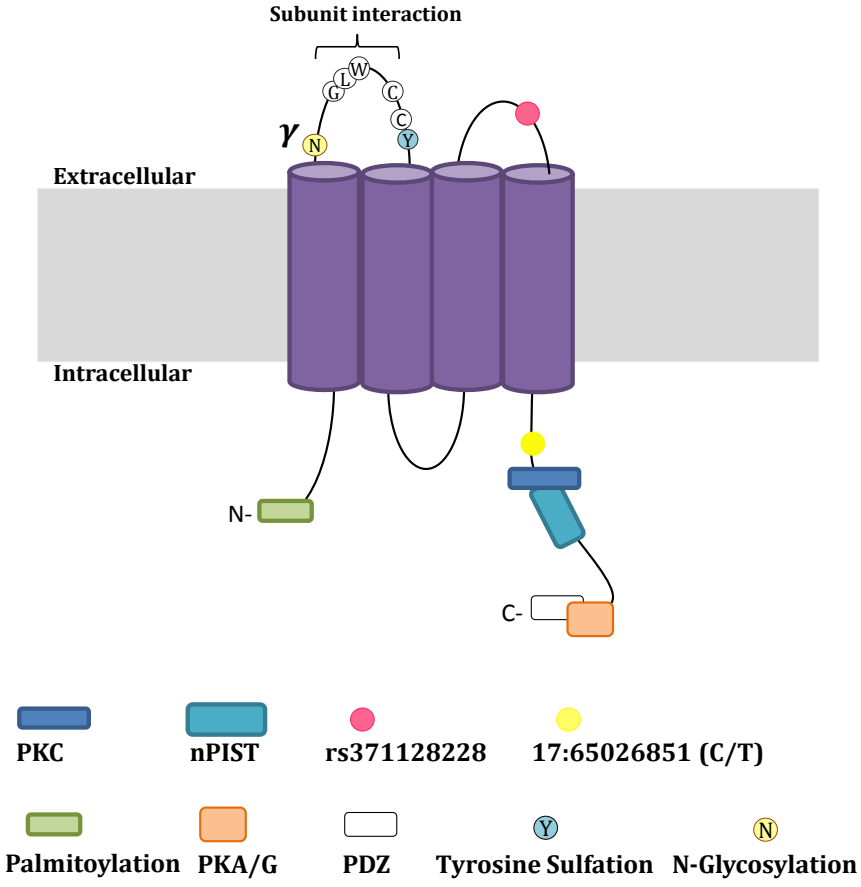


Figure 47. TARP γ4 structure.

The CACNG4 protein contains four transmembrane domains. The amino acid residues shown here are the ones which are highly conserved between mouse, rat and human (Image based on [462]). The locations of the non-synonymous SNVs found using HRM are noted by the red circle in the second extracellular domain and just upstream of the PKC domain.

The variant was completely absent from the control cohort. The allele frequency of rs371128228 does not reach above 0.001 in any of the disease cohorts and is absent in the UCL control cohort. SNV 17:65026851 (C/T), was present in two individuals with BD p=0.7845, one individual with ADS p=0.9635, and one healthy volunteer. No association was seen in the UCL cohort. The variant was not present in the SCZ case cohort.

Preliminary data from the UK10K SCZ cohort was additionally utilised for variant investigation and a Swedish schizophrenia cohort [411] comprising of 2,536 individuals was also investigated for variant frequencies. Further control samples datasets from the exome variant server (EVS) (http://evs.gs.washington.edu/EVS/) European population and the 1000G's (http://www.1000genomes.org/) Utah Residents with Northern and Western European Ancestry (CEU) population were investigated to ensure equal numbers of cases and controls were present (Table 28). The final publically available control database used was the Exome Aggregation Consortium (ExAC) (http://exac.broadinstitute.org) comprising 60,706 unrelated individuals. The European cohort in the ExAC consisted of 33,370 individuals [410]. This cohort contains data from Swedish SCZ and BD studies (12119); these include the 2453 individuals from the previously mentioned Swedish exome cohort. These 2453 individuals were removed from the overall control sample to ensure no overlap occurred as the genotype calls for each individual was previously known. In addition to the remaining 9576 Swedish SCZ and BD individuals, the ExAC cohort contains individuals from SCZ Trios from Taiwan (1505 individuals), SCZ Bulgarian Trios (461). The ExAC data does not currently have individual phenotypes available for each call and so it was impossible to remove these samples from the control cohort. As a result, the calculated p values are somewhat conservative in nature as we are unable to discern whether they are truly arising in the healthy control population or in the cohorts with psychiatric illness. The combined analysis for both *CACNG4* variants is listed in Table 29. These results are calculated to show the combined association between the UCL psychosis cohort versus the UCL healthy volunteers, and the combined cohorts including the calls from the datasets mentioned above. In the combined analysis, two individuals from the UK10K SCZ cohort were heterozygous for the rs371128228 variant. This variant

Gene	Variant	Location (hg19)	Position in gene	Change	Eur MAF	Cohort	Sample number	Genotype calls AA/Aa/aa	MAF	P	Fishers Exact	OR (95% CI)
CACNG4	rs371128228	17:65026648	Exon	G/A	0	BD	1885	1880/5/0	0.001	0.06	0.07	NA
						SCZ	604	603/1/0	0.001	0.19	0.38	NA
						ADS	1225	1221/3/0	0.001	0.11	0.16	NA
						Control	1315	1315/0/0	na			
CACNG4	17:65026851 (C/T)	17:65026851	Exon	C/T	-	BD	1876	1874/2/0	0.001	0.78	0.79	1.40 (0.13-15.4)
						SCZ	674	674/0/0	na	-		NA
						ADS	1227	1226/1/0	0.000	0.96	0.75	1.07 (0.7-17.1)
						Control	1309	1308/1/0	0.000	-		

Table 26. Results of SNVs genotyped from *CACNG*4.

MAF, minor allele frequency, P, Chi Square p value, OR, Odds Ratio, CI, Confidence Interval, NA, Not Applicable.

Variant	Sample	UCL		UK10K SCZ		Swedish SCZ		1000G Eur		EVS Eur		ExAC Eur	
		AA	Aa	AA	Aa	AA	Aa	AA	Aa	AA	Aa	AA	Aa
rs371128228	BD	1885	5	0	0	0	0	0	0	0	0	0	0
	SCZ	604	1	1388	2	2543	0	0	0	0	0	0	0
	ADS	1221	3	0	0	0	0	0	0	0	0	0	0
	Control	1315	0	0	0	0	0	349	0	4228	1	30775	12
17:65026851 (C/T)	BD	1876	2	0	0	0	0	0	0	0	0	0	0
	SCZ	674	0	1390	0	2543	0	0	0	0	0	0	0
	ADS	1227	1	0	0	0	0	0	0	0	0	0	0
	Control	1309	1	0	0	0	0	349	0	4229	0	30826	1

Table 27. Total variant counts for rs371128228 and 17:65026851 (C/T) in *CACNG4*.

UCL, University College London Cohort; UK10K SCZ, UK 10K Schizophrenia cohort; Swedish SCZ, Swedish exome cohort; 1000G Eur, 1000 Genome Project European cohort; EVS, Exome Variant Server; ExAC Eur, Exome Aggregation Consortium European cohort.

Variant		AA	Aa
rs371128228	Case	7643	11
	Control	36665	13
17:65026851 (C/T)	Case	7710	4
	Control	36713	2

Table 28. Total case control sample size for rs371128228 and 17:65026851 (C/T) *CACNG4* variants.

SNP	Cohort	Sample Calls AA/Aa/aa	MAF	P	Fishers Exact	OR (95% CI)
rs371128228	BD+SCZ +ADS	3710/9/0	0.001	0.074	0.061	NA
	UCL control	1315/0/0	NA			
	Combined Psychosis Cohort	7643/11/0	$6.9x10^{-4}$	$1.05x10^{-4}$	NA	4.39 (1.94-9.96)
	Combined Control Cohort	36665/13/0	$8.5x10^{-5}$			
17:65026851 (C/T)	BD+SCZ +ADS	3777/3/0	$7.93x10^{-4}$	0.973	0.7096	1.04 (0.11-10)
	UCL control	1309/1/0	$3.82x10^{-4}$			
	Combined Psychosis Cohort	7710/4/0	$2.59x10^{-4}$	0.001	0.0005	9.52 (1.74-51.99)
	Combined Control Cohort	36713/2/0	$2.72x10^{-5}$			

Table 29. Combined analysis of psychosis group versus Healthy Controls for *CACNAG4* variants.

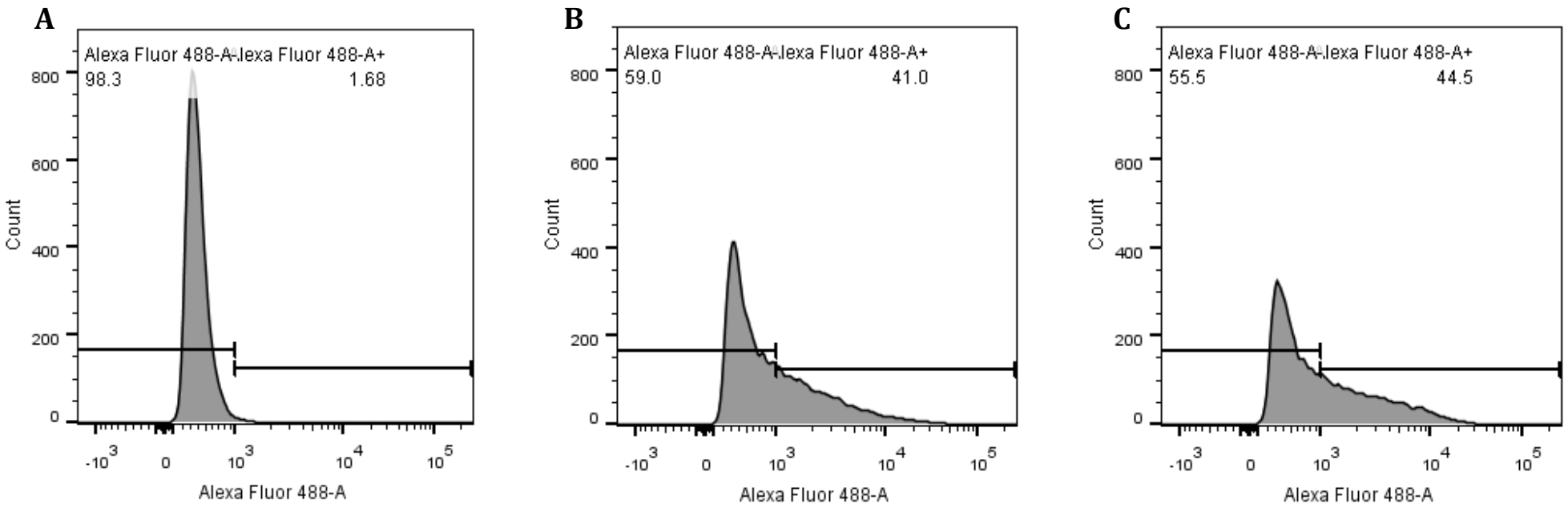
rs371128228 from *CACNG4* is significantly associated with the psychosis group including the samples from the UK10K SCZ samples. MAF, minor allele frequency, OR, Odds Ratio, CI, Confidence Interval, NA, not applicable.

was not present in the Swedish exome data set. Twelve healthy control individuals were heterozygous for this variant. Eleven of these individuals came from the ExAC cohort and one heterozygous individual was present in the EVS dataset. This variant was associated with the combined psychosis group (p=1.05x10^{-4}, OR=4.39). Variant 17:65026851 (C/T) was absent in the additional SCZ databases. It was present in the one European individual from the EXaC cohort of 30826 individuals. Variant 17:65026851 (C/T) is associated with the combined psychosis group (Fishers Exact p=0.0005, OR=9.52).

5.5.4 Effect of rs371128228 on AMPA-R trafficking

Flow cytometry analysis was conducted to determine whether variant rs371128228 has an effect on trafficking of either of the AMPA-Rs, GluR1 or GluR2, to the cell surface. Hek293 cells were co-transfected with wild-type *CACNG4* and *GluR1* or *GluR2*. Additionally, cells were co-transfected with the *CACNG4* pTAGRFP containing variant rs371128228 and *GluR1* or *GluR2*. Both of the AMPA-R plasmids contain super ecliptic pHluorin tag (SEP) which is a modified pH sensitive GFP tag which has been described in the methods section previously (0). The SEP tag is located on the N-terminus of the plasmid and is essentially non-fluorescent under acidic conditions (pH>6.0). When SEP is in a neutral environment, such as at the cell surface, the fluorescent signal increases 100-fold [423-425]. To determine whether rs371128228 has an effect on surface expression of GluR1 or GluR2, fluorescence levels from the Alexa-488 channel of the LSRFortessa flow cytometer, which detects SEP signal, were collected. This data was normalised against a negative control of untransfected cells to determine the fluorescence range of SEP positive cells. This threshold or gate was set separately for each independent replication. An *a priori* minimum of 10000 events (or cells) was collected for each sample. The mean of the fluorescence intensity in each sample was calculated for three replicated independent experiments.

Figure 48 shows the level of fluorescence for the negative control (A), the wild-type *CACNG4/GluR1* cells (B), and rs371128228 *CACNG4/GluR1* (C). A decrease can be seen in the level of fluorescence in the experiments with the mutant allele versus the wild-type. The rs371128228 mutation significantly decreases the level



Negative **Wild-type *CACNG4/GluR1*** **rs371128228 *CACNG4/GluR1***

Figure 48. Histogram of the Fluorescence of SEP for *CACNG4*/GluR1.

A. Untransfected negative controls were used to define the region where cells positive for SEP fluorescence would be detected. Cells above this range are the Alexa Fluor 488-A+ cohort and are located to the right of 10^3 on the x-axis of each histogram. B. Cells transfected with wild-type-*CACNG4* and *GluR1*. C. Cells transfected with rs371128228 *CACNG4* and *GluR1*.

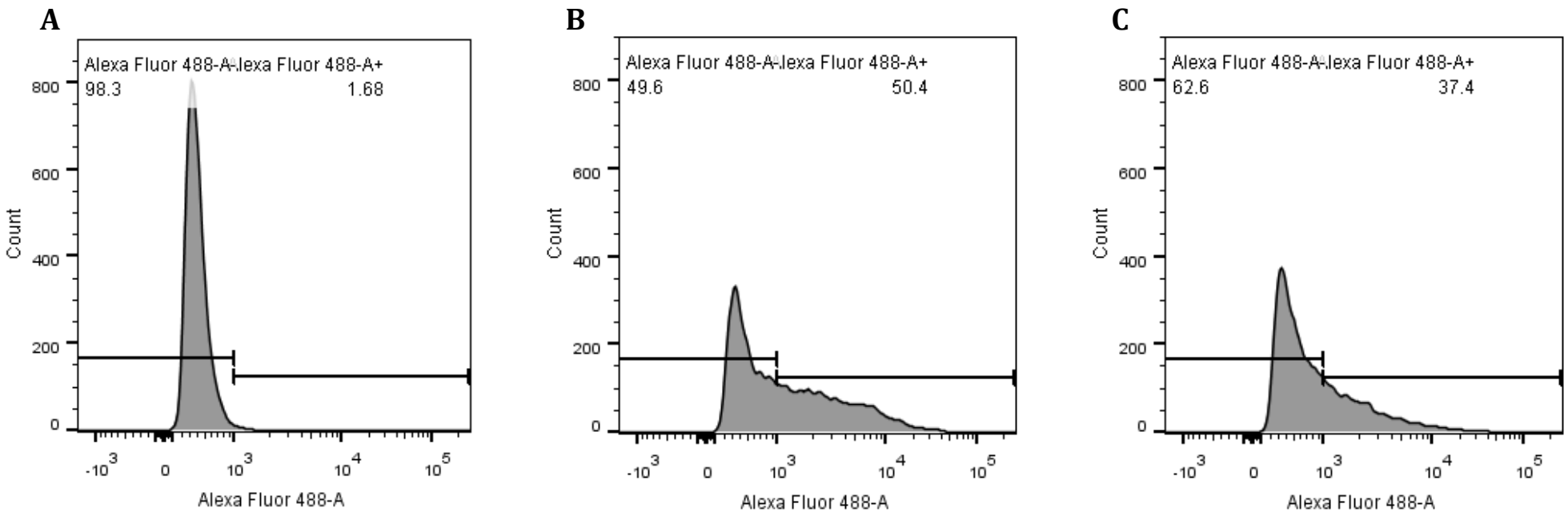


Figure 49. Histogram of the Fluorescence of SEP for *CACNG4/GluR2*.

A. Untransfected negative controls were used to define the region where cells positive for SEP fluorescence would be detected. Cells above this range are the Alexa Fluor 488-A+ cohort and are located to the right of 10^3 the x-axis of on each histogram. B. Cells transfected with wild-type-*CACNG4* and *GluR2*. C. Cells transfected with rs371128228 *CACNG4* and *GluR2*.

Sample	Mean (Alexa Fluor 488-A) N1	N2	N3	Mean	Standard Deviation	Standard Error	T-test (paired, two tailed distribution)
Wild-type *CACNG4/GluR1*	4590	4633	4988	4737.00	218.43	126.11	
Mutated *CACNG4/GluR1*	4286	4348	4228	4287.33	60.01	34.65	0.026
Wild-type *CACNG4/GluR2*	5170	4106	5519	4931.67	736.03	424.95	
Mutated *CACNG4/GluR2*	4141	*	5523	4832.00	977.22	564.20	0.903
Negative control	1656	1811	1985	1817.33	164.59	95.03	

Table 30. Mean Fluorescence from *CACNG4*/AMPA-R experiments.

N1, Replication 1; N2, Replication 2; N3, Replication 3. *No reading for this sample in this replication.

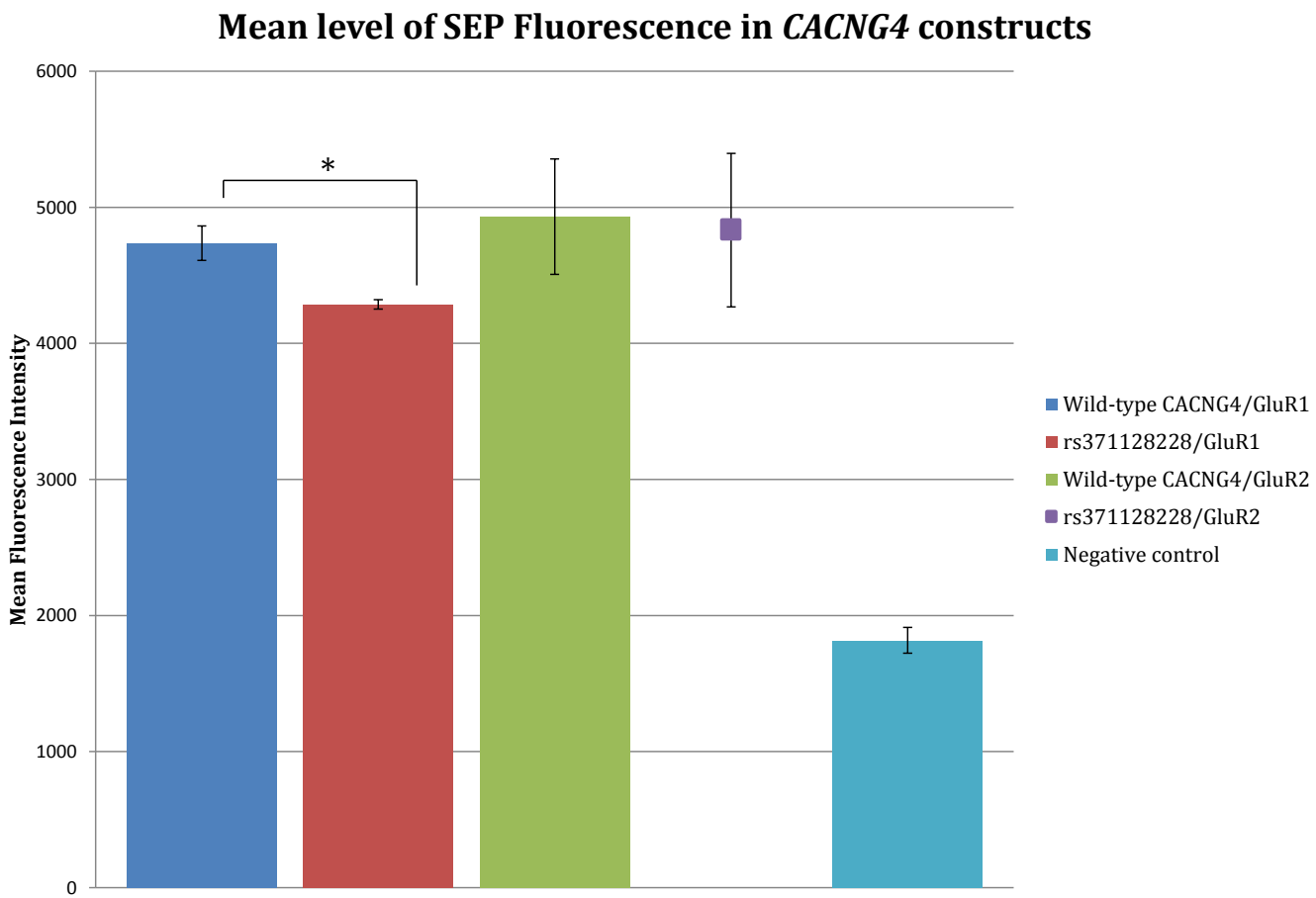


Figure 50. Fluorescence intensity of SEP in HEK293 cells.

The mean fluorescence for each transfection was calculated. Each bar graph represents three independent replicates (n=3). The value for the mutated rs371128228/GluR2 measurement represents the mean of data from two replicates (n=9). *p<0.05

of GluR1 at the cell surface (T-Test, p=0.026) (Table 30). Figure 49 shows the histograms of the same negative control, (A), the wild-type *CACNG4/GluR2* transfected cells (B), and rs371128228 *CACNG4/GluR2* transfected cells (C). The rs371128228 does not show an effect on GluR2 surface expression (T-test, p=0.903). Figure 50 shows the averaged means of fluorescence intensity for all three replications. The second replication of rs371128228 *CACNG4/GluR2* resulted in less than 10,000 events and is not included in the average.

5.6 DISCUSSION

The identification of variations which may confer an underlying susceptibility to the manifestation of psychiatric illness is crucial to understanding disease aetiology and may form the basis for pharmacotherapeutic intervention. The use of HRM analysis to investigate the presence of variation in the *CACNG4* gene proved successful in identifying nine novel variants which were previously unreported at the time of study. Neither rs371128228 nor 17:65026851 (C/T) showed association in either the UCL BD cohort alone, or the combined UCL BD+SCZ+ADS cohort group. Upon addition of the extra datasets both the *CACNG4* variants show an association with the combined psychosis group, (rs371128228, p=, OR=4.39, 17:65026851 (C/T), Fishers Exact p=0.0005, OR=9.52). In the psychosis cohort both variants are extremely rare (rs371128228, MAF =6.9×10^{-4}, 17:65026851 (C/T) MAF=2.59×10^{-4}), and less frequent in the control cohorts. Larger sample sizes are essential for investigating rare variants in disease aetiology. Previous work has supported the view that some genetic risk factors may be common to different psychiatric diagnoses and other illnesses comorbid with BD [39, 145, 463]. Contribution of a single variant to the relative risk of BD, SCZ or ADS is important to delineate, aiding the construction of therapeutic targets designed to target molecular aberrations which may result in response to variation in the *CACNG4* gene. This variant does not explain the signal arising between *CACNG4* and *CACNG5* as previously reported by Curtis *et al.* as there is no LD between the fourth exon of *CACNG4* and the associated variant (Figure 51).

Our criteria for SNV follow up excluded a number of potential variants which can have functional implications. HRM analysis detected the highest number of

variants in the 3'UTR region of *CACNAG4*. Seven exonic variants were reported in *CACNG4*, of these five were synonymous mutations and two were non-synonymous. Six variants were located in the intronic region of the gene, and nine variants were located in the 3'UTR. There are a number of reasons for this apparent burden of variants in the 3'UTR. Firstly, the 3'UTR is not under as much selective pressure as the exonic gene regions and secondly, the 3'UTR is longer than most of the other regions covered by HRM in this gene. 3'UTR variants can affect the stability of mRNA and protein production [464]. mRNA abnormalities have been recently reported in a study using induced pluripotent stem cells (iPSC) to derive neural lines from four related individuals, two brothers with BD and their unaffected parents [465]. These variants additionally provide good follow up candidates for further genotyping in the *CACNG4* gene in psychiatric illness.

Eight novel variants have been identified in in the protein coding region of *CACNG5*; six of these are case-only variants which are associated with both BD and SCZ [390]. Both of the variants in *CACNG4* are located in a region of the gene homologous to the region in *CACNG5* which also harboured case only recurrent variation [390].

Substance abuse is common in individuals with psychiatric disorders. ADS and BD are often co-morbid. Cassidy *et al.*, studying substance abuse histories of 392 patients hospitalised following manic or mixed episodes of BD, report comorbid rates of 14-60% for drug abuse, with a higher rate of substance abuse observed in males with BD versus females. The rate of alcohol abuse was 59.7% in males and 37.8% in females with BD in this study [49]. The direction of this co-morbidity is still uncertain. Do individuals drink because they are depressed, or does their depression lead to drinking? Shared underlying genetic risk factors such as these *CACNG4* variants may explain a part of the co-morbidity seen observed in both BD and ADS cohorts. ADS can be classified into two different types. Alcohol Dependence syndrome type-1 (ADS-I), which has an equal occurrence in both males and females, arises from a genetic predisposition combined with external environmental impacts. Alcohol Dependence syndrome type-2 occurs mainly in the sons of fathers with ADS and is influenced weakly by environmental factors [466]. There are a number of differences between the two types in both severity and

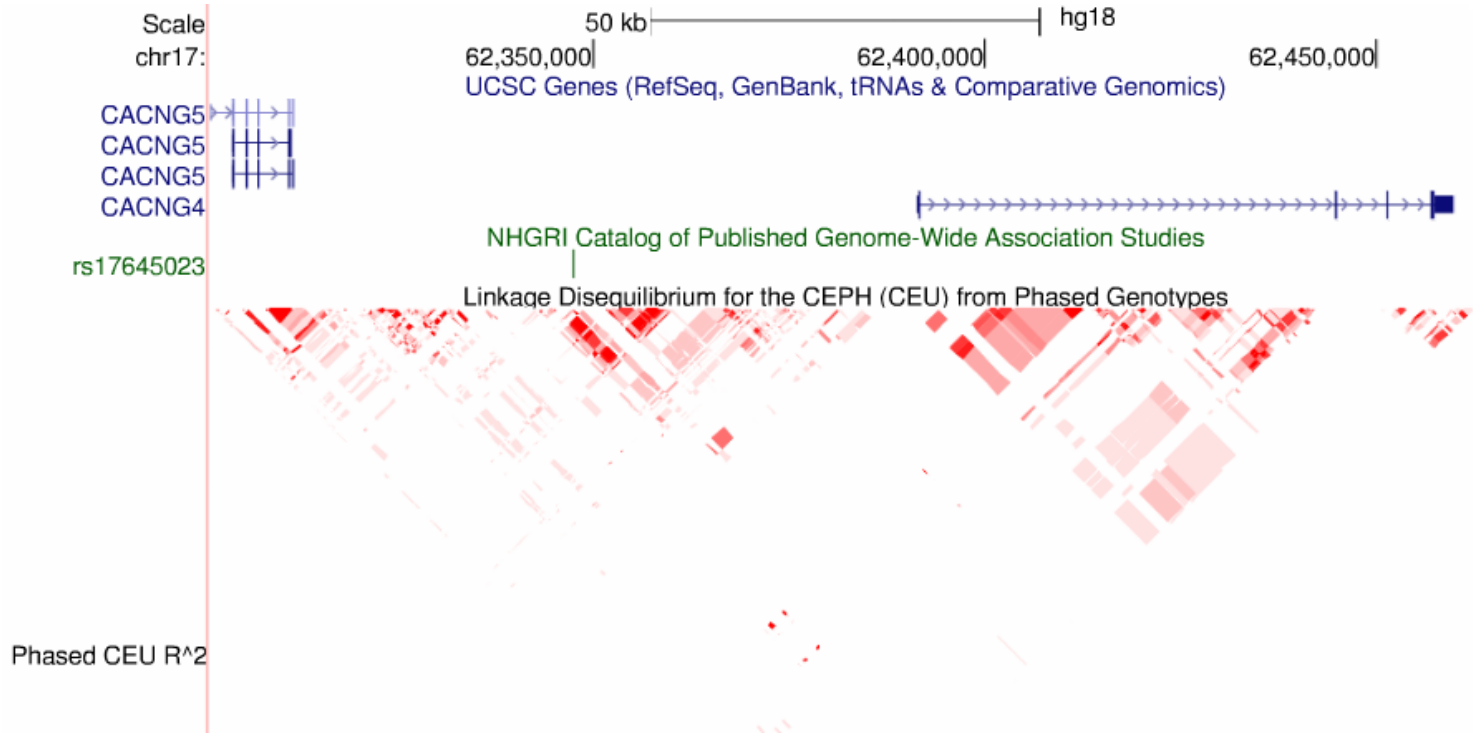


Figure 51. Linkage disequilibrium between *CACNG5* and *CACNG4*.

HapMap project linkage disequilibrium analysis (r^2) between *CACNG5* and *CACNG4* showing the position of the case-case associated variant (rs17645023). Image taken from UCSC genome browser hg18

control of drinking. ADS-I individuals are reported to commence drinking later in life compared to those with ADS-II.

However, their dependence on alcohol increases rapidly throughout their lifetime, while the severity of alcohol dependence recorded in ADS-II individuals remains constant. Characteristic personality traits of individuals with ADS-I have shown they have a higher tendency for harm avoidance and low novelty seeking, and their drinking is often linked to anxiety relief. Characteristic traits of individuals with ADS-II individuals report higher novelty seeking and self-reported drinking to induce euphoria [466]. While these characteristic are not mutually exclusive and both ADS-I and ADS-II are not discrete diseases or separate entities, instead they mark opposite ends of the ADS spectrum. These characteristics are similar to those reported in individuals with BD, and potentially arise due to common variation present in common neurobiological pathways.

AMPA receptors are the primary means through which the CNS mediates fast post-synaptic signalling in the brain. Members of the TARP family of proteins play an integral role in regulating AMPA-R trafficking and anchoring to the PSD. Variations in TARPs have previously shown large detrimental phenotypic effects upon knock out in mice. Slight variation in these genes could lead to a resulting decrease in the level of AMPA-R expression on the PSD. Case-only recurrent variations in genes encoding for TARPs, which play an integral role in synaptic transmission, suggest a role for altered trafficking of AMPA-R to the post-synaptic density in disease susceptibility. *CACNG4* has been implicated in controlling the timing of synaptic transmission, in response to glutamate. Alterations to the transport of these receptors could affect CNS plasticity resulting in abnormal synaptic connectivity.

