Investigating genome-wide transcriptional and methylomic consequences of a balanced t(1;11) translocation linked to major mental illness

Daniel Lawrence McCartney

Doctor of Philosophy

The University of Edinburgh

2016

Declaration

I declare that this thesis has been composed by me, that the work described in this thesis is my own, except where otherwise stated, and that the work described in this thesis has not been submitted for any other degree or professional qualification.

_____

Daniel McCartney

Acknowledgements

Firstly, I'd like to thank my supervisor, Kathy Evans, for her constant support, advice, and encouragement over the past four years. I would also like to thank Rosie Walker, for her guidance and for sharing her extensive knowledge of all things related to statistics and data analysis. I've been extremely lucky to have this level of support from both Kathy and Rosie, and without them this thesis would not be possible. Thanks to David Porteous for providing the opportunity to do this project, and for his encouragement and input throughout the course of this PhD, and to Andrew McIntosh for his support on clinical aspect of the project. I am grateful to my funding body, Mental Health Research UK, for the financial support throughout my studentship. I would also like to thank Cathy Abbott and Andy Sims for their helpful comments and suggestions during my thesis committee meetings.

Susan, Helen, Joyce and Heather have been extremely helpful, keeping things running smoothly in the lab. Thanks in particular to Susan for providing training in several methods, and for her endless support in the lab. Thanks to Stewart and Steve for their assistance with all things computer-related, and to Stewart again for his help with the EPIC array paper. I'm grateful to Pippa Thomson for providing genotype data, Malgorzata Borkowska for the L100P mice, and to Kristin Nicodemus and Lara Neira-Gonzalez for their assistance with the random forest analysis.

Thanks to the other staff and students for making this PhD an enjoyable experience. My office-mates and others have provided great chats both at work and over drinks: thanks to Laura, Pete, Katy, Faith, and Niamh. Thanks to my friends James, Sarah, and Cathy for generally putting up with me, and to Sinéad for the genetics lessons. Finally, I'd like to thank my family for their unending support and encouragement.

Abstract

Schizophrenia, bipolar disorder and major depressive disorder are devastating psychiatric conditions with a complex, overlapping genetic and environmental architecture. Previously, a family has been reported where a balanced chromosomal translocation between chromosomes 1 and 11 [t(1;11)] shows significant linkage to these disorders. This translocation transects three genes: *Disrupted in schizophrenia-1* (*DISC1*) on chromosome 1, a non-coding RNA, *Disrupted in schizophrenia-2* (*DISC2*) antisense to *DISC1*, and a non-coding transcript, *DISC1 fusion partner-1* (*DISC1FP1*) on chromosome 11, all of which could result in pathogenic properties in the context of the translocation. This thesis focuses on the genome-wide effects of the t(1;11) translocation, primarily examining differences in gene expression and DNA methylation, using various biological samples from the t(1;11) family.

To assess the genome-wide effects of the t(1;11) translocation on methylation, DNA methylation was profiled in whole-blood from 41 family members using the Infinium HumanMethylation450 BeadChip. Significant differential methylation was observed within the translocation breakpoint regions on chromosomes 1 and 11. Downstream analysis identified additional regions of differential methylation outwith these chromosomes, while pathway analysis showed terms related to psychiatric disorders and neurodevelopment were enriched amongst differentially methylated genes, in addition to more general terms pertaining to cellular function. Using induced pluripotent stem cell (iPSC) technology, neuronal samples were developed from fibroblasts in a subset of individuals profiled for genome-wide methylation in whole blood (N = 6) with an aim to replicate the significant findings around the breakpoint regions. Here, methylation was profiled using the Infinium HumanMethylation450 BeadChip's successor: the Infinium MethylationEPIC BeadChip. The results from the blood-based study failed to replicate in the neuronal samples, which could be attributed to low statistical power or tissue-specific factors such as methylation quantitative trait loci. The differences in methylation in the most significantly differentially methylated loci were found to be driven by a single individual, rendering further interpretation of the findings from this analysis difficult without additional samples. Cross-tissue

analyses of DNA methylation were performed on blood and neuronal DNA from these six individuals, revealing little correlation between cell types.

DISC1 is central to a network of interacting protein partners, including the transcription factor ATF4, and PDE4; both of which are associated with the cAMP signalling pathway. Haploinsufficiency of DISC1 due to the translocation may therefore be disruptive to cAMP-mediated gene expression. In order to identify transcriptomic effects which may be related to the t(1;11) translocation, genome-wide expression profiling was performed in lymphoblastoid cell line RNA from 13 family members. No transcripts were found to be differentially expressed at the genome-wide significant level. A *post-hoc* power analysis suggested that more samples would be required in order to detect genome-wide significant differential expression. However, imposing a fold-change cut-off to the data identified a number of candidate genes for follow-up analysis, including *SORL1*: a member of the brain-expressed Sortilin gene family. Sortilin genes have been linked to multiple psychiatric disorders including schizophrenia, bipolar disorder and Alzheimer's disease. Follow-up analyses of Sortilin family members were performed in a Disc1 mouse model of schizophrenia, containing an amino acid substitution (L100P). Here, developmental gene expression profiling was performed with an additional aim to optimise and validate work performed by others using this mouse model. However, results from these experiments were variable between two independent batches mice tested. Additional investigation of Sortilin family genes was performed using GWAS data from human samples, using machine learning techniques to identify epistatic interactions linked to depression and brain function, revealing no statistically significant interactions.

The results presented in this thesis suggest a potential mechanism for differential DNA methylation in the context of chromosomal translocations, and suggests mechanisms whereby increased risk of illness is conferred upon translocation carriers through dysregulation of transcription and DNA methylation.

Lay Abstract

Schizophrenia, bipolar disorder and major depression are common psychiatric conditions with both genetic and environmental risk factors. A genetic risk factor for these disorders has been identified in a large Scottish family in the form of a chromosomal rearrangement whereby three genes are disrupted. One of these genes, *DISC1*, has been shown to play multiple roles in brain function and development. The work presented in this thesis aims to investigate widespread effects of this chromosomal rearrangement in the family.

Disease can occur through genetic (sequence-based) and epigenetic (non-sequence-based) mechanisms. One such epigenetic mechanism, DNA methylation, was examined in blood-derived DNA, comparing 17 family members with the above-mentioned chromosomal rearrangement to 24 of those without. Significant differences in DNA methylation were observed on chromosomes 1 and 11, within the regions at which the chromosomal rearrangement occurs. Furthermore, regions containing genes with previously reported functions in the brain and psychiatric illness showed differences in DNA methylation between carriers of the rearrangement compared to those without.

DNA methylation was also examined in stem cell-derived neuronal material from a subset of these individuals, consisting of three carriers of the chromosomal rearrangement to three without. Differences in methylation at nine of the top ten sites were found to be largely driven by a single individual. Unlike the findings observed in blood DNA, there were no differences observed in the regions of the rearrangement in these samples. This may be attributed to differences in sample size or the cell type studied. DNA methylation differences were also studied within individuals, comparing these levels across blood and brain-like cells.

The DISC1 protein has several interacting partners, including proteins involved in the control of gene expression. Individuals with the chromosomal rearrangement have

previously been reported to display half the normal levels of DISC1. This may have downstream effects on the levels of gene expression through the reduced interaction between DISC1 and its partners. To test this, gene expression levels were examined, comparing carriers of the chromosomal rearrangement to those without. Although no significant changes were observed, there was suggestive evidence for a number of genes with disrupted expression. One of these genes, *SORL1*, is a member of the Sortilin gene family. Members of this gene family are abundantly expressed in the brain and have been linked to schizophrenia, bipolar disorder and Alzheimer's disease.

Exploratory work was performed to investigate a relationship between *DISC1* and Sortilin family members. Levels of Sortilin family gene expression were assessed in a mouse containing a mutation in the *Disc1* gene, which had previously been shown to display schizophrenic-like behaviours. Additional work was performed to investigate a genetic interaction between *DISC1* and Sortilin family members in relation to psychiatric illness and brain function. However, no significant findings were observed in either of these analyses.

The work presented in this thesis identifies a relationship between DNA methylation and chromosomal rearrangements, and provides support for genes previously implicated in psychiatric disorders.

Table of Contents

Table of Abbreviations

5-HT	5-Hydroxytrptophan (Serotonin)
A	Adenine
ACC	Anterior cingulate cortex
ADHD	Attention deficit hyperactivity disorder
AMPA	α–amino-3-hydroxy-5-methyl-4-isoxazoleproponionic acid
ATD	Acute tryptophan depletion
BDNF	Brain-derived neurotrophic factor
BLAT	BLAST-like alignment tool
bp	Base pair
C	Cytosine
CAM	Cell adhesion molecule
cAMP	Cyclic adenosine monophosphate
cDNA	Complementary DNA
CNV	Copy number variation
CpG	Cytosine-Guanine dinucleotide motif
CT	Computerised tomography
CV	Coefficient of variation
Cys/C	Cysteine
dH2O	Deionized water
DMEM	Dulbecco's modified eagle medium
DMP	Differentially methylated position
DMR	Differentially methylated region
DMRSE	DMR standard error
DNA	Deoxyribonucleic acid

DNMT	DNA methyltransferase
dNTP	Deoxyribose nucleotide triphosphate
DSM	Diagnostic and Statistical Manual of Mental disorders
E	Embryonic day
EDTA	Ethylenediaminetetraacetic acid
eQTL	Expression quantitative trait locus
ERP	Event-related potential
EWAS	Epigenome-wide association study
F	Phenylalanine
FDR	False discovery rate
FGF	Fibroblast growth factor
fMRI	Functional magnetic resonance imaging
G	Guanine
g	Standard acceleration due to gravity
GCOSE	Genotype combined standard error
GDNF	Glial cell line-derived neurotrophic factor
GO	Gene ontology
GWAS	Genome-wide association study
HF	Hippocampal formation
iPSC	Induced pluripotent stem cell
ISC	International Schizophrenia Consortium
kb	Kilobase
L	Leucine
LCL	Lymphoblastoid cell lines
LD	Linkage disequilibrium

LOD	Logarithim of ODDs
Mb	Megabase
MDD	Major depressive disorder
meQTL	Methylation quantitative trait locus
Met	Methionine
MgCl2	Magnesium Chloride
MHC	Major histocompatibility complex
MRI	Magnetic resonance imaging
mTOR	Mammalian target of rapamycin
NAT	Natural antisense transcript
NMDA	N-Methyl-D-Aspartate
NPC	Neuronal progenitor cell
P	Proline/Postnatal day
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PGC	Psychiatric Genomics Consortium
PSF	Penicillin/Streptomycin/Amphotericin
qRT-PCR	Quantitative real-time PCR
R	Arginine
RIN	RNA integrity number
RNA	Ribonucleic acid
RNAi	RNA interference
RNA-seq	RNA sequencing
RPM	Revolutions per minute
RPMI	Roswell Park Memorial Institute (media)

RSN	Robust spline normalisation
SD	Standard deviation
SEM	Standard error of the mean
Ser/S	Serine
SNP	Single nucleotide polymorphism
SSRI	Selective serotonin reuptake inhibitors
SV	Surrogate variable
SVA	Surrogate variable analysis
T	Threonine/Thymine
TpG	Thymine-Guanine dinucleotide
TPH	Tryptophan hydroxylase
U	Uracil
UCSC	University of California Santa Cruz
Val	Valine
VST	Variance stabilising normalisation
W	Tryptophan
WGBS	Whole-genome bisulphite sequencing
WHO	World Health Organisation
WTCCC	Wellcome Trust Case Control Consortium
WTCRF	Wellcome Trust Clinical Research Facility
Y2H	Yeast-2-hybrid

List of Tables

List of Figures

Chapter 1

Introduction

1 Introduction

1.1 Overview of major mental illness

Schizophrenia, bipolar disorder and major depressive disorder (MDD) are severe psychiatric disorders with a prevalence between approximately 1 and 4% for schizophrenia and bipolar disorder, and approximately 15% for MDD (McGrath *et al.*, 2008; Craddock and Sklar, 2009; Ferrari *et al.*, 2013; Smoller *et al.*, 2013). These disorders are complex with both environmental and genetic risk factors, and contribute to a significant social and economic burden. In England alone, mental illness incurs an estimated annual economic cost of £105 billion, including the cost of care and services, reduced quality of life, and lost productivity at work (Centre for Mental Health, 2010).

1.1.1 Schizophrenia

Schizophrenia is a psychotic disorder characterised by positive symptoms, negative symptoms and cognitive dysfunction. Positive symptoms refer to increases in normal functions and include auditory and visual hallucinations, irrational beliefs (delusions), thought disorder and motor dysfunction. Negative symptoms refer to reductions in normal functions and can include anhedonia (the inability to experience pleasure) and self-neglect. The cognitive effects associated with schizophrenia include impairments in working memory, attention deficits and reduced decision-making abilities (i.e. deficits in executive functioning).

The current edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-V) defines a diagnosis of schizophrenia as presentation for at least one month with two or more symptoms, of which one must be either hallucinations, delusions or disorganised speech. Additional symptoms that can incur a diagnosis include negative symptoms and disorganised or catatonic behaviour (American Psychiatric Association, 2013). Prior to DSM-V, the symptomatic heterogeneity of schizophrenia was addressed by dividing the disorder into clinical subtypes: disorganised, catatonic, paranoid and undifferentiated schizophrenia. However, due to their limited reliability and validity (Korver-Nieberg *et al.*, 2011), subtypes have

been removed and replaced by the introduction of psychopathological dimensions in the current manual. Cognitive dysfunction has not been included as a criterion for a diagnosis of schizophrenia, despite being a common symptom of the disorder (O'Carroll, 2000). The reason for this exclusion was that cognitive deficits are not sufficient as a diagnostic marker for distinguishing between schizophrenia and other psychiatric disorders (Barch *et al.*, 2013). Treatment of schizophrenia involves management of symptoms and prevention of relapse through pharmacological and psychotherapeutic intervention. This can be through antipsychotic administration alone or in conjunction with cognitive, individual, and/or group therapy (Patel *et al.*, 2014). However, recurrence of episodes is common and patients can develop resistance to treatments (Hasan *et al.*, 2012; Elmsley *et al.*, 2013).

1.1.2 Bipolar disorder

Bipolar disorder is characterised by recurrent episodes of elevated mood (mania) and depression. Bipolar disorder is divided into two subtypes: bipolar I and bipolar II. According to DSM-V criteria, a diagnosis of the former type requires presentation of at least one manic episode. The latter type is characterised by presentation of both a depressive and a hypomanic episode, without escalating to a manic episode. The prevalence of bipolar disorder I is equal between males and females. However, gender differences have been reported in bipolar disorder II, with a greater incidence among females (Nivoli *et al.*, 2011). Furthermore, rapid cycling (i.e. four or more episodes in 12 months) is displayed more frequently in females (Lebenluft, 1996). Approximately 75% of patients with an acute manic episode co-present with psychosis (Goodwin and Jamison, 2007). Treatment for bipolar disorder is dependent on the degree of mania or depression presented. Lithium is a well-established treatment for bipolar disorder, functioning as a mood stabiliser. Antipsychotics are also effective in the treatment of mania (Cipriani *et al.*, 2011). In the case of bipolar depression (i.e. depression with a history of mania or hypomania), identification of effective treatments has been more challenging (Geddes and Miklowitz, 2013).

1.1.3 Major depressive disorder

Major depressive disorder (MDD) is a mood disorder characterised by depressive episodes in the absence of mania or hypomania. MDD is more prevalent in females (Piccinelli and Wilkinson, 2000), and has been ranked as the 4th leading cause of disability worldwide by the World Health Organisation (WHO; Murray and Lopez, 1996). Symptoms of MDD include weight changes, sleep disturbances, abnormal motor function, fatigue, feelings of worthlessness or guilt, cognitive deficits, suicidal ideation, a depressed mood, and anhedonia. DSM-V criteria state that to incur a diagnosis of MDD, a patient must present with five or more of these symptoms including at least one of the latter two, almost every day for at least two weeks. This renders MDD a highly heterogeneous disorder in terms of clinical presentation. There are 227 possible combinations of symptoms that can incur a diagnosis of MDD, with some more commonly presented than others (Zimmerman *et al.*, 2015). Recurrence of the disorder is common: at least 50% of patients present with a subsequent episode in their lifetime (American Psychiatric Association, 2000). Treatment of MDD usually involves pharmacological intervention with antidepressant medications, psychotherapy, or a combination of both approaches. However, challenges remain in identifying the most efficient treatments to target a given MDD subtype (Miller and O'Callaghan, 2013)

1.2 Mechanisms of major mental illness

Pharmacological, *post-mortem*, and brain-imaging studies have been key for investigating the mechanisms by which the above disorders might occur. There is strong evidence for aberrant neurotransmission as well as brain structural abnormalities in psychiatric illness which are likely to be involved in its pathophysiology. While *post-mortem* brains are a useful resource for investigating the pathophysiology of psychiatric disorders, they can be subject to numerous potential confounding factors. Tissue quality, medication status, diagnosis, age of death, cause of death and *post-mortem* interval are some of the factors which must be considered to generate robust data from such studies. Nonetheless, these studies have provided valuable insights into the neuropathology of psychiatric illness.

1.2.1 Neurotransmitter dysfunction in major mental illness

Pharmacological intervention of psychiatric disorders is primarily targeted towards neurotransmitter systems in the brain. Neurochemical signaling pathways such as the *N*-Methyl-D-Aspartate (NMDA), dopaminergic, and serotonergic systems have all been implicated in the pathophysiology of schizophrenia, bipolar disorder and MDD. Moreover, current treatments for these illnesses act directly upon these pathways to relieve symptoms.

Early suggestions of neurotransmitter abnormalities in psychiatric illness emerged with the dopamine hypothesis of schizophrenia, which initially arose from the observation that antipsychotic drugs resulted in increased dopamine metabolism in mice (Carlsson and Lindqvist, 1963). It was later noted that antipsychotic drugs directly interacted with dopamine receptors (Creese *et al.*, 1976). Furthermore, the effectiveness of these drugs at relieving symptoms was found to be correlated with their affinity to these receptors (Seeman *et al.*, 1975; Creese *et al.*, 1976). A pathogenic mechanism was suggested whereby psychosis occurred as a result of excess dopamine transmission (Matthysse, 1973). Subsequent research led to a modified hypothesis by Davis *et al*. (1991), who proposed that differential regional distributions of dopamine activity could account for positive and negative symptoms of schizophrenia. Like the original hypothesis, this was narrowly focused on dopamine dysfunction, with no description of the origins of these abnormalities. However, advances in research in the years since then have been able to utilise imaging and genetic studies to provide a higher-resolution representation of dopamine dysfunction in schizophrenia. Howes and Kapur (2009) proposed a new hypothesis whereby dopamine dysfunction results from the interaction of both genetic and environmental factors. This version of the dopamine hypothesis suggests multiple genetic and environmental factors act to cause presynaptic hyperdopaminergia, thus giving rise to psychotic symptoms rather than schizophrenia itself, through aberrantly salient ideas and perceptions stimulated by a normal environment (i.e. delusions and hallucinations). While current antipsychotic drugs act postsynaptically, at the level of the dopamine receptor (Jones and Pilowsky,

2002), normalisation of presynaptic dopamine levels has been suggested as an alternative treatment mechanism based on the current hypothesis (Howes *et al.*, 2012; Bonoldi and Howes, 2014).

Another major class of receptor in the brain is the NMDA receptor. NMDA receptors play essential roles in cellular homeostasis and neurotransmission through their activation by glutamate or glycine (Lakhan *et al.*, 2013). An initial hypothesis of NMDA hypofunction in schizophrenia emerged from the observation that schizophrenia patients had low levels of glutamate in cerebrospinal fluid (Kim *et al.*, 1980). This hypothesis of NMDA hypofunction has been supported by reports of the induction of positive, negative and cognitive symptoms through the antagonism of NMDA receptors using ketamine (Krystal *et al.*, 1994; Gilmour *et al.*, 2012). There is also evidence to suggest a function for NMDA receptors in mood disorders. Ketamine has been shown to relieve symptoms of depression (Zarate *et al.*, 2006) while Michael *et al.* (2003) reported an increase in glutamate levels in the brains of bipolar patients during acute manic episodes. Several mechanisms have been suggested for the role of NMDA receptors in mood disorders. Increased activity of the neurotrophic mammalian target of rapamycin (mTOR) signaling pathway has been reported in rats following administration of ketamine, resulting in increased levels of synaptogenesis (Li *et al.*, 2010). Others have shown that ketamine-mediated blocking of NMDA receptors results in an increased interaction between glutamate and α–amino-3-hydroxy-5-methyl-4-isoxazoleproponionic acid (AMPA) receptors, proposing a mechanism for its antidepressant effects (Andreasen *et al.*, 2013).

In addition to glutamate and dopamine signaling, there is strong evidence for the role of 5-hydroxytryptophan (5-HT; also known as serotonin) in psychiatric illness – particularly in mood disorders. Serotonin is synthesised from tryptophan by action of tryptophan hydroxylase (TPH; Fitzpatrick, 1999). Early reports of a link between serotonin and mood were provided by Coppen (1965), who proposed that deficits in serotonin levels in the brain led to depressive symptoms. Young *et al.* (1985) later observed a lowering of mood in males suffering from acute tryphtophan depletion

(ATD), proposing a pathogenic mechanism for depression via the lowering of serotonin. A widely-used treatment for depression and bipolar depression is the administration of selective serotonin reuptake inhibitors (SSRIs). These work to increase extracellular serotonin levels by preventing their reabsorption into the presynaptic cell (Albert and Benkelfat, 2013; Geddes and Miklowitz, 2013).

There is clear evidence for dysfunction of neurotransmitter systems in psychiatric illness. While current treatments show some promise by targeting individual pathways to alleviate symptoms, a greater understanding of the interactions between the above pathways is likely to be necessary in order to develop more efficient treatment options for these disorders. Furthermore, knowledge of the pathways disrupted in a given patient will be required in order to provide the most efficient treatment.

1.2.2 Brain structural abnormalities

The first image-based reports of brain abnormalities in schizophrenia were by Johnstone *et al.* (1976), who, using computerised tomography (CT) scanning, observed enlarged ventricles in patients. Reduced volumes of frontal and temporal lobes have also been observed in schizophrenia (Turetsky *et al.*, 1995). It has been suggested that the reductions in frontal lobe volume observed in schizophrenia may be related to symptoms of apathy in patients (Roth *et al.,* 2004).

Imaging studies have also reported structural abnormalities in patients with bipolar disorder and MDD (Kempton *et al.*, 2011). Reduced grey matter volumes have been observed in both MDD and bipolar disorder (Lim *et al.*, 1999; Tang *et al.*, 2007). However, treatment with mood stabilisers such as lithium have been shown to affect grey matter volume, rendering the findings for bipolar disorder more difficult to interpret (Moore *et al.*, 2000; Hafeman *et al.*, 2012). A meta-analysis of 20 studies by Lai (2013) reported reduced grey matter volumes in the anterior cingulate cortex (ACC) of MDD patients, while a meta-analysis of schizophrenia brain imaging studies by Glahn *et al.* (2008) identified grey matter abnormalities in multiple brain

regions. Network analysis of the regions containing grey-matter differences revealed four discrete networks, each potentially associated with pathological hallmarks of schizophrenia, such as cognitive deficits. Deficits in white matter volume have also been associated with illness: MRI studies have shown reduced white matter density in both schizophrenia and bipolar disorder (McIntosh *et al.*, 2005), while *post-mortem* studies have revealed increased neuronal density in white matter in both schizophrenia and bipolar disorder (Connor *et al.*, 2009, Connor *et al.*, 2011). These results may suggest that the above disorders are, at least in part, a result of disrupted connectivity in the brain.

Abnormalities in the caudate nucleus have also been reported in schizophrenia. The caudate nucleus is highly innervated by the dopamine system and functions in learning, memory and executive functioning (Grahn *et al.*, 2008). Jernigan *et al.* (1991) reported enlargements of the caudate nucleus in schizophrenia while later studies revealed that this could be modulated by treatment with antipsychotics (Chakos *et al.*, 1994). Enlargement of the caudate nucleus has also been observed in bipolar I (Maller *et al.*, 2014), along with differences in size and shape between affected and unaffected individuals (Ong *et al.*, 2012).

The presence of such structural abnormalities in illness strongly points to a neurodevelopmental origin of these disorders. The neurodevelopmental hypothesis of schizophrenia was first proposed by Weinberger (1987), who suggested the disorder may occur as a result of perturbed brain development. Observations in support of this come from brain imaging, behavioural and animal studies. Structural abnormalities have been reported in the absence of gliosis, suggesting these are not neurodegenerative but rather neurodevelopmental (Harrison, 2000), while cognitive deficits have been observed in individuals who later went on to develop schizophrenia (Fuller *et al.*, 2002). Moreover, illness-associated genetic variation is present in genes related to neurodevelopment, described in greater detail in section 1.3.

1.3 Evidence for a genetic contribution to psychiatric disorders

Family, twin and adoption studies have been useful tools to demonstrate the heritability of psychiatric disorders (Shih *et al.*, 2004). There is a correlation between the risk of schizophrenia and the degree of genetic relatedness between individuals (Figure 1.1; Gottesman, 1991). This risk correlates with the degree of genetic sharing between affected and unaffected individuals: individuals with an affected first degree relative (approximately 50% genetics shared) have a tenfold higher risk of developing the disorder compared to the general population. This risk is decreased to double that of the general population in individuals with an affected cousin (approximately 12.5% genetics shared). This section will summarise the methods utilised in psychiatric research to elucidate the genetic architecture of major mental illness, along with key findings, beginning with evidence from early approaches such as linkage and cytogenetic studies, to the most recent evidence from genome-wide association studies (GWAS).

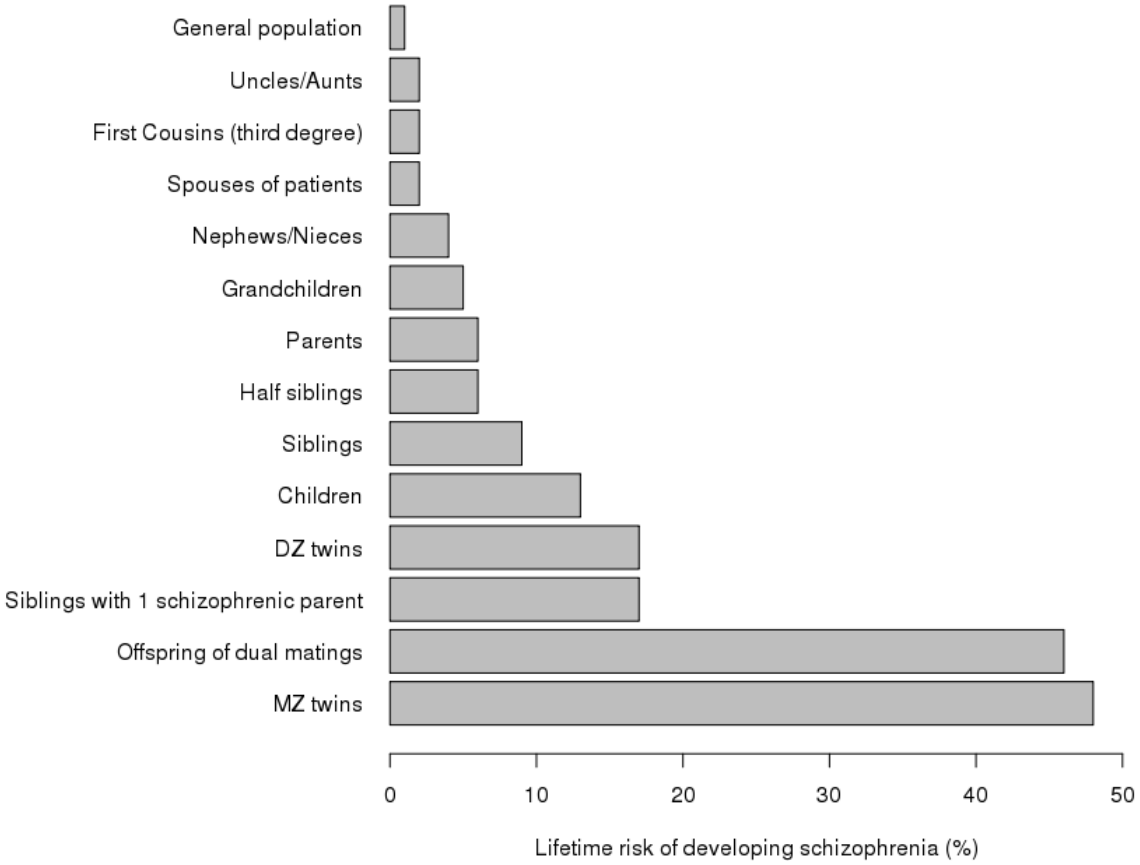


Figure 1.1: Risks for developing schizophrenia.

Figure summarises the relative lifetime risk of schizophrenia in percent (x-axis) with each bar corresponding to the general population and individuals with various degrees of genetic relatedness to a patient suffering from the disorder (y-axis; figure adapted from Gottesman, 1991).

1.3.1 Linkage studies

Genetic linkage analysis is a commonly used method in epidemiology to identify genomic regions that may be inherited or "linked" with a disease or trait among related individuals separated by a low number of meiotic events. Linkage studies have proven useful in identifying the genes underlying disorders with Mendelian patterns of inheritance, such as Huntington's disease (Gusella *et al.*, 1983), but less so for complex psychiatric disorders. Several studies have reported evidence for linkage, but these have generally failed to define specific gene loci and have been largely unsuccessful in yielding replicable results.