Flow cytometry analysis was used to determine whether the *CACNG4* variant, rs371128228, effects cell surface of the AMPA-Rs GluR1 and GluR2. Wild-type *CACNG4* appears to traffic both GluR1 and GluR2 at a similar rate to the cell surface in Hek293 cells as no significant difference was seen in the fluorescence intensities recorded. The rs371128228 variant decreases the level of GluR1 surface expression, but not GluR2. The exact same concentration of DNA was used to transfect the cells in each of the

experimental replicates. The overexpression of the receptors compared to the native TARP/AMPA concentration in the cell would suggest that the results seen would derive primarily from the transfected plasmids rather than the native protein. FACs analysis did not allow us to compare the intracellular concentration of GluR1/GluR2 to the concentration of the AMPA-Rs on the cell surface. The use of the SEP tag was our only indicator in this experiment as to the level of the level of GluR1 or GluR2 receptors at the cell surface. Use of confocal microscopy and the fluorescently labelled *CACNG4*-pTAG-RFP and GluR1/GluR2-SEP would help to further quantify the number of CACNG4/GluR1 or CACNG4/GluR2 complexes at the cell surface.

AMPA receptor trafficking plays an integral role in the regulation of synaptic plasticity in the CNS. Both lithium and sodium valproate have been shown to exert effect on signalling cascade pathways with are known to regulate AMPA-R trafficking [467, 468]. Previous studies have shown that both lithium and sodium valproate attenuate the levels of GluR1 on the membrane surface [469]. Expression arrays have reported a significant increase of *GRIA1*, which encodes GluR1, by lithium in a mouse model previously (T-test p=0.0212, increase in expression=1.18) [470]. This suggests that other proteins involved in the trafficking of AMPA-R, such as TARP-γ4, could be affected by commonly used BD treatments.

Normalised molecular abundances profiles for AMPAR-R and TARPs predict that TARP γ4 plays a role in inner pore complexes of AMPA-R in the thalamus Figure 52 [471]. Research by Schwenk *et al.* provides further evidence for the presence of different AMPA-R/TARP complexes in different brain regions. In the thalamus, GluR2 is the most abundant AMPA-R receptor. GluR2 is the calcium impermeable AMPA-R and is found at highest concentrations in all brain regions studied aside from the cerebellum and brain stem. During development *CACNG4* is widely expressed in neurons throughout neuroaxis [359] and in the adult thalamus, cerebellum and olfactory bulb. Studies focusing on the characteristics suggest that TARP-γ4/GluR1 channels may contribute to calcium entry and consequent synaptic plasticity [472]. Evidence has shown that co-expression of TARPγ4/GluR1 and TARPγ2/GluR1 significantly increases calcium and monovalent cation permeability's compared to GluR1 on its own [454].

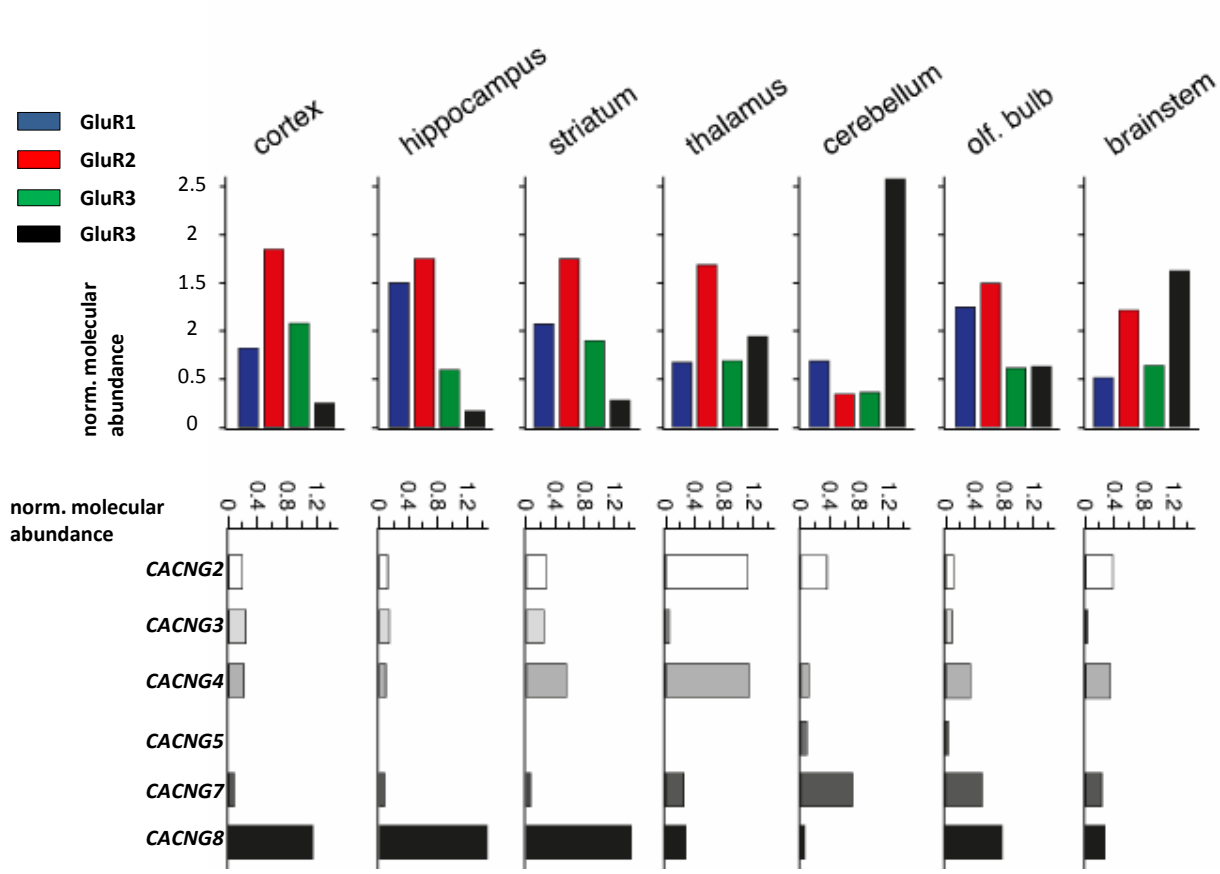


Figure 52. Normalised molecular abundance of GluR's and CACNG proteins for the inner core architecture of GluR's in different brain regions

CACNG4 is found to contribute to the inner core of GluR complexes in the thalamus alongside *CACNG3.* Image taken from [471]

The second mutation in *CACNG4* is located near to the PDZ binding domain in the γ4 protein. The atypical TARP proteins, γ2, γ3, γ4 and γ8, contain identical PDZ- binding motifs at the intracellular C-terminus [473]. This region is has been proposed as the essential binding domain for the targeting of the AMPA-Rs to the synapse [474]. Strikingly, given the identical nature of the domains in all type-I TARPs, deletions in the C termini at the PDZ domain of both γ2 and γ8 have differences in their impact on AMPA-R trafficking. Compared to wild-type trafficking, the frequency and expression of AMPAR-mediated miniature excitatory postsynaptic currents are reduced in the presence of the γ2 with C-terminus deletion. The γ8 C-terminus deletion result in a reduced frequency of AMPAR-mediated miniature excitatory postsynaptic currents but the amplitude of the current remains the same as that of the wild-type γ8 protein.

One must be cautious of proposing putative function which might arise based on evidence from homologous proteins. It is evident that while domains may be conserved, there are still other influencing factors which need to be considered. Evidence for this can be seen in the example of γ3 which does not bind the Arc protein in higher vertebrates. Until the protein structure, and additional functional work is conducted into the role γ4 may have on the trafficking of AMPA-Rs to the PSD one can merely postulate as to how variants in the C-terminus of the protein may impact on AMPA-R scaffolding, or expression at the PSD.

HRM analysis identified two rare non-synonymous variants located in functional regions in the *CACNG4* gene. These variants are associated with a combined mentally ill cohort involving BD and SCZ and a cohort of ADS individuals. The *CACNG4* gene encodes for a transmembrane AMPA receptor regulator. AMPA receptors mediated the majority of fast synaptic currents in the central nervous system and are important for long term potentiation mechanisms and synaptic plasticity. Variants in the regulators of these receptors could have potential knock on effects, reducing the amount of AMPA-R on the post-synaptic density which has been shown in this thesis and have a downstream effect on decreasing or increasing synaptic transmissions.

Chapter 6 Investigation of Variants in Calcium Channel Genes using Next Generation Sequencing Data

6.1 Exonic Regions

6.1.1 Introduction

The mechanisms through which variants in calcium channel genes increase an individual's susceptibility to BD have yet to be elucidated. Many of the signals from BD GWAS studies implicate both individual members of the L-type voltage gated calcium channel gene family and the calcium signalling pathway in pathway analysis [232, 239]. To date there has been a lack of functional variation associated with BD. Variants located in exonic regions which results in changes to the amino acid code may contribute to disease risk by resulting in aberrant protein formation and folding.

Advances in next generation sequencing technology are enabling researchers to conduct high throughput genomic analysis in a shorter time scale and in a more cost effective manner than was previously possible. Whole genome sequencing (WGS) allows for in-depth screening of genetic regions that may not be covered by exome capture techniques. Additionally, variant detection in large genetic regions using techniques such as Sanger sequencing or HRM analysis proves costly and time consuming to investigate.

6.1.2 HYPOTHESIS

Rare variation in protein coding regions of calcium channel genes may disrupt calcium influx into cells. These variants may disrupt the calcium signalling pathway or have an effect on membrane depolarisation in neurons.

6.1.3 AIM

WGS data from 99 individuals with BD selected from the extended UCL BD cohort was obtained. This data was used to screen all members of the L-type family of voltage dependent calcium channel genes for functional variants in exonic regions which may be associated with BD. Variant frequencies were classed using a minus log P (MLP) score of >0.7 generated from a comparison of allele frequency counts from the the 99 BD individuals against the frequency of the variant in the 1000G European cohort. The predicted effects of the variants were investigated using a bioinformatics approach and changes that were classified as deleterious were genotyped in the UCL BD and control cohorts.

6.1.4 RESULTS

Twenty six functional variants were identified from the WGS data from 99 BD individuals in six calcium channel genes, *CACNA1D, CACNA1S, CACNA2D4, CACNB1,* and *CACNB2* (Table 31 and Table 32). These genes have previously been implicated in some manner with BD (see section 1.3.2). Pathway analysis enrichment from the cross disorder PGC meta-analysis, investigating shared genetic variation between MDD, BD, SCZ, ADHD and ASD, implicated *CACNA1D, CACNA1E, CACNA1S, CACNA2D2, CACNA2D4,* and *CACNB2* [268]. *CACNB2* passed the threshold for genome-wide significant in the primary analysis of the five disorders. Additionally, rs11013860 is associated with BD, p=5.15x10^{-5}, in a Han Chinese population [244]. *CACNA1D* was implicated in the pathway analysis in the PGC BD cohort [270] and signal is observed in this gene, in the PGC BD GWAS. A rare deletion in a *CACNA2D*4 has been found in two unrelated late onset bipolar I patients and one control. The break points of this deletion caused the

removal of cache domain in the CACNA2D4 protein which is involved in small molecule recognition in a wide range of proteins (see section 1.3.2.3) [385].

For a variant to qualify for follow up genotyping in the BD cohort it needed to have a minus log P value (MLP) of >0.7 for the Fisher's exact test between 99 BD subjects and the 1000G control cohort [287]. An MLP that only used the European cohort from the 1000G dataset was also calculated. Polyphen2 was utilised to determine the effect the variants may have on protein function. In order for the variants to be considered for further analysis they were required to have a Polyphen2 score of greater than 0.9, which was indicative that they were likely to be damaging. Variant frequencies in the exome variant server (EVS), 1000G dataset and the UK10K dataset gathered for each reported variant. Variant frequencies from the PGC BD GWAS study [270] were reported for each variant if available. Of these variants, only one, rs3850625 in *CACNA1S,* had been previously reported in the Hap Map data on the regional association plots.

Seven variants met the criteria set out above of having an MLP > 0.7 and were denoted as probably damaging by Polyphen2. These are listed in Table 33. One variant, located in *CACNA1D*, rs115066564, did not have an MLP >1, but was predicted as having a PolyPhen2 score of 1, the most detrimental prediction, and so was included in the follow up genotyping assay. Six of these variants had been previously reported by dbSNP, and two, located in *CACNA1D* and *CACNA2D4* respectively, were novel. All eight variants were analysed using Project Hope (http://www.cmbi.ru.nl/hope/method) [461] to determine the effect the amino acid changes may have in the protein structure in response to the mutation.

The *CACNA1D,* variant rs370024486 was predicted to produce an arginine to tryptophan change at position 510. This change results in the loss of the wild-type residue charge. The mutated residue is larger than the wild-type residue which may lead to bumps in the protein folding. In addition to this tryptophan is more hydrophobic than arginine and this can result in the loss of hydrogen bonds and may disrupt correct protein folding.

A second *CACNA1D* variant rs115066564 was predicted to result in the change of an arginine to a histidine at position 930 (R930H). The variant is located in a repeat region of the *CACNA1D* protein sequence which is involved in multimer contact as denoted by the Proteins, Interfaces, Structures and Assemblies (PISA)-database. The histidine amino acid may have a reduced ability to make multimer contact due to its size.

The third variant from the *CACNA1D* gene is chr3:53837454 (G/A). This variant was predicted to result in the change of an aspartic acid to an asparagine at position 1834. Aspartic acid residues are charged, and this is changed to asparagine, a polar amino acid. Amino acids with polar charges may participate in hydrogen bonding and this could result in increased interactions with other molecules and residues.

The final variant to be genotyped in the UCL BD and control cohort was rs373663753. This variant was predicted to result in the change of an arginine to a tryptophan at position 2145. Both amino acid residues are polar in nature and therefore this substitution does not lead to a change in charge. Tryptophan is a larger residue and more hydrophobic than arginine, this can result in the loss of hydrogen bonds and disturb protein folding.

The *CACN2D4* variant, chr12:1963138 (G/T) results in an alanine to valine change at position 849 in the amino acid code. These two amino acids are very similar and do not result in a change to the charge of the amino acids. Valine is larger than alanine and could results in bumps in the protein structure.

Workable genotyping assays could only be developed for four of the eight variants chosen for follow up genotyping. These variants were chr12:1963138 (G/T) from *CACNA2D4* and chr3:53837454 (G/A), rs115066564 and rs373663753 from *CACNA1D.* Genotyping results are listed in Table 34. None of these variants were associated in the BD cohort, although three of the four genotyped variants were present in the case cohort only.

The ExAC browser was used to determine variant frequency for variants which failed to genotype in the BD sample. The *CACNA1D* variant rs370024486 was present in three Europeans out of 4789 which had sequencing data in this region (MAF=0.0003132). It

Gene	Name	Location (hg19)	Amino Acid Change	WGS calls AA/Aa/aa	MAF	1000G MAF	EUR AF	MLP	Eur MLP	UK10K AF	PGC freq	PolyPhen2
CACNA1D	chr3:53752749 (T/G)	chr3:53752749	S507A	98/1/0	0.005	0	0	1.08	0.7	0		Benign 0.219
	rs370024486	chr3:53752758	R510W	98/1/0	0.005	0.000	0	1.08	0.7	0		Probably Damaging 0.963
	rs115066564	chr3:53769508	R930H	98/1/0	0.005	0.001	0.003	0.63	0.31	0		Probably Damaging 1.0
	rs146892408	chr3:53783341	I1121V	98/1/0	0.005	0	0	1.08	0.7	0		Benign 0.001
	chr3:53837454 (G/A)	chr3:53837454	D1834N	98/1/0	0.005	0	0	1.08	0.7	0		Possibly Damaging 0.903
	rs141581705	chr3:53844126	T2018N	98/1/0	0.005	0	0	1.08	0.7	0.0011		Possibly Damaging 0.942
	rs373663753	chr3:53845320	R2145W	98/1/0	0.005	0	0	1.08	0.7	0		Probably Damaging 0.990
CACNA1S	rs149547196	chr1:201009066	P1820S	98/1/0	0.005	0.001	0.001	0.63	0.44	0.005		Benign 0
	rs13374149	chr1:201012484	R1639H	86/13/0	0.065	0.089	0.060	-0.61	0.11	0.043		Benign 0
	rs3850625	chr1:201016296	R1520C	82/16/1	0.090	0.071	0.120	0.49	-0.64	0.124	0.1019	Possibly Damaging 0.85
	rs145910245	chr1:201020165	T1335S	97/2/0	0.010	0.001	0.001	1.43	0.98	0.004		Benign 0.09
	rs147112322	chr1:201042735	T700M	98/1/0	0.005	0	0.000	1.08	0.7	0.000		Benign 0.001
	rs35534614	chr1:201058513	G258D	97/2/0	0.010	0.005	0.010	0.52	-0.2	0.016		Probably Damaging 0.992
	rs12406479	chr1:201079344	A69G	89/9/1	0.055	0.023	0.050	2.23	0.1	0.052		Possibly Damaging 0.874
	rs556751671	chr1:201081375	L31F	98/1/0	0.005	0	0	1.08	0.7	0.000		Probably Damaging 1.0

Table 31. Non-synonymous SNVs in CACNA1S and CACNA1D from whole genome sequencing of 99 BD samples.

Gene	Variant	Location (hg19)	Amino Acid Change	WGS calls AA/Aa/aa	MAF	1000G MAF	EUR AF	MLP	Eur MLP	UK10K AF	PGC freq	PolyPhen2
CACNA2D4	rs61741336	chr12:1906632	P117L	96/3/0	0.015	0	0	3.23	2.09	0		Benign 0.93
	rs55971855	chr12:1908849	F143S	98/1/0	0.005	0.004	0.010	0.23	-0.37	0		Possibly Damaging 0.932
	rs62621429	chr12:1910786	D223N	91/8/0	0.04	0.017	0.030	1.53	0.31	0.028		Probably Damaging 0.96
	12:1949941(G/T)	chr12:1949941	A300S	98/1/0	0.005	0	0	1.08	0.7	0		Benign 0.021
	12:1963138 (C/T)	chr12:1963138	A839V	98/1/0	0.005	0	0	1.08	0.7	0.001		Possibly Damaging 0.816
	rs10735005	chr12:1995403	I813V	1/22/76	0.88	0.927	0.850	-1.79	0.54	0		Benign 0
CACNAB1	rs199798694	chr17:37331645	G67V	97/2/0	0.01	0.002	0.01	0.96	-0.2	0.006		Possibly Damaging 0.838
CACNB2	10:18789795 (G/A)	chr10:18789795	G115R	98/1/0	0.005	0.000	0	1.08	0.7	0		Probably Damaging 1.0
	rs143326262	chr10:18828181	T448I	98/1/0	0.005	0.001	0	0.63	0.7	0.004		Benign 0.36
	rs61733968	chr10:18828486	R550G	92/7/0	0.035	0.007	0.02	2.28	0.78	0.019		Probably Damaging 0.958

Table 32. Non-synonymous SNVs in *CACNA2D4, CACNB1* and *CACNB2* from whole genome sequencing of 99 BD samples.

WGS, Whole Genome Sequencing; MAF, minor allele frequency; EUR, European; AF, Allele Frequency; MLP, Minus Log P.

Gene	Variant	Location (hg19)	Amino Acid Change	WGS calls AA/Aa/aa	MAF	1000G MAF	EUR AF	MLP	Eur MLP	UK10K AF	PGC freq	PolyPhen2
CACNA1D	rs370024486	chr3:53752758	R510W	98/1/0	0.005	0	0	1.08	0.7			Probably Damaging 0.963
	rs115066564	chr3:53769508	R931H	98/1/0	0.005	0.001	0.003	0.63	0.31			Probably Damaging 1.0
	3:53837454 (G/A)	chr3:53837454	D1834N	98/1/0	0.005	0	0	1.08	0.7			Possibly Damaging 0.903
	rs373663753	chr3:53845320	R2145W	98/1/0	0.005	0	0	1.08	0.7			Probably Damaging 0.99
CACNA1S	rs147112322	chr1:201042735	T700M	98/1/0	0.005	0	0	1.080	0.7	0		Probably Damaging 1.0
	rs556751671	chr1:201081375	L31F	98/1/0	0.005	0	0	1.080	0.7	0		Probably Damaging 1.0
CACNA2D4	12:1963138 (G/T)	chr12:1963138	A397V	98/1/0	0.005	0	0	1.080	0.7	0.001		Possibly Damaging 0.816
CACNB2	rs61733968	chr10:18828486	R550G	92/7/0	0.035	0.007	0.02	2.28	0.78	0.019		Probably Damaging 0.96

Table 33. Variants from WGS for follow up genotyping.

WGS, Whole Genome Sequencing, MAF, minor allele frequency, EUR, European, AF, Allele Frequency, MLP, Minus Log P.

Gene	Variant	Location (hg19)	Amino Acid Change	EUR AF	Cohort	Sample number	Genotype calls AA/Aa/aa	MAF	P	Fishers Exact	OR (95% CI)
CACNA2D4	12:1963138 (G/T)	chr12:1963138	A741V	0	BD	1900	1899/1/0	0.0003	0.80	0.65	0.703 (0.04-11.24)
					Control	1335	1334/1/0	0.0004			
CACNA1D	rs115066564	chr3:53769508	R931H	0.001	BD	1894	1891/3/0	0.0008	0.51	0.47	2.1 (0.22-20.2)
					Control	1321	1321/0/0	-			
	3:53837454 (G/A)	chr3:53837454	D1837N	0	BD	1895	1894/1/0	0.0003	0.40	1.00	-
					Control	1321	1321/0/0	-			
	rs373663753	chr3:53845320	R2145W	0	BD	1894	1893/1/0	0.0003	0.40	1.00	-
					Control	1322	1321/0/0	-			

Table 34. Genotyping results for calcium channel variants detected by WGS.

WGS, Whole Genome Sequencing, MAF, minor allele frequency, EUR, European, AF, Allele Frequency, MLP, Minus Log P.

was additionally present in one individual from the Finnish European sample of 40 individuals (MAF=0.012). The *CACNA1S* variant rs147112322 was present in 22 individuals in the ExAC data. Thirteen of these were present in the European (Non-Finnish) population from 33370 individuals (MAF=0.0001948), five were present in the Latino population of 5789 individuals (MAF=0.0004319), and the remaining four individuals were spread amongst the remaining populations. The second *CACNA1S* variant which failed to genotype in our sample was present on the ExAC database in eight individuals. Seven of these were in the European (Non-Finnish) population of 33360 samples (MAF=0.0001049) and one individual in the Latino population of 5788 individuals (MAF=$8.64x10^{-5}$). The *CACNB2* variant was by far the most common variant it was present in 968 individuals in the European (Non-Finnish) (MAF=0.01452) sample, 31, European (Finnish) (MAF=0.0047), 48 Latinos (MAF=0.0041), 33 Africans (MAF=0.0032) and 30 South Asians (MAF=0.0018).

6.1.5 Discussion

WGS identified twenty six variants in the L-type family of calcium channel genes in 99 individuals with BD. Fifteen of these variants reached the cut off rate of an MLP>0.7. Seven variants were classified as deleterious by Polyphen2. A final variant, present in *CACNA1D* (Eur MLP= 0.31), was predicted to have a deleterious score of 1 by Polyphen2 and was included in the analysis. The most common variants found by WGS in the calcium channel gene protein coding regions were deleterious variants with MLP scores >0.7 (34.6%), the next common were benign variants with MLP scores >0.7 (26.9%). Deleterious variants with MLP<0.7 comprised 23.1% of the total variants reported and benign variants were the least commonly reported variants (15.4%).

None of the variants showed an association with BD in the UCL sample. The *CACNA2D4* 12:1963138 (G/T) variant had a lower MAF in the genotyped cohort than the WGS data as it was present in one individual with BD and one control sample. *CACNA1D* variant 3:53837454 (G/A) was not present in any of the healthy volunteers and was present in one individual with BD. rs115066564 was present in

three individuals with BD and present in one healthy control. Finally, variant rs373663753 was present in one individual with BD and no control individuals.

There are a number of limitations which may influence the potential to infer association with rare variants in our sample. One of the main issues is the sample size here is underpowered for the detection of ultra-rare variants in the genome.

It is apparent from this study that WGS is capable of identifying variants with low allele frequencies. All three of the variants in *CACNA1D* are case only variants and it may suggest a role for a burden of rare variants in this gene which may in turn increase an individual's risk of BD. Four of the seven non-synonymous variants in the *CACNA1D* gene were predicted to be deleterious. Ament *et al.* [292] have reported the presence of 19 rare variants in the coding region of the *CACNA1D* gene [292]. This data came from WGS of 3,014 European individuals with BD and 1,717 healthy controls. There was not a significant burden of rare variants in the coding region here. *CACNA1D* is a large gene, encoding a protein 2,161 amino acids long (http://www.uniprot.org/uniprot/Q01668). In contrast, six SNVs were reported for the coding regions of *CACNA1C* a protein of similar length (2,221 amino acids) suggesting a higher prevalence of rare variation in the *CACNA1D* gene. No variants in *CACNA1D* have shown association with the UCL BD cohort in this study.

In an attempt to refine allelic frequencies of each genotyped variant the EXac database was utilised (http://exac.broadinstitute.org/). The EXac database was used to determine variant frequency in other populations, as a comparison to the variants frequency found in the UCL BD cohort. The *CACNA1D* variant, rs115066564, was present in three individuals with BD and one healthy control in the UCL sample. In the EXac database this variant is present in 19 individuals of 8255 (South Asian ancestry), 30 out of 33164 Europeans. This variant has a frequency of $4.5x10^{-4}$ in the European EXac cohort compared to $8x10^{-4}$ in the UCL BD cohort. The second variant 3:53837454 (G/A) has a frequency of $1.5x10^{-5}$ in the European EXac cohort compared to a frequency of $3x10^{-3}$ in the UCL BD cohort. Variant rs373663753 had a frequency of $1.5x10^{-5}$ in the European EXac cohort compared to $3x10^{-3}$ in the UCL BD cohort. The *CACNA2D4* variant was present in

two individuals in the European cohort, an allelic frequency $3.312x10^{-5}$, compared to $3x10^{-4}$ in the UCL BD cases and $4x10^{-4}$ in the UCL controls

Use of the ExAC database allows for the validation of variant calls in addition to comparison of variant frequencies. For example, while the variant rs115066564 has a higher frequency in the UCL case cohort compared to the European control population, it also has a frequency of $1x10^{-3}$ in the South Asian cohort. Variants which failed to genotype in the UCL cohort were screened in the ExAC database to ensure their validity. All variants chosen for genotyping were present on this database.

WGS provides a powerful tool for uncovering rare variation in the genome which has been previously missed by GWAS studies. Our model for variant investigation selected rare variants which were defined as deleterious in nature for a follow up in the UCL BD cohort. These variants did not show an association with BD. The use of the EXaC European cohort data have shown that three of the variants successfully genotyped here have higher frequencies in the BD cohort. This is in addition to the fact that there are cohorts of individuals with BD, SCZ and MDD included in the EXaC sample. The final variant from *CACNA2D4* has a higher frequency in the UCL case cohort, but it is also present in the UCL control cohort.

6.2 Intronic Variants

6.2.1 Introduction

CACNA1C has been repeatedly implicated through GWAS as having an association with BD. In *CACNA1C* the strongest association signal is seen in the third intron of the gene with the markers rs1006737 ($p=7.0\times10^{-8}$) [232, 233], rs4765913 ($p=1.52\times10^{-8}$) [232], rs4765914 ($p=1.52x10^{-8}$) [268] and rs1024582 ($p=1.7\times10^{-7}$) [270]. The third intron of *CACNA1C* is approximately 300kb in length and contains a chromosomal region with high levels of linkage disequilibrium (LD), strong mammalian conservation and multiple ENCODE marked regions (Figure 53). These ENCODE designated sites are of particular importance as they are putative regions of gene expression regulation. The marker with the strongest association is

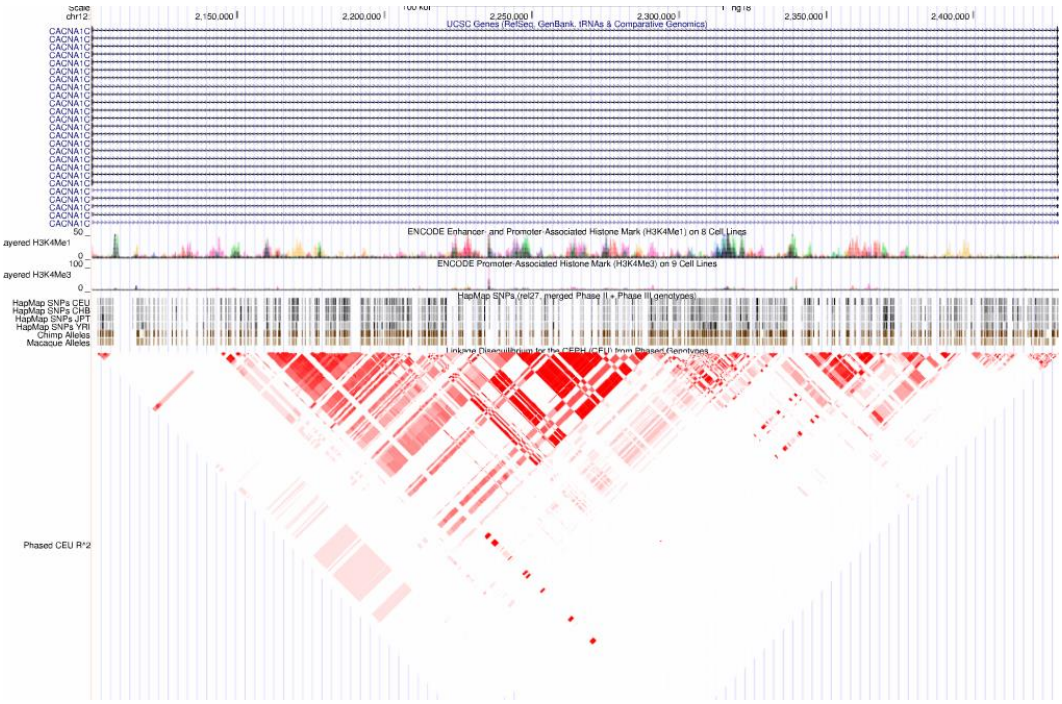


Figure 53. Linkage Disequilibrium in the third intron of *CACNA1C.*

In the third intron of CACNA1C there exists a large region of LD, in which the rs1006737 BD associate variant lies. In this region there are high levels of histone marks, which indicate this region may be involved in gene expression regulation. (Image: UCSC genome browser hg18)

rs1006737 ($p=7.0\times10^{-8}$) is located in this region in *CACNA1C*. The presence of this BD risk allele is associated with lower extraversion, trait anxiety, paranoid ideation and higher harm avoidance [372]. In addition to this the intronic BD risk variant rs1006737 may modulate amygdala function during emotional processing [373]. In the past, investigation of intronic variants has been hampered by the sheer size of genetic regions in which they are located. In addition to this, functional information pertaining to regulatory elements located in these areas was not widely available. This led to the arduous task of inferring functional effect for seemingly innocuous variants. The ENCODE consortium has provided researchers with a means to investigating specific effects of variants based on putative regulatory functions (see section 1.4.2) [475].