1.3.1.1 *Linkage studies of schizophrenia*

More than 30 genome-wide linkage scans have been performed for schizophrenia as well as several meta-analyses. However, many of the results from these studies have been inconsistent. A meta-analysis of 18 studies by Badner and Gershon (2002) reported significant linkage with schizophrenia in regions of chromosomes 8p, 13q and 22q. In a meta-analysis of 20 genome-wide linkage studies, Lewis *et al.* (2003) identified genome-wide significant evidence for linkage on chromosome 2q, with nominally significant results observed at 8p and 22q, among several other regions. A more recent meta-analysis by Ng *et al*. (2009) consisting of 32 independent studies also reported genome-wide significant linkage at chromosome 2q, with additional regions at chromosomes 5q and 2q showing suggestive evidence for linkage with schizophrenia. When they considered studies of European ancestry only, suggestive evidence for linkage was also observed on chromosome 8p.

1.3.1.2 *Linkage studies of bipolar disorder*

Badner and Gershon (2002) performed a meta-analysis of 11 genome-wide linkage scans of bipolar disorder, finding evidence for susceptibility regions at 13q and 22q. In a meta-analysis of 18 studies, Seguardo *et al*. (2003) failed to find genome-wide significant evidence for linkage to bipolar disorder. A more recent meta-analysis by McQueen *et al.* (2005) was successful in identifying genome-wide significant regions

of susceptibility at chromosomes 6q and 8q, using the original genotype data from 11 studies, unlike the former meta-analyses which used summary statistics. A review by Seretti and Mandelli (2008) provided a comprehensive overview of linkage studies of bipolar disorder performed up to December 2007, reporting susceptibility regions on all autosomes along with the X chromosome. The majority of regions reported had been derived from single studies, but some regions have been implicated in multiple studies (e.g. 6q21; Ewald *et al.*, 2002; Park *et al*., 2004), while for others there are reports of both positive and negative findings (e.g. 1q31: Detera-Wadleigh *et al.*, 1999; Ekholm *et al.*, 2003).

Positive findings by linkage analysis depends upon two factors, the existence of loci of major effect in a given family; and either a family of sufficient size to generate a significant finding in itself, or more than one family that are segregating a major effect risk allele at the same locus. The equivocal findings from linkage studies, balanced against the strong evidence for twin and family studies for a strong genetic component implies a high degree of genetic heterogeneity and multiple genetic risk loci.

1.3.2 Association studies

In contrast to linkage studies, which look for evidence of consistent meiotic segregation of a given trait with a specific chromosomal region, association studies look for evidence of allelic distortion between cases and controls in candidate genes, or more commonly now in the era of gene chips, genome-wide. By design and practice, association studies provide higher resolution mapping than linkage analysis for identifying putative risk loci. A further advantage of association studies is the power to detect smaller effect sizes, which may be masked in linkage studies. Association studies depend on comparing a group of unrelated individuals with a phenotype (trait or diagnosis) of interest (cases) with a matched set of individuals without the phenotype (controls). Association is determined based on whether a given variant is present at a statistically significant level in one phenotypic group over another (case vs. control). Such studies can be hypothesis-driven (i.e. candidate gene association studies) or hypothesis-free (i.e. GWAS).

1.3.2.1 *Candidate gene association studies*

Prior to the GWAS era (defined as 2008 or earlier), candidate gene studies tended to be performed on the basis on evidence obtained from linkage and cytogenetic analyses.

A review by Farrell *et al.* (2015) assessed the current status of 25 candidate genes for schizophrenia, based on the number of pre-GWAS era studies as reported by the SZGene database (Allen *et al.*, 2008), or their appearance in four selected reviews on the genetics of schizophrenia (Lohmueller *et al.*, 2003; Owen *et al.*, 2005; Sullivan *et al.*, 2005; Harrison and Weinberger, 2005). The authors reported *COMT* as the gene of interest in the most published candidate gene association studies for schizophrenia (81 publications). *COMT* functions in the degradation of dopamine and maps to chromosome 22q11, a region implicated in schizophrenia by linkage and cytogenetic analyses. A non-synonymous Val108/158Met polymorphism in *COMT* has been reported to be associated with a two to four-fold variation COMT enzyme activity (Lachman *et al.*, 1996; Chen *et al.*, 2003). Harrison and Weinberger (2005) reviewed several studies in which association with schizophrenia was reported for the Val-*COMT* allele (conferring high enzyme activity). Considering the current hypothesis of hyperdopaminergia in schizophrenia (Howes and Kapur, 2009), this finding may appear contradictory. However, a study by Egan *et al.* (2001) reported a negative correlation between Val-*COMT* allele dosage (high activity enzyme) and cognitive function, as well as fMRI evidence for a negative correlation between the Val-*COMT* allele and efficiency of physiological response in the prefrontal cortex during working memory tasks, independent of psychiatric diagnosis. They proposed a mechanism whereby the high activity allele leads to reduced synaptic dopamine in the prefrontal cortex, resulting in deficits in prefrontal function: a neuropathological characteristic of schizophrenia (Weinberger *et al.*, 2001).

A meta-analysis by Seifuddin *et al.* (2012) examining 487 candidate gene association studies on bipolar disorder failed to find significant association with the disorder among 362 genes after correction for multiple testing. However, nominally significant associations ($p < 0.05$) were observed for polymorphisms within four genes: *Brain derived neurotrophic factor* (*BDNF*), *Dopamine receptor D4* (*DRD4*), *D-amino acid*

oxidase activator (*DAOA*), and *Tryptophan hydroxylase 1* (*TPH1*). Several analyses and meta-analyses have been performed on these genes, reporting mixed results.

The BDNF protein plays multiple roles in neural development, maintenance and function (reviewed by Huang and Reichardt, 2001). Upregulation of BDNF has been observed in response to treatment with mood stabilisers, as well as anti-depressant medications, rendering it an attractive functional candidate for bipolar disorder (Fukumoto *et al.*, 2001; Watanabe *et al.*, 2010). However, results from association studies have been inconsistent: several studies have examined association between a non-synonymous variant in *BDNF* (rs6265 - Val66Met) and bipolar disorder, yielding both positive and negative findings (Neves-Pereira *et al.*, 2002; Lohoff *et al.*, 2005; Kunugi *et al.*, 2004). A recent meta-analysis by Gonzáles-Castro *et al.* (2015) failed to find association between bipolar disorder and the Val66Met polymorphism in over 16,000 individuals.

DRD4 encodes a dopamine receptor and is located on chromosome 11p15.5, a region with both positive and negative findings for linkage to bipolar disorder (Serretti and Mandelli, 2008). Due to their roles in neurotransmission, dopamine receptor genes are strong functional candidates for psychiatric disorders (section 1.2.1). A family-based analysis of tandem repeat alleles in exon 3 of *DRD4* reported that the excess transmission of a 4-repeat allele conferred an increased risk of bipolar disorder while the excess non-transmission of a 2-repeat allele conferred a protective effect (Muglia *et al.*, 2002). A meta-analysis of candidate gene association studies of *DRD4* reported significant association between the same 48 bp repeat polymorphism and mood disorders (unipolar and bipolar disorders) among 12 independent samples (López León *et al.*, 2005). Findings for *DRD4* are inconsistent, however. Using a family-based approach, Serretti *et al.* (2002) failed to find significant association between the same 48 bp *DRD4* repeat polymorphism and mood disorders.

DAOA is located on chromosome 13q within the G72/G30 gene complex. Its protein product may be involved in glutamate signaling via the activation of D-amino acid

oxidase (DAAO; Boks *et al.*, 2007). Multiple variants within this gene have been associated with schizophrenia and bipolar disorder in independent studies. Hattori *et al.* (2003), reported a haplotype within *DAOA* that was significantly overtransmitted to individuals with bipolar disorder. Prata *et al.* (2008) also reported a 2 SNP haplotype within *DAOA* showing significant association with bipolar disorder. Williams *et al.* (2006) investigated association between nine tag SNPs at the *DAOA*/*G30* locus, reporting significant association between three variants and bipolar disorder in a British population. The same study reported no significant association between the same variants and schizophrenia. A meta-analysis performed by Shi *et al.* (2008) examined association studies of *DAOA*/*G30* and bipolar disorder and schizophrenia performed before April 2007, and found no significant association between *DAOA*/*G30* and bipolar disorder. However, they reported significant association between schizophrenia and genotype of two variants (rs947267 and rs778293) in an Asian population. A more recent study failed to replicate this finding in a family-based sample, examining these and three additional variants at the *DAOA*/*G30* locus (Müller *et al.*, 2011). However, significant association was reported between rs1935062 and bipolar disorder. In the same study, a meta-analysis was performed and, consistent with Shi *et al.* (2008), reported significant association between *DAOA* and schizophrenia in Asians at rs2391191 (albeit not the same locus as the previously-associated polymorphism), but failed to find significant association between *DAOA* and bipolar disorder.

TPH1 functions in the synthesis of serotonin, rendering it strong candidate gene for mood disorders. It is located on chromosome 11p15.3-14, a region with both positive and negative findings for linkage with bipolar disorder. An A > C intronic variant (A218C, rs1800532) has been the subject of several association studies between *TPH1* and mood disorders (e.g. Bellivier *et al.*, 1998; Viikki *et al.*, 2010). Mandelli *et al.* (2011) reported a nominally significant association between bipolar disorder and a two-marker haplotype containing rs1800532 and rs7933505. However, this did not survive correction for multiple testing. Although there have been also been negative reports for association between this gene and mood disorders (Furlong *et al.*, 1998), a more recent meta-analysis found a significant association between the risk of

developing bipolar disorder and AA homozygosity at rs1800532 in a Caucasian population (Chen *et al.*, 2012).

Candidate gene studies have reported variable results in psychiatric research. These issues with replication may be attributable to factors such as illness heterogeneity or population stratification (Sher, 2001). Furthermore, differences in study designs may also be a contributing factor to discordant findings (Haslam, 2006). The last decade has seen a move from these hypothesis-driven studies to hypothesis-free approaches, examining samples at the genome-wide level to identify putative risk factors for complex disorders.

1.3.2.2 *Genome-wide association studies (GWAS)*

The development of high-throughput genotyping technologies has permitted genome-wide screening of variation which may be associated with complex traits through GWAS. This section will focus on GWAS findings performed to date for schizophrenia, bipolar disorder and MDD, summarising the key findings.

1.3.2.2.1 *GWAS findings for schizophrenia*

The first GWAS performed on schizophrenia involved 25,000 markers genotyped in 325 schizophrenia cases and 320 matched controls (Mah *et al.*, 2006). This study identified *plexin A2* (*PLXNA2*) as a susceptibility gene, although the findings were not significant at the later-suggested level of genome-wide significance to account for the multiple testing burden in GWAS ($p \leq 5 \times 10^{-8}$ as suggested by Pe'er *et al.*, 2008; Barsh *et al.*, 2012). *PLXNA2* is located on chromosome 1q32, within a region previously implicated in schizophrenia by linkage (Gurling *et al.*, 2001). Subsequent GWASs of schizophrenia have generally involved the analysis of >300,000 markers. A study by Sullivan *et al.* (2008) was unable to identify genome-wide significant association with schizophrenia after interrogating genotype at over 490,000 markers across the genome. Failure to detect genome-wide significant association in these early

attempts may well have been an issue of small sample size resulting in low statistical power. O'Donovan *et al.* (2008) analysed a primary sample of schizophrenia cases versus controls as well as two replication samples finding the strongest evidence for association at *zinc finger protein 804A* (*ZNF804A*; meta-analysis $p = 1.61 \times 10^{-7}$). Although not significant at the genome-wide level, when including bipolar disorder in the affected phenotype, genome-wide significant association was attained ($p = 9.96 \times 10^{-9}$). This marked the first study to identify a genome-wide significant variant associated with psychotic illness. A meta-analysis performed by Williams *et al.* (2011) reported genome-wide significant association between this gene and schizophrenia alone, as well as schizophrenia and bipolar disorder combined. *ZNF804A* is a brain-expressed gene of unknown function. Walters *et al*. (2010) reported reduced association between the illness-associated variant and reduced cognitive performance in schizophrenia patients, but not controls. Furthermore, imaging studies have reported association between this variant and brain structural abnormalities (reviewed in Gurung and Prata, 2015).

Shi *et al.* (2009), Steffanson *et al.* (2009) and the International Schizophrenia Consortium (ISC; Purcell *et al.*, 2009) concurrently reported an association between the extended major histocompatibility complex (MHC) and schizophrenia, supporting previous reports of a link between immune system dysfunction and schizophrenia (reviewed in Upthegroves and Barnes, 2014). In their study, Steffanson *et al.* (2009) also reported association between a marker upstream of *neurogranin* (*NRGN*), at chromosome 11q, as well as a marker within *transcription factor 4* (*TCF4*), on chromosome 18q. Both of these genes, along with the MHC region were again implicated in schizophrenia by Steinberg *et al.*, (2011), with the addition of *vaccinia related kinase 2* (*VRK2*), and a second variant approximately 400 kb from the previously-reported *TCF4* risk variant.

The establishment of the psychiatric genomics consortium (PGC) in 2007 brought together over 800 researchers from 36 countries with an aim to identify genetic risk factors in schizophrenia, affective disorders, ADHD and autism. In 2011, the

schizophrenia working group of the PGC published their first GWAS on schizophrenia implicating five new genes (Ripke *et al.*, 2011). This was followed by the identification of a further 14 novel risk loci for schizophrenia by Ripke *et al.* (2013a). Here, among 22 genome-wide significant loci, eight had previously been implicated in schizophrenia and/or bipolar disorder by earlier GWASs, supporting previous reports of a genetic overlap between the two disorders (Lichtenstein *et al.*, 2009). The most recent and largest GWAS of the PGC's schizophrenia working group was published in 2014, in which 83 novel loci among 108 were identified with genome-wide significant association with schizophrenia, (Ripke *et al.*, 2014). Within these loci were 128 independent sites of association. The top hit in this study was a broad region within the MHC ($p = 3.48 \times 10^{-31}$), adding to growing evidence of a relationship between the immune system and schizophrenia. Furthermore, the associated SNPs were found to be significantly enriched for sites that act as enhancers in immune-related tissues. Recently, the relationship between immune function and schizophrenia has been partially explained by variation at the *complement component 4* (C4) genes. The C4 locus consists of two isotypes located in tandem at the MHC: *C4A* and *C4B*, both of which are present in either a long and/or short form. Sekar *et al.* (2016) reported that variation at the C4 locus was associated with differential gene expression levels of *C4A* and *C4B* across multiple brain regions. Furthermore, they reported an increased risk of schizophrenia associated with common C4 structural variants as well as *C4A* expression levels. Of these common structural variants, the short form of *C4B* posed the lowest risk, while carrying tandem copies of the long form of *C4A* was associated with the greatest risk of the disorder. Upon examining *post-mortem* brains, Sekar *et al.* (2016) observed co-localisation of C4 and pre- and postsynaptic markers. Furthermore, they showed that C4 expression led to synapse elimination *in-vivo*. Complement receptors are expressed in the brain by microglia – this study proposed a pathogenic mechanism whereby interaction between neurons and microglia through C4 result in aberrant synaptic pruning, resulting in the reduction in grey matter and synaptic structures observed in schizophrenia (Cannon *et al.*, 2002; Garey *et al.*, 1998).

1.3.2.2.2 *GWAS findings for bipolar disorder*

Significant associations have also been identified by GWAS for bipolar disorder but, to date, fewer loci have attained genome-wide significant association with the disorder than have been identified in schizophrenia. The first studies performed on bipolar disorder failed to find significant associations at the genome-wide level, with the most significant region of association occurring at chromosome 16p12 (WTCCC, 2007). The following year, Baum *et al.* (2008) reported genome-wide significant association with the disorder at *diacylglycerol kinase eta* (*DGKH*) while Sklar *et al.* (2008) failed to report genome-wide significant findings, identifying *myosin 5B* (*MYO5B*) as their top hit. A meta-analysis of Sklar *et al.* (2008) and the WTCCC study by Ferreira *et al.* (2008) reported genome-wide significant associations with bipolar disorder at two genes: *calcium channel, voltage-dependent, L-type, alpha 1C subunit* (*CACNA1C*) and *ankyrin 3* (*ANK3*). Associations have subsequently been identified between *CACNA1C* and schizophrenia and recurrent major depressive disorder (Green *et al.*, 2010). The bipolar disorder working group of the PGC reported association between bipolar disorder and a novel variant in an intronic region of *protein odd oz/ten-m homolog 4* (*ODZ4*), as well as supporting Ferrerira *et al.*'s (2008) finding of association with *CACNA1C* (Sklar *et al.*, 2011). A study by Cichon *et al.* (2011) identified significant association between bipolar disorder and the *neurocan* (*NCAN*) gene, in the MooDS consortium. Although not significant at the genome-wide level, they also identified an intronic SNP within the cell cycle-related gene m*itotic spindle assembly checkpoint protein MAD1* (*MAD1L1*) as the next most significant hit. Significant association has since been reported between *MAD1L1* and schizophrenia in the PGC's most recent GWAS (Ripke *et al.*, 2014). Furthermore, Ruderfer *et al.* (2014) reported genome-wide significant association between *MAD1L1* and schizophrenia and bipolar disorder combined. Recently, Hu *et al.* (2016) reported genome-wide significant association between *MAD1L1* and bipolar disorder alone. A meta-analysis by Chen *et al.* (2013) found support for association between *ANK3* and bipolar disorder along with significant association with *tetratricopeptide repeat and ankyrin repeat containing 1* (*TRANK1*). Using a combined population from the PGC bipolar disorder working group and the MooDS consortium, Mühleisen *et al.* (2014) reported association between bipolar disorder two novel loci: *adenylyl cyclase type 2* (*ADCY2*), and a region between *microRNA 2113* (*MIR2113*) and *POU class 3*

homeobox 2 (*POU3F2*). The same study also supported previous findings, reporting association between bipolar disorder and *ANK3*, *ODZ4* and *TRANK1*. The most recent genome-wide significant findings in bipolar disorder have come from Hou *et al.* (2016). Here, they identified six genome-wide significant loci, of which two were novel: an intergenic region on chromosome 9, and a region within *erb-b2 receptor tyrosine kinase 2* (*ERBB2*). However, limitations to this study include a small sample size and unscreened controls, presenting potential issues with power. The authors also cautioned that the identification of previously-associated loci should not be considered replications, as many of the cases analysed had been included in previous studies.

1.3.2.2.3 *GWAS findings for MDD*

Unlike schizophrenia and bipolar disorder, which have generally been successful in identifying genome-wide significant results by GWAS; it was not until 2015 that genome-wide significant associations were identified for MDD. The CONVERGE consortium reported two loci attaining genome-wide significant association with MDD: one upstream of the chromosome 10 gene *sirtuin 1* (*SIRT1*), and the other in an intronic region of *phospholysine phosphohistidine inorganic pyrophosphate phosphatase* (*LHPP*; Cai *et al.*, 2015). The authors credited their success to the homogeneity of their sample, restricting their analysis to females of Han Chinese ethnicity with severe (hospitalised) recurrent MDD. As mentioned in section 1.1.3, MDD is a highly heterogeneous disorder in terms of clinical presentation. Any combination of five out of nine symptoms are required to incur a diagnosis, according to current DSM-V criteria (i.e. 227 possible combinations). Such heterogeneity may be responsible for noise and a reduction in the statistical power to dissect the underlying genetic architecture of the disorder. Therefore, stratification by gender and clinical presentation, for example, may be the key to identifying more genome-wide significant associations for MDD. More recently, Hyde *et al.* (2016) reported significant association with MDD in a population of European ancestry. Here, they availed of genotype data from over 450,000 consumers of 23andMe and self-reported phenotypes. Upon meta-analysis of 23andMe data and PGC data (Ripke *et al.*, 2013), the authors reported 15 significantly associated loci. While Cai *et al.* (2016) were

successful in identifying genetic associations with MDD through reducing heterogeneity, the success of Hyde *et al*. (2016) was likely attributable to the size of the cohort studied.

As well as individual successes in GWAS, a substantial degree of genetic overlap between these disorders has also been reported. This is best reflected in a GWAS performed by the PGC cross-disorder working group in which four genome-wide significant associations were identified between MDD, bipolar disorder, schizophrenia, ADHD and autism spectrum disorder combined (Smoller *et al.*, 2013). Two of their associations were within calcium-channel genes: the previously-reported gene *CACNA1C*, as well as *calcium voltage-gated channel auxiliary subunit beta 2* (*CACNB2*). The top hit occurred on chromosome 3p and overlapped with findings of a combined bipolar and schizophrenia study (*ITIH3*; Ripke *et al.*, 2011). A fourth region on chromosome 10 overlapped multiple genes, with the strongest signal mapping to an intronic site in *arsenite methyltransferase* (*AS3MT*).

1.3.2.3 *Evidence for a polygenic basis for psychiatric illness*

Associations from GWAS consist of common alleles, each of small effects. It is likely that such alleles, when inherited together, contribute to a polygenic burden of risk of illness. Gottesman and Shields (1967) had previously proposed a polygenic model for the risk of schizophrenia. More recently, polygenic scoring has been useful in assessing the collective effects of risk genes. Purcell *et al.* (2009) demonstrated the first successful application of polygenic scoring to psychiatric GWAS data, calculating polygenic risks for schizophrenia. They reported that although few markers were significantly associated with illness at the individual level, when taken collectively to calculate risk scores, these scores were significantly higher in schizophrenia cases compared to controls. This was also observed in bipolar disorder using polygenic risk scores calculated from schizophrenia GWAS data, indicative of a shared genetic component between the two disorders. More recently, Purcell *et al.* (2014) showed association between rare, deleterious coding variants and schizophrenia. Here, while no individual gene was associated with the disorder, a greater burden of these variants

was observed in cases compared to controls. Although there are limited findings of single gene associations for MDD, polygenic approaches have had some success. Using polygenic risk scores for psychiatric traits from PGC data (Ripke *et al.*, 2013; Sklar *et al.*, 2011; Ripke *et al.*, 2014), Milaneschi *et al.* (2016) reported a significant association between typical MDD and schizophrenia polygenic risk scores in a cohort of 3230 Dutch patients. These findings are consistent with reports of genetic overlap between psychiatric disorders (Smoller *et al.*, 2013). Polygenic scoring has also proven useful in the dissection of the overlapping genetic architecture between psychiatric disorders: Ruderfer *et al.* (2014) created a polygenic risk score capable of significantly discriminating between bipolar disorder and schizophrenia.

Over 90% of disease-associated GWAS variants are located in non-coding regions such as intronic or intergenic regions (Maurano *et al.*, 2012). This may point to an important role for regulatory variants in the pathology of psychiatric disorders. A summary of genome-wide significant association findings to date for schizophrenia, bipolar disorder and MDD is presented in Table 1.1. It is likely that the majority of findings from these studies are not the causal variants but are instead tagging the variant or variants of interest – an important factor to consider in downstream analyses based on these data. Furthermore, it is important to note that the heritability of these disorders will be further explained by rare genetic variation, gene-gene interactions (epistasis), and gene-environment interactions.

Study	Disease(s) of Interest	Findings
Baum *et al.* 2008a	Bipolar disorder	Reported genome-wide significant association between bipolar disorder and *DGHK* in two independent samples.
O'Donovan *et al.*, 2008	Schizophrenia and schizophrenia and bipolar disorder combined	Reported strong evidence for association between *ZNF804A* and schizophrenia, attaining genome-wide significance when both schizophrenia and bipolar disorder were considered.
Ferreira *et al.*, 2008	Bipolar disorder	First study to report genome-wide significant associations with bipolar disorder implicating *CACNA1C* and *ANK3*.
Steffanson *et al.*, 2009	Schizophrenia	Reported genome-wide significant association with schizophrenia at loci in the major histocompatibility complex (MHC), *NRGN*, and *TCF4*.

Study	Disease(s) of Interest	Findings
Rietschel *et al.*, 2011	Schizophrenia	Reported genome-wide significant association between schizophrenia and four SNPs in *AMBRA1*, on chromosome 11.
Shi *et al.*, 2011	Schizophrenia	Two schizophrenia-associated loci identified: one on chromosome 1q24 (*BRP44*) and another on chromosome 8p12 (*LSM1/WHSC1L1*)
Steinberg *et al.*, 2011	Schizophrenia	Found evidence in support of association between schizophrenia and *NRGN* and *TCF4*. Identified two novel loci associated with schizophrenia at *CCDC68* and *VRK2*
Cichon *et al.*, 2011	Bipolar disorder	Reported genome-wide association between bipolar disorder and *NCAN*.

Study	Disease(s) of Interest	Findings
PGC (Ripke *et al.*, 2011)	Schizophrenia and schizophrenia and bipolar disorder combined	Seven schizophrenia-associated loci identified, five of which were novel and mapped to six genes (*MIR137*, *PCGEM1*, *CSMD1*, *MMP16*, *CNNM2* and *NT5C2*). Confirmed association between schizophrenia and *TRIM26* and *CCDC68*. Also reported association between schizophrenia and bipolar disorder and *CACNA1C*, *ANK3* and *ITIH3-ITIH4*, all previously associated with bipolar disorder.
PGC (Sklar *et al.*, 2011)	Bipolar disorder and schizophrenia and bipolar disorder combined	Confirmed evidence for association between bipolar disorder and *CACNA1C*. Identified a novel susceptibility locus at *ODZ4*. Reported association between schizophrenia and bipolar disorder combined and *NEK4, CACNA1C* and a multi-gene region spanning *ITIH-1, -3* and *-4*.

Study	Disease(s) of Interest	Findings
PGC (Smoller *et al.*, 2013)	Autism spectrum disorder, attention deficit-hyperactivity disorder, bipolar disorder, major depressive disorder and schizophrenia	Four genome-wide significant loci identified across all disorders (*ITIH3*, *AS3MT*, *CACNA1C* and *CACNB2*), two of these had previously been associated with schizophrenia and/or bipolar disorder (*ITIH3, CACNA1C*).
Chen *et al.* (2013)	Bipolar disorder	Genome wide significant association with bipolar disorder reported near *TRANK1, LMAN2L* and *PTGFR*. Also provided support for association with *ANK3*.
Ripke *et al.* (2013)	Schizophrenia	13 novel loci identified among 22. The nine remaining loci had previously been associated with schizophrenia and/or bipolar disorder.
Ruderfer *et al.* (2014)	Bipolar disorder and schizophrenia	Identified a novel association between both disorders and (*PIK3C2A*), as well as five previously-identified loci (*TRANK1*, MHC, *MAD1L1*, and *CACNA1C*)

Study	Disease(s) of Interest	Findings
PGC (Ripke *et al.*, 2014)	Schizophrenia	108 genome-wide significant loci consisting of intergenic regions, single genes and multiple genes. The top hit was a broad 400 kb region on chromosome 6, within the MHC
Mühleisen *et al.*, 2014	Bipolar disorder	Five genome-wide significant loci were identified, two of which were novel (*ADCY2* and *MIR2113-POU3F2*). Also confirmed association between bipolar disorder and *ANK3*, *ODZ4* and *TRANK1*.
CONVERGE Consortium (Cai *et al.*, 2015)	MDD	First report of genome-wide significant association in MDD. Two genome-wide significant loci identified on chromosome 10 located 5' of *SIRT1* and within an intron of *LHPP*.
Hou *et al.*, 2016	Bipolar disorder	Two novel genome-wide significant loci were identified: *ERBB2* and an intergenic region on chromosome 9. Reported the first association between *MAD1L1* and bipolar disorder only.

Study	Disease(s) of Interest	Findings
Hyde *et al.*, 2016	MDD	15 novel genome-wide significant loci were identified in a meta-analysis of a previous GWAS of MDD by the PGC (Ripke *et al.*, 2013) and consumer genomic data from 23andMe.

Table 1.1: Summary of genome-wide significant findings for schizophrenia, bipolar disorder and MDD identified by GWAS (2008-2016).

Table summarises GWAS of psychiatric disorders in which genome-wide significant results have been reported based on a p-value threshold of 5 x 10^{-8} (as suggested by Pe'er et al., 2008). "Study" column provides the references to each study, column labelled "Disease(s) of Interest" refers to the disease or diseases under investigation in each study while the column labelled "Findings" summarises the genome-wide significant disease-associated findings of each study.

1.3.3 Chromosomal structural abnormalities

Chromosomal abnormalities such as inversions, duplications, deletions and translocations have been linked to several disorders, including neuropsychiatric disorders (reviewed in Muir *et al.*, 2006, Brand *et al.*, 2014). A relatively frequent deletion at chromosome 22q11 is associated with velocardiofacial and DiGeorge syndromes and occurs in approximately 1 in 4,000 individuals, yielding variable phenotypes (Ou *et al.*, 2008). The most common psychiatric disorder associated with 22q11 deletion syndrome is schizophrenia (Bassett *et al.*, 2005). Among the genes within the affected region is *COMT*, a previously-described risk factor for schizophrenia (section 1.3.2.1).