The National Human Genome Resources Institute (NHGRI) has a catalogue of all SNPs found by GWAS. Approximately 93% of SNP-phenotype associations lie in non-coding regions [399], making it difficult for researchers to imply biological

functioning. Novel SNPs found using the ENCODE methods show strong enrichment for GWAS disease associations. Maurano *et al.* have looked at the localisation of disease associated variants and their results show that 76.6% of these non-coding GWAS SNPs are present in DNAse I hypersensitivity sites (2931 SNPs) or are in complete LD (r^2=1) with SNPs in neighbouring DHS [399]. Individual analysis showed enrichment of GWAS variants in transcription factor binding sites modulated chromatin structure, as previously mentioned.

The data from the ENCODE project can be used to define variants in intronic regions that may have an impact on the regulation of gene expression or transcription.

6.2.2 AIMS

The primary aim of this study was to interrogate whole genome sequencing (WGS) data from 99 individuals with BD to discern variants that may be involved in BD. The third intron of *CACNA1C*, chr12:2,230,353-2,559,413, was screened using WGS data, for variants located in conserved putative regulatory regions as denoted by ENCODE [392], which may also contribute to the risk of an individual developing BD during their lifetime.

6.2.3 HYPOTHESIS

The association signal in the third intron of *CACNA1C* may be arising from variants changing regulatory regions which exert control over gene expression. The chromosomal region displaying strong mammalian conservation may be indicative of a region which is important for gene transcription control as intronic regions do not contribute to the protein code, and are not normally under strong conservation. The other introns in *CACNA1C* do not display the same amount of conservation or contain as many ENCODE marks as intron three. Conserved intronic regions may be indicators of regions in the chromosome where regulatory elements are located. Changes to these regions may not have as detrimental an effect as those seen in protein coding regions, but may be strong enough to ensure that conservation of the base occurs. Alterations to the regulation of gene

expression may result in the excesses or reduced production of the native protein. These characteristics, combined with the evidence from previous BD GWAS studies, make this chromosomal region one of interest in the search for variants which may contribute to the onset of BD in patients.

6.2.4 RESULTS

Next generation sequencing identified a total of 133 variants located in the third intron of *CACNA1C.* The variants in the third intron of *CACNA1C* were selected for genotyping in the expanded BD case control cohort if they met the following four criteria as described in 3.3.3. From the 133 variants, eight variants met the specified criteria and these are listed in Table 35. The *CACNA1C* intron 3 variants rs146482058, rs79398153, rs191953785, rs112312080, rs35469670, 12:2329069 (GG/TGG) and 12:2423175 (G/C) were present in the region of high LD between chr12:2,230,353-2,559,413 and each variant was in an ENCODE region [392]. All eight SNV regions are marked by H3K4me1, H3Kme3, and H3K27ac and active transcriptional enhancers with distinct chromatin signatures [401]. Enrichment for H3K4me1 and H3K27ac at a genetic level distinguishes active enhancers from inactive or poised enhancers [476, 477]. The presence of H3K4me1 and H3K27ac marked chromatin, with low levels of H3K4me3 and an absence of another histone marker, H3K27me3 represent putative human embryonic stem cell (hESC) enhancers and have been shown to localise proximally to genes that are expressed during development in hESCs and in epiblast cells [477]. Additionally, rs112312080 and 12:2423175 (G/C) were found to be present on (DHSs) as listed by the ENCODE project.

Of the SNVs found in the intronic region, rs79398153 was found to be associated with BD (Fishers exact test p=0.015). Eighty-eight heterozygote and three homozygote cases and fourty-four heterozygote controls were detected for this variant. The excess of homozygotes may represent a recessive effect, since only one homozygote would be expected under HWE, but this excess is not statistically significant [478]. None of the other intronic *CACNA1C* SNPs were associated with BD. Imputation was used to determine for rs79398153 in UCL1 cohort using GWAS data showed that it was still significantly associated with BD (P=0.022) [239].

rs79398153 is located in an ENCODE marked region for H3K4me1, H3K27ac and H3Kme3. GWAS data from the UCL cohort is only available for the first 506 BD-I samples recruited

Pairwise linkage disequilibrium analysis of *CACNA1C* intron three SNP rs79398153 with the top four BD GWAS associated variants, rs1006737 [232], rs4765913 [232], rs476590 [479], and rs4765914 [268], were performed (Table 36). This analysis also found no evidence to suggest that the predicted alleles and haplotypes associated with BD in the previous GWAS were in the same haplotypes as the rare BD associated variant (Table 36). The D' scores for each of the individual SNP comparisons seems to suggest high heritability between rs79398153 and each of the SNPs. However, the r^2 estimates for each of the SNV LD analysis range from 0.014-0.032. This can be attributed to the fact that the MAF of rs79398153 in our cohort was 0.032 and therefore was considerably less common than the GWAS significant SNPs. Thus the risk allele of rs79398153 is unlikely to be a proxy for the risk alleles of the GWAS SNPs.

Another intronic variant, rs116947827, located 105bp upstream of rs79398153 was in complete LD in the whole genome sequencing data from the 99 individuals with BD. This variant did not fit the criteria for genotyping in the BD sample as it was located in a repeat region and was not under conservation, having a negative GERP score of -6.69. The ENSEMBL (http://www.ensembl.org/index.html) database was used to calculate linkage disequilibrium scores for both rs79398153 and rs116947827 and the 1000G calls. MAFs in the 1000G data differ in the summary analysis of all populations, rs79398153 (MAF=0.02), and rs116947827 (MAF=0.01). Upon further investigation of the data however, the variants have the same MAF in all European populations. The T allele is not observed in the Finnish population however (Table 38). We postulated that the presence of these variants may have a detrimental effect on the regulation gene expression.

Variant	Location (hg19)	MLP	Eur MLP	GERP scores	Indel/ CNV	Repeat Region	1000G MAF	Encode Region
rs146482058	chr12:2403077	10.85	7.03	0.61	-	-	0	H3K4Me1, H3K4Me3 and H3K27Ac marks
rs79398153	chr12:2295156	1.57	2.16	0.12	-	-	0.01	High H3K4Me1 marks and H3K4Me3 and H3K27Ac marks, Human mRNA
12:2272007 (C/T)	chr12:2272007	3.23	2.09	0.19	-	-	0	H3K4Me1, H3K4Me3 and H3K27Ac marks
rs191953785	chr12:2425097	2.29	2.09	2.32	-	-	0	H3K4Me1, H3K4Me3 and H3K27Ac marks, DNAse hypersensitivity cluster marks
rs112312080	chr12:2354510	2.02	1.56	0.72	-	-	0.001	H3K4Me1, H3K4Me3 and H3K27Ac marks, Human mRNA
rs35469670	chr12:2292743	2.15	1.39	0.45	-	-	0	H3K4Me1, H3K4Me3 and H3K27Ac marks, Human mRNA
12: 2329069 (GG/TGG)	chr12:2329069	2.15	1.39	1.17	-	-	0	H3K4Me1, H3K4Me3 and H3K27Ac marks, Human mRNA
12: 2423175 (G/C)	chr12:2423175	2.15	1.39	1.98	-	-	0	H3K4Me1, H3K4Me3 and H3K27Ac marks, DNAse hypersensitivity cluster marks, Human mRNA

Table 35. *CACNA1C* intronic variants matching study criteria.

MLP, minus log P value; GERP, genomic rate profiling score; Indel, insertion/deletion; CNV, Copy Number Variation.

Gene	Marker 1	Marker 2	D'	r^2	LOD
CACNA1C	rs79398153	rs1006737	0.754	0.016	1.06
	rs79398153	rs1024582	0.729	0.014	0.87
	rs79398153	rs4765913	0.822	0.029	2.12
	rs79398153	rs4765914	0.829	0.032	2.31

Table 36. Pairwise linkage disequilibrium analysis between the BD associated SNP reported here and previous GWAS SNPs.

Gene	Variant	Location (hg19)	Amino Acid Change	Eur 1000G MAF	Cohort	Sample number	Genotype calls AA/Aa/aa	MAF	P	Fishers Exact	OR (95% CI)
CACNA1C	rs146482058	chr12:2403077	-/T	0	BD	1486	1322/134/3	0.048	0.642	0.306	0.94 (0.7-1.2)
					Control	1022	922/96/4	0.051			
	rs79398153	chr12:2295156	C/T	0.016	BD	1498	1380/88/3	0.032	0.025	0.015	1.511 (1.1-2.1)
					Control	1029	985/44/0/	0.021			
	12:2272007 (C/T)	chr12:2272007	C/T	0	BD	1503	1465/10/1	0.004	0.746	0.495	1.167 (0.5-2.9)
					Control	1004	998/5/1	0.003			
	rs.191953785	chr12:2425097	C/T	0.009	BD	1475	1450/25/0	0.008	0.637	0.339	0.868 (0.5-1.6)
					Control	1020	1005/20/0	0.010			
	rs112312080	chr12:2354510	C/T	0.0014	BD	1501	1439/35/0	0.012	0.474	0.298	1.223 (0.7-2.1)
					Control	1023	1008/20/0	0.010			
	rs3546970	chr12:2292743	-/C	0	BD	1497	1394/76/0	0.026	0.317	0.194	1.211 (0.8-1.8)
					Control	1026	982/44/0	0.021			
	12:2329069 (G/T)	chr12:2329069	G/T	0	BD	1498	1457/14/0	0.005	0.642	0.428	1.229 (0.5-2.9)
					Control	1033	1024/8/0	0.004			
	12:2423175 (-/T)	chr12:2423175	-/T	0	BD	1502	1471/4/0	0.001	0.339	0.336	2.785 (.3-24.9)
					Control	1026	1025/1/0	0.001			

Table 37. Tests of association with *CACNA1C* intronic variants with BD.

EUR, European, MAF, minor allele frequency, AF, Allele Frequency, MLP, Minus Log P, OR, Odds Ratio.

Online software PROMO (http://alggen.lsi.upc.es/cgi-bin/promo_v3/promo) was used to determine whether the presence of either rs79398153 or rs116947827 variants resulted in the gain or loss of any TFBS. This analysis was conducted on the region of the *CACNA1C* intron, chr12:2,295,005-chr12:2,295,733, which incorporates both variants. Each of the variants were analysed on their own and in the presence of the other variant. Neither of the variants causes the loss of a TFBS, but rather both SNVs result in the creation two new TFBS as shown in Figure 54. In the presence of rs116947827 alone, PROMO reports the creation of a Yin Yang 1 (YY1) TFBS, and when rs79398153 is present for the analysis a c-Ets-168 TFBS is present. For rs79398153 the same two TFBS, XBP-1 and YY1, are reported for both the variant alone and in conjunction with rs116947827. PROMO predicted that in addition to the creation of four new TFBS, 103 other TFBS were reported from this region when both SNVs were analysed. The additional TFBS are not located in the region of the variants, but span the entire length of the genomic region and are listed in Table 39. This suggests that the genetic region where these variants are located could function as enhancers of gene transcription. Enhancers function through the binding of transcription factors such as YY1.

The creation of the same YY1 TFBS approximately 100bp apart could influence the regulation of gene expression. The YY1 TF is a member of a zinc finger protein family, GLI-Kruppel, which is ubiquitously expressed throughout the human body YY1 TF play a role in both repression and activation of a wide variety of promoters. Tethering of YY1 to the histone deacetylase protein, HDAC3 and HDAC2, result in transcription repression through histone deactylase activity [441, 442].

Changes to the mRNA secondary structure can impact the bio-synthesis of proteins produced through effects on translation [480-482]. A number of different studies have identified mRNA sequences in brain tissue with homology to the third intron of *CACNA1C* region (chr12:2295005-chr12:2295733) and these are described in Table 40. These mRNA sequences have been found in the cerebellum and hippocampus from both foetal and adult brain tissue. Changes to the mRNA structure can be a major factor in the regulation of gene expression. mRNA structure prediction software RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) was used to investigate whether rs79398153 and rs116947827 induce spatial changes to the secondary structure of the mRNA strand.

rs79398153 (C/T)

Wildtype C

```
cagccaggg caCggatcgt ttcctggccg
  =====      GR-alpha [T00337]
     ==== === Pax-5 [T00070]
     ==== === p53 [T00671]
            ====== TFII-I [T00824]
```

Mutation T

```
cagccaggg caTggatcgt ttcctggccg
  =====      GR-alpha [T00337]
     ==== === Pax-5 [T00070]
     ==== === p53 [T00671]
     ==== == XBP-1 [T00902]
          ==== YY1 [T00915]
            ====== TFII-I [T00824]
```

rs116947827 (C/T)

Wildtype C

```
tgcctccac taCgggctgt ggtttttgg
====== AP-2alphaA [T00035]
 ===== GR-alpha [T00337]
           ======= Pax-5 [T00070]
           ======= p53 [T00671]
                    ======== NFI/CTF [T00094]
```

Mutation T

```
tgcctccac taTgggctgt ggtttttgg
====== AP-2alphaA [T00035]
 ===== GR-alpha [T00337]
          ==== YY1 [T00915]
            ======= Pax-5 [T00070]
            ======= p53 [T00671]
            ========= = c-Ets-16 8 [T00115]
                     ======== NFI/CTF [T00094]
```

Figure 54. PROMO transcription factor binding site predictions for CACNA1C intronic variants.

The TFBS highlighted in green indicate new TFBS are predicted to occur in the presence of the new altered alleles by PROMO software. In the presence of mutant T alleles, rs79398153 and rs11947827, four new transcription factor binding sites are created. Two of these are YY1 transcription factor binding sites

Population	rs79398153		rs116947827	
	Allele C	Allele T	Allele C	Allele T
All Populations	0.98	0.02	0.99	0.01
All African	1.00		1.00	
Americans with African ancestry	1.00		1.00	
Luhya	1.00		1.00	
Yoruba	1.00		1.00	
All American	0.95	0.05	0.99	0.01
Colombian	0.98	0.02	0.99	0.01
Mexian Ancestry	0.90	0.10	0.99	0.01
Puerto Ricans	0.96	0.04	0.99	0.01
All Asian	0.99	0.01	1.00	
Han Chinese	1.00	0.01	1.00	
Southern Han Chinese	0.99	0.02	1.00	
Japanese	0.99	0.01	1.00	
All European	0.99	0.01	0.99	0.01
Utah Residents northern and western European ancestry	0.99	0.01	0.99	0.01
Finnish	1.00		1.00	
British	0.98	0.02	0.98	0.02
Iberian	0.93	0.07	0.93	0.07
Tuscan	0.99	0.01	0.99	0.01

Table 38. Allele frequencies from 1000G dataset for rs79398153 and rs116947827.

The 1000G data set has information regarding SNP frequencies from a number of different populations. For the European populations both SNPs are in complete LD. The variant is not present in the African population cohort and has different allele frequencies in the Asian and American Populations.

The prediction of mRNA structure is conducted using minimum free energy (MFE) modelling [483-485]. This is based on the theory that molecule will fold into the lowest energy state. RNAfold uses both MFE predictions (Figure 55) and centroid folding predictions (Figure 56) [486]. A mountain plot representation of the MFE structure, the centroid structure and the thermodynamic structure ensemble of the RNA structures alongside the entropy for each base pair is provided for the *CACNA1C* intron 3 variants (see Figure 57). MFE structure predictions show very little difference between the wild-type mRNA structure (MFE=-288.40), rs79398153 (MFE=-288.30) and rs116947827 (MFE=-288.80). The presence of rs79398153 results in the prediction of stronger binding and an extra loop in the top part of the mRNA structure compared to the wild-type. The predicted mRNA structure of rs116947827 shows additional branching occurring in addition to the loss of some of the RNA loops. While the structural changes are quite pronounced, there was not a reflective change in the MFE value. When both variants are present, there was a slightly higher predicted MFE value (MFE=-291.40). In the double mutant mRNA structure, rs79398153 branching patterns take precedent over the branching pattern from rs116947827. This is a result of the predicted binding strength for rs7938153 which is stronger than that predicted in the rs116947827 structure.

Centroid structural predictions are RNA secondary structures which have minimal base pair distances to all secondary structures in the Boltzmann ensemble. Centroid structure predictions display greater structural changes, as shown in Figure 56, than the MFE structural prediction. Minimum free energy predictions are also provided by RNAfold for the wild-type (MFE=-221.80), rs79398153 (MFE=234.40), rs116947827 (MFE=-206.20) and double mutant (MFE=-187.70) centroid structures. Distinct branch changes can be seen between rs7938153 and the wild-type. This reflects the change which is also observed in the MFE mRNA structures. Unlike the MFE mRNA predictions, the presence of both variants in the centroid structure predictions shows an entirely new loop structure. The double mutant structure has a much lower MFE value than that seen in the wild-type, rs79398153 or rs116947827. Subsequently the mountain plots and positional entropy for each position show differences for the double mutant structures Figure 57. These changes to the mRNA structure predictions provide some evidence that the presence of these SNVs together may impact on gene expression.

TFBS	Number
AP-2	2
C/EBPalpha	1
C/EBPalpha	2
C/EBPbeta	1
CAC-binding protein	6
c-Ets-1	2
c-Ets-1 68	3
c-Jun	7
c-Jun	3
Crx	1
Egr-1	1
Egr-3	1
Elk-1	1
GR	2
HNF-1A	8
HNF-4alpha	1
HNF-4alpha1	3
HNF-4alpha2	3
HOXD10	1
HOXD10	1
HOXD8	4
HOXD8	4
HOXD9	1
HOXD9	1
IRF-2	2
IRF-2	1
LXR-alpha:RXR-alpha	3
LyF-1	3
MyoD	3
NF-1	5
NF-AT1	3
NF-Y	3
POU1F1a	12
POU1F1b	3
POU1F1c	3
SREBP-1c	1
TFII-I	1
Total	103

Table 39. TFBS predicted by PROMO when both rs79398153 and rs116947827 are present.

The presence of rs79398153 and rs116947827 results in the creation of 103 new putative TFBS throughout the genomic region used in *CACNA1C* experiments. These TFBS are in addition to the presence of the two YY1 TFBS which are created at the base position where the variants are altered.

Human mRNA	Tissue	Description	Study
AB209016	Brain	*Homo sapiens* mRNA for Voltage-dependent L-type calcium channel alpha-1C subunit variant protein	Totoki *et al.*, Database Submission [487]
BC146846	Brain, Cerebellum	*Homo sapiens* calcium channel, voltage-dependent, L type, alpha 1C subunit, mRNA (cDNA clone MGC:181949 IMAGE:9056774), complete cds.	Starusberg *et al.* [488]
AJ224873	Brain, Hippocampus	*Homo sapiens* (HLCC105) mRNA for voltage-dependent L-type calcium channel alpha-1C subunit (splice variant).	Soldatov [489]
Z34812	Brain, Hippocampus	*H. sapiens* (HLCC73) mRNA for voltage-dependent L-type Ca channel alpha 1 subunit (splice variant).	Soldatov [489]
Z34813	Brain, Hippocampus	*H. sapiens* (HLCC74) mRNA for voltage-dependent L-type Ca channel alpha 1 subunit (splice variant).	Soldatov [489]
Z34817	Brain, Hippocampus	*H. sapiens* (HLCC86) mRNA for voltage-dependent L-type Ca channel alpha 1 subunit (splice variant).	Soldatov [489]
Z74996	Brain, Hippocampus	*H. sapiens* mRNA for voltage-dependent L-type calcium channel, alpha 1C subunit (splice variant).	Soldatov [489]

Table 40. Human mRNA from brain regions from the UCSC genome browser from the third intron of *CACNA1C*.

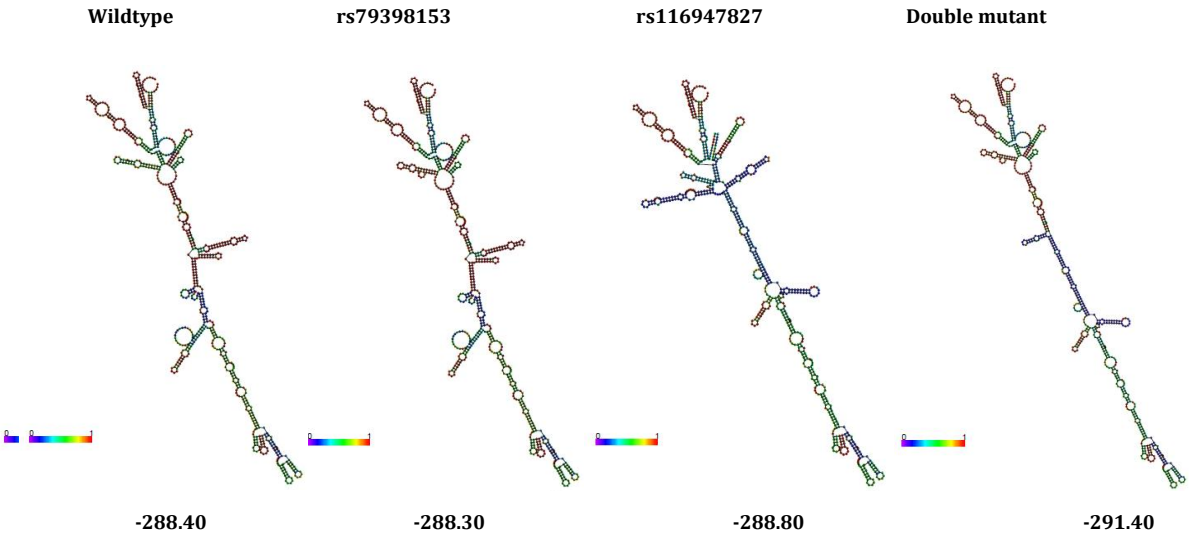


Figure 55. mRNA secondary structures for rs79398153 and rs116947827

This mRNA structure prediction is based on the minimum free energy (MFE) values which are list for each underneath each structure.

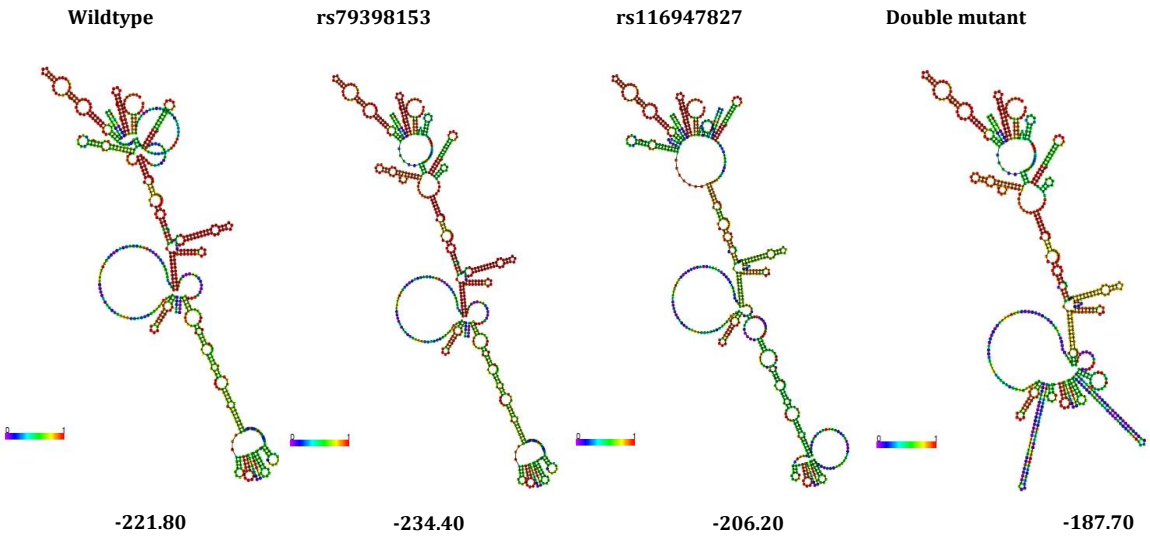


Figure 56. mRNA centroid structural predictions of variants rs79398153 and rs116947827.

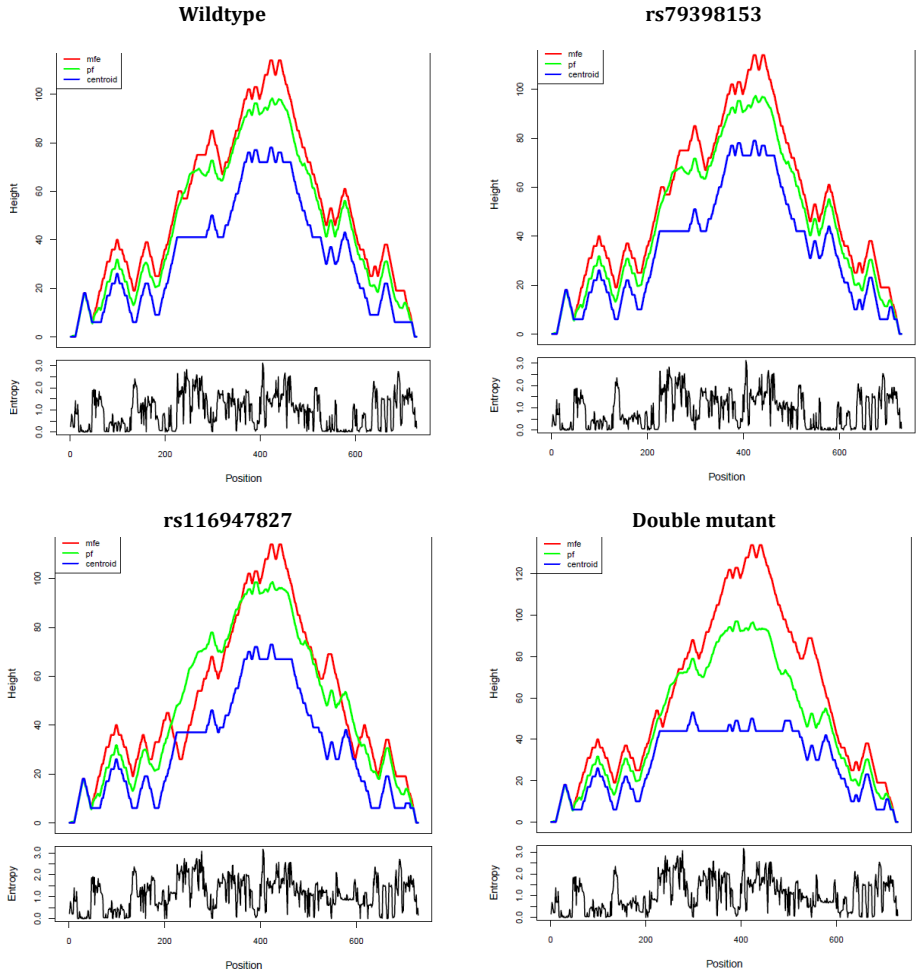


Figure 57. mRNA secondary structures for rs79398153 and rs116947827.

6.2.4.1 Effect of rs79398153 and rs116947827 on luciferase expression

The ability of rs79398153 and rs116947827 to modulate transcription regulation was tested using luciferase reporter gene constructs which are described in section 3.3.6 . Four different vectors containing the DNA sequence from the *CACNA1C* intron 3 (chr12:2295005-chr12:2295733) were created using the pgl3-basic luciferase plasmid. The first plasmid construct contains the wild-type genetic sequence from this region. The second contained the mutated T allele of rs79398153. The third contains the mutated T allele of rs116947827 and a fourth plasmid, contains both mutated alleles from rs79398153 and rs116947827. The inserted DNA sequence was added 35bp upstream of the start codon of the luciferase gene. These clones tested the ability of the genetic sequence to act as a enhancer of luciferase reporter gene activity against the basal level of luciferase expression seen when no additional genetic sequence was inserted. Plasmids were transfected into HEK293 cells and luciferase readings were taken from lysed cells 48 hours following transfection. A control plasmid containing no insert was also used to measure the basal level of luciferase expression seen in each transfection.

No significant difference was observed in the level of luciferase expression with the pgl3-basic vector compared to expression levels in the wild-type and the rare T allele in rs79398153 and rs116947827 constructs. Additionally, no significant difference was observed between the wild-type sequence and the presence of the rare T alleles of rs79398153, and rs116947827 on their own (T-test, rs79398153 p=0.339, rs116947827 p=0.362). A significant difference in luciferase expression was observed when both the rare T alleles of rs79398153 and rs116947827 were present in the pgl3-basic clone. There was a significant change to the level of luciferase expression in the presence of both allelic variants (double mutant p=0.002; see Figure 58).

Clone	Ratio of Renilla to Firefly luciferase expression	Standard Deviation	Standard Error	T-test (paired, two tailed distribution)
pGL3-basic no insert	0.00051	0.00030	0.00010	0.476
Wild-type CACNA1C	0.00057	0.00039	0.00013	NA
rs79398153 T allele	0.00042	0.00023	0.00008	0.339
rs116947827 T allele	0.00040	0.00034	0.00011	0.362
Double Mutant	0.00006	0.00004	0.00001	0.002

Table 41. Student T-test for changes in luciferase expression with *CACNA1C* intronic variants compared to the wild type control.

No significant difference is observed when either of the allelic variants are present on their own. There is a significant change to the level of luciferase expression in the presence of both allelic variants.

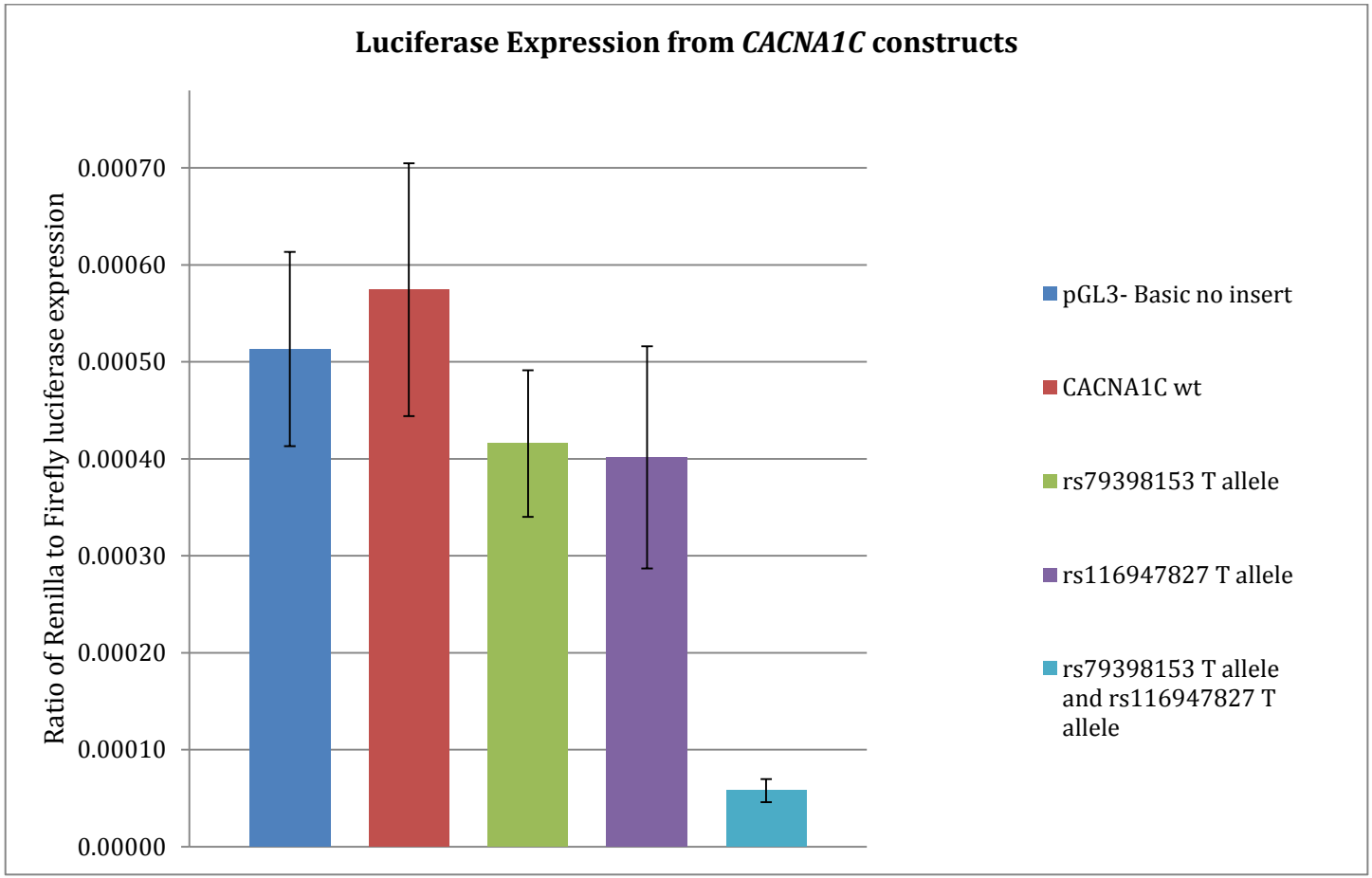


Figure 58. Relative luciferase activity reporter gene assays for the experimental constructs in HEK293 cells.

The Y-axis shows the mean of the renilla to firefly luciferase expression for all replications within an experiment and between experiments. Data expressed as mean SEM of three independent experiments performed in triplicate (n=9) **P< 0.005.

6.2.4.2 DNA-PROTEIN COMPLEX FORMATION BY RS79398153 AND RS116947827 VARIANTS

PROMO analysis predicted the formation of two novel YY1 transcription factor binding sites 105bp from each other in the third intron of *CACNA1C.* Luciferase assay analysis showed that in the presence of both variants there was a 10-fold reduction in luciferase expression. The mRNA structural analysis provides evidence that the presence of both variants leads to changes which may influence gene expression through the creation of novel TFBS. Electromobility shift assay of the allelic variants rs79398153 (Figure 59) and rs116947827 (Figure 60) were conducted using nuclear cell lysate from HEK293 cells to determine whether the variants had a higher binding affinity or increased binding specificity to nuclear proteins (see section 3.3.8).

For variant rs79398135 there was a single band in the presence of the wild type labelled probe and nuclear lysate (Lane 2). The band from the wild-type labelled probe disappeared in the presence of a 200x excess of unlabelled C probe. This was indicative of specific binding. In the presence of the mutated T labelled probe three bands were visible on the gel image (lane 5). The wild-type C unlabelled excess out competes the top band seen from the labelled probe at a 100x fold excess. In comparison the unlabelled T probe does not show as high an affinity for the proteins in the nuclear lysate and did not outcompete the band formed by the T-labelled probe.

The wild-type C allele of rs116947827 results in the formation of two bands, while the mutated T allele results in the formation of a stronger single band. This suggests that the mutated probe binds with greater affinity to one of more nuclear proteins. However, the 200x unlabelled T competitor probe does not result in the outcompeting of the labelled T probe. This may result from one of two things; either the binding of the labelled probe is unspecific and binds to indiscriminate proteins in the nuclear cell lysate which the unlabelled competitor probe may not target, or, the binding of the labelled competitor probe is too strong to be outcompeted at a concentration of 200x excess of the unlabelled competitor probe. This issue could be resolved by incrementally increasing the level of competitor probe from 200x excess up to 1000x excess.

This band was complete abolished in the presence of a 10x-200x excess of the mutated unlabelled rs116947827 T allele probe (Lane 6-10). In contrast the rs116947827 C allele unlabelled probe out competes the band formed by the labelled rs116947827 T allele probe at a 200x excess. This suggests the T allele of rs116947827 has a higher affinity for proteins in the nuclear cell lysate. Similar results were obtained in four independent experiments for each of the allelic variants.

Online software PROMO (http://alggen.lsi.upc.es/cgi-bin/promo_v3/promo) predicted the formation of two YY1 transcription factor binding sites in the region of these variants. A super-shift assay was conducted which included the use of an antibody raised against YY1 as described in 3.3.9. These experiments were used to determine whether one or more of the EMSA bands was the result of probes binding to YY1. Binding of the antibody to the proposed transcription factor should result in a super shift of the observed labelled band because of the increased molecular weight of the antibody/transcription factor/probe complex (see Figure 61). For rs79398153 the wild-type C allele appears to shift in the presence of the YY1 antibody (Lane 3). No band shift is apparent in the presence of the mutated allele (Lane 6). The wild-type C allele of rs116947827 shows a faint band shift in the presence of YY1 (lane 9); however the mutated T allele shows a stronger band formation and shift in the presence of YY1 (lane 12). An additional control of lane containing the YY1 antibody and probe for both wild-type and mutant alleles of rs79398153 and rs116947827 in the absence of nuclear cell lysate should have been added to ensure the results of these experiments were not as a result of artefact. Consistent results for the antibody binding were seen across four replicates of the each EMSA.

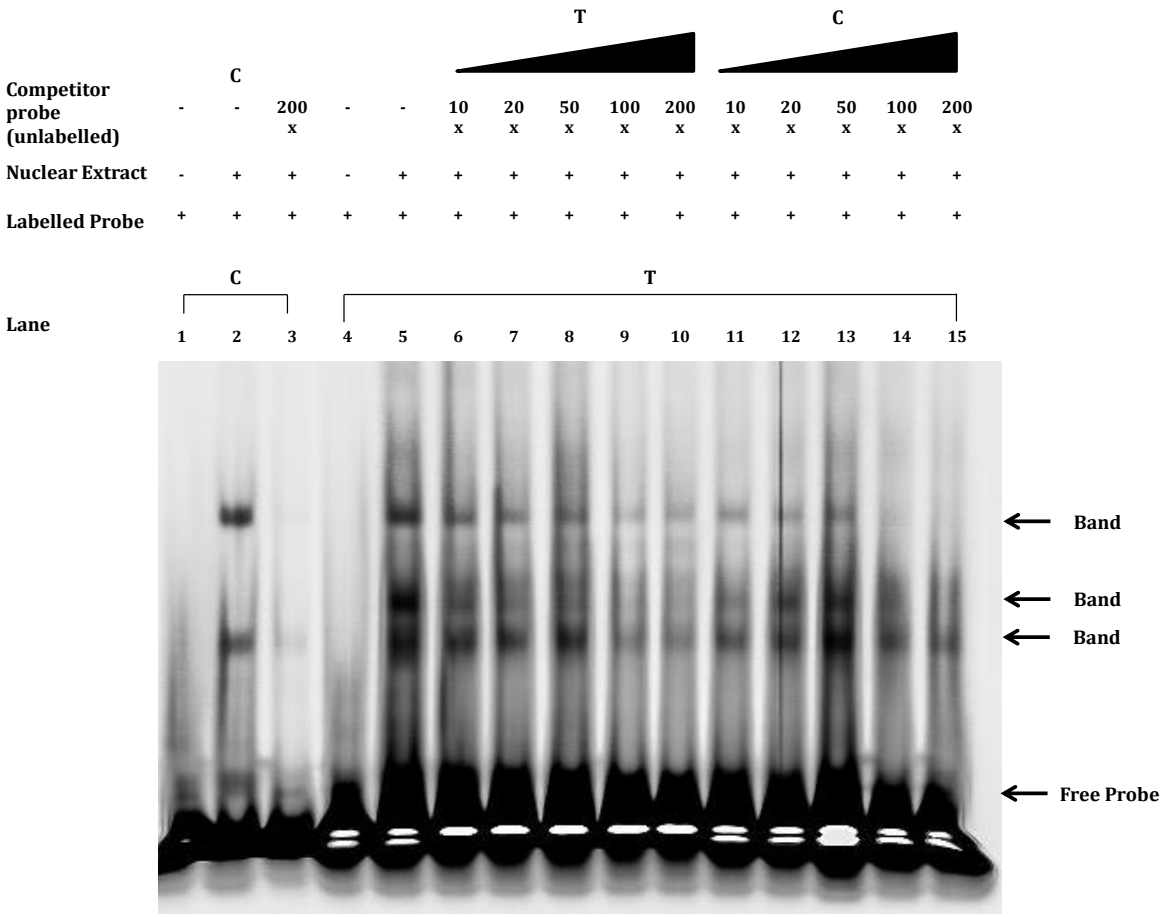


Figure 59. Electomobility shift assay for rs79398153.

EMSAs performed with C (1-3) or T (4-15) labelled probes and HEK293 nuclear cell lysate. The wild-type labelled probe (C) which labels two bands (Lane 2). The 200x excess of unlabelled C probe out competes the top band in the presence of the 200x unlabelled C probe (Lane 3). In contrast the T labelled probe binds three bands of proteins (Lane 5). The top band formed in the presence of the T labelled probe is not competed by the excess of T unlabelled probe suggesting unspecific binding or the need for a larger excess of unlabelled T probe. The top band formed by the T labelled probe is outcompeted by a 100x-200x excess of C-unlabelled probe (Lane 14-15)

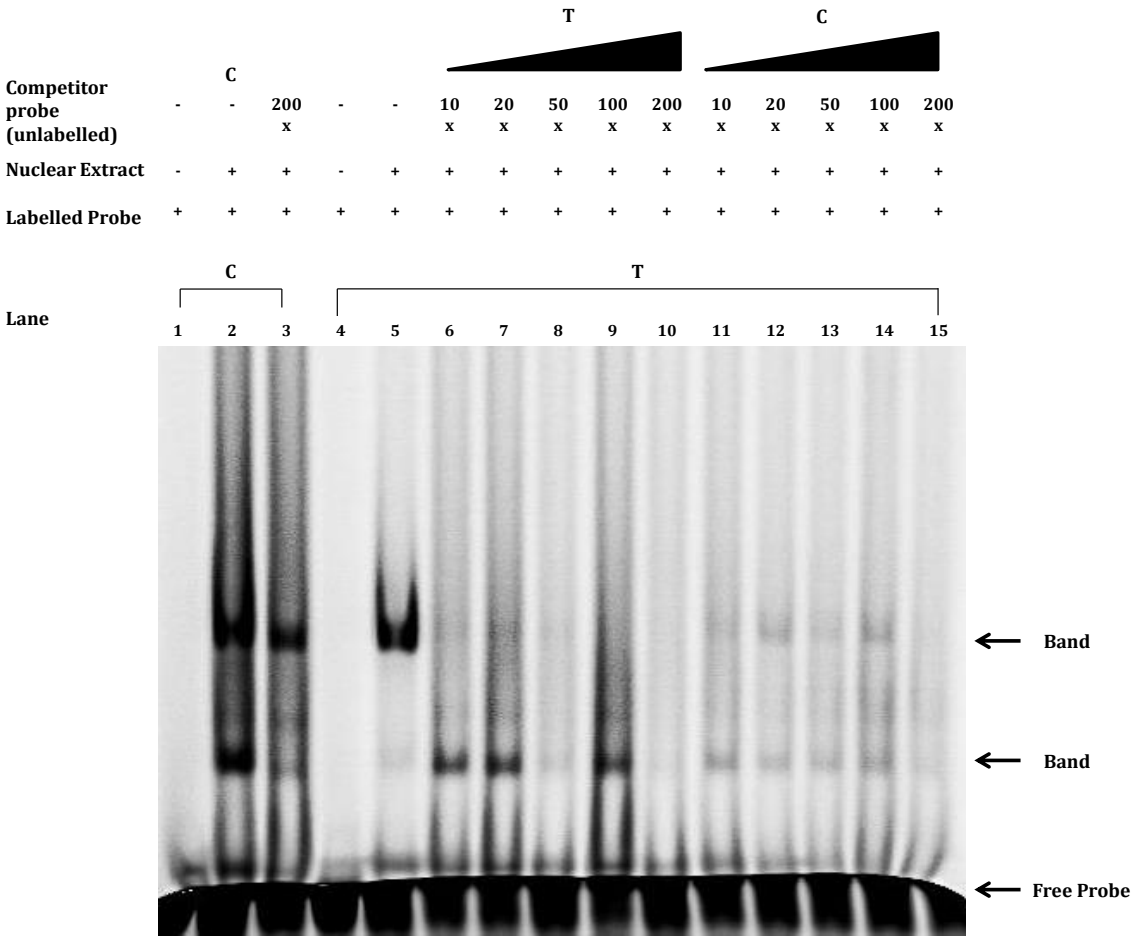


Figure 60. Electromobility shift assay for rs116947827.

EMSAs performed with C (1-3) or T (4-15) labelled probes and HEK293 nuclear cell lysate. The labelled T probe binds a single band of protein, as compared to the C labelled probe and nuclear cell lysate. The T labelled probe is out competed by the 10x unlabelled probe (lanes 6-10). The band formed in the presence of the T labelled probe is not competed out as effectively in the presence of the C-unlabelled probe (lane11-15). In addition to this, the C-labelled band seen in lane 2 is not outcompeted by the unlabelled 200x fold excess, suggesting that this binding is not specific.

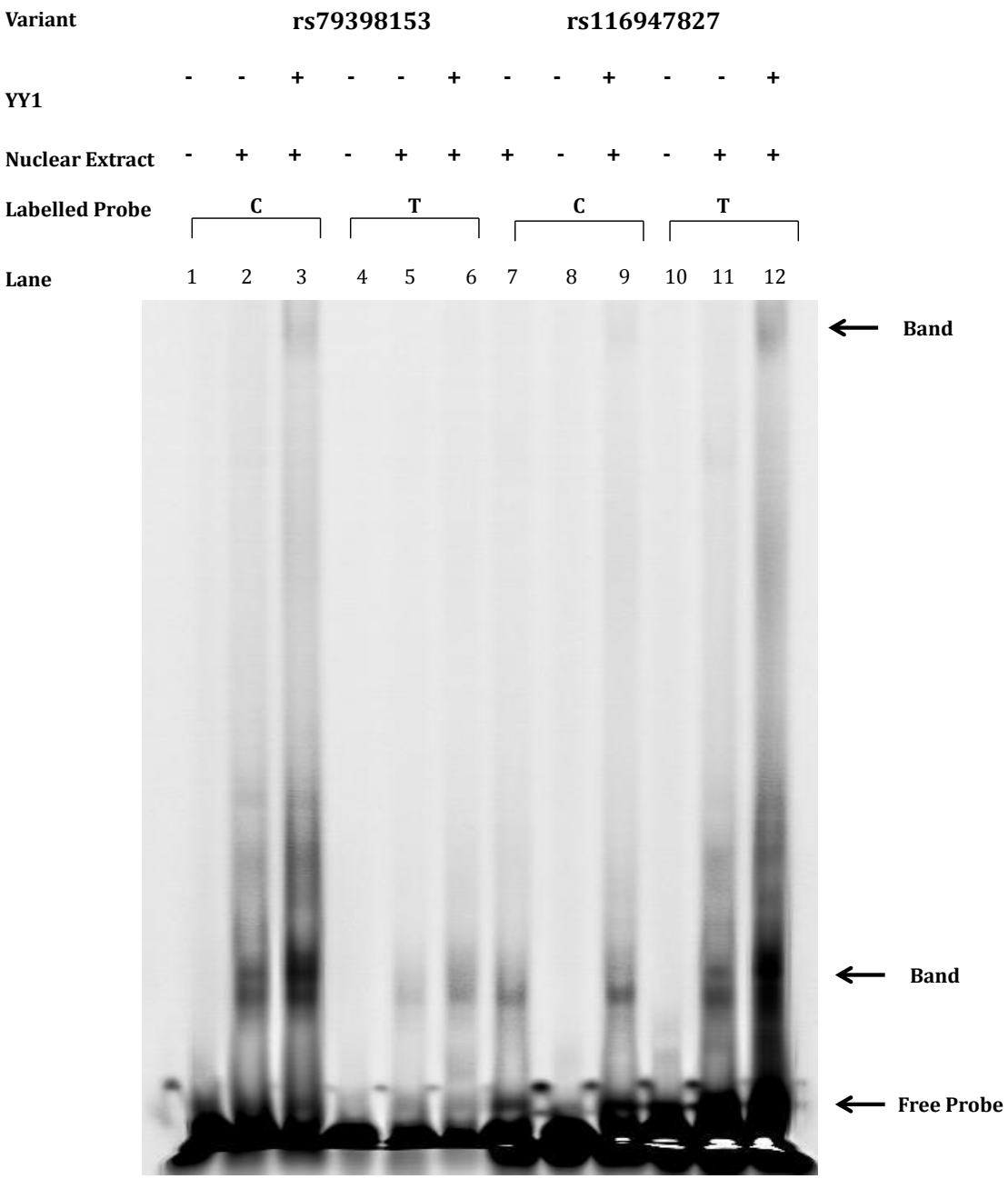


Figure 61. Super shift EMSA assay of *CACNA1C* variants with a YY1 specfic antibody.

In the presence of a YY1 specific antibody, a shift in band size is observed in lane 3, lane 9 and lane 12 suggesting that the anti-YY1 antibody is binding to proteins in the nuclear cell lysate that the labelled probe is binding to.

6.2.5 DISCUSSION

Whole genome sequencing provides the ability to investigate, in depth, a wide range of variation present in an individual. We utilised whole genome data from 99 individuals to investigate whether variants, located in the third region of *CACNA1C* increase an individual's susceptibility to bipolar disorder. The presence of allelic changes in regulatory regions in the intron can have a downstream effect on gene expression and translation. The ENCODE data was utilised to define putative functional elements in this genetic region.

Eight variants were selected for genotyping in the UCL cohort of BD samples and healthy control volunteers. Of the eight variants, one variant, rs79398153, was found to be associated with BD. Upon further investigation of the WGS data a variant 105bp upstream of rs79398153, rs116947827, was found to be in complete LD with rs79398153. *In silico* analysis using PROMO software indicates that these two variants individually created two YY1 TFBS.

The ENCODE track on the UCSC genome browser shows rs79398153 as being located in a conserved region with high H3K4Me1 marks, H3K4Me3 and H3K27Ac marks. The presence of H3K4Me3 is often found close to active promoters. Enrichment for H3K4me1 and H3K27ac at a genetic level distinguishes active enhancers from inactive or poised enhancers [476, 477]. The presence of a combination of H3K4me1 and H3K27ac marked chromatin, with low levels of H3K4me3 and an absence of another histone marker, H3K27me3 is the hallmark of putative human embryonic stem cell (hESC) enhancers and these have been shown to localise proximally to genes that are expressed during development in human embryonic stem cells and in epiblast cells [477]. The presence of these marks indicates that this region may be involved in the regulation of gene expression.

Gene expression reporter assays performed in HEK293 cells showed a significant decrease in the level of gene expression in the presence of both variants. It can be postulated that given the number of new TFBS created when both variants are present that transcription factors may have a higher affinity for binding this region and this in turn impact the regulation of *CACNA1C* gene expression. In addition to this, the presence of these variants are also purported to change mRNA structure.

HEK293 cells were co transfected with the pRL-SV40 and pgl3-basic vector, containing the region of interest, at a ratio of 1:2. *Renilla* expression for this assay is between 2000-5000 times greater than that of the firefly luciferase. As described in section 3.3.7 the pRL-SV40 vector contains the SV40 early enhancer and promoter element to ensure constitutive expression when co transfected in cells with any of the pGL3 vectors. The pGL3-basic vector in contrast does not contain any promoter elements, relying on the insertion of the potential promoter regions for expression of luciferase. The insertion of the *CACNA1C* region drives luciferase expression at a much lower level than that of *Renilla,* which is driven by the SV40 early enhancer/promoter element.

While the pGL3-basic vector is reported to be suitable for investigating the effect putative promoters and enhancer elements have on luciferase expression, the *CACNA1C* region may not be driving gene expression. Giannakis *et al* in 2003 have shown that the mere presence of a DNA insert in a reporter vector can elicit differences in expression compared to the empty control vector through the interaction of external features, such as androgens [490]. However, the reduction of luciferase expression in the presence of both variants was seen consistently over all experimental replicates. Intronic promoters have been previously described to drive expression of mRNA, and have been associated with diseases such as cancer [491]. A number of brain expressed transcripts are reported in this area of the *CACNA1C* gene (Table 40) which this sequence could be acting as a promoter of. The effects reported from these variants could be validated a number of ways, primarily through the use of an alternative reporter assay such as the pCAT3.

Further work on the *CACNA1C* region should include the insertion of the genetic region into the pGL3-promoter vector to determine whether this region alternatively functions as an enhancer element. The pGL3-promoter vector contains the eukaryotic SV40 promoter which drives luciferase expression. The region of interest can then be placed either upstream or downstream of this promoter to determine the effect the region of interest has on enhancing transcription in the presence of an active promoter.

The potential creation of new TFBSs in the presence of both BD associated variants was investigated using electromobility shift assays. The two variants appear to

have contrasting effects on binding of proteins in the nuclear cell lysate. For rs79398153, the wildtype C allele has a higher affinity for proteins in the nucleus than the rare T allele. Additionally, the rare T allele unlabelled probe fails to outcompete the labelled T probe, indicating less specific binding, or the need for a larger excess of unlabelled T probe. On the other hand, the rare T allele of rs116947827 has a higher affinity for binding of nuclear proteins than that of the wildtype C allele. The C allele does show specificity for protein binding, but a larger excess of unlabelled probe is needed to outcompete the band formed by the T labelled probe. The EMSA super shift assay shows a band super shift in the presence of the two alleles which have higher affinities for proteins in the nuclear cell lysate. These band shift, indicate that the labelled probes of the rs79398153 wildtype C allele, and the rs116947827 rare T allele are binding to proteins tagged with the YY1 antibody. This is contradictory to the information provided by the PROMO *in silico* software that reports no YY1 TFBS in the wildtype *CACNA1C* intronic region. The band shifts observed in Figure 61 are quite faint, and not as intense as the band seen lower in the gel, or in the previous EMSA gels. It is possible that the YY1 antibody bound to TFBS in the nuclear proteins and that some of the labelled probe shared homologous sequence for this region, causing faint bands to be seen, rather than a total super shift of the entire band. Further analysis of this region is warranted to determine if there are novel TFBS formed in the presence of these variants. This could be done by using alternative programmes such as R Vista 2.0 which was not originally utilised in this study.

This analysis has identified novel possible BD susceptibility variants in the third intron of *CACNA1C*. The rs79398153 variant appears to be acting independently of the allelic and haplotypic associations found in the previous BD association studies. Independent genetic replication and biological validation of intronic potentially aetiological variants would support the argument for carrying out whole genome sequencing as well as exome sequencing in BD. The biphasic nature of BD makes a compelling argument for the existence of DNA determined pathological switch mechanisms that may manifest themselves in the loss of control of gene expression. It is plausible that the creation of new transcription factor binding sites might lead to the manifestation of BD through environmental interactions with potential newly created TFBS in the intron of *CACNA1C*. These

hypothesis need to be further tested to delineate the exact manner in which these variants lead to an increased risk in developing BD.

Chapter 7 General Discussion

Bipolar disorder is a highly heritable disorder with a prevalence of 1% in the general population. Perturbation of Ca^{2+} metabolism has been strongly implicated in BD. Calcium channel associated genes have been repeatedly implicated in multiple GWAS studies and pathway analysis. At a physiological level, high levels of intracellular Ca^{2+} have been reported in a number of BD individuals in both platelets and lymphocytes [84, 492]. Variants in the anti-apoptotic protein B-cell CLL/lymphoma-2 (*BCL-2)* gene, which effect calcium homeostasis in the endoplasmic reticulum and the mitochondria, have also been associated with higher Ca^{2+} concentration levels in BD-1 lymphomas [493]. The Ca^{2+} signalling pathway is an integral cascade pathway and is the instigator of a variety of cellular mechanisms such as LTP in neurons and initiation of gene transcription. Many of these pathways are important for correct synaptic functioning. Ca^{2+} plays a central role in neurotransmitter release from the presynaptic density and several postsynaptic processes. Previous evidence shows that variants in the α pore of the calcium channel gene result in severe psychiatric illness such as Timothy Syndrome [363].

The L-type family of calcium channel genes and genes encoding other calcium channel subunits were investigated to determine whether DNA variation present in these genes contribute to the genetic risk of BD using HRM analysis and whole genome sequencing. This thesis focused on genetic variants which are present in the protein coding and regulatory region of the calcium channel genes. However, we also utilised whole genome sequencing data to investigate the third intron of *CACNA1C* for variants that may explain the replicated genome-wide association signal reported in this region [232, 261, 270].

Finding variants which contribute to disease risk for complex disorders is challenging. Common variants are likely to have small effect sizes, while variants with larger effect sizes are most likely to be rare in nature. Evolutionary theory holds that deleterious variants by default are likely to be rare due to purifying selection [443, 444]. The discovery sample used the HRM analysis identified both common and rare variants in the BD cohort. However the sample size for follow up

genotyping was not large enough to identify rare variants that may be associated with BD.

The ability to detect an association between rare variants and a disease, is based on a number of factors, one is the relative risk of the variant, as estimated by 1+ λ, where λ is the excess relative risk of disease conferred by the allele [445]. The second is the sample size. Zuk *et al,* have determined that to detect allele association with a 2-fold increase to disease risk, approximately 28,000 samples are needed. For variants with larger relative risk scores, such as a 20-fold increase in relative risk, 260 cases would be needed [445]. Our study on BD was based on a relatively large sample size with approximately 1800 cases and 1300 controls. The initial sample size for the BD WGS included 99 case individuals. To identify more rare disease associated variants with large effect sizes larger sample sizes are needed. Results from studies such as the PGC-BD GWAS [270] emphasise the benefit of having large sample sizes, having identified eight loci passing genome-wide significance.

High-resolution melting curve analysis allowed for the rapid screening of the 5'UTR, promoter region and first exon of the calcium channel genes *CACNA1C, CACNA1D* and *CACNB3*. This technique provided a way to screen the gene for both common and rare variants located in these regions. This technique did not find any variants associated with BD in these three genes.

The functional and regulatory regions of the *CACNG4* gene were screened in full using HRM analysis. This identified two non-synonymous variants, rs371128228 which results in an arginine to glutamine change and 17:65026851 (C/T) which results in an arginine to tryptophan change. Both of these variants are located in the fourth exon of the gene. Neither variant was associated with BD alone (rs371128228, p=0.06, 17:65026851 (C/T) p=0.78). However, both variants were associated with psychiatric illness in a combined sample of approximately 37000 controls and 8000 cases. The combined psychiatric illness sample comprised of 1890 BD samples, 4538 SCZ samples, and 1221 individuals with ADS (rs371128228, p=1.05x10^{-4}, OR=4.39, 17:65026851 (C/T) p=0.001, OR=9.52).

To date there have been no reports of associated variants within the *CACNG4* gene. Variants in other members of this gene family, such as *CACNG2* and *CACNG5,* have been reported to be associated with psychiatric illness. Case only variants in the *CACNG5* gene are present in areas homologous to those harbouring the two rare variants in the *CACNG4* gene [390]. Flow cytometry analysis into the effect of these variants on *CACNG5* trafficking of AMPA-R to the cell surface demonstrate that some of the variants decrease GluR2 levels on the cell surface.

The *CACNG4* variant, rs371128228, disrupt the trafficking of GluR1 to the cell surface (p=0.02). Abnormal AMPA-R trafficking has been previously linked to several neurological disorders suggesting a possible pathogenic effect for these variants [494, 495]. While it is currently unclear exactly how TARPs and AMPA-R interact, studies suggest that TARPs interacts with AMPA-R in the ER to generate a facilitating receptor complex for AMPA-R exit from the ER [351, 496]. Changes in the TARP-γ4 protein structure could impact the efficiency or ability of the TARP to export the AMPA-R from the ER. Additionally the decrease of GluR1 subunits on the cell surface could result from an alteration to the ability of the TARP to anchor to the post-synaptic density. Further functional work to determine the mechanism of action of this variant is needed.

Whole genome sequencing data was used to identify variants in the coding regions of all members of the L-Type family of calcium channel genes. This included *CACNG4* to ensure that none of the variants had been missed using HRM analysis. This identified 26 non-synonymous mutations in the genes *CACNA1D, CACNA1S, CACNA2D4, CACNB1,* and *CACNB2.* Bioinformatic tools, PolyPhen2, and Project Hope, were used to determine whether these variants may be classed as deleterious for follow up genotyping in the UCL cohort. None of these variants showed an association with BD in the UCL samples.

Our study designed an approach to investigate variants from WGS in the third intron of *CACNA1C* where the BD associated variant rs1006737 lies. We first examined the intron for any characteristic signatures which may indicate the reason for its repeated association with BD. We found there is strong LD present in this region in addition to high levels of conservation and high levels of histone marks. In general, intronic regions are not subjected to as much selective pressure

as that seen in exonic regions of gene. We then used the ENCODE data to determine whether there was a reason behind the conservation observed. For example, if this intronic region plays a role in regulation of gene transcription then it might explain the levels of conservation seen. Eight variants were identified that are located in putative regulatory regions as denoted by ENCODE, have allele frequencies larger than those seen in the 1000G European populations and in conserved regions. One of these variants, rs79398153, was associated in the UCL BD cohort (p=0.015, OR=1.511). Further investigation of the WGS data in this region identified a variant, rs116947827, which is in complete LD with the BD associated variants. Use of the ENCODE allowed for the identification of putative functional marks in which these variants are located.