Xu *et al.* (2009) examined *de novo* CNVs in 359 trios from an Afrikaner population, reporting that sporadic cases of schizophrenia had approximately eight times more *de novo* CNVs than unaffected controls ($p = 0.0008$). A similar observation was made by Walsh *et al.* (2008), in that CNVs were present at a significantly higher rate in schizophrenia cases (22/150; $p = 0.0008$) - particularly early-onset cases (15/76; $p = 0.0001$) than in unaffected controls (13/268). Pathway analysis showed these schizophrenia-specific CNV genes were significantly over-represented in neurodevelopmental processes while no significant over-representations were observed for the CNV genes identified in controls.

The role for CNVs in bipolar disorder has been less clear. As previously observed in schizophrenia, Malhotra *et al.* (2009) reported a significantly higher rate of *de novo* CNVs in bipolar cases (8/185) compared to unaffected controls (4/426; $p = 0.009$). Grozeva *et al.* (2010) examined large (>100 kb), rare (frequency < 0.01) CNVs among bipolar disorder cases in the WTCCC finding no significant difference in the burden of CNVs between cases and unaffected controls. They compared their findings to those of Kirov *et al.* (2009) and found that the burden of CNVs in bipolar disorder was significantly less than that observed in schizophrenia. A meta-analysis by Malhotra and Sebat (2012) reported few overlaps between CNVs that contribute to schizophrenia and those that contribute to bipolar disorder and recurrent MDD.

Green *et al.* (2016) examined CNVs among the Bipolar Disorder Research Network sample and found significant evidence for association between bipolar disorder and a CNV at 16p11.2: a region also associated with schizophrenia through CNV analysis (Rees *et al.*, 2014). Similar to Grozeva *et al.* (2010), Green *et al.* (2016) reported lesser association between large CNVs and bipolar disorder compared to the association observed between large CNVs and schizophrenia (Kirov *et al.*, 2009).

Along with inversions, duplications and deletions, there have been multiple reports of chromosomal translocations linked to psychiatric illness. Overhauser *et al.* (1997) reported a female with schizoaffective disorder possessing a balanced translocation between chromosomes 14 and 18 t(14;18)(q11.2;q21.1). The affected region of chromosome 18 contains the gene *MYO5B*. Suggestive association with this gene was reported by Sklar *et al.* (2008) as the top hit in a GWAS of bipolar disorder though it was not genome-wide significant ($p = 1.66 \times 10^{-7}$). Rajkumar *et al.* (2015) cross-referenced the Danish psychiatric case register with the Danish cytogenetic case register leading to the identification of an individual with both bipolar disorder and a translocation between chromosomes 9 and 17 t(9;17)(q33.2;q25.3). Four genes are present at the chromosome 17 breakpoint region of this translocation: *raptor* (*RPTOR*), *endonuclease V* (*ENDOV*), *neuronal pentraxin I* (*NPTX1*) and *ring finger protein 213* (*RNF213*). The same study examined whether these genes were associated with the disorder in the PGC GWAS of bipolar disorder (Sklar *et al.*, 2011). Nominally significant associations were reported for *RPTOR* and *NPTX1*. Following Bonferroni correction, only *NPTX1* remained significant.

A balanced translocation between chromosomes 1 and 11 t(1;11)(q42.1;q14.3) has been linked to schizophrenia, bipolar disorder and recurrent major depressive disorder in a large Scottish pedigree (St. Clair *et al.*, 1990; Blackwood *et al*., 2001). This translocation is the main focus of the work presented in this thesis and is described in greater detail in section 1.4.

Challenges remain in identifying the susceptibility factor(s) associated with structural abnormalities implicated in psychiatric illness. It is possible that a

phenotype may be linked to a single gene disrupted by a structural variant, while others may be linked to a multitude of genes within a CNV or chromosomal rearrangement. Investigation of downstream consequences such as expression or epigenetic changes may be useful in identifying a mechanism whereby structural variants confer increased risk for illness. Moreover, variants in the DNA sequence surrounding such structural abnormalities may play an important role in pathogenesis, should these structural variants be co-inherited with flanking regions containing risk factors.

1.4 The t(1;11) translocation and disrupted in schizophrenia-1

1.4.1 Overview

As discussed in the previous section, there is a range of evidence linking chromosomal structural abnormalities to the risk of developing psychiatric illness. A further example of such structural variation in psychiatric illness is the t(1;11)(q42.1;q14.3) translocation. This was first identified in a single individual through a cytogenetic survey of males in young offenders' institutions (Jacobs *et al.*, 1970). Upon further investigation across four generations of the proband's family, 34 members out of 77 tested were found to carry the translocation. Of these 34 translocation carriers, 16 also had a psychiatric diagnosis in contrast to five translocation non-carriers with a diagnosis (St Clair *et al.,* 1990). The regions surrounding the translocation breakpoints were subsequently cloned, identifying two novel genes on chromosome 1: a large protein-coding gene on the sense strand, and a non-coding RNA on the antisense strand. These were termed *disrupted in schizophrenia-1* and *-2*, respectively (*DISC1* and *DISC2*; Millar *et al.*, 2000). In addition, the translocation also disrupts a non-coding transcript on chromosome 11, termed *boymaw*, or *DISC1 fusion partner-1* (*DISC1FP1*; Zhou *et al.*, 2010; Eykelenboom *et al.*, 2012). Further follow-up was performed by linkage analysis on the family in which additional members were karyotyped and diagnoses ascertained using DSM-IV criteria (Blackwood *et al.*, 2001). Significant linkage was observed between schizophrenia and the translocation with a LOD score of 3.6 between 37 t(1;11) carriers and 50 non-carriers. When considering an affected phenotype to

include schizophrenia, bipolar disorder and recurrent MDD, a maximum LOD score of 7.1 was obtained. The same study also observed significant reductions in P300 event-related potential (ERP) in t(1;11) carriers compared to non-carrying controls: an endophenotype of schizophrenia associated with cognitive function. A recent follow-up of the family sought to confirm linkage between the translocation and illness following recruitment of additional members. Significant linkage was observed between the translocation and a clinical phenotype including schizophrenia, schizoaffective disorder, bipolar disorder and recurrent MDD (LOD = 6.1; Thomson *et al.*, 2016).

1.4.2 DISC1 interactions

A potential mechanism for the role of DISC1 in psychiatric illness was proposed by Millar *et al.* (2005), who identified an interaction between the DISC1 protein and a key regulator of the cAMP signalling pathway, phosphodiesterase 4B (PDE4B). Here, they observed a reduction in this DISC1-PDE4B interaction following induction of cAMP signalling in SH-SY5Y cells. The authors suggested that DISC1 plays a regulatory role in cAMP signalling by releasing PDE4B in the presence of cAMP activity. The same study also reported haploinsufficiency of DISC1 in t(1;11) carriers in patient-derived lymphoblastoid cell lines (LCLs). This might disrupt levels of DISC1-PDE4B interaction resulting in subsequent dysregulation of the cAMP signalling cascade: a pathway implicated in learning, memory and mood (Bauman *et al.*, 2003).

Along with PDE4B, evidence has emerged for several protein-protein interactions involving DISC1. Millar *et al*. (2003) identified several DISC1 protein interactors through a yeast-2-hybrid (Y2H) screen, including activating transcription factor 4 (ATF4). Malavasi *et al.* (2012) later demonstrated ATF4-mediated gene expression was affected by missense variants in DISC1 (R37W and L607F). The DISC1-ATF4 complex was reported by Soda *et al.* (2013) to suppress transcription of *PDE4D9*. Furthermore, they reported dopamine receptor activation resulted in *PDE4D9* upregulation: a process which coincided with DISC1 dissociating from the *PDE4D9*

locus. These findings suggest a role for DISC1 in dopaminergic signalling: a pathway heavily implicated in psychiatric disorders as described in section 1.2.1.

A yeast-2-hybrid study performed by Camargo *et al.* (2007) identified a network of protein interactions centred on DISC1: the DISC1 interactome. This interactome comprised a network of 127 proteins and 158 interactions. Ontology analysis revealed terms relating to intracellular transport, cell cycle and division, and regulation and organisation of the cytoskeleton were over-represented in this network (Camargo *et al*., 2007).

Several of DISC1's protein interactors have themselves been implicated in psychiatric illness and have neurodevelopmental and synaptic functions (Brandon and Sawa, 2011). For example, DISC1 interacts with nuclear distribution element 1 (NDE1) and its orthologue NDE-like 1 (NDEL1; Ozeki *et al.*, 2003; Morris *et al.*, 2003). NDE1 is located on chromosome 16p, a region in which CNVs have been implicated in several neuropsychiatric disorders, including schizophrenia (Ramalingam *et al.*, 2011; Ingason *et al.*, 2011). NDE1 and NDEL1 form a complex with lissencephaly 1 (LIS1) to which the DISC1-PDE4B complex binds, mediating phosphorylation of NDE1 at threonine residue 131 (T131; Bradshaw *et al.*, 2011). Bradshaw *et al.* (2011) also demonstrated that T131 phosphorylation affects neurite outgrowth *in-vitro*, supporting a neurodevelopmental role for DISC1.

1.4.3 DISC1 and neurodevelopment

There are several lines of evidence linking the neurodevelopmental theory of schizophrenia and *DISC1*. Callicott *et al.* (2005) reported significant association between a three SNP haplotype in *DISC1* and schizophrenia. Of the three SNPs

making up this haplotype, a non-synonymous variant (Ser704Cys; rs821616) displayed the strongest evidence for association and was selected for investigation for association with hippocampal formation (HF) structure and function. A significant association was observed between Ser704Cys and HF grey matter volume, with reduced volume observed in Cys homozygotes. Kamiya *et al.* (2005) observed disrupted neuronal migration in the developing cortex when *Disc1* was knocked down by RNAi in mouse embryos. A similar phenotype in embryonic mouse brains was observed when overexpressing a C-terminal truncated form of *Disc1* to model the effects of the t(1;11) translocation on *Disc1*. The same study additionally showed that knockdown of *DISC1* and expression of the truncated form each led in impaired neurite outgrowth *in-vitro*. Duan *et al.* (2007) examined the role of DISC1 in adult hippocampal neurogenesis noting an increase in the speed of neuronal integration following *DISC1* knockdown, and, consequently, altered dendritic development and positioning in adult-born neurons. DISC1 has also been shown to play a key role in the proliferation of neural progenitor cells (NPCs). Mao *et al.* (2009) reported reduced proliferation and migration of adult hippocampal progenitors following *DISC1* knock-down, as well as premature differentiation. These deficiencies in proliferation were found to occur through the Wnt signalling pathway, mediated by a direct interaction between DISC1 and GSK3β. This inhibitory interaction is similar to that between lithium chloride - a widely-used treatment for bipolar disorder – and GSK3β. However, whether or not the inhibitory mechanisms of DISC1 and lithium are similar has yet to be determined. Ishizuka *et al.* (2011) proposed a mechanism whereby DISC1 mediates a transition between neuronal migration and proliferation through phosphorylation at serine 710 (S710) of DISC1. Here, they observed an increase in the interaction between DISC1 and Bardet-Biedle syndrome proteins (BBS1 and BBS4) was correlated with DISC1 S710 phosphorylation. They also demonstrated that the unphosphorylated S710 residue results in a greater affinity between DISC1 and GSK3β, thereby regulating proliferation. Phosphorylation of this residue was shown to result a decreased interaction between DISC1 and GSK3β, coinciding with DISC1-BBS1 interaction and initiation of neuronal migration. Interaction between DISC1 and BBS1 in corticogenesis has previously been demonstrated by Kamiya *et al.* (2008).

1.4.4 DISC1 and synaptic function

In addition to evidence for a neurodevelopmental role of DISC1, multiple lines of evidence exist linking DISC1 to synaptic function, which is hypothesised to be altered in major mental illness (van Spronsen and Hoogenraad, 2010). A study by Cannon *et al.* (2005) reported reduced grey matter density in the prefrontal cortex in carriers of haplotypes within *DISC1* and *translin-associated factor X* (*TSNAX*), which lies upstream of *DISC1*. *Post-mortem* analyses have demonstrated that DISC1 localises to synapses (Kirkpatrick *et al*., 2006), while analysis of rat brain subcellular fractions has shown DISC1 to be present in the postsynaptic density (PSD), where it interacts with Traf2 and Nck-interacting kinase (TNIK) to regulate the levels of PSD proteins (Wang *et al.*, 2011). Additional roles for DISC1 have been identified in regulating the structure of synapses. Following RNAi-mediated knock-down of DISC1, Hayashi-Takagi *et al.* (2010) reported an increase in size and number of dendritic spines. This was found to be related to the interaction between DISC1 and Kalirin-7 (Kal-7), a protein with a well-established role in synaptic development and neurological disorders (Mandela and Ma, 2012). Interaction between Kal-7 and Neuroligin 1 (NL1), a postsynaptic cell adhesion molecule (CAM), was later demonstrated to be regulated by DISC1, with reduced Kal-7-NL1 binding observed in the presence of DISC1 (Owczarek *et al.*, 2015).

1.4.5 DISC1 and mitochondrial trafficking

More recently, DISC1 has been implicated in mitochondrial trafficking, through its interaction with Trafficking Protein, Kinesin Binding 1 (TRAK1; Ogawa *et al.*, 2014). NDE1 and GSK3β have also been shown to interact with TRAK1 in neuronal mitochondrial trafficking complexes (Ogawa *et al.*, 2016). Efficient transport of mitochondria is essential in neurons in order to meet the high-energy requirements of processes such as neurotransmission (Vos *et al.*, 2010). Disruption of mitochondrial trafficking and localisation may therefore be a contributory mechanism for psychiatric disorders (Manji *et al.*, 2012).

1.4.6 The Disc1 L100P mouse

Additional support for the role of DISC1 in psychiatric illness comes from animal studies. Several studies have been published on mutant *Disc1* mouse models, reporting behavioural and neurodevelopmental phenotypes suggestive of psychiatric illness. A study by Clapcote *et al.* (2007) reported two mouse models with missense mutations in *Disc1* (Q31L and L100P). These mice have reduced brain volumes compared to wild-type, supporting a neurodevelopmental role for *Disc1*. Behavioural characterisation of these mice revealed endophenotypes of schizophrenia, including deficits in latent inhibition and acoustic startle response (Geyer and Ellenbroek, 2003). The 31L mutants were found to display depressive-like behaviours, with increased durations of immobility observed in these mice during the forced swim test (David *et al.*, 2003). The schizophrenic-like behaviours were reversed following antipsychotic treatments while depressive behaviours were altered by the antidepressant bupropion. Shoji *et al.* (2012), however, failed to observe these findings in the same model. These inconsistencies may have been due to variable laboratory environments and mixed genetic backgrounds of the mice: both of which are common limitations of animal studies.

Lipina *et al.* (2012) reported rescue of schizophrenic-like behaviours in 100P mice following administration of valproic acid, an anticonvulsant used in the treatment of bipolar disorder. Furthermore, the same study reported a rescuing effect of valproic acid on 13/61 100P-associated gene expression changes in the hippocampus, suggesting these genes might be related to illness through the disruption of *Disc1*. Others have also reported gene dysregulation in the 100P mouse. Brown *et al.* (2011) reported misexpression of presynaptic CAMs Neurexin-1 and -3 (*Nrxn1* and *Nrxn3*) at various stages of brain development in the mouse, supporting a role for *Disc1* in synaptic function. Multiple independent studies have observed hyperlocomotor activity in the 100P mouse (Clapcote *et al.*, 2007; Lipina *et al.*, 2010; Shoji *et al.*, 2012). This phenotype is significantly enhanced by amphetamine in L100P mice compared to wild-type, and is reduced following antipsychotic treatment, suggestive of increased dopamine receptor sensitivity (Lipina *et al.*, 2010). Dopamine dysfunction has also been observed in a *Disc1* mouse model expressing a dominant

negative C-terminal truncated form of human DISC1. Jaaro-Peled *et al.* (2013) reported a reduction in extracellular dopamine levels along with an increase in dopamine receptors in the brains of these mice, compared to wild-type. Structural abnormalities have also been reported in these mice, with mutants possessing enlarged ventricles – a well-established endophenotype of schizophrenia (Hikida *et al.*, 2007; Johnstone *et al.*, 1976). Taken together, the findings from current *Disc1* mouse models provide support for the gene's role in neurodevelopment and signalling.

1.4.7 DISC1 and psychiatric illness

Numerous functional and genetic studies provide support a role for *DISC1* in psychiatric illness. As well as the t(1;11) family, a second family has been reported with a mutation in *DISC1* linked to psychiatric illness. Here, three siblings carrying a 4 bp deletion in DISC1 were reported as suffering from either schizophrenia or schizoaffective disorder, while their mutation-carrying father was reported as unaffected (Sachs *et al.*, 2005). However, the same mutation was not detected in 655 individuals, but was present in two non-psychiatrically screened controls out of 694 (Green *et al.*, 2006).

To date, *DISC1* has not been associated with any psychiatric disorders through GWAS. This has been a source of criticism by some, who have questioned its validity as a risk factor for major mental illness (Sullivan, 2013). Porteous *et al.* (2014) have argued that the lack of genome-wide significant association between *DISC1* and its related disorders is in keeping with the 'common disease; rare variant' hypothesis, citing well-established Alzheimer's disease risk factors *amyloid precursor protein* (*APP*), and *presenelin-1* and *-2* (*PSEN1*, *PSEN2*) as examples - all genes that have to date not been implicated in the disease by GWAS, despite the high penetrance of their mutated forms in the disorder (Van Cauwenberghe *et al.*, 2015). An additional point of debate has been provided by Niculescu (2014), who suggested that the greater genetic diversity observed in biologically important genes (such as *DISC1*) calls for whole gene-based experimental approaches in order to produce replicable

results, rather than SNP-level approaches as seen in GWAS. Furthermore, Thomson *et al*. (2014) have reported additional evidence for common and rare variants in DISC1 associated with recurrent MDD. Of the variants reported, R37W is of particular note as it was among five variants present only in affected individuals. The R37W mutation occurs within a PDE4B binding site (Murdoch *et al.*, 2007) and has been reported to impair DISC1-mediated anterograde mitochondrial trafficking in neurons (Ogawa *et al.*, 2014). A summary of DI SC1 structure, variants and binding sites is presented in Figure 1.2.

1.4.8 The t(1;11) translocation

Due to its status as a protein coding gene with important roles in brain development and function, *DISC1* has been an attractive candidate to explain the pathogenic properties of the t(1;11) translocation. However, it is important to note that *DISC1* is not the only genetic element perturbed by the translocation. It is possible that the psychiatric phenotype in the t(1;11) family occurs not just through disruption of *DISC1*, but through *DISC2, DISCFP1/Boymaw*, generation of fusion transcripts, and/or the co-inheritance of functional or regulatory elements in linkage disequilibrium with the translocation.

The translocation generates fusion transcripts consisting of the 5' region of *DISC1* and the 3' region of *DISC1FP1*. These transcripts encode fusion proteins which localise to mitochondria in t(1;11) family LCLs, with potentially deleterious effects (Eykelenboom *et al.*, 2012). Zhou *et al.* (2010) had previously reported that such transcripts generated insoluble fusion proteins in HEK293 cells. Eykelenboom *et al.* (2012) observed similar insolubility of fusion proteins in COS7 cells. Insolubility of DISC1 may be a pathological mechanism for psychiatric disorders: others have reported DISC1 protein aggregates in *post-mortem* brains of bipolar disorder and schizophrenia patients (Leliveld *et al.*, 2008). Expression of these fusion transcripts has been found to affect protein translation, with reduced levels of synaptic proteins

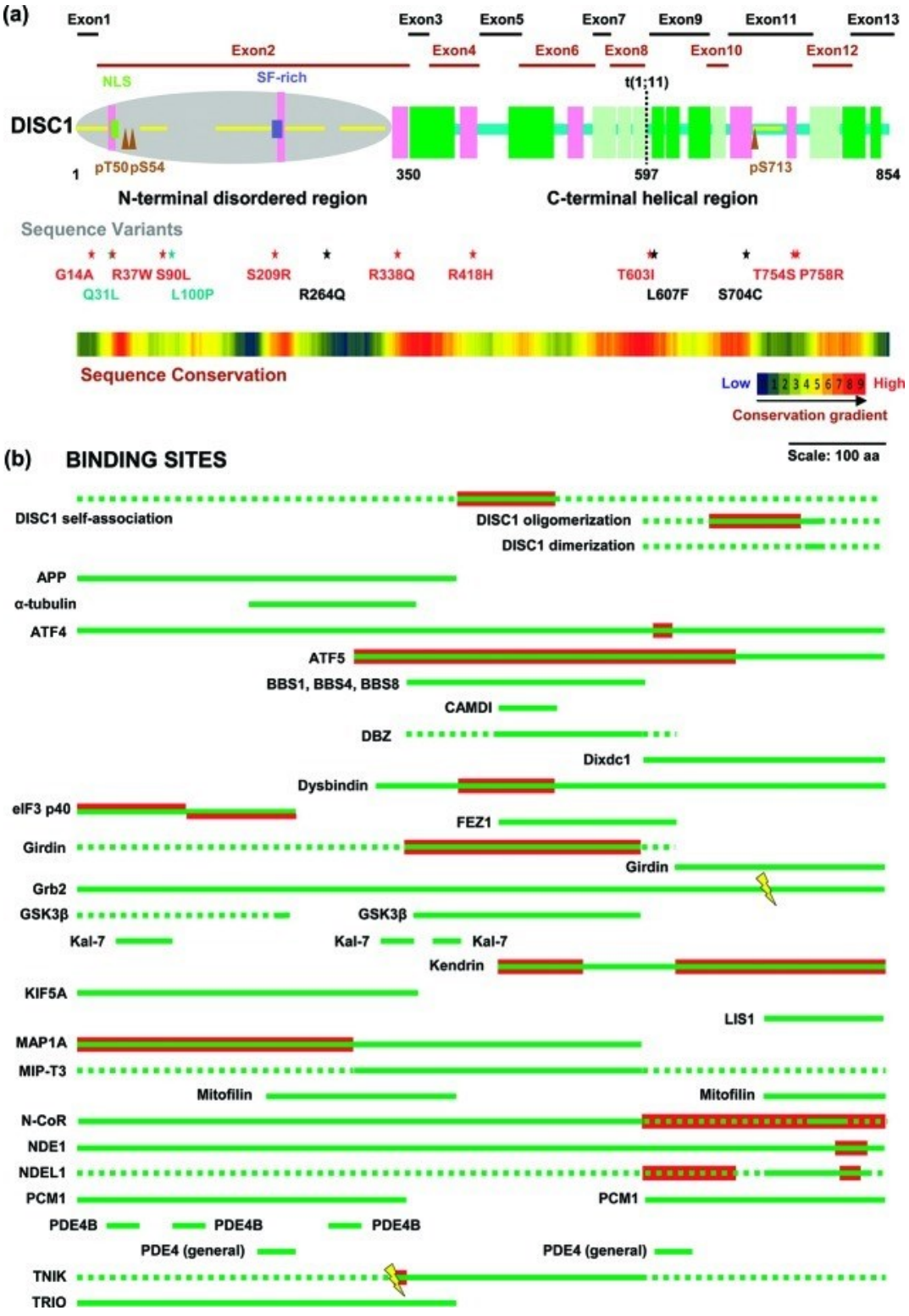


Figure 1.2: DISC1 structure, mutations and binding sites

Figure summarises the genetic and protein structure of DISC1*, including known mutations and sequence conservation (a), along with binding sites for various DISC1 protein interactors (b; Soares* et al.*, 2011)*

observed both *in-vivo* and *in-vitro* (Baohu *et al.*, 2014). Taken together, these findings might indicate a pathogenic role for fusion proteins in t(1;11) carriers.

1.4.9 Recent and ongoing work on the t(1;11) family

Analysis of the t(1;11) family is ongoing. Additional family members have been recruited and have been the subject of linkage, imaging and methylation analyses. Imaging analyses have recently been performed identifying brain structural abnormalities in t(1;11) carriers. Doyle *et al*. (2015) reported reduced cortical thickness in t(1;11) carriers, resembling the phenotypes observed schizophrenia patients. The same study demonstrated a t(1;11) status classification accuracy of 73%, using cortical thickness measures. Furthermore, Whalley *et al.* (2015) reported a reduction in white matter integrity in t(1;11) carriers compared to non-carriers: a phenotype also observed in a group of individuals with psychotic illness compared to healthy controls. In addition to the updated linkage analysis, Thomson *et al.* (2016) reported reduced cortical thickness and gyral folding in t(1;11) carriers. These findings suggest the translocation might impact upon neurodevelopment, conferring an increased risk of illness in t(1;11) carriers. Thomson *et al.* (2016) also reported a reduction in glutamate levels in the dorsolateral prefrontal cortex of t(1;11) carriers: a phenotype also seen in patients with mood disorders (Yüksel and Öngür, 2010).

Study materials from the family include whole blood-derived DNA, LCL-derived RNA, and induced pluripotent stem cell (iPSC)-derived neurons. These have been subjected to expression, methylation and functional studies, some of which will be described further in this thesis.

1.5 The Sortilin gene family

The Sortilin gene family consists of five members: *Sortilin-1*, *Sortilin-related receptor, LDLR class A repeats*, and *Sortilin related VPS10 domain containing receptor 1-3* (*SORT1*, *SORL1*, and *SORCS1-3*, respectively). These genes are

expressed in the developing and mature central nervous system where they play a regulatory role in neuronal viability and function (Nykjaer and Willnow, 2012).

1.5.1 Sortilin genes and psychiatric disorders

Several lines of evidence exist linking Sortilin genes to psychiatric disorders including Alzheimer's disease, schizophrenia, bipolar disorder and MDD. *SORCS2* is located within a region of chromosome 4p, which has been implicated in bipolar disorder through linkage studies (Blackwood *et al.*, 1996; Christoforou *et al.*, 2007). Furthermore, in a GWAS of bipolar disorder, Baum *et al.* (2008a) reported significant association between *SORCS2* and the disorder. In a meta-analysis of bipolar disorder GWAS by Baum *et al.*, (2008a) and the WTCCC (2007), a trend towards significance was reported for *SORCS2* (p = 0.054 Baum *et al.*, 2008b) Ollila *et al.* (2009) attempted to replicate findings by the three GWASs of bipolar disorder performed to date (WTCCC, 2007; Baum *et al.*, 2008a; Sklar *et al.*, 2008), reporting significant association between *SORCS2* and bipolar disorder in a Finnish bipolar family cohort. Furthermore, an analysis by Christoforou *et al.*, 2011 showed significant association between *SORCS2* and bipolar disorder comparing 506 bipolar disorder patients to 633 healthy controls in a Scottish population. *SORCS3* has recently been implicated in MDD by GWAS (Hyde *et al.*, 2016) in a European sample from the customer base of 23andMe, as well as in a combined 23andMe and PGC sample.

SORL1 has been implicated in Alzheimer's disease through GWAS at the genome-wide significant level, as well as through candidate gene approaches (Lambert *et al.*, 2013; Rogaeva *et al.*, 2007). A neuropathological feature of Alzheimer's disease is the presence of β-amyloid plaques arising from the amyloidogenic processing of APP. *SORL1* has been shown to regulate processing of APP, suggesting a potential mechanism for the gene in Alzheimer's disease pathology (Gustafsen *et al.*, 2013). *SORL1* has also been implicated in schizophrenia through an analysis of DNA methylation by Montano *et al.* (2016). Here, hypomethylation was reported in *SORL1* in schizophrenia cases compared to healthy controls. Although Alzheimer's disease and schizophrenia are pathologically distinct, there is a symptomatic overlap between

the two disorders, such as psychosis and cognitive dysfunction. In addition to *SORL1*, a regulatory role for *SORCS1* in APP processing and trafficking has been reported by Reitz *et al.* (2011). Based on these findings, Reitz *et al.* (2013) investigated epistatic interactions amongst Sortilin genes in Alzheimer's disease. Here, they reported interactions between *SORCS1* and *SORCS3*, *SORCS2* and *SORCS3*, and *SORCS1* and *SORCS2* were associated with increased risk of Alzheimer's disease ($p \leq 0.0001$).

Gene expression differences have also been observed in Sortilin family members in psychiatric illness. Buttenschøn *et al.* (2015) reported increased *SORT1* expression in depressed individuals when comparing 152 cases to 216 non-depressed controls. Stelzhammer *et al.* (2013) had previously reported downregulation of *SORT1* in the serum of MDD patients following electroconvulsive therapy, along with *BDNF*. In mice, *Sort1* has been linked to APP processing, with a positive correlation observed between *Sort1* expression and non-amyloidogenic processing of APP (Gustafsen *et al.*, 2013). In addition to the analysis of genetic interactions in Alzheimer's disease, Reitz *et al.* (2011) examined expression of Sortilin genes in *post-mortem* brains of Alzheimer's disease patients compared to healthy controls. Here, they observed reduced levels of *SORCS3* in *post-mortem* brains of Alzheimer's disease patients. Previously, Reitz *et al.* (2011) reported reduced levels of *SORCS1 post-mortem* brains of Alzheimer's disease patients while reduced expression of *SORL1* have also been observed in *post-mortem* brains of Alzheimer's disease patients (Rogaeva *et al.*, 2007).

Mouse models have been generated with knock-out mutations in *Sort1*, *Sorl1, SorCS2* and *SorCS3*. Glerup *et al.* (2013) reported dysfunctional dopaminergic activity in *Sorl1*-deficient mice, along with reduced anxiety: a behavioural phenotype related to ADHD. In contrast, *Sort1* knock-out mice have been shown to display anxiety-like behaviour (Ruan *et al.*, 2016). Deficits in synaptic plasticity relating to long term depression have been reported in *SorCS3* null mice, along with deficits in learning and fear memory modulation: features suggestive of anxiety-related brain

activity (Breiderhoff *et al.*, 2013). *SorCS2* knock-out mice display similar deficits in learning and memory (Glerup *et al.*, 2016). Furthermore, behavioural phenotypes reminscient of ADHD and bipolar disorder were observed in these mice, such as hyperactivity and increased risk taking, respectively (Glerup *et al.,* 2014; Glerup *et al.*, 2016).

Suggestive evidence exists for a regulatory relationship between *DISC1* and Sortilin genes. A study by Wen *et al.* (2014) profiled gene expression in stem cell-derived neurons from individuals with a 4bp *DISC1* frameshift mutation and reported misexpression of *SORCS1*, *SORCS2*, and *SORCS3* in mutant cells compared to wild-type cells. Furthermore, a study of gene expression in the developing brains of 100P *Disc1* mutant mice showed dysregulation of *Sort1* and *SorCS2* (Brown *et al.*, unpublished data). A common function between *DISC1* and a subset of Sortilin family members is APP processing: Shahani *et al.*, (2015) reported a regulatory relationship between *DISC1* and APP processing, as has previously been reported for *SORCS1*, *SORT1* and *SORL1* (Reitz *et al.*, 2011; Gustafsen *et al.*, 2012).

Taken together, evidence from genetic studies of animal models suggest the Sortilin gene family have brain-related functions which, when disrupted, may result in a psychiatric phenotype.

1.6 Environmental factors associated with psychiatric disorders

As well as their use in demonstrating the heritability of psychiatric illness; twin and adoption studies have been useful tools for investigating the role of environmental contributors to these disorders. Evidence for an environmental component to psychiatric illness can perhaps best be seen in monozygotic twins discordant for schizophrenia. Despite being genetically identical, the risk of illness in the co-twin of an affected individual is approximately 50% (Gottesman, 1991; Figure 1.1). Further evidence for an environmental component for psychiatric disorders can be illustrated by the increased risk associated with factors such as stress, degree of urbanisation of

birth, and season of birth (Sullivan, 2005). A commonly-studied mechanism of how the environment may interact with a biological system is DNA methylation: a fundamental epigenetic mark.

DNA methylation typically occurs at the 5-position of a cytosine (C) base when present in the context of a cytosine-guanine dinucleotide (CpG), by action of DNA methyltransferase (DNMT) enzymes (Klose and Bird, 2006). DNA methylation is essential for normal developmental processes such as cell fate specification, X-chromosome inactivation and genomic imprinting, and is also observed in disease (Gopalakrishnan *et al.*, 2008). Disorders such as Russel-Silver syndrome, Prader-Willi syndrome, and Angelman syndrome occur as a result of defective genomic imprinting (Lim and Maher, 2009), whilst Rett syndrome is associated with dysfunction of *Methyl CpG-binding protein 2* (*MECP2*) – a gene whose protein product regulates transcription through binding to methylated DNA (Amir *et al.*, 1999). DNA methylation can be influenced by external factors such as diet, smoking, medication and stress (Lim and Song, 2012; Meaney *et al.,* 2005; Klengel *et al.*, 2014), as well as genetic factors such as sequence variation. Such genetic variants are termed methylation quantitative trait loci (meQTLs) and are associated with both *cis*- and *trans*-effects on methylation at local and independent loci, respectively, acting both in tissue-dependent, and tissue-independent manners (Lemire *et al.*, 2015; Smith *et al*, 2014; Hannon *et al.*, 2015).

A widely-used method in the detection of DNA methylation is bisulphite sequencing. This process involves treatment of DNA with sodium bisulphite, whereby all unmethylated C bases are deaminated to uracil (U), while all methylated C bases remain unchanged. This is followed by amplification by polymerase chain reaction (PCR), resulting in conversion of U bases to thymine (T). Analysis of the DNA sequence is then performed to identify the extent of methylated loci based on the resulting CpG (methylated) and TpG (unmethylated) dinucleotides. In recent years, the development of next-generation sequencing and array-based technologies have allowed researchers to quantify DNA methylation at the genome-wide level. This has

permitted the assessment of the relationship between DNA methylation and various traits and conditions, through epigenome-wide association studies (EWAS).

The greatest advantage of whole-genome bisulphite sequencing (WGBS) is its coverage of all CpG loci. However, the costs associated with this technology have posed a limitation to widespread application of the method (Ziller *et al.*, 2015). A more affordable alternative to WGBS has been array-based technology, whereby methylation levels are assessed by probes targeting bisulphite-converted DNA at loci throughout the genome. The development of array-based options has provided progressively greater coverage of the genome, allowing for an economical alternative to next-generation sequencing methods.

1.6.1 DNA methylation in psychiatric disorders

Several lines of evidence link differential DNA methylation to psychotic illness, mood disorders and stress-related disorders. Monozygotic twin studies have been particularly illustrative of environmental effects as genotype, sex, age and maternal environment are matched between samples (Bell and Spector, 2011). A twin-based EWAS by Dempster *et al.* (2014) examined methylation in buccal cell-derived DNA from individuals with adolescent depression. Although no single site showed differential methylation at the genome-wide significant level, the authors presented a ranked list based on *p*–values and methylation differences between groups. Their top-ranking site mapped to *serine/threonine kinase 32C* (*STK32C*) on chromosome 10, a brain-expressed kinase. Methylation differences at this site were validated by targeted pyrosequencing in DNA derived from buccal cells, and by microarray analysis of *post-mortem* cerebellum DNA from MDD patients. A study by Numata *et al.* (2015) compared methylation in blood-derived DNA between MDD patients and healthy controls, identifying 363 genome-wide significant loci (false discovery rate $q \leq 0.05$). Of the top 100 loci, 84 were nominally significant for differential methylation in an independent sample. Among these sites was *GSK3β*, whose expression was also found to be inversely correlated with methylation in these individuals.

Using blood-derived DNA, Dempster *et al.* (2011) interrogated genome-wide methylation in monozygotic twin pairs that were discordant for either schizophrenia or bipolar disorder. Here, they identified DNA methylation differences associated with both disorders separately and combined, using a weighted *t*-test approach. Of their top 100 ranked loci for differential methylation in psychosis, five genes overlapped with those identified in a study by Mill *et al.* (2008) in which *post-mortem* brain was used as a study tissue. Kinoshita *et al.* (2013) reported 10,747 differentially methylated loci comparing blood-derived DNA from schizophrenia patients to healthy controls ($q \leq 0.05$). A replication sample consisting of three schizophrenia-discordant twin pairs yielded 234 differentially methylated loci in common with the discovery sample, based on unadjusted paired *t*-test *p*-values. Nishioka *et al*. (2013) examined methylation differences in blood between patients with first-episode schizophrenia and age-matched healthy controls identifying 603 sites with significant differential methylation in 589 genes. Ontology analysis showed nucleotide binding and transcription factor binding genes were significantly enriched amongst these results. When comparing their findings with those of Kinoshita *et al.* (2013) and Dempster *et al*. (2011), little overlap was observed. The authors suggested sample size and disease heterogeneity may have contributed to this lack of overlap. A recent large-scale EWAS of schizophrenia by Montano *et al.* (2016) identified 172 sites with differential methylation in both discovery and replication cohorts, including loci within genes previously implicated in schizophrenia through GWAS (Ripke *et al.*, 2014). However, the authors imposed a comparatively liberal *q*-value cut-off of 0.2, increasing the probability of the incidence of type-I errors among these findings, compared to other studies.

The lack of agreement between the studies described here may be attributed to several factors including disease heterogeneity and methodological approaches (Teroganova *et al.*, 2016). Moreover, the use of blood in these studies may also be linked to mixed results. Due to its accessibility, whole blood-derived DNA has been widely-used in EWAS. However, a limitation with the use these samples is their associated cellular

heterogeneity. However, cell compositions can be estimated based on known methylation profiles of cell subtypes from flow-sorted blood (Houseman *et al.*, 2012; Jaffe and Irizarry, 2014). Such estimates can be considered in downstream analyses of differential methylation to reduce or eliminate the confounding effects of tissue heterogeneity.

A major shortcoming in the majority of psychiatry-based EWAS has been the limited access to the primary tissue of interest (i.e. brain tissue). Varying degrees of correlation have been reported between methylation in whole blood and methylation in various brain regions. Hannon *et al.* (2015) reported distinct methylation profiles between blood, cortex and cerebellum, suggesting the use of whole blood-derived DNA in psychiatry-related EWAS may not be a suitable proxy tissue for the brain. A study by Walton *et al.* (2015) reported significant correlation between blood and brain among 4.1% of CpG sites assayed by the Illumina Infinium HumanMethylation450 array - a widely-used platform for EWAS in recent years (Bibikova *et al.*, 2011). However, the authors noted the following limitations to their study: the comparison was made between whole blood and temporal lobe samples biopsied from epilepsy patients, which may not be representative of the healthy brain, or of alternative brain regions. In contrast, others have reported that meQTLs are consistently detected across blood and brain (Smith *et al.*, 2014). Here, they reported overlap of 254/724 SNP-CpG associations in both blood and frontal cortex. Analysis of such meQTL-associated sites in blood may therefore be useful in identifying biologically meaningful results when brain tissue is unavailable for psychiatry-related DNA methylation studies.

Although studies of DNA methylation in psychiatry have generally been performed on blood-derived DNA, there have been a number of reports using brain-derived material (Pidsley and Mill, 2011). A study of DNA methylation in *post-mortem* frontal cortex by Mill *et al.* (2008) reported psychosis-associated differences in DNA methylation among genes enriched in processes pertaining to brain development, stress response and mitochondrial function. Examination of candidate genes for

psychosis showed the *BDNF* Val66Met polymorphism described in section 1.3.2.1 was also associated with DNA methylation levels in these samples. Wockner *et al.* (2014) profiled genome-wide methylation in *post-mortem* prefrontal cortex of schizophrenic patients and healthy controls, identifying 4641 differentially methylated sites in 2929 genes after adjusting for age and *post-mortem* interval ($q \leq 0.05$). Of these genes, 99 were found to be common between this study and a blood-based analysis of DNA methylation in schizophrenia by Nishioka *et al.* (2013). Rather than adjusting for gender in their study, Wockner *et al.* (2014) opted to exclude probes mapping to the sex chromosomes. As sex-specific differences in methylation have been reported at numerous autosomal sites (Hall *et al.*, 2014), the probe-filtering strategy alone may not have eliminated the effects of gender on their results. A study by Numata *et al.* (2014) reported differential DNA methylation at 107 sites in the dorsolateral prefrontal cortex of patients with schizophrenia. Upon analysis of meQTLs in these samples, they reported a number of significant meQTLs at SNPs previously implicated in schizophrenia by GWAS and candidate gene association studies.

Other studies performed on brain tissue have made links between DNA methylation and the neurodevelopmental origin of schizophrenia. A study by Pidsley *et al.* (2014) examined methylation differences in prefrontal cortex and cerebellum samples between schizophrenia patients and healthy controls. Network analysis showed enrichment for neurodevelopmental pathways amongst the genes that were differentially methylated in schizophrenia. Furthermore, the schizophrenia-associated sites in the prefrontal cortex were found to be significantly associated with age in the developing foetal brain. Hannon *et al.* (2015) reported that mQTLs in foetal brain tissue were enriched among loci implicated in schizophrenia by Ripke *et al.* (2014), proposing a potential mechanism relating DNA methylation to the neurodevelopmental hypothesis of schizophrenia.

Disruption of DNA methylation is an attractive hypothesis to explain the environmental component of complex disorders such as psychiatric illness. Although

technologies are ever-improving to detect DNA methylation differences, access to relevant tissue remains a challenge. To this end, animal and cellular models may prove useful in compensating for this shortcoming.

1.7 Cellular models of major mental illness

Although there are well-established protocols for assessing behavioural and physiological characteristics in animal models of psychiatric illness, a limitation is their inability to recapitulate the genetic complexity of these disorders in humans. An alternative approach has been the use of patient-derived cellular models. Due to the scarcity of high quality patient-derived *post-mortem* brain tissue, human-derived *in-vitro* models of major mental illness have been valuable resources for psychiatric researchers. The development of iPSC technology (Takahashi *et al.*, 2007) has been a significant step in modelling neuropsychiatric disorders, among several others (Sterneckert *et al.*, 2014). Psychiatric disorders are generally characterised by symptoms presented by the patient, as there are no established biomarkers that can be used in their diagnosis. Generation of neuronal material derived from patient iPSCs have permitted greater insights into deficits in human neurodevelopment and neuronal function that may be contributing factors in psychiatric disorders.

1.7.1 iPSC models of schizophrenia

Brennand *et al.* (2011) presented one of the earliest iPSC models of schizophrenia. Here, they compared iPSC-derived neurons from four schizophrenia patients to those derived from five healthy controls. They reported decreased connectivity and neurite counts, as well as misexpression of schizophrenia-associated genes in patient samples compared to controls (e.g. *ANK1)*. Furthermore, subsets of patient samples of showed dysregulation of schizophrenia candidate genes including *DISC1* and *ZNF804A* when compared to controls. Lack of consistency across schizophrenia-derived samples may be attributed to the small, heterogeneous sample and small number of individuals in this study. Developmental differences have also been observed in neuronal precursor cells (NPCs) derived from patient iPSCs, in support of a neurodevelopmental origin for schizophrenia. A study by Brennand *et al.* (2015)

examined gene and protein expression in NPCs from four schizophrenia patients. Here, network analysis revealed disruption of cell adhesion pathways at the levels of both gene and protein expression while neuronal maturation pathway disruptions were reported at the gene expression level. These findings may indicate that both of these pathways are perturbed during neurodevelopment in patients with schizophrenia.

Others have developed *DISC1*-based models of iPSC-derived neurons. Wen *et al.* (2014) reported synaptic deficits, reduced *DISC1* expression and morphological differences in iPSC-derived neurons from an American family with a 4 bp frameshift mutation in *DISC1*, described in section 1.4.7 (Sachs *et al*., 2005). Genome editing-mediated repair of the *DISC1* mutation rescued these expression and synaptic deficits. Expression analysis by RNA-seq showed genome-wide transcriptional dysregulation in these neurons. However, an important limitation of this study was the small sample size, which consisted of just one control and two mutant lines. Srikanth *et al*. (2015) also used a gene-editing approach, this time to induce *DISC1* mutations in iPSC-derived neurons. They introduced frameshift mutations in either exon 2 of *DISC1* (an exon common to all *DISC1* isoforms), or in exon 8, within 10 codons of the intron containing the chromosome 1 translocation breakpoint. In the exon 8 mutants, they observed decreased *DISC1* expression, due to nonsense mediated decay of the mutant transcripts. Both mutant forms of *DISC1* displayed defects in Wnt-signalling in NPCs, but not iPSCs. The Wnt pathway has previously been implicated in the pathology of psychiatric disorders (Freyberg *et al.*, 2010). Furthermore, Singh *et al.* (2011) have reported defective Wnt signalling and disrupted neurodevelopment in mice and zebrafish expressing mutant forms of *DISC1*.

1.7.2 iPSC models of bipolar disorder

More recently, iPSC-based studies of bipolar studies have emerged. Chen *et al.* (2014) compared iPSC-derived neurons from three bipolar disorder patients to three controls. They reported significant transcriptional differences between groups at the

neuronal level, but not at the iPSC stage. In addition, positional identity differences were observed between groups: while control neurons expressed transcription factors involved in the regulation of dorsal telencephalic fate, patient cells expressed ventral-associated genes, suggesting illness may relate to disrupted neuronal differentiation during development. Pathway analysis of differentially expressed transcripts in iPSCs revealed an enrichment of genes involved in calcium signalling, in keeping with GWAS findings in bipolar disorder. Moreover, comparison of gene expression data between iPSC-derived neurons from bipolar disorder patients and bipolar disorder patients' *post-mortem* brain samples revealed common dysregulation of transcripts involved in calcium signalling, neurotransmission and cell migration.

A study by Madison *et al.* (2015) compared iPSC-derived neurons from related individuals: two brothers with bipolar disorder and their unaffected parents. Here, proliferation deficits and transcriptional dysregulation were observed in bipolar disorder NPCs. Network analysis of the dysregulated genes suggested disrupted Wnt signalling in the NPCs: a regulatory pathway during NPC proliferation (Valvezan and Klein, 2012). Upon treatment of the bipolar disorder NPCs with a GSK3 inhibitor, the proliferation deficits were rescued. Bavamian *et al.* (2015) assessed the link between bipolar disorder and miR-*34a* using iPSC-derived neurons. This microRNA is predicted to target genes related to bipolar disorder and its expression has been shown to be modulated by mood stabilising drugs (Hunsberger *et al.*, 2013). Dysregulated microRNA expression has been reported in both schizophrenia and bipolar disorder as well as among individuals at high risk of developing bipolar disorder (Moreau *et al.*, 2011; Walker *et al.*, 2015), while *miR-137* has been associated with schizophrenia through GWAS (Ripke *et al.*, 2011). Using a case and control from the study by Madison *et al.* (2015), Bavamian *et al.* (2015) examined miR-*34a* expression during neuronal differentiation finding significant overexpression in the bipolar case, compared to the unaffected control sample. Overexpression of miR-*34a* was shown to downregulate two of its predicted bipolar disorder-associated targets, *ANK3* and *CACNB3*. Furthermore, overexpression of *miR-34a* in developing neurons also resulted in downregulation of pre- and postsynaptic proteins SYP, SYN1 and PSD95. Expression levels of miR-*34a* were

found to be linked to dendritic morphology, with overexpression resulting in increased dendritic length, and inhibition resulting in increased branch numbers. Agostini *et al.* (2011) previously reported a similar phenotype in mice in response to *miR-34a* expression levels.

Cellular models such as those described in this section have provided valuable insights into human neurodevelopment and brain function in the context of major mental illness. However, current limitations posed by iPSC-derived models of neuropsychiatric disorders include small sample sizes, variable culturing conditions and genetic heterogeneity, rendering useful comparisons between studies difficult. Through addressing these issues it is likely that iPSC-derived models will be an effective resource for identifying disrupted pathways and mechanisms related to psychiatric illness.

1.8 Summary

Schizophrenia and affective disorders are devastating illnesses that contribute to a significant social and economic burden, with limited efficient treatment options available. It is evident that these are complex, heterogeneous disorders associated with genetic factors and environmental factors, as well as their interactions. Although it is unlikely that a single causal gene is responsible for these disorders, risk factors are being identified at a growing rate thanks to ever-increasing sample sizes and the development of high-throughput technologies, as well as identification of rare instances of genomic disruption. Moreover, the development of animal and cellular models for these illnesses are likely to be fruitful in identifying disrupted processes, genes and pathways related to these disorders. The work described in this chapter should form the basis of future studies into the biological mechanisms of these disorders in order to understand their associated cellular and molecular dysfunctions, and ultimately, identify targets for more effective treatments.

Chapter 2

Materials and methods

2 Materials and methods

The experiments detailed in this thesis availed of three tissue types obtained from the t(1;11) family: lymphoblastoid cell lines (LCLs), whole blood, and induced pluripotent stem cell (iPSC)-derived neurons. A summary of sample demographics, including tissue type, gender, and translocation carrier status is detailed in Table 2.1. A detailed description of the study was given to all individuals and all participants gave written informed consent. Ethical approval was obtained from the Scotland A Research Ethics committee.

2.1 Polymerase chain reaction (PCR)

A master mix was prepared containing 1X PCR buffer containing 15 mM $MgCl_2$ (Sigma-Aldrich), 0.6 μM dNTPs, 20 μM forward and reverse primer mix, *Taq* DNA polymerase (Sigma-Aldrich), 1 μl DNA sample, and dH_2O to make a total reaction volume of 20 μl. Reactions were placed on a thermocycler and run on the "TD55" program:

TD55 PCR program:

93°C for 1 minute	
93°C for 20 seconds	10 cycles
65°C for 30 seconds (-1°C per cycle)	
72°C for 1 minute	
93°C for 20 seconds	30 cycles
55°C for 30 seconds	
72°C for 1 minute	
72°C for 10 minutes	

Samples were run on a 1-2% (w/v) Agarose/TBE gel and imaged on an ultraviolet transilluminator to check the PCR products alongside a 1 Kb plus DNA ladder (Thermo Fisher).

ID	t(1;11) Status	Gender	LCL	Blood	Neuron	ID	t(1;11) Status	Gender	LCL	Blood	Neuron
18	T	M		●	●	80	N	F		●	
19	T	M		●	●	82	N	F		●	
24	T	F		●	●	85	N	M		●	
28	N	F		●	●	87	N	M		●	
29	N	F		●	●	88	N	M		●	
30	N	F		●	●	89	N	M		●	
40	N	M	●	●		91	N	F		●	
50	T	F	●	●		92	N	M		●	
52	N	F	●	●		94	N	M		●	
61	T	F	●	●		96	N	F		●	
72	T	M	●	●		97	N	M		●	
9	T	M		●		99	N	M		●	
15	T	F		●		100	N	F		●	
26	T	F		●		104	T	M		●	
27	T	F		●		106	N	F		●	
32	T	M		●		107	N	M		●	
41	T	F		●		11	T	M	●		
44	N	F		●		13	T	F	●		
47	N	M		●		16	N	M	●		
49	T	F		●		78	N	M	●		
53	T	M		●		35	T	M	●		
54	N	M		●		39	T	F	●		
55	T	M		●		70	T	F	●		
62	N	F		●		34	N	F	●		

ID	t(1;11) Status	Gender	LCL	Blood	Neuron						
67	T	F		•							
49	T	F		•							
53	T	M		•							

Table 2.1: Summary of t(1;11) family samples

Shown is the linkage ID for each individual, t(1;11) status, where T = translocation carrier and N = non-carrier, sample gender and bullet points summarising whether or not a given individual is present in the lymphoblastoid, blood or iPSC-derived neuronal samples..

2.2 DNA sequencing

2.2.1 Clean-up of PCR products

In order to remove dNTPs and primers, 1 µl of each PCR product with were treated with 3 µl ExoSAP-IT (Affymetrix), adding 1 µl dH_2O for a total reaction volume of 5 µl. Reactions were set up in a 96 well plate and placed on a thermocycler and run under the following conditions:

ExoSAP-IT PCR program:

37°C for 60 minutes

80°C for 20 minutes

2.2.2 Sequencing reaction

To each cleaned PCR product, 1 µl 5X BigDye® sequencing buffer (Applied Biosystems) was added, along with 1 µl BigDye® v3.1 and 1 µl forward primer at a concentration of 3.2 µM. Sequencing reactions were made up to 10 µl by adding 2 µl dH_2O and were placed on a thermocycler to be run under the following conditions:

Seq3-1 PCR Program:

96°C for 1 minute

96°C for 10 seconds
50°C for 5 seconds
60°C for 4 minutes
} 30 cycles

2.2.3 EDTA/ethanol precipitation

Following the reaction, DNA was precipitated by adding 2.5 µl 125 mM EDTA followed by 30 µl 95 % ethanol. The plate was inverted four times and left on the bench at room temperature for 15 minutes. Samples were centrifuged at 3000 RPM for 30 minutes at 8°C. The supernatant was removed by inverting the plate and tapping it on a paper towel. The plate was then briefly centrifuged at 1000 RPM while inverted on a paper towel to remove any remaining liquid. 30 µl 70 % ethanol was added to each well and samples were centrifuged at 3000 RPM for 15 minutes at 8°C. The supernatant was removed as above and DNA pellets were left to dry at room temperature, loosely covering the plate with a paper towel. Samples were sent to Agnes Gallagher at the MRC Human Genetics Unit for sequencing.

2.3 Lymphoblastoid cell culture

EBV-transformed LCLs derived from t(1;11) family members were cultured in RPMI 1640 media (Invitrogen) supplemented with 10 % foetal bovine serum (Invitrogen). Cells were fed by replacing the media on Mondays, Wednesdays and Fridays. Cells were grown in 75 cm^3 flasks for RNA preparation.

2.4 Harvesting of LCLs for RNA extraction

After reaching sufficient confluency in a 75 cm^3 flask, cells were fed for three consecutive days before harvesting. Prior to harvesting, cell counts were obtained by diluting a sample

taken directly from a flask by a factor of 10. Diluted samples were counted on a haemocytometer and, using this count, 10^7 cells were removed from the flask and transferred to a 15 ml tube. Cells were centrifuged at 1200 rpm for 5 minutes and the supernatant was removed using an aspirator vacuum pump. The cell pellet was resuspended in 4 ml PBS (Invitrogen) and spun again at 1200 rpm for 5 minutes and the supernatant was removed as above. The cell pellet was resuspended in 600 µl RLT lysis buffer containing β-mercaptoethanol (Qiagen) and transferred to a 1.5 ml RNase-free tube. Harvested cells were stored at -80°C for RNA extraction at a later date.

2.5 RNA extraction from cultured cells

An equal volume of 70 % ethanol (600 µl) was added to each sample and mixed by pipetting. The sample mixture was transferred to an RNeasy spin column (Qiagen) placed in a 2 ml collection tube. Samples were centrifuged at full speed for 15 seconds to bind the RNA to the spin column. The flow through was discarded and 350 µl buffer RW1 wash buffer was added to each spin column. DNase I incubation mix was prepared by adding 10 µl DNase I to 70 µl Buffer RDD (Qiagen). The mix was added directly onto the membrane of each tube and left at room temperature for 15 minutes. Again, 350 µl buffer RW1 was added to the column and samples were centrifuged at full speed for 15 seconds to wash the spin column membrane-bound RNA, discarding the flow-through. 500 µl buffer RPE was added to each spin column and centrifuged for 15 seconds at full speed to wash the RNA. The flow-through was discarded and this wash step was repeated a final time, centrifuging for 2 minutes to dry the spin column. The spin column was placed in an RNase-free collection tube and 50 µl RNase free water was added to the membrane to elute the RNA. Samples were centrifuged at full speed for 1 minute. This step was repeated adding 30 µl RNase free water for a total eluate of 80 µl. RNA samples were stored at -80°C to await downstream sample preparation procedures. To minimise DNA contamination, the eluate underwent a second DNase treatment using the Ambion® DNA-*free*™ kit (Applied Biosystems). To each sample, 8 µl 10X DNase I buffer was added along with 1 µl rDNase I. Samples were incubated at 37°C for 30 minutes. DNase inactivation reagent was resuspended by gently flicking the tube and 9 µl was added to each reaction. Samples were incubated for 2 minutes at room temperature followed by

centrifugation at full speed for 90 seconds. The RNA-containing supernatant was transferred to a fresh tube with care not to disturb the pelleted inactivation reagent.

2.6 Quality control of RNA

To assess the concentration and quality of extracted RNA, 5 µl of RNA eluate was set aside to be measured on an Agilent 2100 Bioanalyzer (Agilent) at the Wellcome Trust Clinical Research Facility (WTCRF), Western General Hospital, Edinburgh. Measurements of RNA integrity numbers (RIN scores) and concentration in ng/µl were obtained. A RIN threshold of 7.0 was set as the minimum acceptable quality to continue with downstream procedures (Thompson *et al.*, 2007).

2.7 Synthesis of cDNA

Reverse transcription of RNA samples was performed to synthesis cDNA using the Applied Biosystems cDNA synthesis kit (Applied Biosystems). For each sample a reaction was set up containing 4.4 µl $MgCl_2$, 2 µl PCR buffer, 1 µl dNTPs, 0.4 µl RNase inhibitors, 0.5 µl MultiScribe™ reverse transcriptase enzyme and 2 µl random hexamers. RNA was denatured by heating to 65°C for 10 minutes prior to addition to the reactions. The required volume for 500 ng RNA was determined using the concentrations obtained from the Agilent 2100 Bioanalyzer. 500 ng RNA was added to each reaction. RNase-free water was added to make a final reaction volume of 20 µl. Reverse transcriptase-free and sample-free negative control reactions were prepared as above, excluding MultiScribe™ reverse transcriptase and RNA respectively. Reactions were placed on a thermocycler and was incubated for 10 minutes at 25°C, followed by 30 minutes at 48°C, and 5 minutes at 95°C. Samples were then stored at -20°C. To determine whether genomic DNA contamination was present in human samples, 1 µl cDNA samples and the negative controls from the cDNA synthesis reaction underwent PCR using primers spanning exons 7-8 of *WDR1*, alongside a genomic DNA positive control sample. This amplifies a 590 bp region in genomic DNA and a 238 bp region in cDNA. For mouse samples, a primer pair amplifying exons 24 to 25 of *SorCS2* was used, alongside a genomic DNA positive control. This amplifies a 628 bp region in genomic DNA, and a 156 bp region in cDNA.

2.8 Quantitative real-time PCR (qRT-PCR)

qRT-PCR was performed on the ABI 7900-HT system (Applied Biosystems) at the Wellcome Trust Clinical Research Facility, and the Human Genetics Unit, MRC IGMM, Edinburgh.