These variants, rs79398153 and rs116947827, decrease the level of luciferase expression significantly compared to the wild-type sequence (p=0.002). This further implicates their role in regulation of transcription. This could be through the creation of the YY1 transcription factor binding site. The use of the EMSA supershift assay shows support for binding of the YY1 antibody to proteins in the nuclear cell lysate. The YY1 transcription factor is ubiquitously expressed and is involved in either repression or activation of gene expression through the directing of histone deacetylases and histone acetyltransferases to promoters in the genome. Variants which result in the change of gene expression are hard to characterise. Regulatory effects can be far reaching and do not necessarily relate to the gene in which these regions are located. However, given the evidence for disrupted Ca^{2+} levels in individuals with BD, a decrease in the expression of the *CACNA1C* α pore may play a role in aberrant calcium influx. These variants are not in LD with rs1006737, and do not explain the replicated association signal observed at this locus. However our study shows the importance of designing novel approaches to characterising SNVs for genotyping and then functional follow up. The role these variants may have in the aetiology of BD, but perturbations to gene expression may have a detrimental knock-on effect further downstream.

7.1 STUDY LIMITATIONS AND POTENTIAL CONFOUNDS

HRM and whole genome sequencing methods were used for variant investigation both had their advantages and disadvantages. Of the 28 variants identified in *CACNA1C*, *CACNA1D, CACNB3* and *CACNG4* genes using HRM only five of them were A/T or C/G changes. The remaining 25 variants result in base changes which would result in the gain or loss of hydrogen bonds. These bonds play an integral role in the profile of the melt curve produced by the HRM software. Short genetic amplicon regions of 50bp or so would have increased sensitivity in detecting variant changes which do not result in the gain or loss of a hydrogen bond. In amplicons of up to 350bp in length it seems that the HRM software may not call the changes as accurately. HRM allowed for the screening of all BD samples in our cohort. The initial screening is fast and high throughput. These variants can then be validated through Sanger sequencing from the original DNA samples.

The cost of HRM is significantly lower than that of WGS and allows for more individuals to be screened in addition to a shorter turnaround time from the beginning of an experiment to data collection. WGS is high throughput but cost implications often impede the screening of thousands of individuals. WGS allows for probing of the full genetic architecture underlying BD and the sequence depth allows for accurate quality control for base pair calls. In addition to this, regions of the genome which may be difficult to screen using PCR based methods can be investigated using WGS.

Additionally the statistical power of single-variant based association tests for low-frequency variants is reduced unless the sample sizes or variant effect sizes are very large [293]. The sample sizes for follow up genotyping in our study were not large enough to determine whether associations were present between rare variants and BD.

The highest proportions of variants identified in the *CACNG4* gene were present in the 3'UTR region. Variants located in the 3'UTR of *CACNG4* may also play a role in disease susceptibility through their impact on mRNA stability in addition to the variants located in the fourth exon of *CACNG4*. These were not included in our

initial genotyping follow-up but could be potential candidates for further investigation into the role of *CACNG4* in BD.

7.2 FUTURE DIRECTIONS

To further investigate our genetic association findings in the *CACNA1C* gene, genotyping of these variants in larger BD case-control samples is essential. Additionally WGS BD databases could be utilised to screen for these variants. Other variants in intron three of *CACNA1C* that did not meet the criteria in our discovery set could be further genotyped. The rs116947827 variant was present in a repeat region in the third intron of *CACNA1C* and did not fit the criteria for follow up genotyping in the UCL BD sample. However, the luciferase functional assay shows a decrease in gene expression only when both rs7938153 and rs116947827 are present. Haplotype analysis may provide some insight into the LD structure in this region and further highlight variants located in functional regions.

Further functional analysis needs to be conducted to determine how the variants cause the decrease in gene expression and the role of YY1 or other TFBS which are predicted to arise in this region. Recent reports have identified nine genes in the human brain which undergo recursive splicing [497]. Recursive splicing is reported to occurs in genes with long introns and removes the intron from the precursor mRNA through splicing which occurs in a step wise manner from the 5' end of the intron to the 3' end. In vertebrates long introns often contain a recursive splicing exon with 5' and 3' splice sites to aid the step wise splicing and ensuring it occurs in the correct step wise process. The third intron of *CACNA1C* is approximately 330kb in length which is extremely long. It is plausible that this intronic region contains recursive splicing exons which if variants located in these regions disrupted splicing in the precursor mRNA it could result in nonsense-mediated RNA decay.

Investigation of functional variants in *CACNG4* was the first approach at identifying BD associated variants in this TARP in our study. Our criteria for follow up genotyping focused solely on protein coding variants which result in amino acid changes. The presence of multiple variants in the 3'UTR of the *CACNG4* gene may

harbour some potential disease associated changes. Flow cytometry analysis was not conducted on the second BD associated variant in *CACNG4*. If this variant was to also disrupt trafficking of AMPA-Rs it would provide further evidence for the role of this gene in mental illness. This variant is extremely rare and would benefit from follow up analysis in a larger BD sample.

There is a burden of rare case only variants in the *CACNA1D* gene. This gene has been implicated in BD through pathway and WGS sequencing analysis. Follow up of the variants identified in this study in larger cohorts may allow for an insight into the role, if any, these variants may have in risk of BD.

Combined efforts from large consortiums in the field of psychiatric genetics have shown the impact collaborative efforts can have in the search for genetic variants contributing to BD. New analytical approaches to integrating genetic level pathway analysis with biological systems data is crucial in understanding the impact variants identified through GWAS and whole genome sequencing may have at a functional level. By increasing our understanding of the underlying biological mechanisms that lead to bipolar disorder we can provide a platform for the development of targeted personalised treatments, more accurate diagnosis and most importantly, improved quality of life for individuals suffering from bipolar disorder.

References

1. Kessler, R.C., et al., *The global burden of mental disorders: an update from the WHO World Mental Health (WMH) surveys.* Epidemiol Psichiatr Soc, 2009. **18**(1): p. 23-33.
2. Commission, E., *Mental Health Systems in the European Union Member States, Status of Mental Health in Populations and Benefits to be Expected from Investments into Mental Health*, H.P.o.t.E. Union, Editor. 2013.
3. Collins, P.Y., et al., *Grand challenges in global mental health.* Nature, 2011. **475**(7354): p. 27-30.
4. Wittchen, H.U., et al., *The size and burden of mental disorders and other disorders of the brain in Europe 2010.* Eur Neuropsychopharmacol, 2011. **21**(9): p. 655-79.
5. Gustavsson, A., et al., *Cost of disorders of the brain in Europe 2010.* Eur Neuropsychopharmacol, 2011. **21**(10): p. 718-79.
6. FK, G. and J. KR, *Manic-Depressive Illness. Bipolar Disorders and Recurrent Depression*. 2nd edn ed, ed. U.P. Oxford. 2007.
7. Organization, W.H., *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines.* 1992, World Health Organization: Geneva.
8. Perlis, R.H., et al., *Long-term implications of early onset in bipolar disorder: data from the first 1000 participants in the systematic treatment enhancement program for bipolar disorder (STEP-BD).* Biol Psychiatry, 2004. **55**(9): p. 875-81.
9. Grover, S., et al., *Cotard's syndrome: Two case reports and a brief review of literature.* J Neurosci Rural Pract, 2014. **5**(Suppl 1): p. S59-62.
10. Regier, D.A., et al., *Comorbidity of mental disorders with alcohol and other drug abuse. Results from the Epidemiologic Catchment Area (ECA) Study.* JAMA, 1990. **264**(19): p. 2511-8.
11. Cornelius, J.R., et al., *Alcohol and psychiatric comorbidity.* Recent Dev Alcohol, 2003. **16**: p. 361-74.
12. Angst, J. and A. Marneros, *Bipolarity from ancient to modern times: conception, birth and rebirth.* J Affect Disord, 2001. **67**(1-3): p. 3-19.
13. Marneros, A.G., F., *Bipolar Disorders: Mixed States, Rapid Cycling and Atypical Forms*. 2005, Cambridge, United Kingdom.: Cambridge University Press.
14. Fountoulakis, K.N., et al., *Treatment guidelines for bipolar disorder: a critical review.* J Affect Disord, 2005. **86**(1): p. 1-10.
15. Charney, D.S., et al., *Neurobiology of mental illness.*, ed. O.U. Press. 2013.
16. Maciocia, G., *Diagnosis in Chinese Medicine: A Comprehensive Guide*. 2013.
17. Healy, D., *Mania:A Short History of Bipolar*. 2010, JHU Press.
18. Zivanovic, O. and A. Nedic, *Kraepelin's concept of manic-depressive insanity: one hundred years later.* J Affect Disord, 2012. **137**(1-3): p. 15-24.
19. King, J., M. Agius, and R. Zaman, *The Kraepelinian dichotomy in terms of suicidal behaviour.* Psychiatr Danub, 2012. **24 Suppl 1**: p. S117-8.
20. Association, A.P., *Diagnostic and Statistical Manual of Mental Disorders (5th Edition)*. 2013, American Psychiatric Publishing.: Arlington, VA.
21. Cuthbert, B.N. and T.R. Insel, *Toward the future of psychiatric diagnosis: the seven pillars of RDoC.* BMC Med, 2013. **11**: p. 126.

22. Moriyama IM, Loy RM, and R.-S. AHT *History of the statistical classification of diseases and causes of death* 2011.
23. Akiskal, H.S., *Mood Disorders: Clinical Features*. 2005, New York: Lippincott Williams & Wilkins.
24. Regier, D.A., E.A. Kuhl, and D.J. Kupfer, *The DSM-5: Classification and criteria changes.* World Psychiatry, 2013. **12**(2): p. 92-8.
25. Stoll, A.L., et al., *Shifts in diagnostic frequencies of schizophrenia and major affective disorders at six North American psychiatric hospitals, 1972-1988.* Am J Psychiatry, 1993. **150**(11): p. 1668-73.
26. Health, N.C.C.f.M. *Bipolar Disorder The assessment and management of bipolar disorder in adults, children and young people in primary and secondary care* 2014.
27. Rihmer, Z. and K. Kiss, *Bipolar disorders and suicidal behaviour.* Bipolar Disord, 2002. **4 Suppl 1**: p. 21-5.
28. Tondo, L., B. Lepri, and R.J. Baldessarini, *Suicidal risks among 2826 Sardinian major affective disorder patients.* Acta Psychiatr Scand, 2007. **116**(6): p. 419-28.
29. Osby, U., et al., *Excess mortality in bipolar and unipolar disorder in Sweden.* Arch Gen Psychiatry, 2001. **58**(9): p. 844-50.
30. Harris, E.C. and B. Barraclough, *Suicide as an outcome for mental disorders. A meta-analysis.* Br J Psychiatry, 1997. **170**: p. 205-28.
31. Schaffer, A., et al., *International Society for Bipolar Disorders Task Force on Suicide: meta-analyses and meta-regression of correlates of suicide attempts and suicide deaths in bipolar disorder.* Bipolar Disord, 2015. **17**(1): p. 1-16.
32. Angst, F., et al., *Mortality of patients with mood disorders: follow-up over 34-38 years.* J Affect Disord, 2002. **68**(2-3): p. 167-81.
33. Phillips, M.L. and D.J. Kupfer, *Bipolar disorder diagnosis: challenges and future directions.* Lancet, 2013. **381**(9878): p. 1663-71.
34. Association, A.P., *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, (DSM-IV)*. 1994, American Psychiatric Press, Washington, DC
35. Schaffer, A., et al., *International Society for Bipolar Disorders Task Force on Suicide: meta-analyses and meta-regression of correlates of suicide attempts and suicide deaths in bipolar disorder.* Bipolar Disord, 2014.
36. Shih, R.A., P.L. Belmonte, and P.P. Zandi, *A review of the evidence from family, twin and adoption studies for a genetic contribution to adult psychiatric disorders.* Int Rev Psychiatry, 2004. **16**(4): p. 260-83.
37. Simonsen, C., et al., *Neurocognitive dysfunction in bipolar and schizophrenia spectrum disorders depends on history of psychosis rather than diagnostic group.* Schizophr Bull, 2011. **37**(1): p. 73-83.
38. Ellison-Wright, I. and E. Bullmore, *Anatomy of bipolar disorder and schizophrenia: a meta-analysis.* Schizophr Res, 2010. **117**(1): p. 1-12.
39. Purcell, S.M., et al., *Common polygenic variation contributes to risk of schizophrenia and bipolar disorder.* Nature, 2009. **460**(7256): p. 748-52.
40. Markham, J.A. and J.I. Koenig, *Prenatal stress: role in psychotic and depressive diseases.* Psychopharmacology (Berl), 2011. **214**(1): p. 89-106.
41. Haukvik, U.K., et al., *Pre- and perinatal hypoxia associated with hippocampus/amygdala volume in bipolar disorder.* Psychol Med, 2014. **44**(5): p. 975-85.
42. Torrey, E.F., et al., *Seasonality of births in schizophrenia and bipolar disorder: a review of the literature.* Schizophr Res, 1997. **28**(1): p. 1-38.

43. Hultman, C.M., et al., *Prenatal and perinatal risk factors for schizophrenia, affective psychosis, and reactive psychosis of early onset: case-control study.* BMJ, 1999. **318**(7181): p. 421-6.
44. Brown, A.S., et al., *Further evidence of relation between prenatal famine and major affective disorder.* Am J Psychiatry, 2000. **157**(2): p. 190-5.
45. Uher, R., *Gene-environment interactions in severe mental illness.* Front Psychiatry, 2014. **5**: p. 48.
46. Nosarti, C., et al., *Preterm birth and psychiatric disorders in young adult life.* Arch Gen Psychiatry, 2012. **69**(6): p. E1-8.
47. Garno, J.L., et al., *Impact of childhood abuse on the clinical course of bipolar disorder.* Br J Psychiatry, 2005. **186**: p. 121-5.
48. Post, R.M., et al., *Verbal abuse, like physical and sexual abuse, in childhood is associated with an earlier onset and more difficult course of bipolar disorder.* Bipolar Disord, 2015. **17**(3): p. 323-30.
49. Cassidy, F., E.P. Ahearn, and B.J. Carroll, *Substance abuse in bipolar disorder.* Bipolar Disord, 2001. **3**(4): p. 181-8.
50. Feingold, D., et al., *The association between cannabis use and mood disorders: A longitudinal study.* J Affect Disord, 2014. **172C**: p. 211-218.
51. Kroon, J.S., et al., *Incidence rates and risk factors of bipolar disorder in the general population: a population-based cohort study.* Bipolar Disord, 2013. **15**(3): p. 306-13.
52. Mann-Wrobel, M.C., J.T. Carreno, and D. Dickinson, *Meta-analysis of neuropsychological functioning in euthymic bipolar disorder: an update and investigation of moderator variables.* Bipolar Disord, 2011. **13**(4): p. 334-42.
53. Heinrichs, R.W. and K.K. Zakzanis, *Neurocognitive deficit in schizophrenia: a quantitative review of the evidence.* Neuropsychology, 1998. **12**(3): p. 426-45.
54. Seidman, L.J., et al., *Neuropsychological performance and family history in children at age 7 who develop adult schizophrenia or bipolar psychosis in the New England Family Studies.* Psychol Med, 2013. **43**(1): p. 119-31.
55. Bora, E., M. Yücel, and C. Pantelis, *Neurocognitive markers of psychosis in bipolar disorder: a meta-analytic study.* J Affect Disord, 2010. **127**(1-3): p. 1-9.
56. Reichenberg, A., et al., *Neuropsychological function and dysfunction in schizophrenia and psychotic affective disorders.* Schizophr Bull, 2009. **35**(5): p. 1022-9.
57. Zanelli, J., et al., *Specific and generalized neuropsychological deficits: a comparison of patients with various first-episode psychosis presentations.* Am J Psychiatry, 2010. **167**(1): p. 78-85.
58. Dickerson, F., et al., *Cognitive functioning in recent onset psychosis.* J Nerv Ment Dis, 2011. **199**(6): p. 367-71.
59. Rieder, R.O., et al., *Computed tomographic scans in patients with schizophrenia, schizoaffective, and bipolar affective disorder.* Arch Gen Psychiatry, 1983. **40**(7): p. 735-9.
60. Kupferschmidt, D.A. and K.K. Zakzanis, *Toward a functional neuroanatomical signature of bipolar disorder: quantitative evidence from the neuroimaging literature.* Psychiatry Res, 2011. **193**(2): p. 71-9.
61. Strakowski, S.M., et al., *Ventricular and periventricular structural volumes in first- versus multiple-episode bipolar disorder.* Am J Psychiatry, 2002. **159**(11): p. 1841-7.

62. Schneider, M.R., et al., *Neuroprogression in bipolar disorder.* Bipolar Disord, 2012. **14**(4): p. 356-74.
63. Beyer, J.L., et al., *Hyperintense MRI lesions in bipolar disorder: A meta-analysis and review.* Int Rev Psychiatry, 2009. **21**(4): p. 394-409.
64. Hallahan, B., et al., *Structural magnetic resonance imaging in bipolar disorder: an international collaborative mega-analysis of individual adult patient data.* Biol Psychiatry, 2011. **69**(4): p. 326-35.
65. Bearden, C.E., et al., *Three-dimensional mapping of hippocampal anatomy in unmedicated and lithium-treated patients with bipolar disorder.* Neuropsychopharmacology, 2008. **33**(6): p. 1229-38.
66. Sassi, R.B., et al., *Reduced left anterior cingulate volumes in untreated bipolar patients.* Biol Psychiatry, 2004. **56**(7): p. 467-75.
67. Jackowski, A.P., et al., *The involvement of the orbitofrontal cortex in psychiatric disorders: an update of neuroimaging findings.* Rev Bras Psiquiatr, 2012. **34**(2): p. 207-12.
68. Hafeman, D.M., et al., *Effects of medication on neuroimaging findings in bipolar disorder: an updated review.* Bipolar Disord, 2012. **14**(4): p. 375-410.
69. Bourne, C., et al., *Neuropsychological testing of cognitive impairment in euthymic bipolar disorder: an individual patient data meta-analysis.* Acta Psychiatr Scand, 2013. **128**(3): p. 149-62.
70. Janowsky, D.S., et al., *A cholinergic-adrenergic hypothesis of mania and depression.* Lancet, 1972. **2**(7778): p. 632-5.
71. Janowsky, D.S., D.H. Overstreet, and J.I. Nurnberger, *Is cholinergic sensitivity a genetic marker for the affective disorders?* Am J Med Genet, 1994. **54**(4): p. 335-44.
72. Cousins, D.A., K. Butts, and A.H. Young, *The role of dopamine in bipolar disorder.* Bipolar Disord, 2009. **11**(8): p. 787-806.
73. Emilien, G., et al., *Dopamine receptors--physiological understanding to therapeutic intervention potential.* Pharmacol Ther, 1999. **84**(2): p. 133-56.
74. van Enkhuizen, J., et al., *The catecholaminergic-cholinergic balance hypothesis of bipolar disorder revisited.* Eur J Pharmacol, 2015. **753**: p. 114-26.
75. Salvi, V., et al., *The use of antidepressants in bipolar disorder.* J Clin Psychiatry, 2008. **69**(8): p. 1307-18.
76. Artigas, F., et al., *Acceleration of the effect of selected antidepressant drugs in major depression by 5-HT1A antagonists.* Trends Neurosci, 1996. **19**(9): p. 378-83.
77. Wang, J., S.K. Michelhaugh, and M.J. Bannon, *Valproate robustly increases Sp transcription factor-mediated expression of the dopamine transporter gene within dopamine cells.* Eur J Neurosci, 2007. **25**(7): p. 1982-6.
78. Eickholt, B.J., et al., *Effects of valproic acid derivatives on inositol trisphosphate depletion, teratogenicity, glycogen synthase kinase-3beta inhibition, and viral replication: a screening approach for new bipolar disorder drugs derived from the valproic acid core structure.* Mol Pharmacol, 2005. **67**(5): p. 1426-33.
79. Suhara, T., et al., *D1 dopamine receptor binding in mood disorders measured by positron emission tomography.* Psychopharmacology (Berl), 1992. **106**(1): p. 14-8.

80. Rao, J.S., et al., *Dysregulated glutamate and dopamine transporters in postmortem frontal cortex from bipolar and schizophrenic patients.* J Affect Disord, 2012. **136**(1-2): p. 63-71.
81. Rao, J.S., et al., *Increased excitotoxicity and neuroinflammatory markers in postmortem frontal cortex from bipolar disorder patients.* Mol Psychiatry, 2010. **15**(4): p. 384-92.
82. Sanacora, G., et al., *Targeting the glutamatergic system to develop novel, improved therapeutics for mood disorders.* Nat Rev Drug Discov, 2008. **7**(5): p. 426-37.
83. Hashimoto, K., A. Sawa, and M. Iyo, *Increased levels of glutamate in brains from patients with mood disorders.* Biol Psychiatry, 2007. **62**(11): p. 1310-6.
84. Dubovsky, S.L., et al., *Intracellular calcium signalling in peripheral cells of patients with bipolar affective disorder.* Eur Arch Psychiatry Clin Neurosci, 1994. **243**(5): p. 229-34.
85. Berk, M., et al., *Pathways underlying neuroprogression in bipolar disorder: focus on inflammation, oxidative stress and neurotrophic factors.* Neurosci Biobehav Rev, 2011. **35**(3): p. 804-17.
86. Hornung, J.P., *The human raphe nuclei and the serotonergic system.* J Chem Neuroanat, 2003. **26**(4): p. 331-43.
87. Berton, O. and E.J. Nestler, *New approaches to antidepressant drug discovery: beyond monoamines.* Nat Rev Neurosci, 2006. **7**(2): p. 137-51.
88. Pittenger, C. and R.S. Duman, *Stress, depression, and neuroplasticity: a convergence of mechanisms.* Neuropsychopharmacology, 2008. **33**(1): p. 88-109.
89. Krishnan, V. and E.J. Nestler, *The molecular neurobiology of depression.* Nature, 2008. **455**(7215): p. 894-902.
90. Berridge, M.J., *Calcium signalling remodelling and disease.* Biochem Soc Trans, 2012. **40**(2): p. 297-309.
91. Berridge, M.J., *Dysregulation of neural calcium signaling in Alzheimer disease, bipolar disorder and schizophrenia.* Prion, 2013. **7**(1): p. 2-13.
92. Giegling, I., et al., *Genetic findings in schizophrenia patients related to alterations in the intracellular Ca-homeostasis.* Prog Neuropsychopharmacol Biol Psychiatry, 2010. **34**(8): p. 1375-80.
93. Berridge, M.J., *Calcium signalling and psychiatric disease: bipolar disorder and schizophrenia.* Cell Tissue Res, 2014. **357**(2): p. 477-92.
94. Jagannath, A., S.N. Peirson, and R.G. Foster, *Sleep and circadian rhythm disruption in neuropsychiatric illness.* Curr Opin Neurobiol, 2013. **23**(5): p. 888-94.
95. Berridge, M.J., *Calcium regulation of neural rhythms, memory and Alzheimer's disease.* J Physiol, 2014. **592**(Pt 2): p. 281-93.
96. Suh, B.C. and B. Hille, *Recovery from muscarinic modulation of M current channels requires phosphatidylinositol 4,5-bisphosphate synthesis.* Neuron, 2002. **35**(3): p. 507-20.
97. Brietzke, E., et al., *Comparison of cytokine levels in depressed, manic and euthymic patients with bipolar disorder.* J Affect Disord, 2009. **116**(3): p. 214-7.
98. Abbas, A.K., Lichtman, A.H.,Pillai, S., *Cellular and molecular immunology*. 7th Ed ed. 2012, Philadelphia, PA Elsevier Saunders.
99. Miller, A.H., et al., *Cytokine targets in the brain: impact on neurotransmitters and neurocircuits.* Depress Anxiety, 2013. **30**(4): p. 297-306.

100. Watson, S., et al., *Hypothalamic-pituitary-adrenal axis function in patients with bipolar disorder.* Br J Psychiatry, 2004. **184**: p. 496-502.
101. Knable, M.B., et al., *Molecular abnormalities of the hippocampus in severe psychiatric illness: postmortem findings from the Stanley Neuropathology Consortium.* Mol Psychiatry, 2004. **9**(6): p. 609-20, 544.
102. Rosenblat, J.D., et al., *Inflamed moods: a review of the interactions between inflammation and mood disorders.* Prog Neuropsychopharmacol Biol Psychiatry, 2014. **53**: p. 23-34.
103. Citrome, L., *Treatment of bipolar depression: making sensible decisions.* CNS Spectr, 2014. **19 Suppl 1**: p. 4-11; quiz 1-3, 12.
104. Yildiz, A., et al., *Efficacy of antimanic treatments: meta-analysis of randomized, controlled trials.* Neuropsychopharmacology, 2011. **36**(2): p. 375-89.
105. CADE, J.F., *Lithium salts in the treatment of psychotic excitement.* Med J Aust, 1949. **2**(10): p. 349-52.
106. Martinowich, K., R.J. Schloesser, and H.K. Manji, *Bipolar disorder: from genes to behavior pathways.* J Clin Invest, 2009. **119**(4): p. 726-36.
107. Agam, G., et al., *Myo-inositol-1-phosphate (MIP) synthase: a possible new target for antibipolar drugs.* Bipolar Disord, 2002. **4 Suppl 1**: p. 15-20.
108. Shaltiel, G., et al., *Valproate decreases inositol biosynthesis.* Biol Psychiatry, 2004. **56**(11): p. 868-74.
109. Chang, P., et al., *The antiepileptic drug valproic acid and other medium-chain fatty acids acutely reduce phosphoinositide levels independently of inositol in Dictyostelium.* Dis Model Mech, 2012. **5**(1): p. 115-24.
110. Chang, P., M.C. Walker, and R.S. Williams, *Seizure-induced reduction in PIP3 levels contributes to seizure-activity and is rescued by valproic acid.* Neurobiol Dis, 2014. **62**: p. 296-306.
111. Kim, H.J. and S.A. Thayer, *Lithium increases synapse formation between hippocampal neurons by depleting phosphoinositides.* Mol Pharmacol, 2009. **75**(5): p. 1021-30.
112. Grutzendler, J., N. Kasthuri, and W.B. Gan, *Long-term dendritic spine stability in the adult cortex.* Nature, 2002. **420**(6917): p. 812-6.
113. Hedgepeth, C.M., et al., *Activation of the Wnt signaling pathway: a molecular mechanism for lithium action.* Dev Biol, 1997. **185**(1): p. 82-91.
114. Wada, A., *Lithium and neuropsychiatric therapeutics: neuroplasticity via glycogen synthase kinase-3beta, beta-catenin, and neurotrophin cascades.* J Pharmacol Sci, 2009. **110**(1): p. 14-28.
115. Mason, J.L., et al., *Insulin-like growth factor (IGF) signaling through type 1 IGF receptor plays an important role in remyelination.* J Neurosci, 2003. **23**(20): p. 7710-8.
116. Wang, J.F., C. Bown, and L.T. Young, *Differential display PCR reveals novel targets for the mood-stabilizing drug valproate including the molecular chaperone GRP78.* Mol Pharmacol, 1999. **55**(3): p. 521-7.
117. Moore, G.J., et al., *A longitudinal study of the effects of lithium treatment on prefrontal and subgenual prefrontal gray matter volume in treatment-responsive bipolar disorder patients.* J Clin Psychiatry, 2009. **70**(5): p. 699-705.
118. Lyoo, I.K., et al., *Lithium-induced gray matter volume increase as a neural correlate of treatment response in bipolar disorder: a longitudinal brain imaging study.* Neuropsychopharmacology, 2010. **35**(8): p. 1743-50.

119. Shim, S.S., et al., *Effects of 4-weeks of treatment with lithium and olanzapine on long-term potentiation in hippocampal area CA1.* Neurosci Lett, 2012. **524**(1): p. 5-9.
120. Chuang, D.M., et al., *Neuroprotective effects of lithium in cultured cells and animal models of diseases.* Bipolar Disord, 2002. **4**(2): p. 129-36.
121. Rowe, M.K. and D.M. Chuang, *Lithium neuroprotection: molecular mechanisms and clinical implications.* Expert Rev Mol Med, 2004. **6**(21): p. 1-18.
122. Chen, G., et al., *Enhancement of hippocampal neurogenesis by lithium.* J Neurochem, 2000. **75**(4): p. 1729-34.
123. O'Leary, O.F., R.M. O'Connor, and J.F. Cryan, *Lithium-induced effects on adult hippocampal neurogenesis are topographically segregated along the dorso-ventral axis of stressed mice.* Neuropharmacology, 2012. **62**(1): p. 247-55.
124. Manji, H.K., G.J. Moore, and G. Chen, *Bipolar disorder: leads from the molecular and cellular mechanisms of action of mood stabilizers.* Br J Psychiatry Suppl, 2001. **41**: p. s107-19.
125. Müller-Oerlinghausen, B. and U. Lewitzka, *Lithium reduces pathological aggression and suicidality: a mini-review.* Neuropsychobiology, 2010. **62**(1): p. 43-9.
126. Serra, G., et al., *Memantine: New prospective in bipolar disorder treatment.* World J Psychiatry, 2014. **4**(4): p. 80-90.
127. 185, N.c.g. *Bipolar disorder: the assessment and management of bipolar disorder in adults, children and young people in primary and secondary care* 2014.
128. Samuels, B.A., et al., *Serotonin 1A and Serotonin 4 Receptors: Essential Mediators of the Neurogenic and Behavioral Actions of Antidepressants.* Neuroscientist, 2014.
129. Dubovsky, S.L., *Mania.* Continuum (Minneap Minn), 2015. **21**(3 Behavioral Neurology and Neuropsychiatry): p. 737-55.
130. Yatham, L.N., et al., *Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) collaborative update of CANMAT guidelines for the management of patients with bipolar disorder: update 2013.* Bipolar Disord, 2013. **15**(1): p. 1-44.
131. Meyer, T.D. and M. Hautzinger, *Cognitive behaviour therapy and supportive therapy for bipolar disorders: relapse rates for treatment period and 2-year follow-up.* Psychol Med, 2012. **42**(7): p. 1429-39.
132. **Wray, N.** and **P.** V**isscher**. ***Estimating trait heritability***. **1(1):29** 2008.
133. Craddock, N. and I. Jones, *Genetics of bipolar disorder.* J Med Genet, 1999. **36**(8): p. 585-94.
134. Tsuang, M. and S. Faraone, *The Genetics of Mood Disorders*. 1990, Baltimore: John Hopkins University Press.
135. Craddock, N., M.C. O'Donovan, and M.J. Owen, *The genetics of schizophrenia and bipolar disorder: dissecting psychosis.* J Med Genet, 2005. **42**(3): p. 193-204.
136. Sullivan, P.F., M.C. Neale, and K.S. Kendler, *Genetic epidemiology of major depression: review and meta-analysis.* Am J Psychiatry, 2000. **157**(10): p. 1552-62.
137. Barnett, J.H. and J.W. Smoller, *The genetics of bipolar disorder.* Neuroscience, 2009. **164**(1): p. 331-43.