2.8.1 Reference gene selection:

Reference genes were selected from a panel of ubiquitously-expressed housekeeping gene assays. From this panel, the most stably expressed genes across all samples were selected by GeNorm analysis using Biogazelle qBase+ software. Selections were made based on the GeNorm M- and GeNorm V- results. The GeNorm M result is a plot of the reference gene expression stability across all samples. The GeNorm M - value is derived by a stepwise process, determining the most stable reference genes to use based on the average pairwise variation between each individual gene and all other reference genes, eliminating the worst-performing gene at each step until just two remain. GeNorm-V determines the minimum optimal number of reference genes to use by comparing normalisation factors derived from n versus n+1 reference genes among samples (Vandesompele *et al.*, 2002). A summary of reference genes used is presented in Table 2.2.

Catalogue no.	Gene (Symbol)	Species	Manufacturer
ge-DD-6	ATP Synthase, H^+ Transporting, Mitochondrial F1 Complex *(ATP5B)*	Human	Primerdesign
ge-DD-7	Ubiquitin C (*UBC*)	Human	Primerdesign
ge-DD-8	TATA Box Binding Protein (*TBP*)	Human	Primerdesign
ge-DD-9	Nuclear Cap Binding Protein Subunit 2, 20kDa (*NCBP2*)	Human	Primerdesign
ge-DD-10	Glyceraldehyde-3-Phosphate Dehydrogenase (*GAPDH*)	Human	Primerdesign
ge-DD-11	Cytochrome C-1 (*CYC1*)	Human	Primerdesign
HK-DD-hu-300	Succinate Dehydrogenase Complex, Subunit A, Flavoprotein (*SDHA*)	Human	Primerdesign
4333761T	Ribosomal Protein, Large, P0 (*RPLP0*)	Human	Applied Biosystems
Mm00446968_m1	Hypoxanthine Phosphoribosyltransferase 1 (*Hprt*)	Mouse	Applied Biosystems
Mm01143545_m1	Hydroxymethylbilane Synthase (*Hmbs*)	Mouse	Applied Biosystems
HK-DD-mo-600	Succinate Dehydrogenase Complex, Subunit A, Flavoprotein *(Sdha*)	Mouse	Primerdesign
HK-DD-mo-600	Ribosomal Protein, Large, P0 (*Rplp0*)	Mouse	Primerdesign
DD-mo-600	Peptidylprolyl Isomerase D (*Ppid*)	Mouse	Primerdesign
HK-DD-mo-600	Ubiquitin C (*Ubc*)	Mouse	Primerdesign

Table 2.2: Reference genes used in qRT-PCR experiments.

Table summarises each reference gene, including gene name, catalogue number, species and manufacturer.

2.8.2 Experimental setup:

Each reaction consisted of 5 µl master mix (Applied Biosystems), 0.5 µl probe and 4.5 µl cDNA. Samples were loaded in triplicate on a 384-well PCR plate which was sealed with an adhesive plastic sheet and run on the ABI 7900-HT. For each probe, a standard curve was generated by preparing a serial dilution of pooled cDNA.

qRT-PCR Program:

96°C for 10 seconds
60°C for 4 minutes
} 40 cycles

2.8.3 Data analysis:

Raw data were exported from SDS v2.3 software (Applied Biosystems) to Microsoft Office® Excel for downstream analysis. Outlying triplicates were defined as those outwith 1 cycle threshold (Ct) value of the remaining two triplicates. Outliers were removed and log quantity means were calculated from raw Ct values based on the standard curve for the probe being analysed.

Log quantities were obtained by subtracting the *y*-axis intercept from each Ct value and dividing the result from the slope of the standard curve (*x* coefficient). The antilog of the resulting log quantity was calculated by raising 10 to the power of the log quantity for each cell. Sample quantity means and standard deviations were calculated for each triplicate to compare gene expression between groups.

For each sample, a normalisation factor was obtained by calculating the geometric mean of the reference genes tested for each sample. Aw data for the genes of interest were normalised by dividing the quantity mean of each sample by its corresponding normalisation factor. Means and standard errors were calculated for each group for comparison.

2.8.4 Statistical analysis

Normality was assessed using a Shapiro-Wilk test. Normally distributed data were compared using an unpaired student's *t*-test. In cases where data were non-normal but showed homogenous variance as determined by a Levene's test, a Mann-Whitney U test was performed on rank-transformed data.

2.9 Microarray-based gene expression analysis of LCL RNA

RNA samples prepared from t(1;11) family LCLs were hybridised to the Human HT-12 v4 Expression BeadChip (Illumina, San Diego, CA) at the Wellcome Trust Clinical Research Facility, Edinburgh. Samples from 13 individuals were included on the

microarrays: eight t(1;11) carriers and five non-carrying controls. Samples were labelled and amplified by Helen Torrance. Microarray expression data were preprocessed and analysed in R (R Core Team ,2012).

2.9.1 Raw data input and quality control

Raw microarray data were loaded into an R environment using the *lumiR()* function available in the package *lumi*. Samples were spread across two slides. To control for technical variation, a within-slide technical replicate was present along with a between-slide replicate (KK108). Reproducibility of data from these replicates was determined using the Pearson's correlation coefficient by means of the *cor.test()* function, within the pre-loaded *stats* package in R.

2.9.2 Data transformation, normalisation and filtering

As one of the prerequisites for data undergoing linear regression is homoscedasticity, or equal variance across samples, raw data were transformed by variance stabilizing transformation (VST) prior to normalisation. This was performed using the *lumiT()* function in *limma*. In order to minimize systematic variation which may affect measured expression levels, normalisation was carried out. Data were normalised by robust spline normalisation (RSN), using *limma*'s *lumiN()* function. Probes with a detection *p*-value of ≥ 0.05 were removed from the dataset. This ensured only probes with a high signal-to-background ratio were used in the final analysis, with the additional aim to reduce the burden of multiple tests.

2.9.3 Data analysis

2.9.3.1 *Identification of differentially expressed transcripts*

Differential gene expression between t(1;11) carriers and non-carriers was identified by linear regression of the normalised, filtered dataset. Surrogate variables were identified in using the *sva* package in R, which were used alongside sample gender as covariates. Linear regression was performed using the *limma*'s *lmFit()* function. The output of this function was used to calculate t-statistics, F-statistics and log of odds for each probe using the *eBayes()* function in *limm*a. A transcript was defined as differentially expressed if it was associated with an unadjusted *p*-value of ≤ 0.05 combined with an absolute fold-change (FC) of ≥ 1.2.

2.9.3.2 *Gene ontology analysis*

Gene ontology (GO) analysis was performed using GOrilla (Gene Ontology enRIchment anaLysis and visuaLisAtion tool; Eden *et al.*, 2009). GOrilla identifies significant enrichment for GO terms using a minimal hypergeometric score (mHG), assigning significance to pathways enriched for genes present towards the top of a *p*-value ranked list (Eden *et al*., 2007). All transcripts tested for differential expression were ranked by *p*-value and submitted to GOrilla for analysis. In the case of multiple probes per gene, the lowest *p*-value associated with that gene was submitted.

2.10 Methylation analysis in whole blood and iPSC-derived neuronal DNA

500 ng genomic DNA was treated with sodium bisulphite using the EZ-96 DNA Methylation Kit (Zymo Research, Irvine, California), following the manufacturer's instructions at the WTCRF, Western General Hospital, Edinburgh. Blood-derived samples were analysed using the Infinium HumanMethylation450 BeadChip and iPSC-derived samples were analysed using the Infinium HumanMethylationEPIC BeadChip (Illumina Inc., San Diego, CA). Both chips were run at the WTCRF.

During preprocessing, data were analysed in the form of β-values (calculated by dividing the methylated signal by the sum of the methylated and unmethylated signal plus an offset of 100 for each probe), which occur as a value between 0-1, with 1 representing 100 % methylation at that site. *β*-values were Logit transformed to *M*-values for analysis of differential methylation as recommended by Du *et al*. (2010).

2.10.1 Sample information

The whole-blood-based study consisted of 41 individuals: 17 t(1;11) carriers and 24 non-carrying relatives spread across six slides of the Infinium HumanMethylation450k platform. The iPS-derivative-based study consisted of three differentiations each from six individuals: three t(1;11) carriers and three non-carrying controls. Samples were spread across three slides. Samples were assigned to slides such that, as far as possible, group and gender were counter-balanced across slides.

2.10.2 Identification of potentially cross-hybridising on the Illumina Infinium HumanMethylationEPIC BeadChip

Using a protocol described by Chen *et al.* (2013), all possible probe sequences on the array were aligned to four bisulphite-converted reference genomes representing a forward methylated sequence, reverse methylated sequence, forward unmethylated sequence and reverse unmethylated sequence. These were obtained through modifying the hg19 human genome sequence downloaded on 11th January 2016. To generate the methylated genomes, all non-CpG cytosines were converted to thymines (T). To generate the unmethylated genomes, all C bases were converted to T bases.

Several Infinium Type II assays contain "R" nucleotides, representing adenine or guanine depending on whether an underlying cytosine is methylated or unmethylated in bisulphite-converted DNA. All possible Type II probe sequences were generated and added to a list of Type I probe sequences. This list constituted the query sequences to undergo alignment to a four reference genomes using the BLAST-like alignment tool (BLAT; Kent, 2002). Probes with an off-target match of ≥ 47 bp, along with a match at

the signal-generating end base were deemed potentially cross-hybridising based on the criteria of Chen *et al.* (2013).

2.10.3 Quality control assessments of methylation data

Raw intensity (.idat) files were read into R using the *minfi* package (Aryee *et al.*, 2014), which was used to perform initial quality control assessments. Filtering of poor-performing samples and sites was performed. Samples were removed from the dataset if ≥ 1% sites had a detection *p*-value of ≥ 0.05. Probes were removed from the dataset if: (i) the CpG site (and/or the site of single base extension, at the base before the cytosine, for Type I probes) overlapped a single nucleotide polymorphism (SNP) present in any individual based on whole-genome sequence data from the individuals tested; (ii) they were predicted to cross-hybridise (Chen *et al.*, 2013; McCartney *et al.*, 2016); (iii) they had more than five samples with a beadcount of less than 3; or (iv) ≥ 1 samples had a detection *p*-value of ≥ 0.05 (Schalkwyk *et al.*, 2013; Pidsley *et al.*, 2013).

2.10.4 Normalisation of the methylation data

Data normalised by 14 normalisation methods were compared with the raw data to assess each method's performance at reducing technical variation between arrays. This technique was first described by Pidsley *et al.* (2013) and all required functions are available in the *wateRmelon* and *minfi* packages in R (Schalkwyk *et al*., 2013; Aryee *et al.*, 2014; Table 2.3). Three performance metrics were used to assess each method's ability to reduce technical variation: differentially methylated region standard error (DMRSE), genotype combined standard error (GCOSE), and Seabird.

The DMRSE metric is obtained by calculating the standard deviation of methylation at known imprinting differentially methylated regions (iDMRs). The array contains 220 probes which provide a measure for this metric where methylation levels are expected to be around 50% (β = 0.5). The *dmrse_row* function was used to call the between-sample standard error of β - values for all 220 probes. Lower values for the DMRSE metric indicate better performance of a given normalisation strategy.

The GCOSE method takes advantage of 65 probes present on the array which detect highly polymorphic CpG target sites. At these positions, β-values are expected to follow a tri-modal distribution corresponding to allele *A* homozygotes (e.g. CG/CG), *AB* heterozygotes (e.g. CG/CX) and *BB* homozygotes (e.g. CX/CX) representing high, medium and low methylation levels corresponding to each genotype, respectively. Peak width at each genotype is positively correlated with technical variability. The GCOSE metric is calculated by k-means clustering to define each genotype, noting sum of squares and the number of samples per cluster for 65 probes. In the absence of technical variation, the sum of squares should be zero. Sums of squares for each cluster were summed across 65 probes and divided by the sum of samples. To obtain a standard-error like metric, these three values were divided by the square root of the total number of samples and averaged to obtain a single metric. As with the DMRSE metric, lower values are indicative of better performance of a given normalisation method.

The Seabird metric estimates X-chromosome probes based on the expected methylation levels observed in female individuals (approximately 50 %) by means of a receiver operator characteristic (ROC) analysis. Normalisation methods associated with an area under the curve (AUC) closer to 1 are deemed to perform better. The metric of 1-AUC was used to remain in keeping with the "lower is better" scoring system of the DMRSE and GCOSE metrics.

To determine the best-performing normalisation method, each metric was ranked and the mean rank was obtained. The mean rank was then ranked to obtain a final score for each normalisation strategy (rank of ranks). The method with the lowest rank of ranks was selected to normalise the raw data for downstream analysis.

Normalisation Function:Package	Description	Reference
preprocessSWAN():minfi	Subset within array normalisation	Maksimovic *et al.*, 2012
preprocessNoob():minfI	Background correction method with dye-bias normalisation	Triche *et al.*, 2013
betaqn():wateRmelon	Beta mixture quantile normalisation	Teschendorff *et al.*, 2013;
dasen():wateRmelon	Same as nasen but type I and type II backgrounds are equalised first	Schalkwyk *et al.*, 2013; Pidsley *et al.*, 2013
nasen():wateRmelon	Quantile normalises methylated and unmethylated intensites separately each for type I and type II probes, then calculates betas	
nanet():wateRmelon	Quantile normalises methylated and unmethylated intensities together, then calculates betas	
naten():wateRmelon	Quantile normalises methylated and unmethylated intensities separately, then calculates betas	
nanes():wateRmelon	quantile normalises methylated and unmethylated intensities separately for type I probes, and together for type II probes	
danes():wateRmelon	Background equalisation of type I and type II probes, quantile normalisation of methylated and unmethylated intensities separately	
danet():wateRmelon	Background equalisation of type I and type II probes, quantile normalisation of methylated and unmethylated intensities together, followed by calculation of betas	
danen():wateRmelon	Background equalisation only, no normalisation	
daten1():wateRmelon	Type I and type II background are equalised first (smoothed only for methylated), quantile normalises methylated and unmethylated intensities separately, then calculates betas.	
daten2():wateRmelon	Type I and type II background are equalised first (smoothed for methylated and unmethylated), quantile normalises methylated and unmethylated intensities separately, then calculates betas.	
fuks():wateRmelon	Peak-based correction	Dedeurwaerder *et al.*, 2011

Table 2.3: Summary of normalisation strategies compared in the genome-wide DNA methylation analyses.

Table summarises the normalisation methods tested, including the function for the normalisation method and its corresponding R package, a brief description of the method and references.

2.10.5 Estimation of cell proportions in whole blood

Estimated cell counts for B-lymphocytes, granulocytes, monocytes, natural killer cells, $CD4^+$ T-lymphocytes and $CD8^+$ T-lymphocytes were generated using the *estimateCellCounts()* function in *minfi*. This function implements Jaffe and Irizarry's modified version of Houseman's algorithm (Jaffe and Irizarry, 2014, Houseman *et al.*, 2012). Between-group differences in cell composition were assessed using an unpaired Student's t-test. A p-value of ≤ 0.05 was deemed to represent a significant between-group difference.

2.10.6 Data Analysis

Methylation data were analysed at both the probe level (differentially methylated positions) and the regional level, consisting of multiple probes with correlated methylation signal (differentially methylated regions). Gene ontology analysis was performed on the single probe data to identify ontological terms enriched for differentially methylated genes.

2.10.6.1 *Identification of differentially methylated positions*

Linear regression was performed on the data to identify differentially methylated positions (DMPs) between groups using *limma*'s *lmFit()* function. For each probe, a t-statistic, F-statistic and log-odd of differential methylation was determined using *limma*'s *eBayes()* function. Multiple testing correction was implemented controlling for the false discovery rate (FDR; Benjamini and Hochberg, 1995). DMPs were deemed significant if they had an associated FDR q - value of ≤ 0.05.

2.10.6.2 *Identification of differentially methylated regions*

The probe lasso algorithm was used to determine differentially methylated regions (DMRs) in the DNA methylation data (Butcher and Beck, 2015). This function is available in the R package *ChAMP* (Morris *et al.*, 2014). Probe lasso identifies a DMR

if it detects three probes with differential methylation within a maximum user-defined "lasso radius", set by default at 2000 bp.

2.10.6.3 *Gene ontology analysis*

GOrilla was used to identify to identify functions and pathways enriched for differentially methylated genes in each study. As multiple probes on the array are annotated to a single gene ID, the lowest *p*-value observed for a probe in each gene associated with that gene for GO analysis.

2.10.6.4 *DNA Methylation age calculation*

Using the DNA methylation age calculator (http://dnamage.genetics.ucla.edu/; Horvath, 2013), β-values methylation data were uploaded for all individuals profiled. From these data, DNA methylation age was calculated based on levels at 353 probes.

2.11 Culturing of iPS-derivatives

Frozen lines of neuronal precursor cells (NPCs), previously generated by Dr. Ellen Grunewald from iPS cells derived from fibroblast biopsies from the t(1;11) family were used in this study. Reagents and recipes for cell culture media used at the various stages of growth and differentiation are detailed in Table 2.4. Cell culture was performed by Helen Torrance Dr. Kirsty Millar, Susan Anderson, and myself.

	RosV2F10	Default	Default + Forskolin	Default + BDNF + GDNF
PSF	500 µl	500 µl	500 µl	500 µl
Glutamax	500 µl	250 µl	250 µl	250 µl
N2	500 µl	250 µl	250 µl	250 µl
B27	50 µl	100 µl	100 µl	100 µl
FGF	25 µl			
Heparin		50 µl	50 µl	50 µl
Forskolin			50 µl	
BDNF				25 µl
GDNF				25 µl
Advanced DMEM F12	To 50 ml	To 50 ml	To 50 ml	To 50 ml

Table 2.4: Media recipes for NPC and neuronal culturing.

Table summarises the relative volumes of reagents required to make media for use in NPC culture and subsequent neuronal differentiation.

2.11.1 Maintenance of neuronal precursor cells

NPCs were seeded in a single well of a 6-well plate containing 3 ml RosV2F10F10 media. Prior to seeding, each well of a 6-well plate was pre-coated with 1 ml RosV2F10 containing 10 µl Matrigel and incubated at 4°C for a minimum of 12 hours and a maximum of 7 days. Cells were fed every Monday, Wednesday and Friday by removing 1.5 ml media and replacing with 1.5 ml fresh media. Cells were passaged at a 1:2 split ratio when 100 % confluency was reached. When passaging, media was removed and set aside (conditioned media). Cell dissociation was initiated by adding 1ml Accutase dissociation reagent. After 30 seconds 1 ml conditioned media was reintroduced to the well to neutralise the dissociation reagent. Cells were dissociated by flushing the wells with the neutralised Accutase using a P1000 pipette. Cells were transferred to a falcon tube and centrifuged at 1200 rpm for 3 minutes. While spinning, media was removed from the Matrigel-coated plates using an aspirator vacuum pump and 1 ml fresh RosV2F10 was added to each well. Following centrifugation, the supernatant was removed and the cell pellet was suspended in

conditioned media at a volume of 1 ml per well. To each well, 1 ml of resuspended cells was added.

2.11.2 Neuronal differentiation and maintenance

When 100 % confluency was reached in at least 8 wells across two 6-well plates, neuronal differentiation was initiated in 12-well plates. Neurons underwent differentiation for five weeks prior to harvesting.

2.11.2.1 *Initiation of neuronal differentiation*

Two 12-well plates were coated with 500 µl Default media containing 5 µl Matrigel and incubated at 4°C for a minimum of 12 hours. At least 2 hours prior to seeding, the contents of the wells were removed and replaced with 500 µl Default media containing 5 µl Laminin and incubated at 37°C (or 4°C for long-term storage up to 7 days). NPCs were dissociated and pelleted as above. The pellet was resuspended in 10 ml Default media and cells were counted. The appropriate volume containing 24 x 10^6 cells was taken from the cell suspension and transferred to a 50 ml falcon tube. Default media was added to make a total volume of 12 ml. Media was removed from the two pre-coated 12-well plates with an aspirator vacuum pump and 500 µl fresh media was added to each well. To this, 500 µl cell suspension was added, seeding 1 x 10^6 cells per well.

2.11.2.2 *Maintenance of differentiating neurons*

During differentiation, cells were fed on Mondays, Wednesdays and Fridays. Extra feedings were performed on Tuesdays and Thursdays if deemed necessary based on the colour of the culture media. Neurons were differentiated over 5 weeks. Cells were fed with Default media in week 1, Default + Forskolin media in weeks 2-3, and Default + BDNF + GDNF media in weeks 4-5. For the first feed, 500 µl was gently added to each well, tilting the plate to avoid disrupting the cells. Subsequent feedings involved removing 750 µl media and adding 750 µl fresh media. Following seeding,

differentiating neurons were fed with Default media until the following Monday. Neurons were harvested the Monday following the fifth week of differentiation.

2.11.2.3 *Harvesting neurons for downstream DNA extraction*

Cell culture media were removed from 4 wells of the 12-well plate and set aside. Cell dissociation was initiated by adding 500 µl Accutase dissociation reagent for 30 seconds. The dissociation reaction was neutralised by adding 500 µl conditioned media and the neurons were dissociated by flushing the wells with a pipette, as with the NPCs. Neurons were transferred to a 15 ml falcon and centrifuged at 2500 rpm for 5 minutes. The supernatant was removed using an aspirator vacuum pump and the pellet was washed with 7 ml PBS. The cells were centrifuged again at 2500 rpm for 5 minutes and the supernatant removed. The neurons were resuspended in 200 µl PBS and stored at -20°C for DNA extraction at a later date.

2.11.3 DNA extraction from neurons

DNA was extracted using a DNeasy kit (Qiagen). 180 µl Buffer ATL and 20 µl proteinase K were added to the to the 200 µl PBS containing cell pellets followed by incubation of samples at 56°C for 10 minutes to lyse cells. Following vortexing for 15 seconds, 200 µl Buffer AL was added and mixed by vortexing. This was followed by 200 µl 100% ethanol. Following vortexing, mixtures were transferred to a spin column in a 2 ml collection tube and centrifuged at 6000 x *g* for 1 minute after which flow-through and collection tubes were discarded. Spin columns were transferred to new 2 ml collection tubes and 500 µl Buffer AW1 was added. Samples were centrifuged at 6000 x *g* for 1 minute. Spin columns were transferred to new 2 ml collection tubes and 500 µl Buffer AW2 was added. Samples were centrifuged for 3 minutes at 20,000 x *g*. Spin columns were placed in a clean 1.5 ml microcentrifuge tube and 200 µl Buffer AE was placed directly on the column. Following a 1 minute incubation at room temperature, samples were eluted by centrifuging for 1 minute at 6000 x *g*. DNA eluates were stored at -20°C for later use.

2.12 Collection of L100P and wild-type mouse samples

2.12.1 Mouse strains

L100P heterozygote mice were obtained from Malgorzata Borkowska (Centre for Integrative Physiology, University of Edinburgh). The mice were the products of crosses between L100P homozygote mice RIKEN BRC (Tsukuba, Japan) and C57BL/6J.

2.12.2 Dissection of embryonic and postnatal brains.

Whole embryonic brains were harvested at 13.5, 15.5 and 18.5 days post conception (DPC). These timings were calculated from the point a vaginal plug was observed in a female mouse, indicative of mating, beginning at 0.5 DPC. Pregnant dams were sacrificed by cervical dislocation for embryonic sample collection. Pups were extracted from the uterus using sterile scissors and placed in a petri dish containing cooled PBS. Whole brains were removed and placed in a tube containing pre-chilled RNAlater (Thermo Fisher Scientific). Whole brains were also collected from mice at postnatal days 1, 7 and 20. Hippocampi were dissected from adult mice under a dissection microscope, following permeation of samples with RNAlater.

2.12.3 RNA extraction from mouse brain tissue

Samples were transferred from RNAlater to homogeniser tubes containing ceramic beads. 1 ml Trizol® reagent (Thermo Fisher Scientific) was added to the tubes and samples were homogenised on a precellys 24 homogeniser (Bertin Technologies). The homogenate was transferred to a microcentrifuge tube and incubated at room temperature for 5 minutes, followed by the addition of 200 µl chloroform. Tubes were shaken vigorously for 20 seconds and centrifuged at 10,000 x *g* for 18 minutes at 4°C. The aqueous phase then underwent RNA extraction following the procedure described in section 2.5. The remaining phases, containing protein and DNA, were stored at -80°C for future use.

2.12.4 Genomic DNA extraction from mouse tissue

Ear notches from adult mice were used to extract DNA for genotyping. For embryonic and samples up to post-natal day 20, tail tips were used. 300µl 50mM NaOH was added to the ear notches and tail tips at room temperature followed by incubation at 95°C for 15 minutes to break down the tissues. Samples were mixed by vortex and pulse spun in a centrifuge to collect any condensation. 50µl 1M Tris-HCl was added and samples vortexed for 15 seconds. Samples centrifuged at 13,000rpm for 6 minutes and the supernatant as collected for genotyping. Samples were stored at -20°C until required.

2.12.5 Sex determination

Adult mice were sexed by visual inspection. Sex was determined in the remaining mice by PCR amplification of exons 9 and 10 of *Jarid1c* on the X-chromosome and *Jarid1d* on the Y-chromosome (Clapcote and Roder, 2005). This results in two products of 302 and 331 bp in male samples, and a single product of 331 bp in females.

2.12.6 L100P genotyping

Genomic DNA was amplified by PCR using the following primers: forward 5'-AGA CCA GGC TAC ATG AGA AGC-3', and reverse 5'-AAG CTG GAA GTG AAG GTG TCT-3' (Clapcote *et al.*, 2007. PCR products were sequenced as described in section 2.2.

2.13 Genetic association analysis

2.13.1 Sample information

The Generation Scotland cohort was used for the genetic association analysis (Smith *et al.*, 2013). Genotype data were available from 19,994 individuals for 561,125 SNPs.

2.13.2 Selection of markers

2.13.2.1 *Quality control*

For each gene, markers from subset of control individuals (N=100, 50% male) including markers within approximately 1 Mb of the gene's flanking regions were uploaded to Haploview. Markers were filtered out of the analysis if they had (*i*) a Hardy-Weinberg *p*-value of ≥ 0.001, (*ii*) a genotype % of ≤ 94 %, (*iii*) at least one Mendelian error, or (*iv*) a minor allele frequency of ≥ 0.05.

2.13.2.2 *Defining the gene region*

Linkage Disequilibrium (LD) blocks were drawn based on the solid spine method, whereby the first and last marker of a block are in tight LD with intermediate markers (D' ≥ 0.8), but intermediate markers may not necessarily be in LD with one another. LD blocks were manually combined if adjacent blocks had a multiallelic D' of ≥ 0.8 to generate blocks up to a maximum length of 500 kb. Singleton markers were manually assigned to individual blocks and adjacent blocks were joined if they met the D' requirements. The gene region was defined as the range between the 5'-most marker of the block containing the gene's TSS, and the 3'-most marker of the block containing the end of the transcript, based on UCSC Genome Browser's hg19 coordinates (GRCh37).

2.13.2.3 *Identification of tagging SNPs*

All markers within the LD-block defined gene region for 4700 non-depressed individuals were submitted to Haploview's Tagger to identify tagging SNPs using an R^2 cut-off of ≥ 0.8.

2.13.2.4 *Random forest implementation*

The *Random Jungle* program was used to perform random forest analysis on SNP lists (Schwarz *et al.*, 2010). Files were prepared in .ped and .map formats containing tagging SNPs for the genes to be tested for association and phenotype in *PLINK* (Purcell *et al.*, 2007), using the following command in a Linux terminal window:

plink --recode --bfile [binary pedfile name] *--extract* [tag SNP filename.txt] *--out* [output filename]

The output file was converted to a .raw format displaying minor allele counts (0, 1 and 2) for each marker in each individual using the following command:

plink --recodeA --file [filename] *--out* [output file]

The .raw file was read into R for further analysis. As the random forest algorithm cannot handle missing data, "NA" values related to each marker were converted to the median minor allele count in the sample set for that marker using *randomForest*'s *na.roughfix()* function. A copy of this file was created and the phenotype was permuted using the base function *sample()*. Random forest analysis was performed on the original .raw file and the permuted .raw file 100 and 1000 times, respectively, using the following command:

rjungle –f [file] *–i 4 –ntree 1000 –z n –u –D PHENOTYPE –U 63 –o* [Outputfile.$i]

This command performs random forest on the .raw file using 1000 trees for each of iteration (*-ntree* argument). Bootstrap aggregating (bagging) was performed without replacement (*-u* argument) and the raw permutation importance measure was selected as the method for estimating variable importance (*-i* argument). If the outcome variable was continuous, an additional argument "*–y 3*" was added to select the appropriate base classifier.

Empirical *p*-values were calculated based from the real and permuted VIMs for each SNP using the following calculation:

$$P_{SNPX} = \text{(Permuted VIM}_{SNPX} > \text{Median Real VIM}_{SNPX}) / 1000$$

SNPs with an empirical *p*-value ≤ 0.05 were selected for analysis of epistatic interactions.

2.13.3 Identification of gene-gene interactions

Allele counts of genes selected for the epistasis study were read into R. For pairwise tests, a table of all possible two-way combinations from the gene set was generated using the *combn()* function available in the *utils* package. Combinations were considered if they involved SNPs within at least two genes. Logistic and linear regression was performed depending on whether the phenotype was categorical or quantititave, respectively. For each SNP combination, a full model was created containing the additive terms of each gene, along with the an interaction term using the *glm()* function available in the *stats* package. The top 20 principal components as estimated by GCTA (Yang *et al.*, 2011) were fitted to account for population structure.