138. Song, J., et al., *Bipolar disorder and its relation to major psychiatric disorders: a family-based study in the Swedish population.* Bipolar Disord, 2015. **17**(2): p. 184-93.
139. Lichtenstein, P., et al., *Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study.* Lancet, 2009. **373**(9659): p. 234-9.
140. Wray, N.R. and I.I. Gottesman, *Using summary data from the danish national registers to estimate heritabilities for schizophrenia, bipolar disorder, and major depressive disorder.* Front Genet, 2012. **3**: p. 118.
141. Heun, R. and W. Maier, *The distinction of bipolar II disorder from bipolar I and recurrent unipolar depression: results of a controlled family study.* Acta Psychiatr Scand, 1993. **87**(4): p. 279-84.
142. Andreasen, N.C., et al., *Familial rates of affective disorder. A report from the National Institute of Mental Health Collaborative Study.* Arch Gen Psychiatry, 1987. **44**(5): p. 461-9.
143. Coryell, W., et al., *A family study of bipolar II disorder.* Br J Psychiatry, 1984. **145**: p. 49-54.
144. Baldessarini, R.J., et al., *Age at onset versus family history and clinical outcomes in 1,665 international bipolar-I disorder patients.* World Psychiatry, 2012. **11**(1): p. 40-6.
145. Lee, S.H., et al., *Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs.* Nat Genet, 2013. **45**(9): p. 984-94.
146. Consortium, C.-D.G.o.t.P.G., *Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis.* Lancet, 2013. **381**(9875): p. 1371-9.
147. Colvert, E., et al., *Heritability of Autism Spectrum Disorder in a UK Population-Based Twin Sample.* JAMA Psychiatry, 2015. **72**(5): p. 415-23.
148. McGuffin, P., et al., *The heritability of bipolar affective disorder and the genetic relationship to unipolar depression.* Arch Gen Psychiatry, 2003. **60**(5): p. 497-502.
149. Kieseppä, T., et al., *High concordance of bipolar I disorder in a nationwide sample of twins.* Am J Psychiatry, 2004. **161**(10): p. 1814-21.
150. Edvardsen, J., et al., *Heritability of bipolar spectrum disorders. Unity or heterogeneity?* J Affect Disord, 2008. **106**(3): p. 229-40.
151. Bienvenu, O.J., D.S. Davydow, and K.S. Kendler, *Psychiatric 'diseases' versus behavioral disorders and degree of genetic influence.* Psychol Med, 2011. **41**(1): p. 33-40.
152. Kerem, B., et al., *Identification of the cystic fibrosis gene: genetic analysis.* Science, 1989. **245**(4922): p. 1073-80.
153. Botstein, D. and N. Risch, *Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease.* Nat Genet, 2003. **33 Suppl**: p. 228-37.
154. Anderson, C.A., et al., *Synthetic associations are unlikely to account for many common disease genome-wide association signals.* PLoS Biol, 2011. **9**(1): p. e1000580.
155. Reich, T., P.J. Clayton, and G. Winokur, *Family history studies: V. The genetics of mania.* Am J Psychiatry, 1969. **125**(10): p. 1358-69.
156. Baron, M., *X-linkage and manic-depressive illness: a reassessment.* Soc Biol, 1991. **38**(3-4): p. 179-88.
157. Mendlewicz, J., et al., *Color blindness linkage to bipolar manic-depressive illness. New evidence.* Arch Gen Psychiatry, 1979. **36**(13): p. 1442-7.

158. Mendlewicz, J., P. Linkowski, and J. Wilmotte, *Linkage between glucose-6-phosphate dehydrogenase deficiency and manic-depressive psychosis.* Br J Psychiatry, 1980. **137**: p. 337-42.
159. Del Zompo, M., et al., *Linkage between X-chromosome markers and manic-depressive illness. Two Sardinian pedigrees.* Acta Psychiatr Scand, 1984. **70**(3): p. 282-7.
160. Baron, M., *Linkage between an X-chromosome marker (deutan color blindness) and bipolar affective illness. Occurrence in the family of a lithium carbonate-responsive schizo-affective proband.* Arch Gen Psychiatry, 1977. **34**(6): p. 721-5.
161. Leckman, J.F., et al., *New data do not suggest linkage between the Xg blood group and bipolar illness.* Arch Gen Psychiatry, 1979. **36**(13): p. 1435-41.
162. Gershon, E.S., et al., *Color blindness not closely linked to bipolar illness. Report of a new pedigree series.* Arch Gen Psychiatry, 1979. **36**(13): p. 1423-30.
163. Mendlewicz, J., et al., *Polymorphic DNA marker on X chromosome and manic depression.* Lancet, 1987. **1**(8544): p. 1230-2.
164. Egeland, J.A., et al., *Bipolar affective disorders linked to DNA markers on chromosome 11.* Nature, 1987. **325**(6107): p. 783-7.
165. Fisher, R., *Statistical Methods for Research Workers*. 5th ed. 1932, Edinburgh: Oliver and Boyd.
166. Gershon, E.S. and J.A. Badner, *Incorporation of molecular data and redefinition of phenotype: new approaches to genetic epidemiology of bipolar manic depressive illness and schizophrenia.* Dialogues Clin Neurosci, 2001. **3**(1): p. 63-71.
167. Badner, J.A. and E.S. Gershon, *Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia.* Mol Psychiatry, 2002. **7**(4): p. 405-11.
168. Levinson, D.F., et al., *Genome scan meta-analysis of schizophrenia and bipolar disorder, part I: Methods and power analysis.* Am J Hum Genet, 2003. **73**(1): p. 17-33.
169. Lewis, C.M., et al., *Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia.* Am J Hum Genet, 2003. **73**(1): p. 34-48.
170. Segurado, R., et al., *Genome scan meta-analysis of schizophrenia and bipolar disorder, part III: Bipolar disorder.* Am J Hum Genet, 2003. **73**(1): p. 49-62.
171. Craddock, N., et al., *Mathematical limits of multilocus models: the genetic transmission of bipolar disorder.* Am J Hum Genet, 1995. **57**(3): p. 690-702.
172. Green, E. and N. Craddock, *Brain-derived neurotrophic factor as a potential risk locus for bipolar disorder: evidence, limitations, and implications.* Curr Psychiatry Rep, 2003. **5**(6): p. 469-76.
173. Craddock, N. and L. Forty, *Genetics of affective (mood) disorders.* Eur J Hum Genet, 2006. **14**(6): p. 660-8.
174. Sklar, P., et al., *Family-based association study of 76 candidate genes in bipolar disorder: BDNF is a potential risk locus. Brain-derived neutrophic factor.* Mol Psychiatry, 2002. **7**(6): p. 579-93.
175. Neves-Pereira, M., et al., *The brain-derived neurotrophic factor gene confers susceptibility to bipolar disorder: evidence from a family-based association study.* Am J Hum Genet, 2002. **71**(3): p. 651-5.
176. Geller, B., et al., *Linkage disequilibrium of the brain-derived neurotrophic factor Val66Met polymorphism in children with a prepubertal and early adolescent bipolar disorder phenotype.* Am J Psychiatry, 2004. **161**(9): p. 1698-700.

177. Oswald, P., et al., *Non-replication of the brain-derived neurotrophic factor (BDNF) association in bipolar affective disorder: a Belgian patient-control study.* Am J Med Genet B Neuropsychiatr Genet, 2004. **129B**(1): p. 34-5.
178. Skibinska, M., et al., *Association analysis of brain-derived neurotrophic factor (BDNF) gene Val66Met polymorphism in schizophrenia and bipolar affective disorder.* World J Biol Psychiatry, 2004. **5**(4): p. 215-20.
179. Hong, C.J., et al., *Association study of a brain-derived neurotrophic-factor genetic polymorphism and mood disorders, age of onset and suicidal behavior.* Neuropsychobiology, 2003. **48**(4): p. 186-9.
180. Nakata, K., et al., *Association study of the brain-derived neurotrophic factor (BDNF) gene with bipolar disorder.* Neurosci Lett, 2003. **337**(1): p. 17-20.
181. Barbosa, I.G., et al., *Increased BDNF levels in long-term bipolar disorder patients.* Rev Bras Psiquiatr, 2013. **35**(1): p. 67-9.
182. Chumakov, I., et al., *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, 2002. **99**(21): p. 13675-80.
183. Hattori, E., et al., *Polymorphisms at the G72/G30 gene locus, on 13q33, are associated with bipolar disorder in two independent pedigree series.* Am J Hum Genet, 2003. **72**(5): p. 1131-40.
184. Chen, Y.S., et al., *Findings in an independent sample support an association between bipolar affective disorder and the G72/G30 locus on chromosome 13q33.* Mol Psychiatry, 2004. **9**(1): p. 87-92; image 5.
185. Schumacher, J., et al., *Examination of G72 and D-amino-acid oxidase as genetic risk factors for schizophrenia and bipolar affective disorder.* Mol Psychiatry, 2004. **9**(2): p. 203-7.
186. Williams, N.M., et al., *Variation at the DAOA/G30 locus influences susceptibility to major mood episodes but not psychosis in schizophrenia and bipolar disorder.* Arch Gen Psychiatry, 2006. **63**(4): p. 366-73.
187. Zuliani, R., et al., *Genetic variation in the G72 (DAOA) gene affects temporal lobe and amygdala structure in subjects affected by bipolar disorder.* Bipolar Disord, 2009. **11**(6): p. 621-7.
188. Berrettini, W.H., *Are schizophrenic and bipolar disorders related? A review of family and molecular studies.* Biol Psychiatry, 2000. **48**(6): p. 531-8.
189. Edenberg, H.J., et al., *Initial genomic scan of the NIMH genetics initiative bipolar pedigrees: chromosomes 3, 5, 15, 16, 17, and 22.* Am J Med Genet, 1997. **74**(3): p. 238-46.
190. Kelsoe, J.R., et al., *A genome survey indicates a possible susceptibility locus for bipolar disorder on chromosome 22.* Proc Natl Acad Sci U S A, 2001. **98**(2): p. 585-90.
191. Detera-Wadleigh, S.D., et al., *A high-density genome scan detects evidence for a bipolar-disorder susceptibility locus on 13q32 and other potential loci on 1q32 and 18p11.2.* Proc Natl Acad Sci U S A, 1999. **96**(10): p. 5604-9.
192. Papolos, D.F., et al., *Bipolar spectrum disorders in patients diagnosed with velo-cardio-facial syndrome: does a hemizygous deletion of chromosome 22q11 result in bipolar affective disorder?* Am J Psychiatry, 1996. **153**(12): p. 1541-7.
193. Grossman, M.H., B.S. Emanuel, and M.L. Budarf, *Chromosomal mapping of the human catechol-O-methyltransferase gene to 22q11.1----q11.2.* Genomics, 1992. **12**(4): p. 822-5.

194. *No association between bipolar disorder and alleles at a functional polymorphism in the COMT gene. Biomed European Bipolar Collaborative Group.* Br J Psychiatry, 1997. **170**: p. 526-8.
195. Kunugi, H., et al., *No evidence for an association of affective disorders with high- or low-activity allele of catechol-o-methyltransferase gene.* Biol Psychiatry, 1997. **42**(4): p. 282-5.
196. Lachman, H.M., et al., *Lack of association of catechol-O-methyltransferase (COMT) functional polymorphism in bipolar affective disorder.* Psychiatr Genet, 1997. **7**(1): p. 13-7.
197. Kirov, G., et al., *Low activity allele of catechol-O-methyltransferase gene associated with rapid cycling bipolar disorder.* Mol Psychiatry, 1998. **3**(4): p. 342-5.
198. Craddock, N., S. Davé, and J. Greening, *Association studies of bipolar disorder.* Bipolar Disord, 2001. **3**(6): p. 284-98.
199. Shifman, S., et al., *A highly significant association between a COMT haplotype and schizophrenia.* Am J Hum Genet, 2002. **71**(6): p. 1296-302.
200. Shifman, S., et al., *COMT: a common susceptibility gene in bipolar disorder and schizophrenia.* Am J Med Genet B Neuropsychiatr Genet, 2004. **128B**(1): p. 61-4.
201. Ozelius, L., et al., *Human monoamine oxidase gene (MAOA): chromosome position (Xp21-p11) and DNA polymorphism.* Genomics, 1988. **3**(1): p. 53-8.
202. Preisig, M., et al., *Association between bipolar disorder and monoamine oxidase A gene polymorphisms: results of a multicenter study.* Am J Psychiatry, 2000. **157**(6): p. 948-55.
203. Brunner, H.G., et al., *Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A.* Science, 1993. **262**(5133): p. 578-80.
204. Black, G.C., et al., *Dinucleotide repeat polymorphism at the MAOA locus.* Nucleic Acids Res, 1991. **19**(3): p. 689.
205. Hinds, H.L., et al., *Characterization of a highly polymorphic region near the first exon of the human MAOA gene containing a GT dinucleotide and a novel VNTR motif.* Genomics, 1992. **13**(3): p. 896-7.
206. Hotamisligil, G.S. and X.O. Breakefield, *Human monoamine oxidase A gene determines levels of enzyme activity.* Am J Hum Genet, 1991. **49**(2): p. 383-92.
207. Lim, L.C., et al., *Evidence for a genetic association between alleles of monoamine oxidase A gene and bipolar affective disorder.* Am J Med Genet, 1995. **60**(4): p. 325-31.
208. Rubinsztein, D.C., et al., *Genetic association between monoamine oxidase A microsatellite and RFLP alleles and bipolar affective disorder: analysis and meta-analysis.* Hum Mol Genet, 1996. **5**(6): p. 779-82.
209. Kawada, Y., et al., *Possible association between monoamine oxidase A gene and bipolar affective disorder.* Am J Hum Genet, 1995. **56**(1): p. 335-6.
210. Eslami Amirabadi, M.R., et al., *Monoamine oxidase a gene polymorphisms and bipolar disorder in Iranian population.* Iran Red Crescent Med J, 2015. **17**(2): p. e23095.
211. Craddock, N., et al., *No evidence for allelic association between bipolar disorder and monoamine oxidase A gene polymorphisms.* Am J Med Genet, 1995. **60**(4): p. 322-4.

212. Nöthen, M.M., et al., *Association analysis of the monoamine oxidase A gene in bipolar affective disorder by using family-based internal controls.* Am J Hum Genet, 1995. **57**(4): p. 975-8.
213. Parsian, A. and R.D. Todd, *Genetic association between monoamine oxidase and manic-depressive illness: comparison of relative risk and haplotype relative risk data.* Am J Med Genet, 1997. **74**(5): p. 475-9.
214. van Praag, H.M., J. Korf, and J. Puite, *5-Hydroxyindoleacetic acid levels in the cerebrospinal fluid of depressive patients treated with probenecid.* Nature, 1970. **225**(5239): p. 1259-60.
215. Briley, M.S., et al., *Tritiated imipramine binding sites are decreased in platelets of untreated depressed patients.* Science, 1980. **209**(4453): p. 303-5.
216. Owens, M.J. and C.B. Nemeroff, *Role of serotonin in the pathophysiology of depression: focus on the serotonin transporter.* Clin Chem, 1994. **40**(2): p. 288-95.
217. Tharoor, H., et al., *Study of the association of serotonin transporter triallelic 5-HTTLPR and STin2 VNTR polymorphisms with lithium prophylaxis response in bipolar disorder.* Psychiatr Genet, 2013. **23**(2): p. 77-81.
218. Uher, R., et al., *Serotonin transporter gene moderates childhood maltreatment's effects on persistent but not single-episode depression: replications and implications for resolving inconsistent results.* J Affect Disord, 2011. **135**(1-3): p. 56-65.
219. Aas, M., et al., *Serotonin transporter gene polymorphism, childhood trauma, and cognition in patients with psychotic disorders.* Schizophr Bull, 2012. **38**(1): p. 15-22.
220. Malhotra, D. and J. Sebat, *CNVs: harbingers of a rare variant revolution in psychiatric genetics.* Cell, 2012. **148**(6): p. 1223-41.
221. Zhang, D., et al., *Singleton deletions throughout the genome increase risk of bipolar disorder.* Mol Psychiatry, 2009. **14**(4): p. 376-80.
222. Bergen, S.E., et al., *Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder.* Mol Psychiatry, 2012. **17**(9): p. 880-6.
223. Priebe, L., et al., *Genome-wide survey implicates the influence of copy number variants (CNVs) in the development of early-onset bipolar disorder.* Mol Psychiatry, 2012. **17**(4): p. 421-32.
224. Malhotra, D., et al., *High frequencies of de novo CNVs in bipolar disorder and schizophrenia.* Neuron, 2011. **72**(6): p. 951-63.
225. Grozeva, D., et al., *Rare copy number variants: a point of rarity in genetic risk for bipolar disorder and schizophrenia.* Arch Gen Psychiatry, 2010. **67**(4): p. 318-27.
226. McQuillin, A., et al., *Analysis of genetic deletions and duplications in the University College London bipolar disorder case control sample.* Eur J Hum Genet, 2011. **19**(5): p. 588-92.
227. Green, E.K., et al., *Copy number variation in bipolar disorder.* Mol Psychiatry, 2016. **21**(1): p. 89-93.
228. Hunter, D.J., D. Altshuler, and D.J. Rader, *From Darwin's finches to canaries in the coal mine--mining the genome for new biology.* N Engl J Med, 2008. **358**(26): p. 2760-3.
229. Pearson, T.A. and T.A. Manolio, *How to interpret a genome-wide association study.* JAMA, 2008. **299**(11): p. 1335-44.

230. Baum, A.E., et al., *A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder.* Mol Psychiatry, 2008. **13**(2): p. 197-207.
231. Consortium, W.T.C.C., *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.* Nature, 2007. **447**(7145): p. 661-78.
232. Ferreira, M.A., et al., *Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder.* Nat Genet, 2008. **40**(9): p. 1056-8.
233. Sklar, P., et al., *Whole-genome association study of bipolar disorder.* Mol Psychiatry, 2008. **13**(6): p. 558-69.
234. Schulze, T.G., *Genetic research into bipolar disorder: the need for a research framework that integrates sophisticated molecular biology and clinically informed phenotype characterization.* Psychiatr Clin North Am, 2010. **33**(1): p. 67-82.
235. Smith, E.N., et al., *Genome-wide association study of bipolar disorder in European American and African American individuals.* Mol Psychiatry, 2009. **14**(8): p. 755-63.
236. Scott, L.J., et al., *Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry.* Proc Natl Acad Sci U S A, 2009. **106**(18): p. 7501-6.
237. Hattori, E., et al., *Preliminary genome-wide association study of bipolar disorder in the Japanese population.* Am J Med Genet B Neuropsychiatr Genet, 2009. **150B**(8): p. 1110-7.
238. Djurovic, S., et al., *A genome-wide association study of bipolar disorder in Norwegian individuals, followed by replication in Icelandic sample.* J Affect Disord, 2010. **126**(1-2): p. 312-6.
239. Sklar, P., et al., *Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4.* Nat Genet, 2011. **43**(10): p. 977-83.
240. Smith, E.N., et al., *Genome-wide association of bipolar disorder suggests an enrichment of replicable associations in regions near genes.* PLoS Genet, 2011. **7**(6): p. e1002134.
241. Yosifova, A., et al., *Genome-wide association study on bipolar disorder in the Bulgarian population.* Genes Brain Behav, 2011. **10**(7): p. 789-97.
242. Cichon, S., et al., *Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder.* Am J Hum Genet, 2011. **88**(3): p. 372-81.
243. Pereira, A.C., et al., *Genetic association and sequencing of the insulin-like growth factor 1 gene in bipolar affective disorder.* Am J Med Genet B Neuropsychiatr Genet, 2011. **156**(2): p. 177-87.
244. Lee, M.T., et al., *Genome-wide association study of bipolar I disorder in the Han Chinese population.* Mol Psychiatry, 2011. **16**(5): p. 548-56.
245. Chen, D.T., et al., *Genome-wide association study meta-analysis of European and Asian-ancestry samples identifies three novel loci associated with bipolar disorder.* Mol Psychiatry, 2011.
246. Mühleisen, T.W., et al., *Genome-wide association study reveals two new risk loci for bipolar disorder.* Nat Commun, 2014. **5**: p. 3339.
247. Xu, W., et al., *Genome-wide association study of bipolar disorder in Canadian and UK populations corroborates disease loci including SYNE1 and CSMD1.* BMC Med Genet, 2014. **15**: p. 2.

248. Schulze, T.G., et al., *Two variants in Ankyrin 3 (ANK3) are independent genetic risk factors for bipolar disorder.* Mol Psychiatry, 2009. **14**(5): p. 487-91.
249. Kuo, P.H., et al., *Identification of novel loci for bipolar I disorder in a multi-stage genome-wide association study.* Prog Neuropsychopharmacol Biol Psychiatry, 2014. **51**: p. 58-64.
250. Dudbridge, F. and A. Gusnanto, *Estimation of significance thresholds for genomewide association scans.* Genet Epidemiol, 2008. **32**(3): p. 227-34.
251. Shea, J., et al., *Comparing strategies to fine-map the association of common SNPs at chromosome 9p21 with type 2 diabetes and myocardial infarction.* Nat Genet, 2011. **43**(8): p. 801-5.
252. Goldstein, D.B., *Common genetic variation and human traits.* N Engl J Med, 2009. **360**(17): p. 1696-8.
253. Park, J.H., et al., *Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants.* Proc Natl Acad Sci U S A, 2011. **108**(44): p. 18026-31.
254. Yang, J., et al., *Common SNPs explain a large proportion of the heritability for human height.* Nat Genet, 2010. **42**(7): p. 565-9.
255. Wang, Z., et al., *Myosin Vb mobilizes recycling endosomes and AMPA receptors for postsynaptic plasticity.* Cell, 2008. **135**(3): p. 535-48.
256. Shin, O.H., et al., *Evolutionarily conserved multiple C2 domain proteins with two transmembrane regions (MCTPs) and unusual Ca2+ binding properties.* J Biol Chem, 2005. **280**(2): p. 1641-51.
257. Devlin, B. and K. Roeder, *Genomic control for association studies.* Biometrics, 1999. **55**(4): p. 997-1004.
258. Reich, D.E. and D.B. Goldstein, *Detecting association in a case-control study while correcting for population stratification.* Genet Epidemiol, 2001. **20**(1): p. 4-16.
259. Zheng, G., B. Freidlin, and J.L. Gastwirth, *Robust genomic control for association studies.* Am J Hum Genet, 2006. **78**(2): p. 350-6.
260. Browning, S.R. and B.L. Browning, *Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering.* Am J Hum Genet, 2007. **81**(5): p. 1084-97.
261. Green, E.K., et al., *Replication of bipolar disorder susceptibility alleles and identification of two novel genome-wide significant associations in a new bipolar disorder case-control sample.* Mol Psychiatry, 2013. **18**(12): p. 1302-7.
262. Watanabe, E., et al., *Distribution of a brain-specific proteoglycan, neurocan, and the corresponding mRNA during the formation of barrels in the rat somatosensory cortex.* Eur J Neurosci, 1995. **7**(4): p. 547-54.
263. Zhou, X.H., et al., *Neurocan is dispensable for brain development.* Mol Cell Biol, 2001. **21**(17): p. 5970-8.
264. van Gorp, W.G., et al., *Cognitive impairment in euthymic bipolar patients with and without prior alcohol dependence. A preliminary study.* Arch Gen Psychiatry, 1998. **55**(1): p. 41-6.
265. Ali, S.O., et al., *A preliminary study of the relation of neuropsychological performance to neuroanatomic structures in bipolar disorder.* Neuropsychiatry Neuropsychol Behav Neurol, 2000. **13**(1): p. 20-8.
266. Secolin, R., et al., *Family-based association study for bipolar affective disorder.* Psychiatr Genet, 2010. **20**(3): p. 126-9.

267. Trejo, J.L., E. Carro, and D.J. Burks, *Experimental models for understanding the role of insulin-like growth factor-I and its receptor during development.* Adv Exp Med Biol, 2005. **567**: p. 27-53.
268. Smoller, J.W., et al., *Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis.* Lancet, 2013. **381**(9875): p. 1371-9.
269. Green, E.K., et al., *The bipolar disorder risk allele at CACNA1C also confers risk of recurrent major depression and of schizophrenia.* Mol Psychiatry, 2010. **15**(10): p. 1016-22.
270. Group, P.G.C.B.D.W., *Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4.* Nat Genet, 2011. **43**(10): p. 977-83.
271. Curtis, D., et al., *Case-case genome-wide association analysis shows markers differentially associated with schizophrenia and bipolar disorder and implicates calcium channel genes.* Psychiatr Genet, 2011. **21**(1): p. 1-4.
272. Hashimoto, K., et al., *Impairment of AMPA receptor function in cerebellar granule cells of ataxic mutant mouse stargazer.* J Neurosci, 1999. **19**(14): p. 6027-36.
273. Kavanagh, D.H., et al., *The ENCODE project: implications for psychiatric genetics.* Mol Psychiatry, 2013. **18**(5): p. 540-2.
274. Park, J.H., et al., *Estimation of effect size distribution from genome-wide association studies and implications for future discoveries.* Nat Genet, 2010. **42**(7): p. 570-5.
275. Consortium, S.W.G.o.t.P.G., *Biological insights from 108 schizophrenia-associated genetic loci.* Nature, 2014. **511**(7510): p. 421-7.
276. Ripke, S., et al., *Genome-wide association analysis identifies 13 new risk loci for schizophrenia.* Nat Genet, 2013. **45**(10): p. 1150-9.
277. Crowley, J.J. and K. Sakamoto, *Psychiatric genomics: outlook for 2015 and challenges for 2020*, W.D.a.L. Duncan, Editor. 2015: Current Opinion in Behavioral Sciences. p. 102-107.
278. Consortium, T.N.a.P.A.S.o.t.P.G., I.I.B.D.G.C. (IIBDGC), and I.I.B.D.G.C. IIBDGC, *Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways.* Nat Neurosci, 2015.
279. Schork, A.J., et al., *All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs.* PLoS Genet, 2013. **9**(4): p. e1003449.
280. Beyene, J., et al., *Pathway-based analysis of a genome-wide case-control association study of rheumatoid arthritis.* BMC Proc, 2009. **3 Suppl 7**: p. S128.
281. Lehner, T., G. Senthil, and A.M. Addington, *Convergence of Advances in Genomics, Team Science, and Repositories as Drivers of Progress in Psychiatric Genomics.* Biol Psychiatry, 2015. **77**(1): p. 6-14.
282. Chen, Y.C., et al., *A hybrid likelihood model for sequence-based disease association studies.* PLoS Genet, 2013. **9**(1): p. e1003224.
283. Cruceanu, C., et al., *Family-based exome-sequencing approach identifies rare susceptibility variants for lithium-responsive bipolar disorder.* Genome, 2013. **56**(10): p. 634-40.
284. Kerner, B., et al., *Rare Genomic Variants Link Bipolar Disorder with Anxiety Disorders to CREB-Regulated Intracellular Signaling Pathways.* Front Psychiatry, 2013. **4**: p. 154.

285. Strauss, K.A., et al., *A population-based study of KCNH7 p.Arg394His and bipolar spectrum disorder.* Hum Mol Genet, 2014. **23**(23): p. 6395-406.
286. Georgi, B., et al., *Genomic view of bipolar disorder revealed by whole genome sequencing in a genetic isolate.* PLoS Genet, 2014. **10**(3): p. e1004229.
287. Fiorentino, A., et al., *Analysis of ANK3 and CACNA1C variants identified in bipolar disorder whole genome sequence data.* Bipolar Disord, 2014.
288. Thomas, G.M. and R.L. Huganir, *MAPK cascade signalling and synaptic plasticity.* Nat Rev Neurosci, 2004. **5**(3): p. 173-83.
289. Chen, G. and H.K. Manji, *The extracellular signal-regulated kinase pathway: an emerging promising target for mood stabilizers.* Curr Opin Psychiatry, 2006. **19**(3): p. 313-23.
290. Xia, Z. and D.R. Storm, *Role of signal transduction crosstalk between adenylyl cyclase and MAP kinase in hippocampus-dependent memory.* Learn Mem, 2012. **19**(9): p. 369-74.
291. Davis, S., et al., *The MAPK/ERK cascade targets both Elk-1 and cAMP response element-binding protein to control long-term potentiation-dependent gene expression in the dentate gyrus in vivo.* J Neurosci, 2000. **20**(12): p. 4563-72.
292. Ament, S.A., et al., *Rare variants in neuronal excitability genes influence risk for bipolar disorder.* Proc Natl Acad Sci U S A, 2015. **112**(11): p. 3576-81.
293. Lee, S., et al., *Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies.* Am J Hum Genet, 2012. **91**(2): p. 224-37.
294. Han, F. and W. Pan, *A data-adaptive sum test for disease association with multiple common or rare variants.* Hum Hered, 2010. **70**(1): p. 42-54.
295. Wu, M.C., et al., *Rare-variant association testing for sequencing data with the sequence kernel association test.* Am J Hum Genet, 2011. **89**(1): p. 82-93.
296. Tanabe, T., et al., *Primary structure of the receptor for calcium channel blockers from skeletal muscle.* Nature, 1987. **328**(6128): p. 313-8.
297. Takahashi, M., et al., *Subunit structure of dihydropyridine-sensitive calcium channels from skeletal muscle.* Proc Natl Acad Sci U S A, 1987. **84**(15): p. 5478-82.
298. Dolphin, A.C., *Calcium channel diversity: multiple roles of calcium channel subunits.* Curr Opin Neurobiol, 2009. **19**(3): p. 237-44.
299. Carafoli, E., et al., *Generation, control, and processing of cellular calcium signals.* Crit Rev Biochem Mol Biol, 2001. **36**(2): p. 107-260.
300. Richards, M., et al., *The HOOK-domain between the SH3 and the GK domains of Cavbeta subunits contains key determinants controlling calcium channel inactivation.* 2007: Channels (Austin).
301. Arikkath, J. and K.P. Campbell, *Auxiliary subunits: essential components of the voltage-gated calcium channel complex.* Curr Opin Neurobiol, 2003. **13**(3): p. 298-307.
302. Berridge, M.J., *Neuronal calcium signaling.* Neuron, 1998. **21**(1): p. 13-26.
303. Bading, H., *Nuclear calcium signalling in the regulation of brain function.* Nat Rev Neurosci, 2013. **14**(9): p. 593-608.
304. Stuart, G.J. and B. Sakmann, *Active propagation of somatic action potentials into neocortical pyramidal cell dendrites.* Nature, 1994. **367**(6458): p. 69-72.
305. Westenbroek, R.E., M.K. Ahlijanian, and W.A. Catterall, *Clustering of L-type Ca2+ channels at the base of major dendrites in hippocampal pyramidal neurons.* Nature, 1990. **347**(6290): p. 281-4.

306. Berridge, M.J. and R.F. Irvine, *Inositol trisphosphate, a novel second messenger in cellular signal transduction.* Nature, 1984. **312**(5992): p. 315-21.
307. Catterall, W.A., *Voltage-gated calcium channels.* Cold Spring Harb Perspect Biol, 2011. **3**(8): p. a003947.
308. Xie, J. and D.L. Black, *A CaMK IV responsive RNA element mediates depolarization-induced alternative splicing of ion channels.* Nature, 2001. **410**(6831): p. 936-9.
309. Hardingham, G.E., et al., *Distinct functions of nuclear and cytoplasmic calcium in the control of gene expression.* Nature, 1997. **385**(6613): p. 260-5.
310. West, A.E., E.C. Griffith, and M.E. Greenberg, *Regulation of transcription factors by neuronal activity.* Nat Rev Neurosci, 2002. **3**(12): p. 921-31.
311. Bading, H. and M.E. Greenberg, *Stimulation of protein tyrosine phosphorylation by NMDA receptor activation.* Science, 1991. **253**(5022): p. 912-4.
312. Bading, H., D.D. Ginty, and M.E. Greenberg, *Regulation of gene expression in hippocampal neurons by distinct calcium signaling pathways.* Science, 1993. **260**(5105): p. 181-6.
313. Hardingham, G.E., F.J. Arnold, and H. Bading, *A calcium microdomain near NMDA receptors: on switch for ERK-dependent synapse-to-nucleus communication.* Nat Neurosci, 2001. **4**(6): p. 565-6.
314. Deisseroth, K., E.K. Heist, and R.W. Tsien, *Translocation of calmodulin to the nucleus supports CREB phosphorylation in hippocampal neurons.* Nature, 1998. **392**(6672): p. 198-202.
315. Shaywitz, A.J. and M.E. Greenberg, *CREB: a stimulus-induced transcription factor activated by a diverse array of extracellular signals.* Annu Rev Biochem, 1999. **68**: p. 821-61.
316. Soderling, T.R., *The Ca-calmodulin-dependent protein kinase cascade.* Trends Biochem Sci, 1999. **24**(6): p. 232-6.
317. Kwok, R.P., et al., *Nuclear protein CBP is a coactivator for the transcription factor CREB.* Nature, 1994. **370**(6486): p. 223-6.
318. Vo, N. and R.H. Goodman, *CREB-binding protein and p300 in transcriptional regulation.* J Biol Chem, 2001. **276**(17): p. 13505-8.
319. Lonze, B.E. and D.D. Ginty, *Function and regulation of CREB family transcription factors in the nervous system.* Neuron, 2002. **35**(4): p. 605-23.
320. Chawla, S., et al., *CBP: a signal-regulated transcriptional coactivator controlled by nuclear calcium and CaM kinase IV.* Science, 1998. **281**(5382): p. 1505-9.
321. Burgoyne, R.D. and J.L. Weiss, *The neuronal calcium sensor family of Ca2+-binding proteins.* Biochem J, 2001. **353**(Pt 1): p. 1-12.
322. Carrión, A.M., et al., *DREAM is a Ca2+-regulated transcriptional repressor.* Nature, 1999. **398**(6722): p. 80-4.
323. Mellström, B. and J.R. Naranjo, *Ca2+-dependent transcriptional repression and derepression: DREAM, a direct effector.* Semin Cell Dev Biol, 2001. **12**(1): p. 59-63.
324. Houdusse, A. and C. Cohen, *Structure of the regulatory domain of scallop myosin at 2 A resolution: implications for regulation.* Structure, 1996. **4**(1): p. 21-32.
325. Ledo, F., et al., *The DREAM-DRE interaction: key nucleotides and dominant negative mutants.* Biochim Biophys Acta, 2000. **1498**(2-3): p. 162-8.

326. Rhoads, A.R. and F. Friedberg, *Sequence motifs for calmodulin recognition.* FASEB J, 1997. **11**(5): p. 331-40.
327. Reuter, H., *Properties of two inward membrane currents in the heart.* Annu Rev Physiol, 1979. **41**: p. 413-24.
328. Tsien, R.W., et al., *Multiple types of neuronal calcium channels and their selective modulation.* Trends Neurosci, 1988. **11**(10): p. 431-8.
329. Carbone, E. and H.D. Lux, *A low voltage-activated, fully inactivating Ca channel in vertebrate sensory neurones.* Nature, 1984. **310**(5977): p. 501-2.
330. Williams, M.E., et al., *Structure and functional expression of an omega-conotoxin-sensitive human N-type calcium channel.* Science, 1992. **257**(5068): p. 389-95.
331. Nowycky, M.C., A.P. Fox, and R.W. Tsien, *Three types of neuronal calcium channel with different calcium agonist sensitivity.* Nature, 1985. **316**(6027): p. 440-3.
332. Olivera, B.M., et al., *Calcium channel diversity and neurotransmitter release: the omega-conotoxins and omega-agatoxins.* Annu Rev Biochem, 1994. **63**: p. 823-67.
333. Mintz, I.M., et al., *P-type calcium channels blocked by the spider toxin omega-Aga-IVA.* Nature, 1992. **355**(6363): p. 827-9.
334. Randall, A. and R.W. Tsien, *Pharmacological dissection of multiple types of Ca2+ channel currents in rat cerebellar granule neurons.* J Neurosci, 1995. **15**(4): p. 2995-3012.
335. Lacerda, A.E., et al., *Normalization of current kinetics by interaction between the alpha 1 and beta subunits of the skeletal muscle dihydropyridine-sensitive Ca2+ channel.* Nature, 1991. **352**(6335): p. 527-30.
336. De Waard, M., M. Pragnell, and K.P. Campbell, *Ca2+ channel regulation by a conserved beta subunit domain.* Neuron, 1994. **13**(2): p. 495-503.
337. Pragnell, M., et al., *Calcium channel beta-subunit binds to a conserved motif in the I-II cytoplasmic linker of the alpha 1-subunit.* Nature, 1994. **368**(6466): p. 67-70.
338. De Waard, M., et al., *Properties of the alpha 1-beta anchoring site in voltage-dependent Ca2+ channels.* J Biol Chem, 1995. **270**(20): p. 12056-64.
339. Witcher, D.R., et al., *Association of native Ca2+ channel beta subunits with the alpha 1 subunit interaction domain.* J Biol Chem, 1995. **270**(30): p. 18088-93.
340. Altier, C., et al., *The Cavβ subunit prevents RFP2-mediated ubiquitination and proteasomal degradation of L-type channels.* Nat Neurosci, 2011. **14**(2): p. 173-80.
341. Waithe, D., et al., *Beta-subunits promote the expression of Ca(V)2.2 channels by reducing their proteasomal degradation.* J Biol Chem, 2011. **286**(11): p. 9598-611.
342. Buraei, Z. and J. Yang, *The ß subunit of voltage-gated Ca2+ channels.* Physiol Rev, 2010. **90**(4): p. 1461-506.
343. Klugbauer, N., et al., *Molecular diversity of the calcium channel alpha2delta subunit.* J Neurosci, 1999. **19**(2): p. 684-91.
344. Wycisk, K.A., et al., *Mutation in the auxiliary calcium-channel subunit CACNA2D4 causes autosomal recessive cone dystrophy.* Am J Hum Genet, 2006. **79**(5): p. 973-7.
345. Jay, S.D., et al., *Structural characterization of the dihydropyridine-sensitive calcium channel alpha 2-subunit and the associated delta peptides.* J Biol Chem, 1991. **266**(5): p. 3287-93.

346. Dolphin, A.C., *The α2δ subunits of voltage-gated calcium channels.* Biochim Biophys Acta, 2013. **1828**(7): p. 1541-9.
347. Walker, D. and M. De Waard, *Subunit interaction sites in voltage-dependent Ca2+ channels: role in channel function.* Trends Neurosci, 1998. **21**(4): p. 148-54.
348. Humphries, M.J., *Integrin structure.* Biochem Soc Trans, 2000. **28**(4): p. 311-39.
349. Cantí, C., et al., *The metal-ion-dependent adhesion site in the Von Willebrand factor-A domain of alpha2delta subunits is key to trafficking voltage-gated Ca2+ channels.* Proc Natl Acad Sci U S A, 2005. **102**(32): p. 11230-5.
350. Chu, P.J., H.M. Robertson, and P.M. Best, *Calcium channel gamma subunits provide insights into the evolution of this gene family.* Gene, 2001. **280**(1-2): p. 37-48.
351. Tomita, S., et al., *Functional studies and distribution define a family of transmembrane AMPA receptor regulatory proteins.* J Cell Biol, 2003. **161**(4): p. 805-16.
352. Kang, M.G., et al., *Biochemical and biophysical evidence for gamma 2 subunit association with neuronal voltage-activated Ca2+ channels.* J Biol Chem, 2001. **276**(35): p. 32917-24.
353. Moss, F.J., et al., *The novel product of a five-exon stargazin-related gene abolishes Ca(V)2.2 calcium channel expression.* EMBO J, 2002. **21**(7): p. 1514-23.
354. Ferron, L., et al., *The stargazin-related protein gamma 7 interacts with the mRNA-binding protein heterogeneous nuclear ribonucleoprotein A2 and regulates the stability of specific mRNAs, including CaV2.2.* J Neurosci, 2008. **28**(42): p. 10604-17.
355. Coombs, I.D. and S.G. Cull-Candy, *Transmembrane AMPA receptor regulatory proteins and AMPA receptor function in the cerebellum.* Neuroscience, 2009. **162**(3): p. 656-65.
356. Vandenberghe, W., R.A. Nicoll, and D.S. Bredt, *Stargazin is an AMPA receptor auxiliary subunit.* Proc Natl Acad Sci U S A, 2005. **102**(2): p. 485-90.
357. Osten, P. and Y. Stern-Bach, *Learning from stargazin: the mouse, the phenotype and the unexpected.* Curr Opin Neurobiol, 2006. **16**(3): p. 275-80.
358. Qiao, X. and H. Meng, *Nonchannel functions of the calcium channel gamma subunit: insight from research on the stargazer mutant.* J Bioenerg Biomembr, 2003. **35**(6): p. 661-70.
359. Kato, A.S., et al., *New transmembrane AMPA receptor regulatory protein isoform, gamma-7, differentially regulates AMPA receptors.* J Neurosci, 2007. **27**(18): p. 4969-77.
360. Din, N., et al., *The function of GluR1 and GluR2 in cerebellar and hippocampal LTP and LTD is regulated by interplay of phosphorylation and O-GlcNAc modification.* J Cell Biochem, 2010. **109**(3): p. 585-97.
361. Ziff, E.B., *TARPs and the AMPA receptor trafficking paradox.* Neuron, 2007. **53**(5): p. 627-33.
362. Tomita, S., et al., *Stargazin modulates AMPA receptor gating and trafficking by distinct domains.* Nature, 2005. **435**(7045): p. 1052-8.
363. Splawski, I., et al., *Ca(V)1.2 calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism.* Cell, 2004. **119**(1): p. 19-31.
364. Paşca, S.P., et al., *Using iPSC-derived neurons to uncover cellular phenotypes associated with Timothy syndrome.* Nat Med, 2011. **17**(12): p. 1657-62.

365. Yarotskyy, V., et al., *Roscovitine binds to novel L-channel (CaV1.2) sites that separately affect activation and inactivation.* J Biol Chem, 2010. **285**(1): p. 43-53.
366. Weiner, M., L. Warren, and J.G. Fiedorowicz, *Cardiovascular morbidity and mortality in bipolar disorder.* Ann Clin Psychiatry, 2011. **23**(1): p. 40-7.
367. Laursen, T.M., et al., *Increased mortality among patients admitted with major psychiatric disorders: a register-based study comparing mortality in unipolar depressive disorder, bipolar affective disorder, schizoaffective disorder, and schizophrenia.* J Clin Psychiatry, 2007. **68**(6): p. 899-907.
368. Brugada, P. and J. Brugada, *Right bundle branch block, persistent ST segment elevation and sudden cardiac death: a distinct clinical and electrocardiographic syndrome. A multicenter report.* J Am Coll Cardiol, 1992. **20**(6): p. 1391-6.
369. Cordeiro, J.M., et al., *Accelerated inactivation of the L-type calcium current due to a mutation in CACNB2b underlies Brugada syndrome.* J Mol Cell Cardiol, 2009. **46**(5): p. 695-703.
370. Antzelevitch, C., et al., *Loss-of-function mutations in the cardiac calcium channel underlie a new clinical entity characterized by ST-segment elevation, short QT intervals, and sudden cardiac death.* Circulation, 2007. **115**(4): p. 442-9.
371. Rajakulendran, S., D. Kaski, and M.G. Hanna, *Neuronal P/Q-type calcium channel dysfunction in inherited disorders of the CNS.* Nat Rev Neurol, 2012. **8**(2): p. 86-96.
372. Roussos, P., et al., *The CACNA1C and ANK3 risk alleles impact on affective personality traits and startle reactivity but not on cognition or gating in healthy males.* Bipolar Disord, 2011. **13**(3): p. 250-9.
373. Jogia, J., et al., *The impact of the CACNA1C gene polymorphism on frontolimbic function in bipolar disorder.* Mol Psychiatry, 2011. **16**(11): p. 1070-1.
374. Bhat, S., et al., *CACNA1C (Cav1.2) in the pathophysiology of psychiatric disease.* Prog Neurobiol, 2012. **99**(1): p. 1-14.
375. Small, S.A., et al., *A pathophysiological framework of hippocampal dysfunction in ageing and disease.* Nat Rev Neurosci, 2011. **12**(10): p. 585-601.
376. Tesli, M., et al., *Association analysis of ANK3 gene variants in nordic bipolar disorder and schizophrenia case-control samples.* Am J Med Genet B Neuropsychiatr Genet, 2011. **156B**(8): p. 969-74.
377. Franke, B., et al., *Genetic variation in CACNA1C, a gene associated with bipolar disorder, influences brainstem rather than gray matter volume in healthy individuals.* Biol Psychiatry, 2010. **68**(6): p. 586-8.
378. Perrier, E., et al., *Initial evidence for the role of CACNA1C on subcortical brain morphology in patients with bipolar disorder.* Eur Psychiatry, 2011. **26**(3): p. 135-7.
379. Kempton, M.J., et al., *Effects of the CACNA1C risk allele for bipolar disorder on cerebral gray matter volume in healthy individuals.* Am J Psychiatry, 2009. **166**(12): p. 1413-4.
380. Tesli, M., et al., *No evidence for association between bipolar disorder risk gene variants and brain structural phenotypes.* J Affect Disord, 2013.
381. Moskvina, V., et al., *Gene-wide analyses of genome-wide association data sets: evidence for multiple common risk alleles for schizophrenia and bipolar disorder and for overlap in genetic risk.* Mol Psychiatry, 2009. **14**(3): p. 252-60.

382. Murakami, M., et al., *Modified behavioral characteristics following ablation of the voltage-dependent calcium channel beta3 subunit.* Brain Res, 2007. **1160**: p. 102-12.
383. Klassen, T., et al., *Exome sequencing of ion channel genes reveals complex profiles confounding personal risk assessment in epilepsy.* Cell, 2011. **145**(7): p. 1036-48.
384. Jan, W.C., et al., *Exploring the associations between genetic variants in genes encoding for subunits of calcium channel and subtypes of bipolar disorder.* J Affect Disord, 2014. **157**: p. 80-6.
385. Van Den Bossche, M.J., et al., *Identification of a CACNA2D4 deletion in late onset bipolar disorder patients and implications for the involvement of voltage-dependent calcium channels in psychiatric disorders.* Am J Med Genet B Neuropsychiatr Genet, 2012. **159B**(4): p. 465-75.
386. Prabhu, S. and I. Pe'er, *Ultrafast genome-wide scan for SNP-SNP interactions in common complex disease.* Genome Res, 2012. **22**(11): p. 2230-40.
387. Letts, V.A., et al., *The mouse stargazer gene encodes a neuronal Ca2+-channel gamma subunit.* Nat Genet, 1998. **19**(4): p. 340-7.
388. Chen, L., et al., *Stargazin regulates synaptic targeting of AMPA receptors by two distinct mechanisms.* Nature, 2000. **408**(6815): p. 936-43.
389. Noebels, J.L., et al., *Stargazer: a new neurological mutant on chromosome 15 in the mouse with prolonged cortical seizures.* Epilepsy Res, 1990. **7**(2): p. 129-35.
390. Lin, Y., *Identification and characterisation of rare CACNG5 genetic variants in bipolar disorder and schizophrenia.*, in *Division of Psychiatry*. 2015, University College London: London, UK.
391. Drummond, J.B., et al., *Transmembrane AMPA receptor regulatory protein (TARP) dysregulation in anterior cingulate cortex in schizophrenia.* Schizophr Res, 2013. **147**(1): p. 32-8.
392. Consortium, T.E.P., *The ENCODE (ENCyclopedia Of DNA Elements) Project.* Science, 2004. **306**(5696): p. 636-640.
393. Birney, E., et al., *Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.* Nature, 2007. **447**(7146): p. 799-816.
394. Felsenfeld, G., et al., *Chromatin structure and gene expression.* Proc Natl Acad Sci U S A, 1996. **93**(18): p. 9384-8.
395. Thurman, R.E., et al., *The accessible chromatin landscape of the human genome.* Nature, 2012. **489**(7414): p. 75-82.
396. Gerstein, M.B., et al., *Architecture of the human regulatory network derived from ENCODE data.* Nature, 2012. **489**(7414): p. 91-100.
397. Dong, X., et al., *Modeling gene expression using chromatin features in various cellular contexts.* Genome Biol, 2012. **13**(9): p. R53.
398. Sanyal, A., et al., *The long-range interaction landscape of gene promoters.* Nature, 2012. **489**(7414): p. 109-13.
399. Maurano, M.T., et al., *Systematic localization of common disease-associated variation in regulatory DNA.* Science, 2012. **337**(6099): p. 1190-5.
400. Barski, A., et al., *High-resolution profiling of histone methylations in the human genome.* Cell, 2007. **129**(4): p. 823-37.
401. Heintzman, N.D., et al., *Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.* Nat Genet, 2007. **39**(3): p. 311-8.

402. Hon, G.C., R.D. Hawkins, and B. Ren, *Predictive chromatin signatures in the mammalian genome.* Hum Mol Genet, 2009. **18**(R2): p. R195-201.
403. Heintzman, N.D., et al., *Histone modifications at human enhancers reflect global cell-type-specific gene expression.* Nature, 2009. **459**(7243): p. 108-12.
404. Dunham, I., et al., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.
405. Endicott, J. and R.L. Spitzer, *A diagnostic interview: the schedule for affective disorders and schizophrenia.* Arch Gen Psychiatry, 1978. **35**(7): p. 837-44.
406. McGuffin, P., A. Farmer, and I. Harvey, *A polydiagnostic application of operational criteria in studies of psychotic illness. Development and reliability of the OPCRIT system.* Arch Gen Psychiatry, 1991. **48**(8): p. 764-70.
407. Association, A.P., *Diagnostic and Statistical Manual of Mental Disorders, 3rd Edition, (DSM-III)*. 1980, American Psychiatric Press, Washington, DC.
408. Sharp, S.I., et al., *Genetic association of the tachykinin receptor 1 TACR1 gene in bipolar disorder, attention deficit hyperactivity disorder, and the alcohol dependence syndrome.* Am J Med Genet B Neuropsychiatr Genet, 2014. **165B**(4): p. 373-80.
409. Fiorentino, A., S.I. Sharp, and A. McQuillin, *Association of rare variation in the glutamate receptor gene SLC1A2 with susceptibility to bipolar disorder and schizophrenia.* Eur J Hum Genet, 2014.
410. Consortium, E.A. [cited 2015 9th April]; Available from: http://exac.broadinstitute.org.
411. Purcell, S.M., et al., *A polygenic burden of rare disruptive mutations in schizophrenia.* Nature, 2014. **506**(7487): p. 185-90.
412. Koressaar, T. and M. Remm, *Enhancements and modifications of primer design program Primer3.* Bioinformatics, 2007. **23**(10): p. 1289-91.
413. Untergasser, A., et al., *Primer3--new capabilities and interfaces.* Nucleic Acids Res, 2012. **40**(15): p. e115.
414. Fernandez-Rachubinski, F., et al., *Incorporation of 7-deaza dGTP during the amplification step in the polymerase chain reaction procedure improves subsequent DNA sequencing.* DNA Seq, 1990. **1**(2): p. 137-40.
415. Graham, F.L., et al., *Characteristics of a human cell line transformed by DNA from human adenovirus type 5.* J Gen Virol, 1977. **36**(1): p. 59-74.
416. Wittwer, C.T., et al., *High-resolution genotyping by amplicon melting analysis using LCGreen.* Clin Chem, 2003. **49**(6 Pt 1): p. 853-60.
417. Staden, R., *The Staden sequence analysis package.* Mol Biotechnol, 1996. **5**(3): p. 233-41.
418. Schug, J. and G.C. Overton, *Modeling transcription factor binding sites with Gibbs Sampling and Minimum Description Length encoding.* Proc Int Conf Intell Syst Mol Biol, 1997. **5**: p. 268-71.
419. Aslanidis, C. and P.J. de Jong, *Ligation-independent cloning of PCR products (LIC-PCR).* Nucleic Acids Res, 1990. **18**(20): p. 6069-74.
420. Kopec, C.D., et al., *Glutamate receptor exocytosis and spine enlargement during chemically induced long-term potentiation.* J Neurosci, 2006. **26**(7): p. 2000-9.
421. Everts, I., C. Villmann, and M. Hollmann, *N-Glycosylation is not a prerequisite for glutamate receptor function but Is essential for lectin modulation.* Mol Pharmacol, 1997. **52**(5): p. 861-73.
422. Higy, M., T. Junne, and M. Spiess, *Topogenesis of membrane proteins at the endoplasmic reticulum.* Biochemistry, 2004. **43**(40): p. 12716-22.

423. Miesenböck, G., D.A. De Angelis, and J.E. Rothman, *Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins.* Nature, 1998. **394**(6689): p. 192-5.
424. Sankaranarayanan, S., et al., *The use of pHluorins for optical measurements of presynaptic activity.* Biophys J, 2000. **79**(4): p. 2199-208.
425. Ashby, M.C., K. Ibaraki, and J.M. Henley, *It's green outside: tracking cell surface proteins with pH-sensitive GFP.* Trends Neurosci, 2004. **27**(5): p. 257-61.
426. Hellman, L.M. and M.G. Fried, *Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions.* Nat Protoc, 2007. **2**(8): p. 1849-61.
427. Musso, M., et al., *Betaine, dimethyl sulfoxide, and 7-deaza-dGTP, a powerful mixture for amplification of GC-rich DNA sequences.* J Mol Diagn, 2006. **8**(5): p. 544-50.
428. Bernstein, H.S. and S.R. Coughlin, *Pombe Cdc5-related protein. A putative human transcription factor implicated in mitogen-activated signaling.* J Biol Chem, 1997. **272**(9): p. 5833-7.
429. Glass, C.K. and M.G. Rosenfeld, *The coregulator exchange in transcriptional functions of nuclear receptors.* Genes Dev, 2000. **14**(2): p. 121-41.
430. Kim, K., et al., *Domains of estrogen receptor alpha (ERalpha) required for ERalpha/Sp1-mediated activation of GC-rich promoters by estrogens and antiestrogens in breast cancer cells.* Mol Endocrinol, 2003. **17**(5): p. 804-17.
431. Wang, W., et al., *Transcriptional activation of E2F1 gene expression by 17beta-estradiol in MCF-7 cells is regulated by NF-Y-Sp1/estrogen receptor interactions.* Mol Endocrinol, 1999. **13**(8): p. 1373-87.
432. Dyson, N., *The regulation of E2F by pRB-family proteins.* Genes Dev, 1998. **12**(15): p. 2245-62.
433. O'Connor, R.J., et al., *The p107 tumor suppressor induces stable E2F DNA binding to repress target promoters.* Oncogene, 2001. **20**(15): p. 1882-91.
434. Stülke, J., et al., *PRD--a protein domain involved in PTS-dependent induction and carbon catabolite repression of catabolic operons in bacteria.* Mol Microbiol, 1998. **28**(5): p. 865-74.
435. **Lester, J.T.**, *Interactions between ICP4 and the Cellular Transcription Machinery that Mediate HSV-1 Gene Expression*, in *School of Medicine*. 2009, University of Pittsburgh.
436. Massari, M.E., et al., *Characterization of ABF-1, a novel basic helix-loop-helix transcription factor expressed in activated B lymphocytes.* Mol Cell Biol, 1998. **18**(6): p. 3130-9.
437. Perez-Villamil, B., P.T. Schwartz, and M. Vallejo, *The pancreatic homeodomain transcription factor IDX1/IPF1 is expressed in neural cells during brain development.* Endocrinology, 1999. **140**(8): p. 3857-60.
438. Schwartz, P.T., et al., *Pancreatic homeodomain transcription factor IDX1/IPF1 expressed in developing brain regulates somatostatin gene transcription in embryonic neural cells.* J Biol Chem, 2000. **275**(25): p. 19106-14.
439. Gordon, S., et al., *Transcription factor YY1: structure, function, and therapeutic implications in cancer biology.* Oncogene, 2006. **25**(8): p. 1125-42.
440. Donohoe, M.E., et al., *Targeted disruption of mouse Yin Yang 1 transcription factor results in peri-implantation lethality.* Mol Cell Biol, 1999. **19**(10): p. 7237-44.

441. Yang, W.M., et al., *Isolation and characterization of cDNAs corresponding to an additional member of the human histone deacetylase gene family.* J Biol Chem, 1997. **272**(44): p. 28001-7.
442. Yang, W.M., et al., *Transcriptional repression by YY1 is mediated by interaction with a mammalian homolog of the yeast global regulator RPD3.* Proc Natl Acad Sci U S A, 1996. **93**(23): p. 12845-50.
443. Goldstein, D.B., et al., *Sequencing studies in human genetics: design and interpretation.* Nat Rev Genet, 2013. **14**(7): p. 460-70.
444. Kryukov, G.V., L.A. Pennacchio, and S.R. Sunyaev, *Most rare missense alleles are deleterious in humans: implications for complex disease and association studies.* Am J Hum Genet, 2007. **80**(4): p. 727-39.
445. Zuk, O., et al., *Searching for missing heritability: designing rare variant association studies.* Proc Natl Acad Sci U S A, 2014. **111**(4): p. E455-64.
446. Kato, A.S., et al., *TARPs differentially decorate AMPA receptors to specify neuropharmacology.* Trends Neurosci, 2010. **33**(5): p. 241-8.
447. Fukaya, M., et al., *Spatial diversity in gene expression for VDCCgamma subunit family in developing and adult mouse brains.* Neurosci Res, 2005. **53**(4): p. 376-83.
448. Nomura, T., et al., *Cerebellar long-term depression requires dephosphorylation of TARP in Purkinje cells.* Eur J Neurosci, 2012. **35**(3): p. 402-10.
449. Bowie, D. and M.L. Mayer, *Inward rectification of both AMPA and kainate subtype glutamate receptors generated by polyamine-mediated ion channel block.* Neuron, 1995. **15**(2): p. 453-62.
450. Kamboj, S.K., G.T. Swanson, and S.G. Cull-Candy, *Intracellular spermine confers rectification on rat calcium-permeable AMPA and kainate receptors.* J Physiol, 1995. **486 (Pt 2)**: p. 297-303.
451. Feldmeyer, D., et al., *Neurological dysfunctions in mice expressing different levels of the Q/R site-unedited AMPAR subunit GluR-B.* Nat Neurosci, 1999. **2**(1): p. 57-64.
452. Swanson, G.T., S.K. Kamboj, and S.G. Cull-Candy, *Single-channel properties of recombinant AMPA receptors depend on RNA editing, splice variation, and subunit composition.* J Neurosci, 1997. **17**(1): p. 58-69.
453. Soto, D., et al., *Stargazin attenuates intracellular polyamine block of calcium-permeable AMPA receptors.* Nat Neurosci, 2007. **10**(10): p. 1260-7.
454. Kott, S., et al., *Comparative analysis of the pharmacology of GluR1 in complex with transmembrane AMPA receptor regulatory proteins gamma2, gamma3, gamma4, and gamma8.* Neuroscience, 2009. **158**(1): p. 78-88.
455. Milstein, A.D. and R.A. Nicoll, *Regulation of AMPA receptor gating and pharmacology by TARP auxiliary subunits.* Trends Pharmacol Sci, 2008. **29**(7): p. 333-9.
456. Chen, L., et al., *Impaired cerebellar synapse maturation in waggler, a mutant mouse with a disrupted neuronal calcium channel gamma subunit.* Proc Natl Acad Sci U S A, 1999. **96**(21): p. 12132-7.
457. Zhang, W., et al., *Structural basis of arc binding to synaptic proteins: implications for cognitive disease.* Neuron, 2015. **86**(2): p. 490-500.
458. Link, W., et al., *Somatodendritic expression of an immediate early gene is regulated by synaptic activity.* Proc Natl Acad Sci U S A, 1995. **92**(12): p. 5734-8.