Two-SNP interaction: $Phenotype \sim \beta_1 SNP_1 + \beta_2 SNP_2 + \beta_3(SNP_1 * SNP_2) + PC1\text{-}20$

Three-SNP interaction: $Phenotype \sim \beta_1 SNP_1 + \beta_2 SNP_2 + \beta_3 SNP_3 + \beta_4(SNP_1 * SNP_2) + \beta_5(SNP_1 * SNP_3) + \beta_6(SNP_2 * SNP_3) + \beta_7(SNP_1 * SNP_2 * SNP_3) + PC1\text{-}20$

A reduced model was also created, omitting the interaction term:

Two-SNP interaction: $Phenotype \sim \beta_1 SNP_1 + \beta_2 SNP_2 + PC1\text{-}20$

Three-SNP interaction: $Phenotype \sim \beta_1 SNP_1 + \beta_2 SNP_2 + \beta_3 SNP_3 + \beta_4 (SNP_1 * SNP_2) + \beta_5 (SNP_1 * SNP_3) + \beta_6 (SNP_2 * SNP_3) + PC1\text{-}20$

The full and reduced models were compared using a likelihood ratio test (LRT) to assess whether the interaction term significantly affects the goodness-of-fit of the full model. A chi–square *p*–value threshold of ≤ 0.05 was implemented to define a significant difference between the goodness-of-fit of each model.

Chapter 3

Analysis of gene expression in t(1;11) family lymphoblastoid cell lines

3 Analysis of gene expression in t(1;11) family lymphoblastoid cell lines

3.1 Introduction

Several lines of evidence support a role for DISC1 in transcriptional dysregulation. Malavasi *et al.* (2012) reported that DISC1 had a dose-dependent inhibitory effect on ATF4-mediated gene expression. DISC1 has also been linked to cAMP-signalling through its interaction with PDE4B (Millar *et al.*, 2005), which may result in dysregulation of cAMP-responsive genes in translocation carriers where *DISC1* expression is altered. Haploinsufficiency of DISC1 has been reported in t(1;11) LCL samples (Millar *et al.*, 2005). Taken together, these findings suggest that t(1;11)-mediated reductions in DISC1 protein expression may result in altered expression of other genes. Another mechanism whereby the translocation might impact upon gene expression is through its impact on *DISC2* - the non-coding RNA antisense to *DISC1*. Natural antisense transcripts (NATs), such as *DISC2,* have been implicated in multiple processes, including transcriptional regulation and disease (Faghihi and Wahlestedt, 2009). Furthermore, *DISC2* has been proposed to play a regulatory role in *DISC1* expression (Millar *et al.,* 2004; Chubb *et al.*, 2008). In the context of the t(1;11) translocation, like *DISC1*, *DISC2* expression may also be disrupted. On chromosome 11, the translocation directly disrupts the non-coding transcript *Boymaw*, a brain-expressed NAT. Zhou *et al.* (2010) proposed that *Boymaw* has a regulatory effect on *CHORDC1*, a gene with which it partially overlaps on the antisense strand. In addition to the direct disruption of the genes on these chromosomes, fusion proteins generated by the t(1;11) translocation have been proposed to result in aberrant interactions with binding partners of DISC1, potentially affecting its downstream effects on gene expression (Eykelenboom *et al.*, 2012). A further mechanism through which the translocation might impact upon gene expression is through the passive transmission of regulatory variants, i.e. eQTLs, on the derived chromosomes.

This chapter describes an analysis of differential gene expression in LCL-derived RNA from 13 individuals from the t(1;11) family. The underlying hypothesis of this work was

that the t(1;11) translocation might have far-reaching effects on gene expression, not only through the direct disruption of the breakpoint genes and their regulatory functions described above, but also through downstream effects due to the disruption of putative regulatory roles of these genes, and factors in linkage disequilibrium with the translocation.

3.2 Sample quality control measures

3.2.1 RNA integrity

High quality, intact total RNA is a necessity for obtaining reliable information from gene expression experiments. The quality of LCL-derived RNA samples from thirteen t(1;11) family members was assessed on the Agilent 2100 Bioanalyser by Audrey Duncan at the Wellcome Trust Clinical Research Facility (WTCRF), Edinburgh. The output - RNA integrity number (RIN) - was used as a metric for sample quality. A study by Thompson *et al.*, (2007) observed a substantial increase in false positives in microarray analysis of rat liver RNA samples with a RIN value of ≤ 7. This value was therefore set as the minimum acceptable measure of sample quality. RIN values of the LCL samples tested were high, ranging from 9.7 to 10 (Figure 3.1).

3.2.2 Assessment of genomic DNA contamination

To ensure successful DNase treatment during RNA preparation, synthesised cDNA samples underwent PCR analysis with an assay that amplifies a region spanning exons 7-8 of the *WD repeat domain 1* gene (*WDR1*) on chromosome 4. A genomic DNA positive control sample and negative controls for both the reverse transcription reaction (non-enzyme and non-template controls), and a negative PCR control (non-template control) were also subject to amplification. All cDNA PCR products were the expected size of 238 bp corresponding to the spliced region, while the larger, intron-containing product of 590 bp was observed in the genomic DNA control samples. This indicated that the LCL-derived cDNA samples used in this study were free of genomic DNA (Figure 3.2).

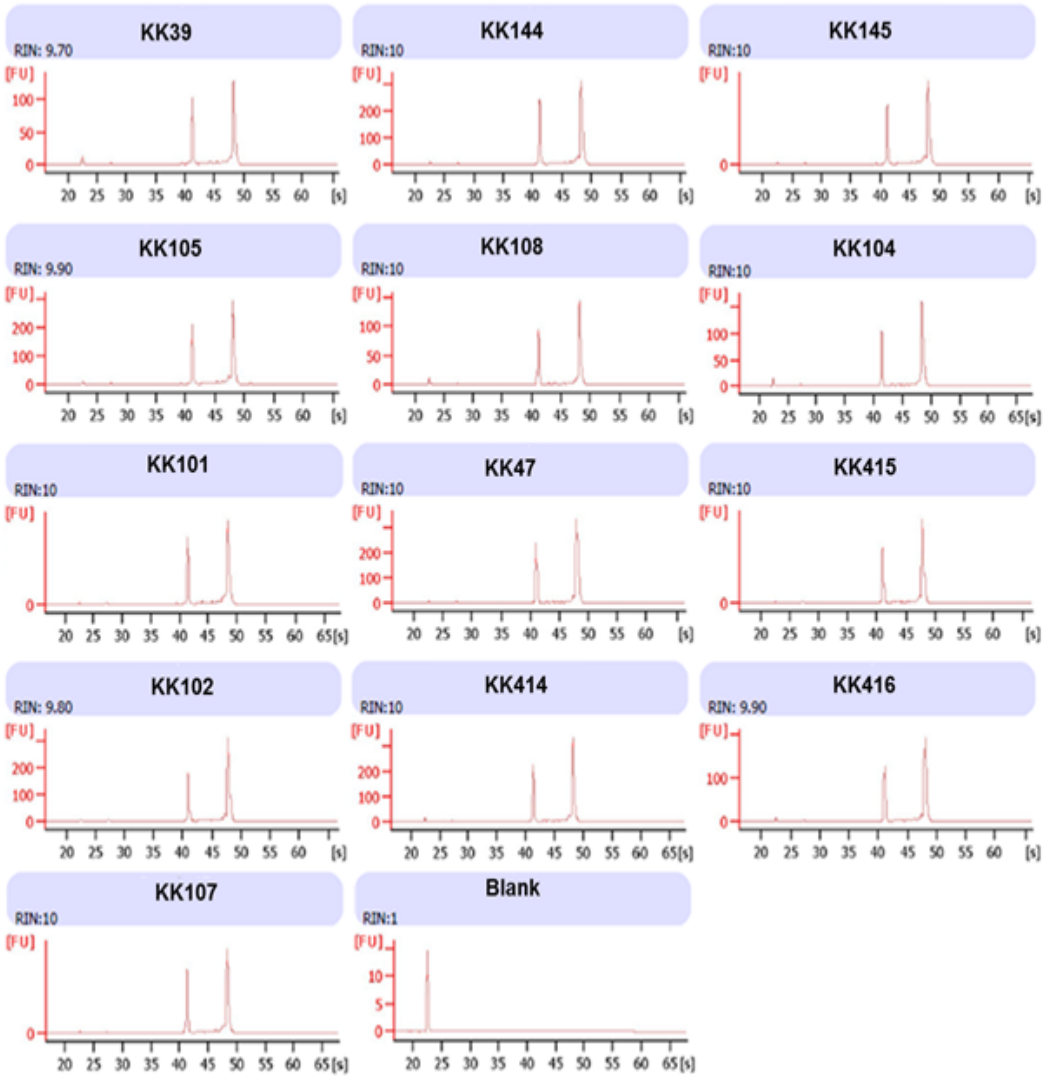


Figure 3.1: Agilent Bioanalyser's electropherogram output for visualising RNA integrity.

Electropherograms are displayed for thirteen RNA samples from t(1;11) family LCLs. The associated RIN value accompanies the sample ID. Time in seconds is shown on the x-axis while fluorescence units are presented on the y-axis. The left and right peaks in each sample prefixed with "KK" represent the 18S and 28S ribosomal subunits respectively. The peak in the sample labelled "Blank" represents the marker

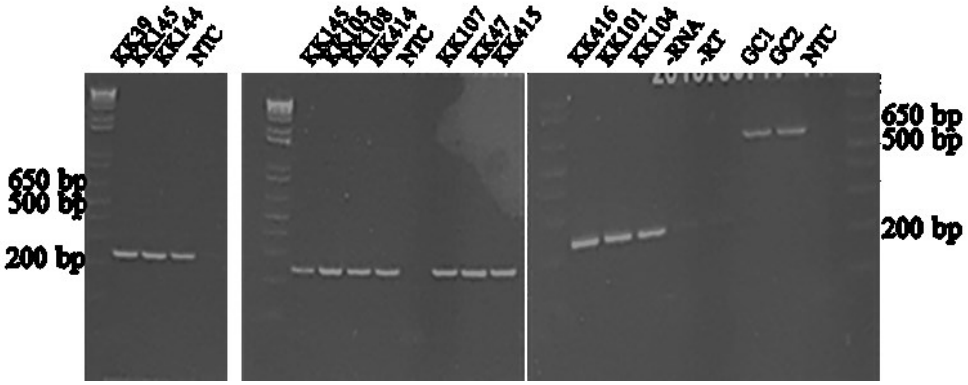


Figure 3.2: Assessment of genomic DNA contamination in t(1;11) RNA.

Shown are the PCR amplicons of WDR1 *exons 7-8 run for t(1;11) cDNA samples, two genomic DNA control samples, and negative non-template, non-enzyme and non-RNA controls. Lower bands (238 bp) consist correspond to cDNA samples while upper bands (590 bp) correspond to genomic DNA control samples. In each lane, cDNA samples are presented (prefixed "KK"), non-template controls (NTC), non-RNA controls (-RNA), non-reverse transriptase controls (-RT), and genomic DNA controls (prefixed "GC").*

3.3 Microarray-based gene expression analysis of t(1;11) LCLs

3.3.1 Quality control measures

The quality of the microarray data was assessed using both internal, array-based controls and custom measures available in the R package *lumi*. Probe hybridisation, staining, sample integrity, probe-binding specificity, sample variance and sample reproducibility were all assessed.

3.3.1.1 *Internal control probes*

The Illumina HumanHT-12 v4 expression array has five sets of internal controls. These allow the user to assess sample integrity and ensure hybridisation and biotin-labelling steps were successful. Figure 3.3 provide a summary of the findings from each set of control probes.

Three sets of hybridisation controls are present on the array: Cy3-labelled hybridisation controls, low stringency controls, and high stringency controls. The Cy-3-labelled controls consist of three pairs of probes targeting oligonucleotides present at low, medium

and high concentrations on the array to determine successful hybridisation. If successful hybridisation has occurred, such probes would be expected to yield a signal gradient consistent with their targets' concentration on the array (i.e. low, medium and high signal). Signal from these probes suggested hybridisation was successful as the expected gradient of low, intermediate and high signal was observed for the corresponding probes (Figure 3.3A).

Probe hybridisation stringency is determined by two sets of control probes. There are eight low-stringency control probes present on each array: four containing probes with two mismatched bases (mm2), and four containing the corresponding perfect-match sequence (pm). Signal from mm2 probes should be significantly lower than that of the pm probes under appropriate levels of hybridisation stringency. The mean signal observed in the pm probes was 4880 units - significantly higher than that observed in the mm2 probes, at 640 units (Welch's t-test $p = 3.68 \times 10^{-28}$; Figure 3.3B). Signal generation was assessed by two probes hybridizing to biotin-labelled targets in the hybridisation buffer. A positive hybridisation signal was observed for these probes, indicating successful staining (Figure 3.3C).

Two sets of internal controls are present on the array to determine sample integrity. Each array contains 770 negative control probes with permuted sequences non-specific to the human genome. Any signal detected from these probes would be deemed an artefact due to background, non-specific binding, or cross-hybridisation. Background signal intensity was calculated from the mean signal intensity of 82.6 units for these probes. Noise in the array was determined by the standard deviation of the signal of the negative control probes (Figure 3.3D). Seven control probes targeting housekeeping genes provide a measure of the integrity of each sample on the array. It would be expected that the signal for these probes be significantly higher than the average signal obtained across all genes on the array if a sample were intact. The average signal intensity obtained from the housekeeping probes across all samples was 7099 units while the average signal intensity across all genes was 212. This indicated the samples were of high quality (Welch's *t*-test $p = 1.62 \times 10^{-16}$; Figure 3.3E).

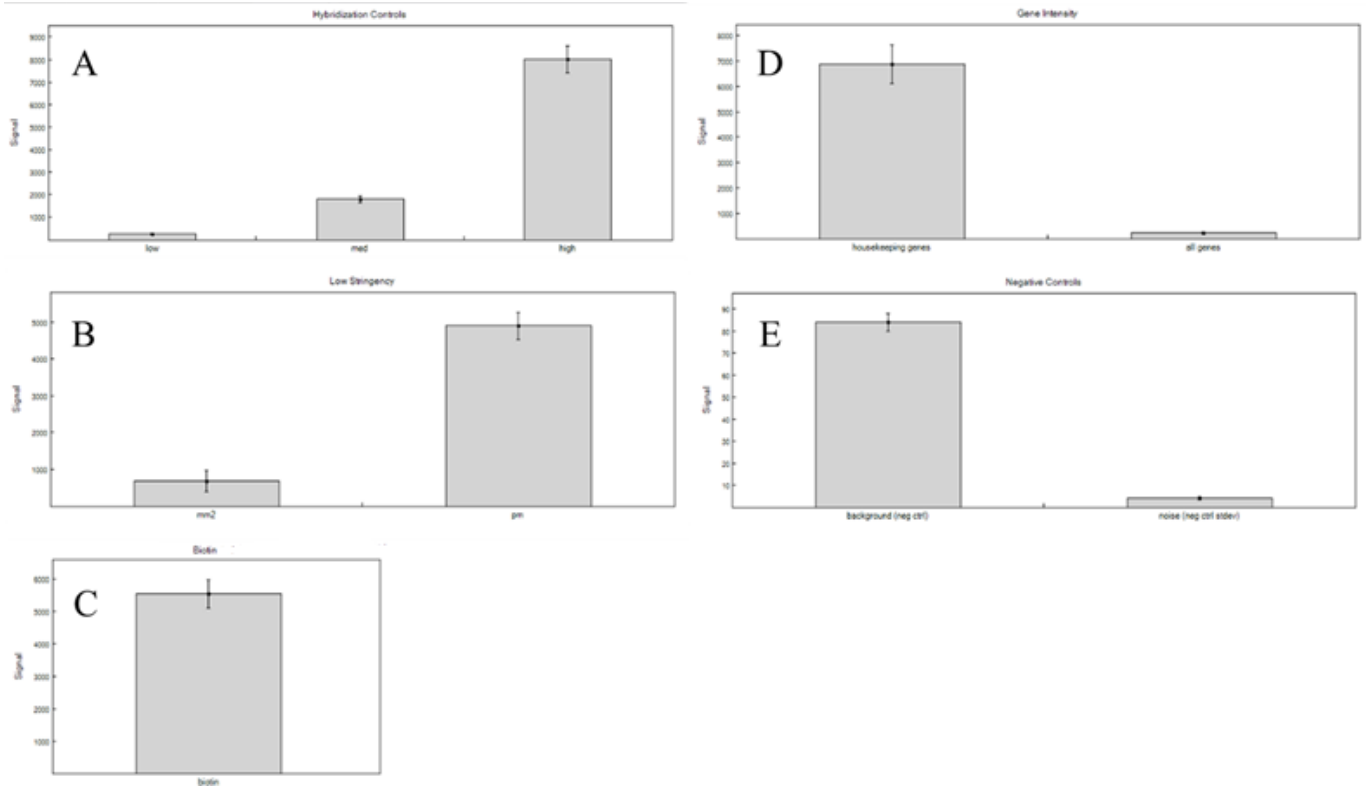


Figure 3.3: Illumina HT-12 control probe summary provided by GenomeStudio software.

Shown is a screenshot of control probe plots as displayed in GenomeStudio for signal from (A) hybridisiation controls, (B) low stringency controls consisting of mismatched (left) and perfectly matched (right) probes, (C) biotin staining controls, (D) housekeeping gene control probes, and (E) negative controls. Control probe types are displayed on the x-axes and are represented by bars.

3.3.1.2 *Technical replicates*

In order to assess the degree of reproducibility and inter-chip variability of the HumanHT-12 array, a technical replicate was used. One sample selected at random, a t(1;11) carrier sample (KK108), was used as a technical replicate. This sample was run in triplicate on one chip, one replicate being run on array position K. A between-chip replicate was also run in position K on the second chip.

Reproducibility was assessed by pairwise comparisons of the replicate samples both within and between chips. High correlation was observed between within-chip replicate samples, and between-chip replicates (Pearson's $R^2 \geq 0.995$; Figure 3.4). The coefficient of variation (CV) was calculated for the within-chip triplicates to determine the variation of 47,323 probes across each array. The median CV was 3.5% and more than 99.9% of probes had a CV of < 20% between the three replicates. This was repeated on the two inter-chip samples. A marginal increase in variation was observed between chips. The median CV was 4.7 % and 98.4 % of probes had a CV of <20 % between replicates. These results suggested both within- and between-chip variability was negligible.

3.3.1.3 *Confirmation of sample gender*

To identify samples whose gender might have been incorrectly annotated, multidimensional scaling was performed on Y-chromosome probes, with the expectation that samples would cluster by gender. Samples were annotated "M" and "F" corresponding to males and females, respectively, based on their annotation. Two visible clusters were present corresponding to gender, suggesting all samples were annotated correctly for gender (Figure 3.5A). To support this, hierarchical clustering analysis of expression data from Y-chromosome probes with CV ≥ 0.2 was performed. Two clusters were present corresponding to males and females (Figure 3.5B).

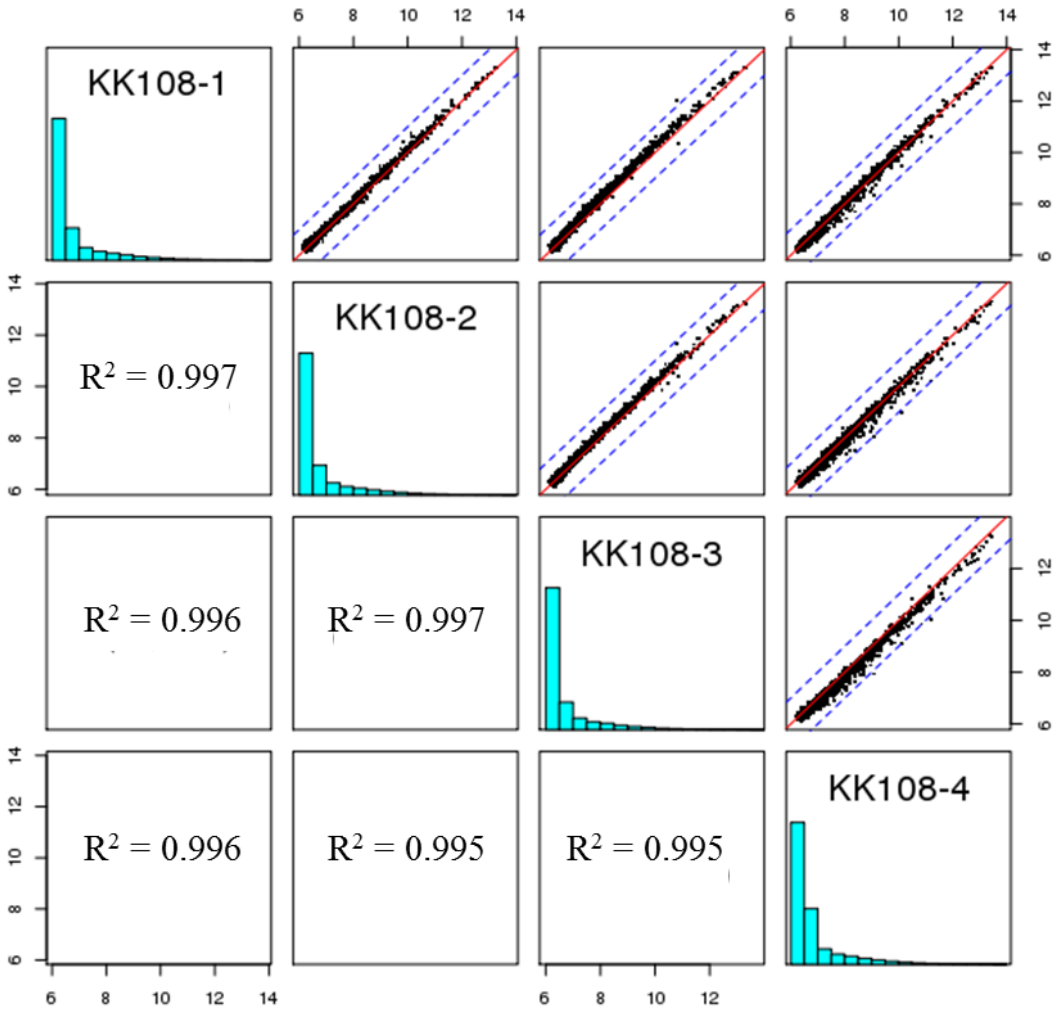


Figure 3.4: Pairwise analysis of technical replicate samples.

Shown are all pairwise comparisons between four technical replicate samples (KK108). Density plots of signal for each technical replicate are shown in the diagonal. Below the diagonal are Pearson's correlation coefficients (R^2) for each pairwise comparison. Above the diagonal are scatter plots of log2-transformed intensities. Both x– and y-axes present log_2-transformed signal intensities.

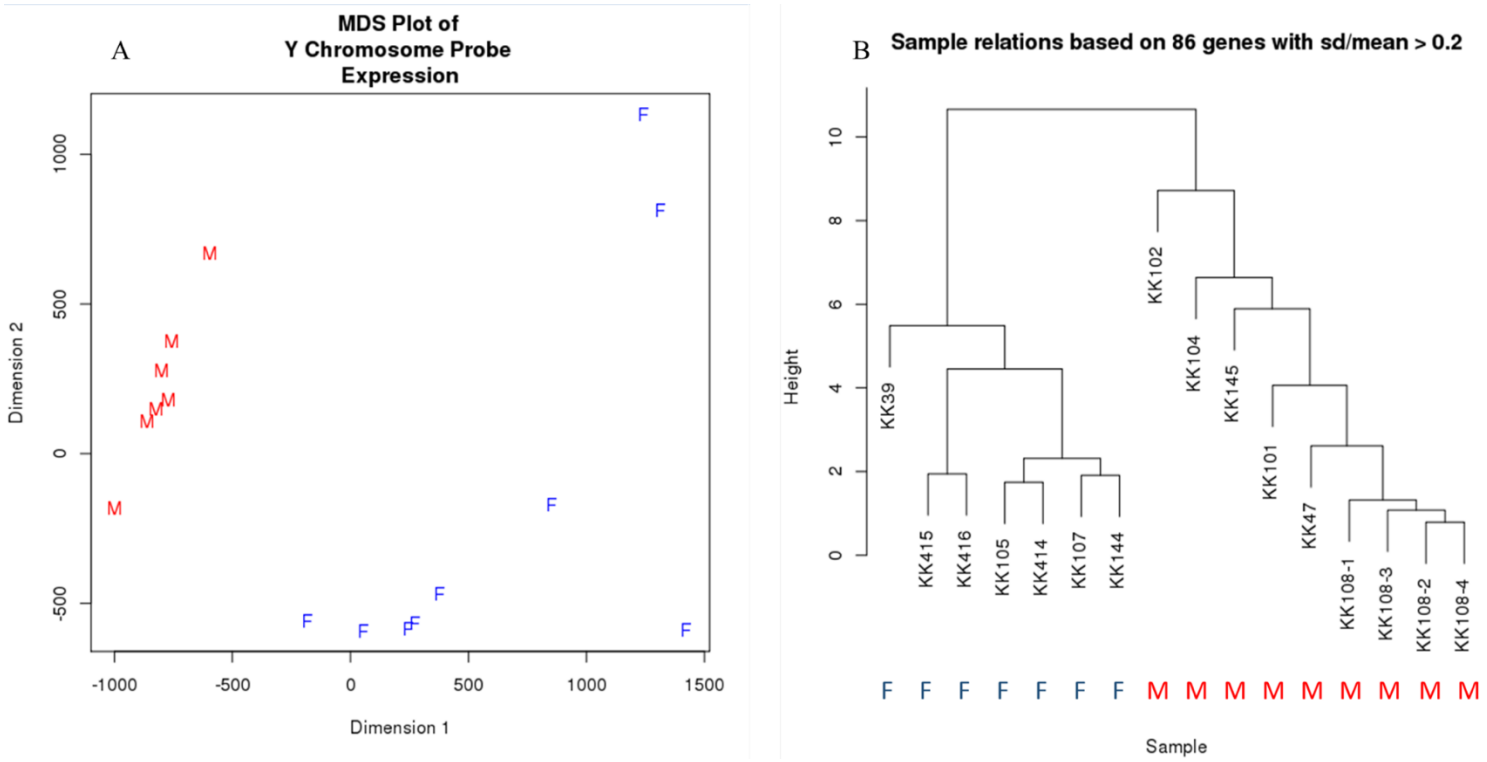


Figure 3.5: MDS and hierarchical clustering of Y-chromosome expression.

Shown in Figure 3.6A is a multidimensional scaling plot of expression data for Y-chromosome probes in thirteen LCL-derived RNA samples. The relationship between samples was assessed in two dimensions (x- and y-axes). Shown in Figure 3.5B is a dendrogram based on expression data from variable Y-chromosome probes (CV >0.2; Figure 3.5B). Male ("M") samples and female ("F") samples are color-coded red and blue, respectively. Dendrogram height is representative of variability between samples. Sample IDs are prefixed with "KK". The four technical replicates (KK108) are suffixed with numbers 1-4.

3.3.1 Data preprocessing

Due to the high-dimensionality and noise associated with array-based data, a critical step prior to analysis is data preprocessing. Data were first transformed to reduce within-array variance and normalised to remove background noise. Probes were then filtered based on their detection *p*–value, thus removing the burden of unnecessary multiple tests, and to ensure the dataset consisted of signal from informative probes only.

3.3.1.1 *Data transformation*

As variance generally increases with signal intensity, transformation is a common step prior to the analysis of microarray data. In order to stabilise the variance to meet the assumptions of downstream statistical tests, Log transformation is commonly used. As this can cause problems such as inflation of variance in low-signal probes, Lin *et al.* (2008) have designed a method of variance stabilisation based on variance stabilising normalisation (Rocke and Durbin, 2001). Variance stabilising transformation (VST) uses within-array technical replicates to model the relationship of the mean and variance in each sample. Samples were transformed by VST (Figure 3.6).

3.3.1.2 *Data normalisation*

In order to remove systematic sources of variation such as hybridisation or levels of RNA, robust spline normalisation (RSN) was performed on the data. RSN combines features of quantile normalisation and loess normalisation (Du *et al.*, 2008). Hierarchical cluster analysis was performed on the data pre- and post-normalisation (Figure 3.7). A reduction in variability was observed across the data as demonstrated by a reduction in the height of the dendrogram (3.7B).

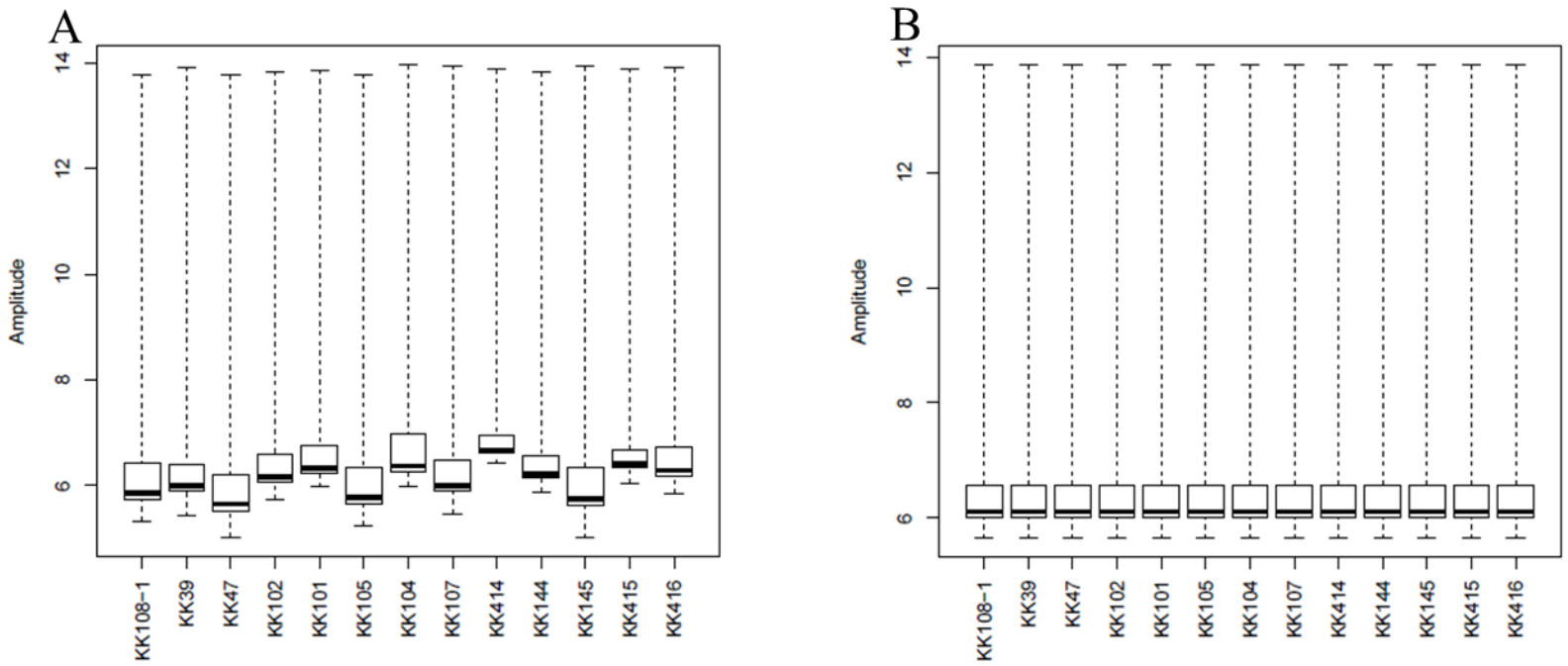


Figure 3.6: Distribution of gene expression signal pre- and post-variance stabilising transformation.

Box and whisker plots are shown corresponding to each sample (x-axis) displaying the distribution of background-subtracted signal (y-axis) before variance stabilising transformation (A), and after transformation (B).Thick horizontal lines in boxes represent the median signal per sample, upper and lower bounds of the boxes represent the third and first quartiles, respectively, and whiskers extend to the minimum and maximum signal detected in each sample.

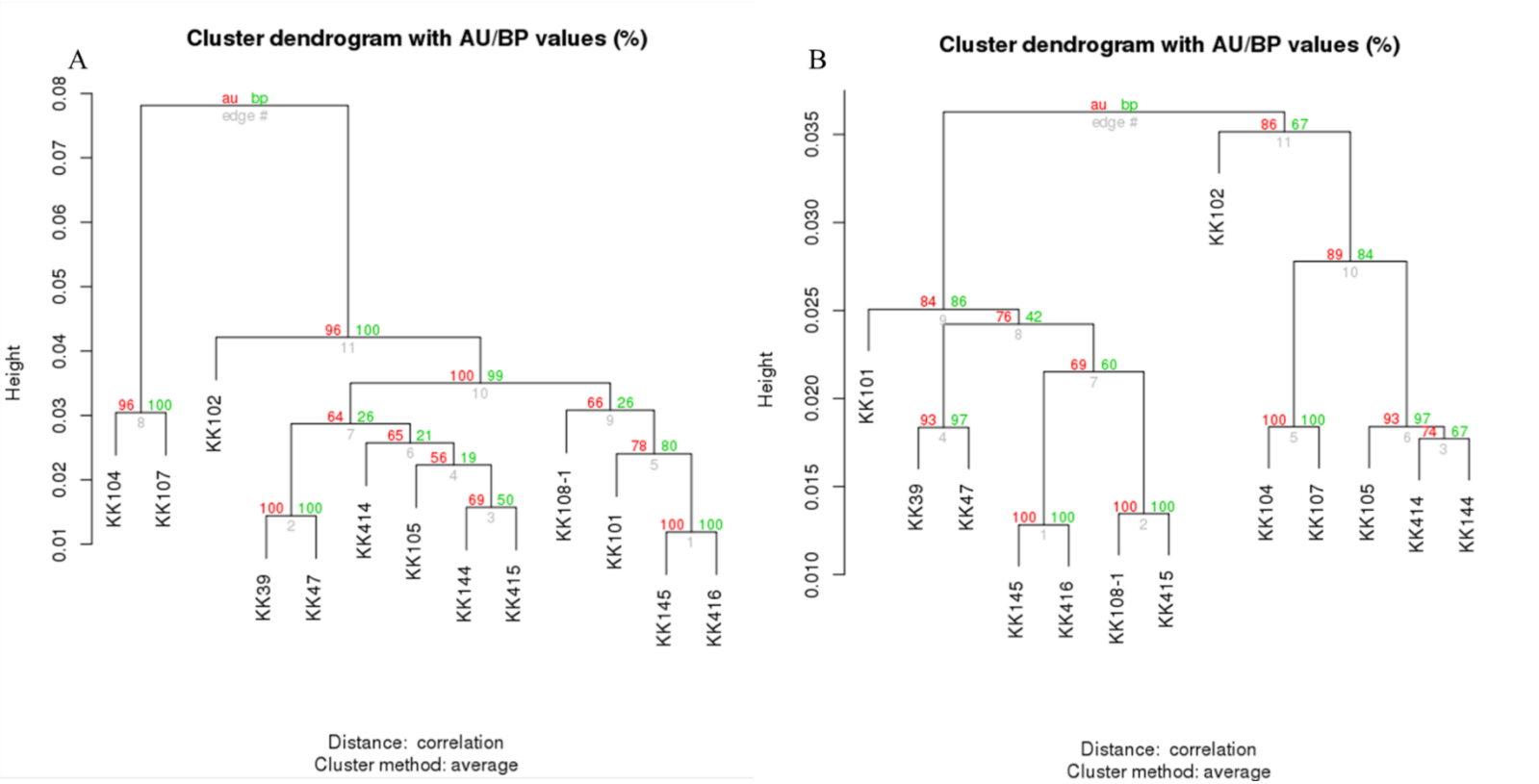


Figure 3.7: Dendrogram of gene expression data for 13 lymphoblastoid samples before and after normalisation.

Cluster dendrograms are presented pre- transformation and normalisation (A) and post- transformation and normalisation (B). *The height of the dendrogram (y-axis) represents sample variability. Red values at each node correspond to the approximately unbiased (au) p-value while green values correspond to the bootstrap probability (bp) percentage p–values. au values > 95 clusters that are strongly supported by the data. Grey digits indicate edge numbers.*

3.3.1.3 *Probe filtering*

Data filtering was performed to ensure all probes tested for differential expression mapped to genes detected above background on the array. This has the additional benefit of reducing the burden of multiple testing. Probes with a detection *p*–value of ≥ 0.05 were removed from the dataset prior to differential expression analysis. Before probe filtering, a total of 47,323 probes were detected on the array. Post-probe filtering, data from 29,497 probes remained for analysis of differential gene expression.

3.3.2 Array-based identification of differentially expressed transcripts

To identify latent, unmodeled sources of variation in the data, surrogate variable analysis (SVA) was performed, identifying one significant surrogate variable. Differential gene expression was assessed between t(1;11) carriers and non-carriers by linear regression fitting gender and the surrogate variable identified by SVA. No transcripts were found to be significantly differentially-expressed after correction for multiple testing (Benjamini-Hochberg FDR $q \geq 0.05$ for all sites). The genomic inflation factor (λ) was calculated from the data showing deflation of *p*-values for differential expression between t(1;11) carriers and non-carriers ($\lambda = 0.93$; Figure 3.8). This might suggest the study was underpowered to detect genome-wide significant results. Therefore, probes nominally significant for differential expression in t(1;11) carriers were considered for further analysis ($p \leq 0.05$), and a fold-change cut-off threshold of ±1.2 was applied to these probes to define differential expression (n = 303; Figure 3.9; Table 3.1).

Among the 303 genes that met these criteria, the largest negative fold-change in expression seen in t(1;11) carriers was observed at *ependymin related 1*, on chromosome 7 (*EPDR1*: $p = 6.32 \times 10^{-5}$, FC = -2.62). The greatest positive fold-change in expression seen in carriers was also observed on a chromosome 7 gene, *GTPase, IMAP family member 6* (*GIMAP6*: $p = 5.02 \times 10^{-3}$, FC = 1.70). The lowest *p*-value for differential expression was associated with *interleukin-17 receptor B*, on chromosome 3 (*IL17RB*; $p = 3.52 \times 10^{-6}$, FC = -2.58).

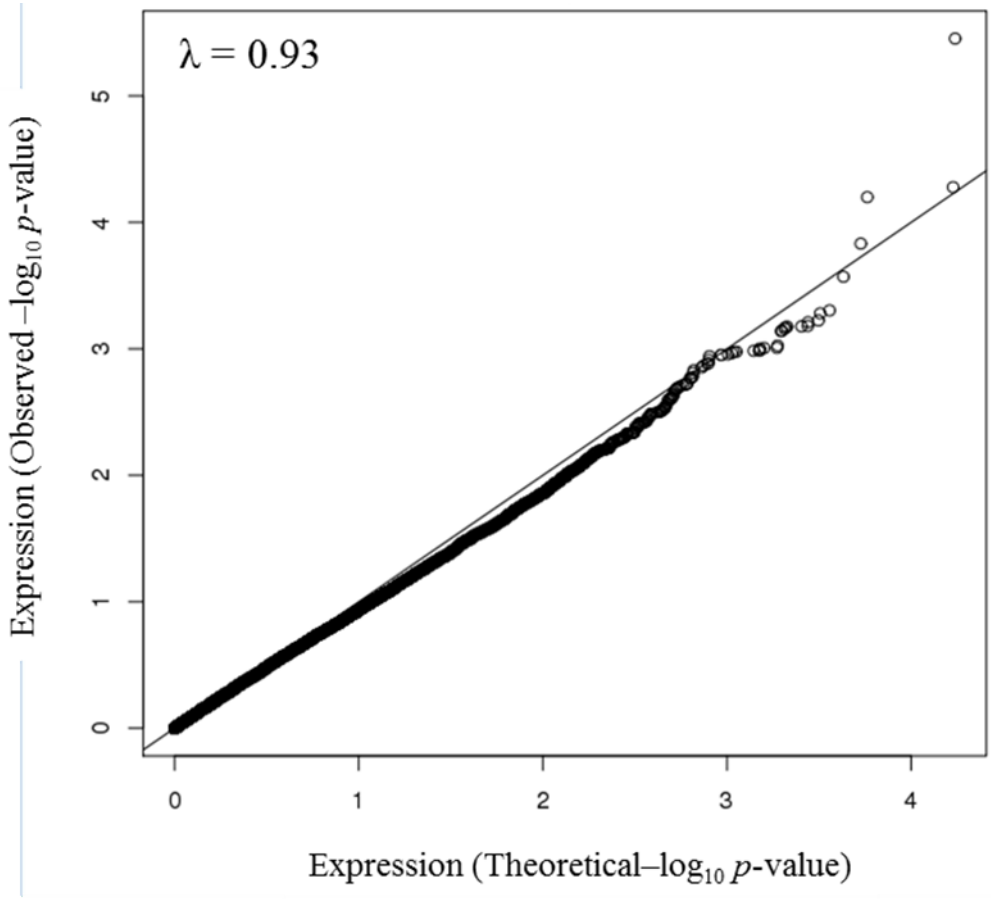


Figure 3.8: Quantile-quantile plot of *p*-values for differential gene expression between t(1;11) carriers and non-carriers.

Shown are observed unadjusted –log$_{10}$ p-values (y-axis; circular points) plotted over the expected distribution of –log$_{10}$ p-values under the null (x-axis; solid diagonal line). The genomic inflation factor λ is presented in the upper-left corner of the plot.

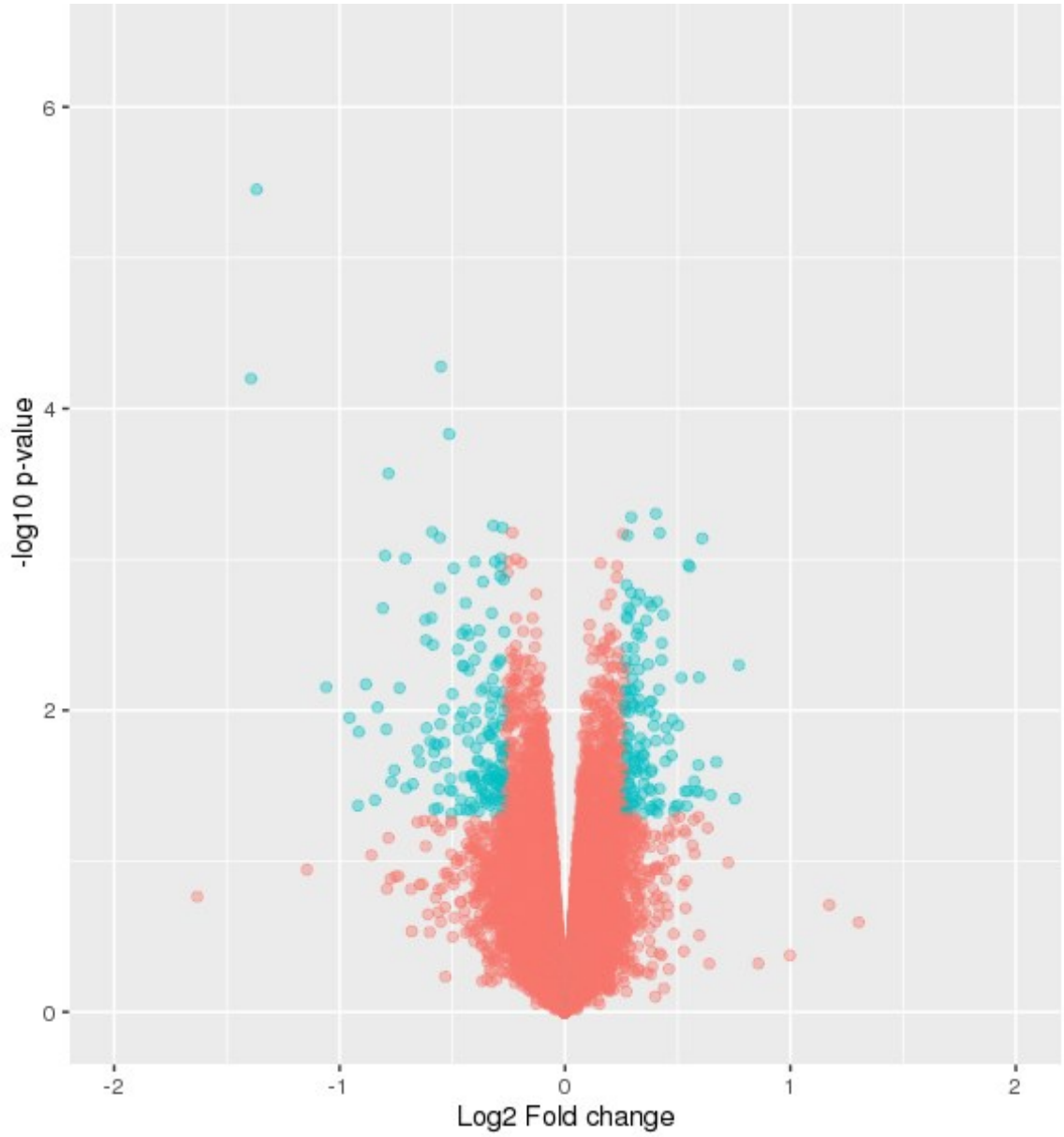


Figure 3.9: Volcano plot of gene expression data comparing t(1;11) carriers to non-carriers.

Each point represents a transcript measured by the Illumina HT12 array in this experiment. The x-axis displays the log_2 fold-change of a given transcript while the y-axis presents the $-log_{10}$ p – value. Blue points represent probes that have met the both the p-value cut-off of 0.05 and absolute fold-change of 1.2. Red points represent probes that have failed to meet at least one of these criteria.

ProbeID	Chromosome	Gene	Fold-Change	t	Differential Expression *p*-value
ILMN_1767523	3	*IL17RB*	-2.58	-9.89	3.52×10^{-6}
ILMN_2098446	18	*PMAIP1*	-1.46	-7.10	5.27×10^{-5}
ILMN_1675797	7	*EPDR1*	-2.63	-6.94	6.32×10^{-5}
ILMN_2046470	14	*DAAM1**	-1.43	-6.21	0.0001
ILMN_1659270	5	*OTP*	-1.72	-5.73	0.0003
ILMN_1709484	15	*BLM*	1.32	5.26	0.0005
ILMN_1708936	9	*EXOSC3*	1.23	5.22	0.0005
ILMN_1724181	4	*IL15*	-1.25	-5.13	0.0006
ILMN_1789106	1	*IPP*	-1.21	-5.10	0.0006
ILMN_3251423	3	*CHDH*	-1.50	-5.05	0.0007

Table 3.1: Differentially expressed genes in t(1;11) carrier LCL RNA.

Summary of genes meeting an absolute expression fold-change ≥ 1.2 in t(1;11) carriers with an associated p-*value ≤ 0.05 for differential expression. In order of column appearance are the HT12 array probe's associated Illumina ID, the probe's target chromosome, gene name, fold-change in t(1;11) carriers, differential expression t-statistic, and unadjusted* p-*value for differential expression between t(1;11) carriers and non-carriers. Genes flagged with a red asterisk (*) are those with multiple non-overlapping probes meeting the criteria for differential expression in t(1;11) carriers. Shown are the top 10 differentially expressed genes ranked by* p-*value. The total list of genes meeting the* p-*value and fold-change cut-off is presented in Appendix I (Table A1).*

During the probe filtering step, a probe was retained for analysis if it was expressed above background in at least one sample (detection $p \leq 0.05$). As this may have led to the inclusion or exclusion of probes expressed in a group-specific manner, a Fisher's exact test was performed to determine the relationship between probe detection and translocation carrier status. Of the 47,323 probes present on the array, 17,826 were not expressed in any samples, while 13,237 were expressed above background in all samples. The expression levels of the remaining 16,260 probes were assessed for association with t(1;11) carrier status: a total of 170 probes showed significant association (Fisher's exact test $p \leq 0.05$). Of these, one probe was expressed above background level only in those samples without the t(1;11) chromosome (Fisher's exact test $p = 0.0008$). This probe corresponded to a region the chromosome 15 gene *ATPase, class V, type 10A* (*ATP10A*). Of these 170 probes, only two met the both the fold-change and *p*–value criteria for differential gene expression: the chromosome 3 gene *peroxisome proliferator-activated receptor, gamma* (*PPARG*, $p = 0.007$, FC = -1.28), and the chromosome 1 gene, *coagulation factor V* (*F5*, $p = 0.01$, FC = -1.25).

Among the genes detected by multiple probes on the HT-12 expression array, 16 were significantly differentially expressed in t(1;11) carriers. These comprised 15 pairs of probes corresponding to 15 genes, and one set of three probes which mapped to the chromosome 1 gene *ATPase, Na+/K+ transporting, beta 1 polypeptide* (*ATP1B1*, $p \leq 0.03$, average FC = -1.56). For all of these gene sets, directions of expression fold-change were consistent among their associated probes.

With regards to the translocation-affected chromosomes, 1 and 11, 30 and 17 genes were differentially expressed, respectively. *DISC1* was not assessed for differential expression as none of the four probes present on the array mapping to the gene were detected above background in these samples, and were therefore removed at the probe-filtering stage. A summary of the *DISC1*-targeting probes on the HumanHT-12 array, relative to the t(1;11) breakpoint is presented in Figure 3.10. Among the differentially expressed genes on chromosome 11 was *SORL1*, a member of the brain-expressed

Sortilin gene family. *SORL1* was downregulated in t(1;11) carriers (p = 0.007, FC = -1.29).

3.3.3 Gene ontology analysis of differentially expressed transcripts

To determine which, if any, molecular functions, biological processes or cellular components were over-represented amongst the above differentially expressed genes, gene ontology analysis was performed on the data. A list of genes ranked by differential expression p-value was submitted to GOrilla for this analysis (n = 16,078 genes). In total, 26 terms were significantly over-represented amongst differentially expressed genes ($q \leq 0.05$; Table 3.2). The most significantly enriched term overall was "nuclear part" (q = 0.0055), in the GO component class. The most significantly enriched GO function was "purine nucleoside binding" (q = 0.0127), while "regulation of protein metabolic process" was the most over-represented GO process (q = 0.0276)

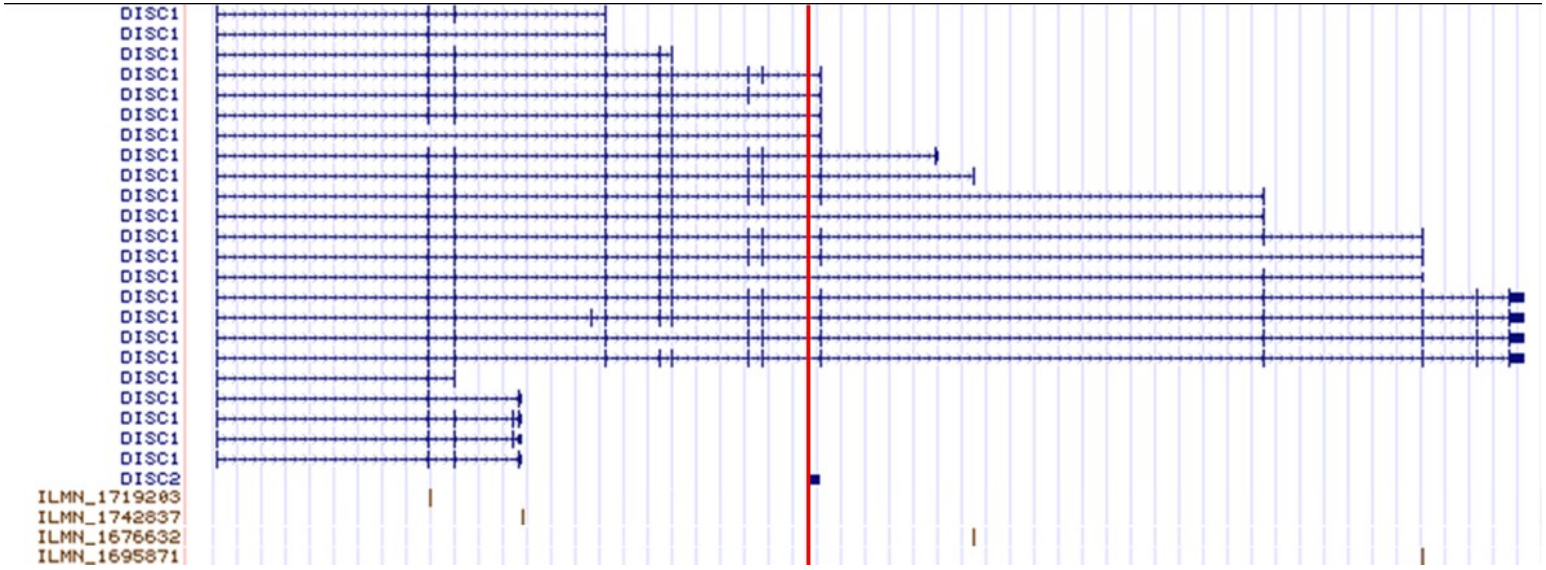


Figure 3.10: Locations of Illumina HT12-v4 probes mapping to *DISC1.*

Shown is a UCSC genome browser screenshot of DISC1 isoforms*,* DISC2 *(blue) and the binding sites Illumina HT12-v4 probes mapping to the region (brown). The vertical red line represents the t(1;11) chromosome 1 breakpoint. Transcripts are represented by horizontal blue lines with each exon indicated by transections.*

GO term	Description	Class	Enrichment	Enrichment *p*-value	FDR *q*-value
GO:0044428	Nuclear part	Component	221/642	3.28×10^{-6}	0.0055
GO:0001883	Purine nucleoside binding	Function	197/1294	1.24×10^{-5}	0.0127
GO:0032550	Purine ribonucleoside binding	Function	197/1294	1.08×10^{-5}	0.0148
GO:0032553	Ribonucleotide binding	Function	199/1294	1.84×10^{-5}	0.0151
GO:0017076	Purine nucleotide binding	Function	198/1294	3.19×10^{-5}	0.0163
GO:0035639	Purine ribonucleoside triphosphate binding	Function	194/1294	2.86×10^{-5}	0.0167
GO:0032549	Ribonucleoside binding	Function	198/1294	8.44×10^{-6}	0.0173
GO:0032555	Purine ribonucleotide binding	Function	197/1294	2.57×10^{-5}	0.0175
GO:0005654	Nucleoplasm	Component	163/640	6.91×10^{-5}	0.0192
GO:0001882	Nucleoside binding	Function	200/1294	4.84×10^{-6}	0.0198
GO:0043232	Intracellular non-membrane-bounded organelle	Component	197/878	3.71×10^{-5}	0.0206
GO:0044464	Cell part	Component	598/640	6.60×10^{-5}	0.022
GO:0044424	Intracellular part	Component	990/1178	5.33×10^{-5}	0.0222
GO:0097367	Carbohydrate derivative binding	Function	228/1346	5.11×10^{-5}	0.0232
GO:0051246	Regulation of protein metabolic process	Process	46/149	4.06×10^{-6}	0.0276
GO:0043228	Non-membrane-bounded organelle	Component	197/878	3.71×10^{-5}	0.0309
GO:0030554	Adenyl nucleotide binding	Function	165/1294	8.61×10^{-5}	0.032
GO:0032559	Adenyl ribonucleotide binding	Function	164/1294	7.92×10^{-5}	0.0324
GO:0016462	Pyrophosphatase activity	Function	105/1382	0.0001	0.0338
GO:0005524	ATP binding	Function	161/1294	0.0001	0.0348
GO:0016818	Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	Function	105/1382	0.0001	0.0356
GO:0016817	Hydrolase activity, acting on acid anhydrides	Function	105/1382	0.0001	0.0362
GO:1901265	Nucleoside phosphate binding	Function	277/1549	0.0002	0.0368
GO:0000166	Nucleotide binding	Function	277/1549	0.0001	0.0369
GO:0051247	Positive regulation of protein metabolic process	Process	43/214	3.02×10^{-6}	0.041
GO:1901363	Heterocyclic compound binding	Function	369/896	0.0002	0.0456

Table 3.2: Summary of GO terms enriched for differentially methylated genes in t(1;11) carriers.

Shown in order of column appearance are the GO identifiers, the GO terms, the enrichment p-values, and enrichment q-values for genes showing differential expression in t(1;11) carrier LCLs.

3.4 Analysis of differential gene expression in t(1;11) LCLs by qRT-PCR

3.4.1 Selection of genes for qRT-PCR validation

Prior to the commencement of this PhD, a study was carried out by Dr. Kirsty Millar and Prof. David Porteous in collaboration with Dr. Miguel Camargo (Merck, Sharp and Dohme), to investigate differential gene expression in an independent growth of cells from the same individuals as above. Gene expression was profiled using a Rosetta/Merck Human RSTA Custom Affymetrix 2.0 microarray (Affymetrix). Results from this experiment were validated by qRT-PCR by an MSc student, Xu Tang (XT).

Expression levels of 14 genes were examined by XT by qRT-PCR. These genes were selected on the basis of evidence for dysregulation in the L100P *Disc1* mouse model (Clapcote *et al.*, 2007; Brown *et al.*, unpublished data), and for localisation to CNVs implicated in schizophrenia (Walsh *et al.*, 2008; Xu *et al.*, 2008). Differential expression was confirmed in eight of the 14 genes by XT (Table 3.3). Amongst the eight genes validated by XT was *HIPK2*, which was also found to be differentially expressed in t(1;11) carriers in the microarray experiment described in section 3.3 ($p = 0.0068$; FC = 1.25; Table 3.1). There was no evidence for significant differential expression of the remaining seven genes in the microarray analysis described in section 3.3. However, a member of the Sortilin gene family, *SORL1*, was found to be downregulated in t(1;11) carriers. Due to prior evidence of dysregulation of Sortilin gene family members in the L100P *Disc1* mouse (Brown *et al.*, unpublished data), *SORL1* was selected as an additional gene for qRT-PCR validation, along with the eight genes previously investigated by XT.