459. Guzowski, J.F., et al., *Environment-specific expression of the immediate-early gene Arc in hippocampal neuronal ensembles.* Nat Neurosci, 1999. **2**(12): p. 1120-4.
460. Fromer, M., et al., *De novo mutations in schizophrenia implicate synaptic networks.* Nature, 2014. **506**(7487): p. 179-84.
461. Venselaar, H., et al., *Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces.* BMC Bioinformatics, 2010. **11**: p. 548.
462. Chen, R.S., et al., *Calcium channel gamma subunits: a functionally diverse protein family.* Cell Biochem Biophys, 2007. **47**(2): p. 178-86.
463. O'Brien, N.L., et al., *The functional GRM3 Kozak sequence variant rs148754219 affects the risk of schizophrenia and alcohol dependence as well as bipolar disorder.* Psychiatr Genet, 2014. **24**(6): p. 277-8.
464. Zhao, W., et al., *Massively parallel functional annotation of 3' untranslated regions.* Nat Biotechnol, 2014. **32**(4): p. 387-91.
465. Madison, J.M., et al., *Characterization of bipolar disorder patient-specific induced pluripotent stem cells from a family reveals neurodevelopmental and mRNA expression abnormalities.* Mol Psychiatry, 2015. **20**(6): p. 703-17.
466. Cloninger, R.S., S.Bohman, M., *Type I and Type II Alcoholism: An Update*. 1996: Alcohol Health & Research World.
467. Coyle, J.T. and R.S. Duman, *Finding the intracellular signaling pathways affected by mood disorder treatments.* Neuron, 2003. **38**(2): p. 157-60.
468. Du, J., et al., *Structurally dissimilar antimanic agents modulate synaptic plasticity by regulating AMPA glutamate receptor subunit GluR1 synaptic expression.* Ann N Y Acad Sci, 2003. **1003**: p. 378-80.
469. Du, J., et al., *Modulation of synaptic plasticity by antimanic agents: the role of AMPA glutamate receptor subunit 1 synaptic expression.* J Neurosci, 2004. **24**(29): p. 6578-89.
470. McQuillin, A., M. Rizig, and H.M. Gurling, *A microarray gene expression study of the molecular pharmacology of lithium carbonate on mouse brain mRNA to understand the neurobiology of mood stabilization and treatment of bipolar affective disorder.* Pharmacogenet Genomics, 2007. **17**(8): p. 605-17.
471. Schwenk, J., et al., *Regional diversity and developmental dynamics of the AMPA-receptor proteome in the mammalian brain.* Neuron, 2014. **84**(1): p. 41-54.
472. Shelley, C., et al., *Influence Of The TARP Gamma-4 On Homomeric GluR1 AMPAR Channels*. 2009: Poster, Biophysical Society 53rd Annual Meeting.
473. Tomita, S., et al., *Bidirectional synaptic plasticity regulated by phosphorylation of stargazin-like TARPs.* Neuron, 2005. **45**(2): p. 269-77.
474. Nicoll, R.A., S. Tomita, and D.S. Bredt, *Auxiliary subunits assist AMPA-type glutamate receptors.* Science, 2006. **311**(5765): p. 1253-6.
475. Consortium, E.P., *The ENCODE (ENCyclopedia Of DNA Elements) Project.* Science, 2004. **306**(5696): p. 636-40.
476. Creyghton, M.P., et al., *Histone H3K27ac separates active from poised enhancers and predicts developmental state.* Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21931-6.
477. Rada-Iglesias, A., et al., *A unique chromatin signature uncovers early developmental enhancers in humans.* Nature, 2011. **470**(7333): p. 279-83.
478. Curtis, D., *Consideration of plausible genetic architectures for schizophrenia and implications for analytic approaches in the era of next generation sequencing.* Psychiatr Genet, 2013. **23**(1): p. 1-10.

479. Liu, Y., et al., *Meta-analysis of genome-wide association data of bipolar disorder and major depressive disorder.* Mol Psychiatry, 2011. **16**(1): p. 2-4.
480. Gaspar, P., et al., *mRNA secondary structure optimization using a correlated stem-loop prediction.* Nucleic Acids Res, 2013. **41**(6): p. e73.
481. de Smit, M.H. and J. van Duin, *Control of translation by mRNA secondary structure in Escherichia coli. A quantitative analysis of literature data.* J Mol Biol, 1994. **244**(2): p. 144-50.
482. Hall, M.N., et al., *A role for mRNA secondary structure in the control of translation initiation.* Nature, 1982. **295**(5850): p. 616-8.
483. Zuker, M. and P. Stiegler, *Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information.* Nucleic Acids Res, 1981. **9**(1): p. 133-48.
484. Mathews, D.H., et al., *Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure.* J Mol Biol, 1999. **288**(5): p. 911-40.
485. Mathews, D.H., *Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization.* RNA, 2004. **10**(8): p. 1178-90.
486. Ding, Y., C.Y. Chan, and C.E. Lawrence, *RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble.* RNA, 2005. **11**(8): p. 1157-66.
487. Totoki, Y., et al., *Homo sapiens protein coding cDNA*. 2005: Published Only in Database (2005).
488. Strausberg, R.L., et al., *Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.* Proc Natl Acad Sci U S A, 2002. **99**(26): p. 16899-903.
489. Soldatov, N.M., *Molecular diversity of L-type Ca2+ channel transcripts in human fibroblasts.* Proc Natl Acad Sci U S A, 1992. **89**(10): p. 4628-32.
490. Giannakis, G., et al., *Aberrant cryptic responsiveness of the pCAT 3- and pGL3-promoter reporter vectors.* Biotechniques, 2003. **35**(2): p. 332-9.
491. Relle, M., et al., *Intronic promoters and their noncoding transcripts: a new source of cancer-associated genes.* 2014, Molecular Carcinogenesis. p. pg 117-124.
492. Emamghoreishi, M., et al., *High intracellular calcium concentrations in transformed lymphoblasts from subjects with bipolar I disorder.* Am J Psychiatry, 1997. **154**(7): p. 976-82.
493. Uemura, T., et al., *Bcl-2 SNP rs956572 associates with disrupted intracellular calcium homeostasis in bipolar I disorder.* Bipolar Disord, 2011. **13**(1): p. 41-51.
494. Hsia, A.Y., et al., *Plaque-independent disruption of neural circuits in Alzheimer's disease mouse models.* Proc Natl Acad Sci U S A, 1999. **96**(6): p. 3228-33.
495. Hsieh, H., et al., *AMPAR removal underlies Abeta-induced synaptic depression and dendritic spine loss.* Neuron, 2006. **52**(5): p. 831-43.
496. Bedoukian, M.A., A.M. Weeks, and K.M. Partin, *Different domains of the AMPA receptor direct stargazin-mediated trafficking and stargazin-mediated modulation of kinetics.* J Biol Chem, 2006. **281**(33): p. 23908-21.
497. Sibley, C.R., et al., *Recursive splicing in long vertebrate genes.* Nature, 2015. **521**(7552): p. 371-5.

Appendix I

The tables below outline the classifications of depression, mania, hypomania with and without psychosis as defined by the ICD-10 manual for clinical descriptions and diagnostic guidelines. In addition to the eight categories of bipolar disorder outlined below two more are present in the ICD-10, F31.8 other bipolar affective disorders and F31.9 Bipolar affective disorders, unspecified. These are not included in the table as there are no symptom guidelines for these subtypes.

F30.0 Hypomania	F30.1 Mania without psychotic symptoms	F30.2 Mania with psychotic symptoms
A.The mood is elevated or irritable to a degree that is definitely abnormal for the individual concerned and sustained for at least four consecutive days.	A. A mood which is predominantly elevated, expansive or irritable and definitely abnormal for the individual concerned. This mood change must be prominent and sustained for at least a week (unless it is severe enough to require hospital admission).	A. The episode meets the criteria for mania without psychotic symptoms (F30.1)
B. At least three of the following must be present, leading to some interference with personal functioning in daily living:	B. At least three of the following must be present (four if the mood is merely irritable), leading to severe interference with personal functioning in daily living:	B. The episode does not simultaneously meet the criteria for schizophrenia (F20) or schizo-affective disorder, manic type (F25.0).
(1) increased activity or physical restlessness;	(1) Increased activity or physical restlessness;	C. Delusions or hallucinations are present, other than those listed as typical schizophrenic in F20 G1.1b, c and d in the ICD-10 (i.e. delusions other than those that are completely impossible or culturally inappropriate and hallucinations, that are not in the third person or giving a running commentary). The commonest examples are those with grandiose, self-referential, erotic or persecutory content.
(2) increased talkativeness;	(2) Increased talkativeness ('pressure of speech');	
(3) difficulty in concentration or distractibility;	(3) Flight of ideas or the subjective experience of thoughts racing;	
(4) decreased need for sleep;	(4) Loss of normal social inhibitions resulting in behaviour which is inappropriate to the circumstances;	
(5) increased sexual energy;	(5) Decreased need for sleep;	
(6) mild spending sprees, or other types of reckless or irresponsible behaviour;	(6) Inflated self-esteem or grandiosity;	
	(7) Distractibility or constant changes in activity or plans;	
(7) increased sociability or over-familiarity.	(8) Behaviour which is foolhardy or reckless and whose risks the subject does not recognize e.g. spending sprees, foolish enterprises, reckless driving;	
	(9) Marked sexual energy or sexual indiscretions.	

Table 42. ICD-10 criteria for categorising a manic episode [7].

F32 .0 Mild Depressive episodes	**F32.1 Moderate Depressive episodes**	**F32.2 Severe Depressive episodes without psychotic symtpoms**	**F32.3 Severe depressive episode with psychotic symptoms**
A. The depressive episode should last for at least 2 weeks, and meet the criteria set out in F32 for general depressive episode	A. The general criteria for depressive episode (F32) must be met.	The general criteria for depressive episode (F32) must be met.	A. The general criteria for depressive episode (F32) must be met.
B. At least two of the following three symptoms must be present: 1)A depressed mood to a degree that is definitely abnormal for the individual, present for most of the day and almost every day, largely uninfluenced by circumstances, and sustained for at least 2 weeks. (2) loss of interest or pleasure in activities that are normally pleasurable; (3) decreased energy or increased fatigability.	B. At least two of the three symptoms listed for F32.0, criterion B, must be present. C. Additional symptoms from F32.0, criterion C, must be present, to give a total of at least six.	B. All three of the symptoms in criterion B, F32.0, must be present. C. Additional symptoms from F32.0, criterion C, must be present, to give a total of at least eight. D. There must be no hallucinations, delusions, or depressive stupor	B. The criteria for severe depressive episode without psychotic symptoms (F32.2) must be met with the exception of criterion D. C. The criteria for schizophrenia (F20.-) or schizoaffective disorder, depressive type (F25.1) are not met.

Table 43. ICD-10 criteria for categorising a depressive episode [7].

F32 .0 Mild Depressive episodes	F32.1 Moderate Depressive episodes	F32.2 Severe Depressive episodes without psychotic symtpoms	F32.3 Severe depressive episode with psychotic symptoms
C. An additional symptom or symptoms from the following list should be present, to give a total of at least four: (1) loss of confidence and self-esteem; (2) unreasonable feelings of self-reproach or excessive and inappropriate guilt; (3) recurrent thoughts of death or suicide, or any suicidal behaviour; (4) complaints or evidence of diminished ability to think or concentrate, such as indecisiveness or vacillation; (5) change in psychomotor activity, with agitation or retardation (either subjective or objective); (6) sleep disturbance of any type;			D. Either of the following must be present: (1) delusions or hallucinations, other than those listed as typically schizophrenic in F20, criterion G1(1)b, c, and d (i.e. delusions other than those that completely impossible or culturally inappropriate and hallucinations that are not in the third person or giving a running commentary); the commonest examples are those with depressive, guilty, hypochondriacal, nihilistic, self-referential, or persecutory content; (2) depressive stupor.

Table 44. ICD-10 criteria for categorising a depressive episode [1].

F31.0 Bipolar affective disorder, current episode hypomanic	**F31.1 Bipolar affective disorder, current episode manic without psychotic symptoms**	**F31.2 Bipolar affective disorder, current episode manic with psychotic symptoms**	**F31.3 Bipolar affective disorder, current episode moderate or mild depression**
A. The current episode meets the criteria for hypomania (F30.0).	A. The current episode meets the criteria for mania without psychotic symptoms (F30.1).	A. The current episode meets the criteria for mania with psychotic symptoms (F30.2).	A. The current episode meets the criteria for a depressive episode of either mild (F32.0) or moderate severity (F32.1).
B. There has been at least one other affective episode in the past, meeting the criteria for hypomanic or manic episode (F30.-), depressive episode (F32.-) or mixed affective episode (F38.00).	B. There has been at least one other affective episode in the past, meeting the criteria for hypomanic or manic episode (F30.-), depressive episode (F32.-) or mixed affective episode (F38.00).	B. There has been at least one other affective episode in the past, meeting the criteria for hypomanic or manic episode (F30.-), depressive episode (F32.-) or mixed affective episode (F38.00). A fifth character may be used to specify whether the psychotic symptoms are congruent or incongruent with the mood: F31.20 with mood congruent psychotic symptoms. F31.21 with mood incongruent psychotic symptoms	B. There has been at least one other affective episode in the past, meeting the criteria for hypomanic or manic episode (F30.-), or mixed affective episode (F38.00).

Table 45. ICD-10 classifications of Bipolar Disorder.

31.4 Bipolar affective disorder, current episode severe depression without psychotic symptoms	**F31.5 Bipolar affective disorder, current episode severe depression with psychotic symptoms**	**F31.6 Bipolar affective disorder, current episode mixed**	**F31.7 Bipolar affective disorder, currently in remission**
A. The current episode meets the criteria for a severe depressive episode without psychotic symptoms (F32.2).	A. The current episode meets the criteria for a severe depressive episode with psychotic symptoms (F32.3).	A. The current episode is characterized by either a mixture or a rapid alternation (i.e. within a few hours) of hypomanic, manic and depressive symptoms.	A. The current state does not meet the criteria for depressive or manic episode in any severity, or for any other mood disorder in F3 (possibly because of treatment to reduce the risk of future episodes).
B. There has been at least one well authenticated hypomanic or manic episode (F30.-) or mixed affective episode (F38.00) in the past.	B. There has been at least one well authenticated hypomanic or manic episode (F30.-) or mixed affective episode (F38.00) in the past.	B. Both manic and depressive symptoms must be prominent most of the time during a period of at least two weeks.	B. There has been at least one well authenticated hypomanic or manic episode
	A fifth character may be used to specify whether the psychotic symptoms are congruent or incongruent with the mood: F31.50 with mood congruent psychotic symptoms. F31.51 with mood incongruent psychotic symptoms	C. There has been at least one well authenticated hypomanic or manic episode (F30.-), depressive (F32.-) or mixed affective episode (F38.00) in the past.	(F30.-) in the past and in addition at least one other affective episode (hypomanic or manic (F30.-), depressive (F32.-), or mixed (F38.00)).

Table 46. ICD-10 classifications of Bipolar Disorder.

F38.00 Mixed affective episode
A. The episode is characterized by either a mixture or a rapid alternation (i.e. within a few hours) of hypomanic, manic and depressive symptoms.
B. Both manic and depressive symptoms must be prominent most of the time during a period of at least two weeks.
C. No previous hypomanic, depressive or mixed episodes.

Table 47. ICD-10 classification of mixed affective episode

Appendix II

Buffers and Reagents for Materials and Methods Section

II.I DNA EXTRACTION FROM BLOOD SAMPLES

10X RBC Lysis Buffer: 100mM NaCl, 100mM EDTA. For 1 litre: 80.2g NH4Cl; 8.40g NaHCO3 3.7g EDTA. Make up to 1 litre. Store the RBC at 4°C for 6 months.

Cell Lysis Solution: 10 mM Tris-HCl pH 8.0, 25 mM EDA, 0.5% SDS. For 1 litre, 10 ml 1M Tris-HCL, 50 ml 0.5M EDTA, 25ml 4M 20% SDS and made up to one litre using deionised water.

Proteinase K enzyme (20 mg/ml) (Sigma-Aldrich, UK)

Proteinase K buffer: 50mM Tris.HCl (pH8.0), 50mM EDTA, 100mM NaCl. For one litre, 25ml 2M Tris, 100ml 0.5M EDTA, and 25ml 4M NaCl and made up to one litre using deionised water.

Protein Precipitation solution: 5M Ammonium Acetate (pH5.2). For one litre, dissolve 385.4 g of ammonium acetate in one litre of deionised water. 70% ethanol (Sigma-Aldrich, UK)

Low EDTA TE: 10 mM Tris, 0.1 mM EDTA. For one litre add 20ml 0.5M Tris-HCL, (pH8.0) and 200µl 0.5M EDTA (pH8.0) to 979.80ml of deionised water.

II.II AGROSE GELS

1% Agrose Gel: 0.9g of Agrose dissolved in 90ml of 1x TBE Buffer (Sigma-Aldrich, UK)

Ethidium Bromide (Sigma-Aldrich, UK)

Gel Buffer: 500ml of 1x TBE Buffer.

Gel Combs

II.III CELL CULTURE

Complete growth medium: Dulbecco's modified Eagle's medium (DMEM, Sigma-Aldrich, UK) supplemented with 10% FBS (Sigma-Aldrich, UK) and 1% Penicllin – Streptomycin (Sigma-Aldrich, UK).

Trypsin-EDTA solution (Sigma-Aldrich, UK)

1 X Phosphate buffered saline (PBS) (Sigma-Aldrich, UK)

Freezing medium: 1ml of DMEM supplemented with 20% FBS, 10% DMSO and 1% PenStrep

II.IV TRANSIENT TRANSFECTION

Lipofectamine 2000 (Invitrogen Life technologies, UK)

Dulbecco's modified Eagle's medium (DMEM, Sigma-Aldrich, UK) supplemented with 10% FBS (Sigma-Aldrich, UK) and 1% Penicllin–Streptomycin (Sigma-Aldrich, UK).

Opti-MEM® I Reduced Serum Medium (Thermo Fisher Scientific, UK)

II.V PURIFICATION OF PCR PRODUCTS

MicroCLEAN: 0.5 M NaCl, 1 mM Tris HCl pH 8.0, 0.1 mM EDTA, 20 % w/v PEG8000, 1.75 mM MgCl2. All the contents were added to 50 ml of PCR

water and the solution was heated gently for the contents to dissolve. The solution was filter sterilized using a 0.45 µM filter.

II.VI MINI AND MIDI PREP

Library Efficiency® DH5α™ Competent Cells (Life Technologies, UK)

LB Broth: For one litre add 10g NaCL, 10g Trypton and 5g Yeast to 800ml of deionised water. Once all of the reagents have dissolved add deionised water up to 1 litre. Autoclave and store at 4°C.

LB Agar: For one litre add 10g NaCL, 10g Trypton and 5g Yeast to 800ml of deionised water. Once all of the reagents have dissolved add deionised water up to 1 litre. Add 20g Agarose and autoclave.

SOC Broth (Life Technologies, UK)

Kanamycin (50µg/ml): Add 0.5g of sodium ampicillin in sufficient H2O to make a final volume of 10ml to make a stock solution of 100mg/ml. Sterilise using a 0.22µm filter and aliquot out. Aliquots should be stored at -20°C. Dilute this 1:100 when adding to LB Broth or Agar for a final concentration of 50µg/ml

Ampicillin (100µg/ml): Add 1g of sodium ampicillin in sufficient H2O to make a final volume of 10ml to make a stock solution of 100mg/ml. Sterilise using a 0.22µm filter and aliquot out. Aliquots should be stored at -20°C. Dilute this 1:100 when adding to LB Broth or Agar for a final concentration of 100µg/ml

II.VII EMSA

Cytoplasmic Lysis buffer: 10mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.6, 1 mM EDTA (ethylene diamine tetraacetic acid), 0.1 mM EGTA (ethylene glycol tetraacetic acid),

10 nM KCl, 1 mM dithiothreitol (DTT), 1 mM vanadate, 1 tablet of 'complete' protease inhibitor (Roche Diagnostics Ltd, UK).

Nuclear Lysis buffer: 20mM HEPES pH 7.6, 0.2mM EDTA, 0.1mM EGTA, 25% glycerol, 0.42M NaCl, 1mM DTT, 1mM vanadate and 1/2 tablet 'complete' protease inhibitor.

Nonidet P-40 (NP-40, Sigma-Aldrich, UK)

1 X Phosphate buffered saline (PBS) (Sigma-Aldrich, UK)
The cytoplasmic and the nuclear lysis buffers were filter sterilized using 0.2 µm membrane filter and stored at 4°C.

4% Polyacrylamide Gel: 5ml of 40% Acrylamide/Bis-acrylamide solution (Sigma-Aldrich, UK), 2.5ml 10X TBE, 80µl TEMED (Tetramethylethylenediamine), 300µl of 10 % APS (w/v) and 42.12ml of deionised water.

Binding buffer: 2µl of 10x Parker Buffer (40mM HEPES (pH7.6), 100mM of KCL, 2mM EDTA, 1mM DTT, (pH 7.5) and deionised H20), 2µl 2.5mM DTT, 1µl 50ng/µl of Poly(dI.dC) (Sigma-Aldrich, UK), 1µl of 5mM MgCl2, 1µl of 0.05% NP-40. To this 1µl of each probe of interest was added and up to 2µg of nuclear proteins.

Running Buffer: 0.5x TBE Buffer, 50ml of 10xTBE Buffer, 950ml of deionised water.

Appendix III

Primers used for HRM analysis and genotyping of variants in *CACNA1C*, *CACNA1D*, *CACNB3*, and *CACNG4* in Chapter 5

Primer	Primer Seq	Amplicon Size
CACNA1C Prom F1	gctccctttgacgtcatgc	313bp
CACNA1C Prom R1	ctctggcaggagctgtttct	
CACNA1C 5'UTR F1	cctctgcagaaacagctcct	329bp
CACNA1C 5'UTR R1	aaccccaaaaactgctcctt	
CACNA1C 5'UTR F2	ggtgctcagttcttggaagg	249bp
CACNA1C 5'UTR R2	cattgaccatggaccaaaaa	
CACNA1C Exon1 F	gggggtgttttcacatttctt	129bp
CACNA1C Exon1 R	aggaacccggtggaaaag	

Table 48. Primers used for HRM analysis of *CACNA1C*

Primer	Primer Seq	Amplicon Size
CACNA1D Prom F1	gatgtgagctccggctgcc	129bp
CACNA1D Prom R1	ctgcaaaggagaactggcttg	
CACNA1D Prom F2	cggatgtgagctccggctg	391bp
CACNA1D Prom R2	gtccctgcccttattctcgc	
CACNA1D Exon1 F	ctgttatttgtccccgtccc	247bp
CACNA1D Exon1 R	ctggaggaccatgacaggat	

Table 49. Primers used for HRM analysis of *CACNA1D*

Primer	Primer Seq	Amplicon Size
CACNB3 Prom F1	aatccttggcagggaagtg	359bp
CACNB3 Prom R1	gaccgtgggagacagaagag	
CACNB3 Prom F2	agctgccgtcctgtcattag	396bp
CACNB3 Prom R2	gatccaggcagggaggag	
CACNB3 5'UTR F1	gcggagtagttcctgggttg	238bp
CACNB3 5'UTR R1	gccggctcctacaacaatag	
CACNB3 5'UTR F2	gcgggcactattgttgtagg	257bp
CACNB3 5'UTR R2	cgagagagcgcgaaggag	
CACNB3 Exon1 F	ctcccccatgtatgacgac	66bp
CACNB3 Exon1 R	ggtgtcaccccaacttccc	
CACNAB3 M13 F	cacgacgttgtaaaacgacgcggagtagttcctgggttg	339bp
CACNAB3 M13 R	ggataacaatttcacacaggggtgtcaccccaacttccc	

Table 50. Primers used for HRM analysis of *CACNB3*

Primer	Primer Seq	Amplicon Size
CACNG4 5'UTR F1	agctcagaggagcgcacag	202bp
CACNG4 5'UTR R1	agcgctggggctcaaactc	
CACNG4 5'UTR F2	cctcggagtttgagccccag	222bp
CACNG4 5'UTR R2	tgagcgagaaggcggcgaag	
CACNG4 Exon1 F	caccatggtgcgatgcgac	260bp
CACNG4 Exon1 R	gcgcgcgcgtgtgtatatg	
CACNG4 Exon2 F	acggatgagtgaacccagac	148bp
CACNG4 Exon2 R	ccacagaaagagagtccagg	
CACNG4 Exon3 F	ctcactgccctctgctttgtg	209bp
CACNG4 Exon3 R	accacacgagacaggcactc	
CACNG4 Exon4 F1	tccacctgttccctacactg	204bp
CACNG4 Exon4 R1	gccacaatgaaagacagagc	
CACNG4 Exon4 F2	caggtgacccgagtgacaag	221bp
CACNG4 Exon4 R2	gtggaccttgagctggacct	
CACNG4 Exon4 F3	cgtcttcctcttctccttatg	207bp
CACNG4 Exon4 R3	aagtcatgcacctgcaggaag	
CACNG4 Exon4 F4	gaagatcacaggggccatc	237bp
CACNG4 Exon4 R4	cttgaggatcacatgccatc	
CACNG4 3'UTR F1	ctttctctccgctccagcctc	354bp
CACNG4 3'UTR R1	ggcaagattcgtttccaaga	
CACNG4 3'UTR F2	gctgaaggctgactttgtc	262bp
CACNG4 3'UTR R2	ctgacaagagcccagagacc	
CACNG4 3'UTR F3	cgaatcttgccagaaaaacg	272bp
CACNG4 3'UTR R3	acagcgcaaaggccaaag	
CACNG4 3'UTR F4	catctctccctctctccaag	319bp
CACNG4 3'UTR R4	tgctccttgggtctatgaca	
CACNG4 3'UTR F5	cttcagagtctggagggttc	285bp
CACNG4 3'UTR R5	gcgttgatttgcaggaagtt	
CACNG4 3'UTR F6	ggagacctaagggtgaacag	388bp
CACNG4 3'UTR R6	ttcttcttccaagcctgcac	
CACNG4 3'UTR F7	tccagccattcttttcttcc	271bp
CACNG4 3'UTR R7	gacatcatcactaggccgaac	
CACNG4 3'UTR F8	ctcagctacctgctcacag	230bp
CACNG4 3'UTR R8	gctgggcacacttggacatg	
CACNG4 3'UTR F9	catcagggttggcatcatct	272bp
CACNG4 3'UTR R9	gtgttccaatgaatcctacc	
CACNG4 3'UTR F10	cccgctgccccagaagtttc	265bp
CACNG4 3'UTR R10	gggagagaaggtgggcacag	
CACNG4 3'UTR F11	cccgctgcatctcagacctg	222bp
CACNG4 3'UTR R11	ccctaactgcccacaaagac	

Table 51. Primers used for HRM analysis of *CACNG4*

Variant	Primer
CACNA1C Prom F1	GAAGGTGACCAAGTTCATGCTCGGCGCGGCGCGGCGG
CACNA1C Prom F2	GAAGGTCGGAGTCAACGGATTCGGCGCGGCGCGGCGA
CACNA1C Prom R	CTGGGCGCGGAGGCAGGAA

Table 52.Primers for Genotyping *CACNA1C* promoter SNP1

Variant	Primer
CACNG4 rs371128228 F1	GAAGGTGACCAAGTTCATGCTCGAGCGCCGTCGCCG
CACNG4 rs371128228 F2	GAAGGTGACCAAGTTCATGCTCGAGCGCCGTCGCCA
CACNG4 rs371128228 R	CTTCCTCTTCTCCTTATGCCSGGAT
CACNG4 17:65026851 (C/T) F1	GAAGGTGACCAAGTTCATGCTAGGTGACCCGAGTGACAAGCA
CACNG4 17:65026851 (C/T) F2	GAAGGTGACCAAGTTCATGCTAGGTGACCCGAGTGACAAGCT
CACNG4 17:65026851 (C/T) R	CAAAGTAAAAAGACCAGCCGTAGTTGTA

Table 53. Primers for Genotyping variants from the fourth exon of *CACNG4*

Appendix IV

Primers used for genotyping variants in Chapter 6

Primer	Primer Sequence
CACNA1C_rs146482058_F1	GAAGGTGACCAAGTTCATGCTAAGAGTTCTCTGTCACCTTAAGAAAAAAA
CACNA1C_rs146482058_F2	GAAGGTCGGAGTCAACGGATTAAGAGTTCTCTGTCACCTTAAGAAAAAAT
CACNA1C_rs146482058_R	GTGTCTGCCAGTGCTTACACTATAATTTA
CACNA1C_rs79398153_F2	GAAGGTGACCAAGTTCATGCTTGGTTTTCTCAGCCAGGGCAC
CACNA1C_rs79398153_F2	GAAGGTCGGAGTCAACGGATTGGTTTTCTCAGCCAGGGCAT
CACNA1C_rs79398153_R	GAACCGAGTGAGGTGCCTGGAA
CACNA1C_rs191953785_F1	GAAGGTGACCAAGTTCATGCTGTTGTTGTTCTCTTGCCCGATG
CACNA1C_rs191953785_F2	GAAGGTCGGAGTCAACGGATTCTGTTGTTGTTCTCTTGCCCGATA
CACNA1C_rs191953785_R	GTCAAAGTCTCTTAATGTCACTGAAGCTT
CACNA1C_rs112312080_F1	GAAGGTGACCAAGTTCATGCTGTCAATGAAAATTTAAACCCGGGTGG
CACNA1C_rs112312080_F2	GAAGGTCGGAGTCAACGGATTAGTCAATGAAAATTTAAACCCGGGTGA
CACNA1C_rs112312080_R	CATTGAGAACCCTTCAGCACATGCAT
CACNA1C_rs113414207 _F1	GAAGGTGACCAAGTTCATGCTGGACAGACCCCCCCCCG
CACNA1C_rs113414207 _F2	GAAGGTCGGAGTCAACGGATTGGACAGACCCCCCCCCC
CACNA1C_rs113414207 _R	CTTCATTTGTGTGGGACCGTGACAT
CACNA1C_12: 2329069 (G/T)_F1	GAAGGTGACCAAGTTCATGCTGAAAGCCTCCCTCCTGGGC
CACNA1C_12: 2329069 (G/T)_F2	GAAGGTCGGAGTCAACGGATTGGAAAGCCTCCCTCCTGGGA
CACNA1C_12: 2329069 (G/T)_R	AGTCTGGAGGCCTGGTGACCTA
CACNA1C_12: 2423175 (G/C)_F1	GAAGGTGACCAAGTTCATGCTGGAAAGGCCGTGTGCCC
CACNA1C_12: 2423175 (G/C)_F2	GAAGGTCGGAGTCAACGGATTCTGGAAAGGCCGTGTGCCG
CACNA1C_12: 2423175 (G/C)_R	CCAGGGAGGCTCAGGAAAATATCAA

Table 54. Primers for Genotyping *CACNA1C* intronic variants identified in WGS data

Primer	Primer Sequence
CACNA2D4_A741V_F1	GAAGGTGACCAAGTTCATGCTCATGGAAGCCTACTGGACAGC
CACNA2D4_A741V_F2	GAAGGTCGGAGTCAACGGATTCCATGGAAGCCTACTGGACAGT
CACNA2D4_A741V_R	AACCTGGACCTACTCGGACATGTT
CACNA1D_R931H_F1	GAAGGTGACCAAGTTCATGCTGTTCCGGAAGGAGTGGCTGC
CACNA1D_R931H_F2	GAAGGTCGGAGTCAACGGATTGTTCCGGAAGGAGTGGCTGT
CACNA1D_R931H_R	GCCGCAGAGGACCCCATCC
CACNA1D_D1837N_F1	GAAGGTGACCAAGTTCATGCTGGAGCTGTTCATCTCCTGAGTC
CACNA1D_D1837N_F2	GAAGGTCGGAGTCAACGGATTGGGAGCTGTTCATCTCCTGAGTT
CACNA1D_D1837N_R	TCCTCTCTCCCAACTGCAGGTC
CACNA1D_R2145W_F1	GAAGGTGACCAAGTTCATGCTGTAGCTCATAGTCCTGCCG
CACNA1D_R2145W_F2	GAAGGTCGGAGTCAACGGATTCCTGTAGCTCATAGTCCTGCCA
CACNA1D_R2145W_R	GGGATGTGGGCCCCCTCTC

Table 55. Primers for Genotyping *CACNA2D4* and *CACNA1D* intronic variants identified in WGS data

Publications