	Microarray Results		qRT-PCR Validation (Xu Tang)	
Gene (*Symbol*)	***P*-value**	**Fold-Change**	***P*-value**	**Fold-Change**
Cortactin (*CTTN*)	0.0004	1.56	0.06	2.82
Abnormal spindle homologue, microencephaly associated (*ASPM*)	0.001	-1.03	0.46	-2
Phosphoinositide-3-kinase, catalytic, Delta polypeptide (*PIK3CD*)	0.002	1.41	0.1	1.48
Polo-like kinase 1 (*PLK1*)	0.003	-1.13	0.37	-1.84
Cell division cycle 25 homologue B (*CDC25B*)	0.003	-1.03	0.46	-1.65
Protein tyrosine kinase 2, Beta (*PTK2B*)	0.02	1.18	0.27	1.28
Fragile histidine triad gene (*FHIT*)	0.01	-2.11	0.05	1.49
Synaptic vesicle glycoprotein 2B (*SV2B*)	0.01	7.7	0.02	8.44
Nuclear factor of kappa light polypeptide gene enhancer in B-cells 2 (*NFKB2*)	0.0005	1.87	0.02	1.72
Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, Alpha (*NFKBIA*)	0.02	1.56	0.03	1.45
Discs large homologue-associated protein (*DLGAP1*)	0.003	-3.38	0.02	-5.58
Homeodomain-interacting protein kinase 2 (*HIPK2*)	0.005	1.82	0.0009	1.55
MyoD family inhibitor domain containing (*MDFIC*)	0.009	1.79	0.04	1.51
Neurexin-3, Alpha (*NRXN3*)	0.02	5.07	0.03	2.67

Table 3.3: Genes previously assessed for differential expression in t(1;11) lymphoblastoid samples.

Genes are presented with corresponding p – values and fold changes for differential expression for the qRT-PCR validation of the original microarray results, and the qRT-PCR validation performed by Xu Tang. Results are colour-coded for whether a gene was significantly differentially expressed ($p \leq 0.05$) in the microarray analysis only (yellow) or both microarray and validation analyses (blue).

3.4.2 Reference gene selection for qRT-PCR

A panel of eight reference assays was used to select the optimum number of reference genes to which expression of the genes of interest would be normalised. Raw expression data from the thirteen LCL-derived cDNA samples was input into the geNorm program for analysis (Vandesompele *et al.*, 2002). In order of increasing stability, based on the geNorm-M value, were *TBP*, *ATP5B*, *CYC1*, *NCBP2*, *HNRNPD*, *RPLP0*, *UBC*, and *GAPDH* (Figure 3.11). The optimum number of reference genes was determined to be two, based on the geNorm-V value of ≤ 0.15 for V 2/3. This indicated that the use of three or more reference genes for data normalisation would not be significantly better than using two genes (Figure 3.11). Based on these results, the two most stably-expressed reference genes from this panel (*GAPDH* and *UBC*) were selected as reference genes.

3.4.3 Analysis of differential gene expression by qRT-PCR

Expression levels were examined in nine genes in t(1;11) LCL cDNA samples using Taqman® gene expression assays. Raw data were normalised to *UBC* and *GAPDH* as determined by geNorm analysis. Translocation carriers were compared to non-carriers using linear regression, covarying for gender. Significant differential expression was observed for *HIPK2* ($p = 0.002$, FC = 1.61), and *SORL1* ($p = 0.045$, FC = -1.50). The directions of effect observed at these genes were consistent with those observed in the microarray analysis (Figure 3.12; Table 3.4).

In addition, significant differential expression was observed for *DLGAP1* ($p = 0.022$, FC = 4.30; Figure 3.13), and *SV2B* ($p = 6.02 \times 10^{-5}$, FC = -4.36; Figure 3.13). *SV2B* was not found to be differentially expressed in the microarray analysis described in section 3.3 ($p = 0.48$), while *DLGAP1* expression was not detected above background on the array.

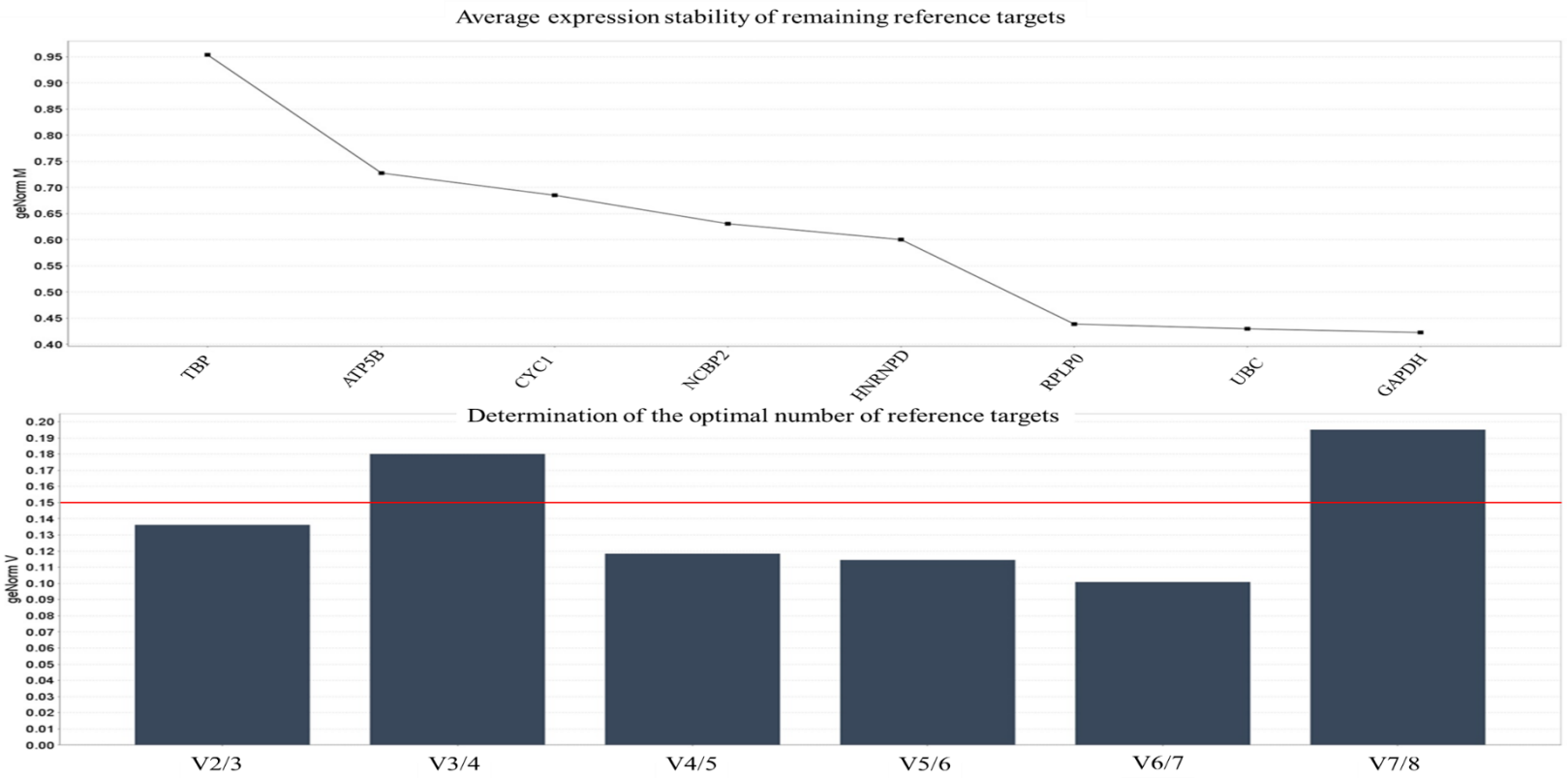


Figure 3.11: Identification of the most stable reference gene across samples using geNorm.

Shown are geNorm M (line) and geNorm V (bar) plots. The geNorm M plot presents genes in order of increasing expression stability from left to right while the geNorm V plot presents the average pairwise variation between gene sets. The red horizontal line represents the geNorm V threshold of 0.15 below which gene sets are stably expressed.

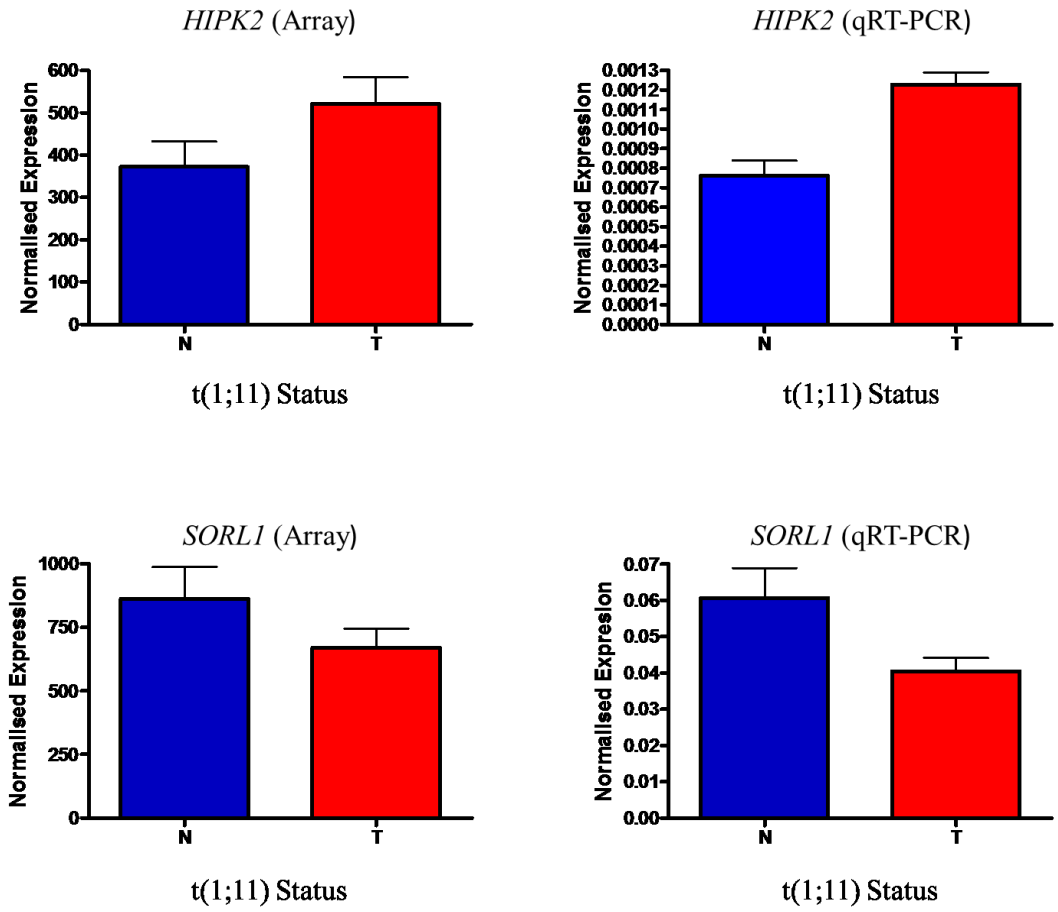


Figure 3.12: *HIPK2* and *SORL1* expression in t(1;11) family LCL samples as measured by microarray and qRT-PCR.

Shown are normalised expression values (y-axes) plotted against t(1;11) carrier status (x-axes; "N" = non-carrier, "T" = t(1;11) carrier) for HIPK2 *and* SORL1.

	Microarray Results		qRT-PCR Validation	
Gene (*Symbol*)	*P*-value	Fold-Change	*P*-value	Fold-Change
Fragile histidine triad gene (*FHIT*)	0.254	1.23	0.196	-1.43
Nuclear factor of kappa light polypeptide gene enhancer in B-cells 2 (*NFKB2*)	0.398	-1.14	0.085	-1.46
Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, Alpha (*NFKBIA*)	0.818	-1.04	0.214	-1.57
MyoD family inhibitor domain containing (*MDFIC*)	0.894	-1.02	0.506	1.21
Neurexin-3, Alpha (*NRXN3*)	0.456	1.05	*1*	1.35
Synaptic vesicle glycoprotein 2B (*SV2B*)	0.482	1.12	0.022	4.30
Discs large homologue-associated protein (*DLGAP1*)	NA	NA	6.02×10^{-5}	-4.36
Homeodomain-interacting protein kinase 2 (*HIPK2*)	0.007	1.25	0.002	1.61
Sortilin-related receptor, L(DLR class) A (*SORL1*)	0.007	-1.29	0.045	-1.50

Table 3.4: Genes assessed for differential expression by both microarray and qRT-PCR analyses in the current study.

Genes are presented with corresponding p–values and fold changes for differential expression for the microarray results, and the qRT-PCR validation. Yellow cells contain data for genes that were originally found to be differentially expressed by microarray and qRT-PCR analyses in an independent batch of LCLs, but not the current study. Blue cells contain data for genes that were originally found to be differentially expressed in the independent batch of LCLs by microarray analysis and qRT-PCR, with differential expression observed by at least one method in the current study. Orange cells contain data for SORL1 – a gene selected for validation by qRT-PCR in the current study based on its functional relevance in psychiatric illness. "NA" indicates genes that were not detected by either the microarray or qRT-PCR assay used in these samples. Italicised p*-values correspond to those obtained using a Mann-Whitney U test whilst non-italicised* p*-values correspond to those obtained using a two-tailed Welch's* t*-test, depending on the distribution of the data for that gene.*

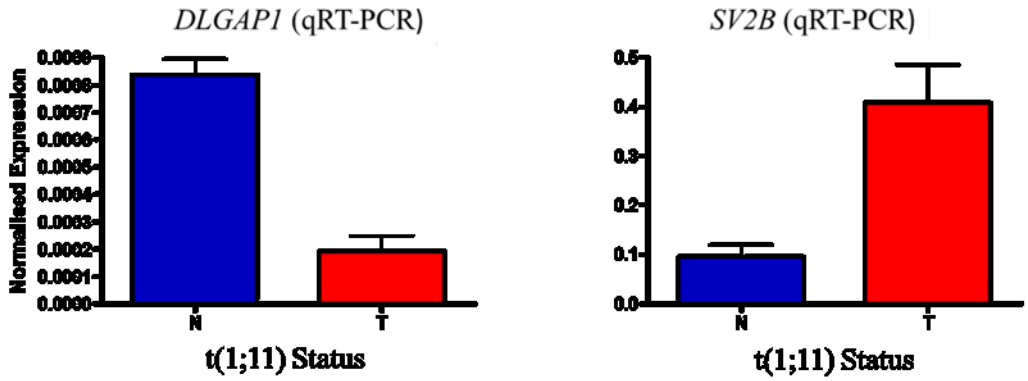


Figure 3.13: *DLGAP1* and *SV2B* expression in t(1;11) family LCL samples as measured by qRT-PCR.

Shown are normalised expression values (y-axes) plotted against t(1;11) carrier status (x-axes; "N" = non-carrier, "T" = t(1;11) carrier) for DLGAP1 *and* SV2B.

3.5 Discussion

Gene expression was assessed in LCL-derived RNA, comparing levels in t(1;11) carriers to those in non-carriers. The work described in this chapter was based on the hypothesis that the t(1;11) translocation might affect gene expression not only at *DISC1*, as previously observed (Millar *et al.*, 2005), but also elsewhere in the genome. This may be directly, through disruption of the genes at the chromosome 1 and 11 breakpoints; or indirectly, through transmission of variants on the derived chromosomes that affect gene expression.

It is important to consider the limitations of this analysis prior to interpreting the findings. A major drawback to the current study is the use of LCL-derived RNA. Cell-line derived samples may be prone to technical confounders, as well as biological aberrations randomly accumulated during the culturing process.

Thirteen samples were profiled for gene expression. Prior to analysis, sample quality was assessed at the RNA level, and the cDNA level following reverse-transcription of RNA. RNA was found to be intact while cDNA samples were found to be free of genomic DNA contamination. Upon analysis of differential gene expression between t(1;11) carriers and non-carriers, no differences were observed at the genome-wide significant level. Deflation of *p*-values was observed which might suggest that, at 13 samples, the study was underpowered to detect genome-wide significant differences in gene expression. To determine this, a *post-hoc* power calculation was performed using the R package, *pwr* (Champely, 2017). This revealed that at the current sample size, the study had 80 % power to detect an effect size of 3.26 at an alpha of 0.000001 (Bonferroni $p = 0.05$). To attain 80% power at the fold-change and significance thresholds used in this study, a further seven t(1;11) samples would be required. The results should therefore be interpreted with caution. As it is unlikely that LCLs will be generated from additional t(1;11) family members, any functional follow-up of these findings should be preceded by targeted validation, such as qRT-PCR, in an independent batch of samples.

Three hundred and three genes showed nominally significant differential expression in t(1;11) carriers ($p \leq 0.05$) at an arbitrary absolute fold-change cut-off of 1.2. Using this less conservative threshold to define differential expression, the most significantly differentially expressed gene was *IL17RB* ($p = 3.52 \times 10^{-6}$, FC = -2.58). It is of interest that a study by Wen *et al*. (2014) observed upregulation of this gene in iPSC-derived neurons from patients with a 4 bp frameshift mutation in *DISC1* ($p = 0.0001$, FC = 4.14). Downregulation of this gene has, however, been reported by Kim *et al.* (2014) in cases of suicide among schizophrenia patients.

Among the 303 differentially expressed transcripts, 23 were also reported as dysregulated in the study by Wen *et al.* (2014). This study compared samples in triplicate, from two individuals with a *DISC1* frameshift and one without. Considering the small sample size, the number of significantly differentially expressed genes in their study (n = 2012) might suggest the authors treated each triplicate as an individual data point. Such an approach might result in inflation of significant results for differential expression due to the high degree of correlation expected between data points corresponding to a given individual. This caveat aside, a particularly noteworthy example of an overlapping gene between this study and the t(1;11) study is *FABP5.* This gene was upregulated both in t(1;11) carriers ($p = 0.0036$, FC = 1.35) and in *DISC1* mutant neurons reported by Wen *et al.* (2014; $p = 1.69 \times 10^{-5}$, FC = 3.52). *FABP5* is a member of a gene family encoding fatty acid binding transport proteins (FABPs). Shimamoto *et al*. (2014) reported upregulation of *FABP5* in schizophrenia *post-mortem* brain. Furthermore, there is evidence for a role of *FABP5* in regulating hippocampal cognitive function (Yu *et al.*, 2014), while depletion of FABPs has been observed in the synaptosomes of aged mouse brains (Pu *et al.*, 1999). Other members of this family were also upregulated in t(1;11) carriers: two probes corresponding to *FABP5L2*, and one probe corresponding to *FABP5L9*. The findings of differential *FABP* expression here support existing evidence for a role for these genes in psychiatric illness, and may indicate a potential mechanism by which illness occurs in t(1;11) individuals through abnormal fatty acid transport. Such an effect may have

detrimental results in the brain by disrupting normal fatty acid levels and signalling processes (Bazinet and Layé, 2014).

The greatest positive fold-change in expression amongst the 303 differentially expressed genes was observed in *GIMAP6*. This gene is a member of the *GIMAP* family, members of which function in the regulation of cell survival (Kruken *et al.*, 2004). A study by Gregg *et al*. (2008) reported upregulation of *GIMAP6* in children with early-onset autism compared to the general population. Given the reported overlap in the genetic architecture of autism and other psychiatric disorders (Smoller *et al.*, 2013), these findings may support *GIMAP6* as a risk gene for pathogenesis in the t(1;11) family. The greatest negative fold-change in expression reported in t(1;11) carriers was observed in *EPDR1*. In zebrafish, the proposed homologue of this gene regulates aggressive behaviour, and functions as a neuronal cell adhesion molecule (CAM). Neuronal CAMs have previously been implicated in schizophrenia and bipolar disorder (Hargreaves *et al.*, 2014). Should downregulation of *EPDR1* occur in the brains of t(1;11) individuals, it is possible the CAM pathway might be affected, thus conferring increased risk of illness in these individuals. Further work is required, however, to determine the function of *EPDR1* in humans. Assessment of whether these fold-changes are observed at the protein levels for these genes would further clarify whether cell survival and/or CAM pathways are disrupted in t(1;11) individuals. To this end, the recently-generated iPSC-derived neurons may prove a useful resource.

Millar *et al.* (2005), previously reported lower levels of *DISC1* expression in t(1;11) LCLs. Four probes are present on the HT-12 array to measure *DISC1* expression. In the above study, these probes had detection *p*–values > 0.05 in all samples and were therefore filtered prior to analysis of differential expression. Millar *et al.* (2005) quantified *DISC1* expression LCL samples by qRT-PCR. It is likely that the qRT-PCR assays used were more sensitive to *DISC1* expression levels than the probes on the HT-12 array.

Although *DISC1* was not detected above background in any sample, a correlation was observed between translocation carrier status and detection above background for 170 genes. Of these, *ATP10A* was exclusively detected in non-carriers of the translocation (Fisher's exact test $p = 0.0008$). *ATP10A* is maternally imprinted in the brain, and is located at chromosome 15q11-q13, an autism susceptibility region (Bolton *et al.*, 2004). Its putative function involves maintenance of cell membrane integrity which may suggest a regulatory role for the gene in neurotransmission, when expressed in the brain (Herzing *et al.*, 2001).

Upregulation of *MAPK1* was observed in translocation carriers ($p = 0.0025$, FC = 1.21). *MAPK1* encodes ERK2, a key component of MAPK signalling pathway. DISC1 has previously been implicated in ERK signaling where it is thought to play a role in the pathogenesis of MDD, and regulation of gene expression in astrocytes (Hashimoto *et al.*, 2006; Wang *et al.*, 2016). In the MAPK signalling cascade, ERK activation results in CREB phosphorylation and binding to cAMP response elements to regulate transcription. It is therefore possible that the genome wide differences in expression may occur, at least in part, through the downstream effects of altered ERK2 levels in these individuals.

Nine genes were selected for validation by qRT-PCR based on the findings from this microarray experiment, and findings from a previous study of expression in t(1;11) LCLs. Of these nine genes, four were differentially expressed in t(1;11) carriers: *HIPK2*, *SORL1*, *SV2B* and *DLGAP1*. Of these genes, *HIPK2* and *SORL1* were the only ones to show differential expression both in the microarray and qRT-PCR analyses presented above.

Upregulation of *HIPK2* had previously been reported in t(1;11) carriers by Xu Tang and Miguel Camargo. Overexpression of *HIPK2* has been associated with apoptotic cell death (Bracaglia et al., 2009). Apoptosis has been reported in cases of bipolar disorder and schizophrenia (Benes, 2006). If overexpressed in the brains of t(1;11)

individuals, it is possible that *HIPK2* may mediate apoptosis, contributing to an increased risk of illness.

Downregulation of *SORL1* was observed in t(1;11) carriers on the expression array, a finding which was validated by qRT-PCR analysis. *SORL1* is a member of the Sortilin gene family which has been shown to play a role in processing of amyloid precursor protein (APP). APP processing can occur via the amyloidogenic pathway by the action of BACE1 – or β-secretase - and subsequent γ–secretase-mediated processing, or the non-amyloidogenic pathway via α–secretase and subsequent γ–secretase-mediated processing. *ADAM17*, an α–secretase-encoding gene, was also downregulated in t(1;11) carriers. The amyloidogenic pathway results in the formation of neurotoxic Amyloid-β plaques, a key feature of Alzheimer's disease pathology. Conversely, the non-amyloidogenic pathway has been associated with neurotrophic effects through the secretion of the APP ectodomain (sAPP-α; Hartl *et al.*, 2013). Others have demonstrated a negative correlation between expression of *SORL1,* and both amyloidogenic and non-amyloidogenic processing of APP (Gustafsen *et al.*, 2013). DISC1 has been shown to interact with APP in neurodevelopment (Young-Pearse *et al.*, 2011), and there is evidence for a role of DISC1 in APP processing (Shahani *et al.*, 2015). Taken together, these findings suggest a possible link between the t(1;11) translocation and the APP processing pathway through transcriptional dysregulation of multiple constituent genes in the pathway. The Illumina HT-12 expression array contains two probes to measure expression of the *BACE1* gene. Neither of these probes were expressed above background in any of the LCL samples (ILMN_2320349 detection *p*–value = 0.17, ILMN_1797804 detection *p*–value = 0.16). This would suggest that BACE1 activity, and therefore, Amyloid-β is not present in these samples. A cellular model from the t(1;11) family expressing all members of the APP processing pathway would be required to determine whether the translocation is linked to APP processing through altered expression and/or interaction between the above genes. Neurons derived from t(1;11) family iPSCs have recently become available and these may be a useful resource to investigate this further. To determine whether *DISC1* and the APP processing pathway genes might interact in cognitive or depressive phenotypes, an *in-silico* approach was taken with an aim to identify epistatic

interactions between these genes in an unrelated sample of Scottish ancestry. This work is described in Chapter 7.

Of the remaining seven genes tested for differential expression by qRT-PCR, dysregulation of *DLGAP1* and *SV2B* was observed in t(1;11) carriers. *SV2B* was downregulated in t(1;11) carriers, a finding which was not observed in the microarray analysis (p=0.48). The HT-12 array contains three probes to measure *SV2B* expression. These probes map to last exon of all *SV2B* transcript variants reported on UCSC genome browser (http://genome.ucsc.edu), with the exception of a UCSC transcript (uc002bqt.3). The Taqman® assay used to measure *SV2B* expression by qRT-PCR (Hs00916046_m1) maps to an exon boundary common to all reported isoforms, including uc002bqt.3. It is therefore possible that differential expression of *SV2B* in t(1;11) LCLs is specific to this isoform, which would not be detected by the HT-12 array. *SV2B* encodes a synaptic vesicle protein and functions in regulating presynaptic calcium levels (Wan *et al.*, 2010). If misexpressed in the brains of t(1;11) carriers, *SV2B* might contribute to defective neurotransmission.

DLGAP1 expression was not detected in t(1;11) family LCLs by the HT-12 array. It was, however, found to be upregulated in t(1;11) carriers by qRT-PCR. The HT-12 array contains two probes to measure expression of all reported *DLGAP1* isoforms. Similarly, the Taqman® assay used to measure *DLGAP1* expression (Hs00191052_m1) was also designed to measure expression of all known isoforms. It is possible that *DLGAP1* expression levels in LCLs are too low for detection by the HT-12 array, and is only measurable in these samples by more sensitive methods, such as qRT-PCR. *DLGAP1* is located on chromosome 18p, a region in which Pickard *et al.* (2005) reported an inversion associated with schizophrenia and bipolar disorder in a Danish family. Although not directly disrupted by the inversion, they proposed *DLGAP1* expression might be altered by the inversion through the disruption of nearby regulatory elements. DLGAP1 interacts with PSD-95 at the post-synaptic density (Kim *et al.*, 1997). PSD-95 functions in synaptic regulation of neurotransmitter receptors and adhesion of pre- and post-synaptic terminals (Ziff, 1997). DLGAP1, among other

interactors, may function to anchor ion-channel/PSD-95 complexes to the postsynaptic density (Kim *et al.*, 1997). This might suggest dysregulation of *DLGAP1* in the brain, like *SV2B*, might affect neurotransmission, rendering both genes attractive functional candidates for pathogenesis in the family.

Gene ontology analysis was performed on the data to identify over-represented terms amongst the most differentially expressed genes which might indicate processes disrupted by the t(1;11) translocation relevant to pathogenesis. There was no enrichment of differentially expressed genes for brain-specific terms. This is perhaps not unexpected, given the use of LCLs. The most significant over-represented term was nuclear part ($q = 0.005$).

A major limitation to this study is the use of LCL-derived RNA. In the context of a translocation linked to major mental illness, the optimum tissue would be t(1;11) family-derived neuronal material. The small sample size also posed a challenge, probably rendering the study underpowered to detect any genome-wide significant differences in expression. In addition, cell passage numbers were unavailable for these samples. Grafodatskaya *et al.* (2010) reported randomly distributed differential methylation patterns in high passage LCLs, which could impact upon gene expression. Moreover, spontaneous aneuploidy is a known feature of long term cell culture, including LCLs (Miyai *et al.*, 2008; Shirley *et al.*, 2012). It is therefore recommended that subsequent work involving these cell lines should involve karyotyping as a quality control measure. Recently, iPSC-derived neuronal samples have become available from a subset of t(1;11) family members. Analysis of gene expression in these samples is ongoing. Identification of differentially expressed genes common to both iPSC-derivatives and LCLs may inform upon tissue-agnostic effects of the t(1;11) translocation on gene expression. However, findings from the iPSC-derivatives are likely to provide a more physiologically relevant representation of differential gene expression in the brains of t(1;11) carriers.

Chapter 4

Analysis of DNA methylation in t(1;11) family whole blood-derived samples

4 Analysis of DNA methylation in t(1;11) family whole blood-derived samples

4.1 Overview

DNA methylation is a fundamental epigenetic modification which plays a key regulatory role in the development and establishment of cellular identity. Its capacity to be modulated by non-genetic factors has rendered it an attractive candidate to assess the impact of environmental effects on complex traits. In addition to environmental factors, genetic factors can also regulate levels of DNA methylation (Lemire *et al.*, 2015). DNA methylation is thought to play a regulatory role in gene expression through remodelling of chromatin structure (Lewis and Bird, 1991; Hashimshony *et al.*, 2003). Aberrant DNA methylation has been observed in numerous disorders, including psychiatric illness (Gopalakrishnan *et al.*, 2008; Grayson and Guidotti, 2013). The work presented in this chapter was performed with the aim of determining whether the t(1;11) translocation was associated with differential DNA methylation in the family, and whether these differences might be correlated with diagnosis in individuals carrying the translocation.

Methylation of whole blood-derived DNA was profiled in 17 t(1;11) carriers and 24 non-carrying relatives using the Infinium HumanMethylation450 BeadChip (Illumina®, San Diego, California; 450k array). The 450k array interrogates DNA methylation at 485,577 sites across the genome (Bibikova *et al.*, 2011). Two comparisons of DNA methylation were performed: an analysis between t(1;11) carriers and non-carriers; and an between t(1;11) carriers with a psychotic disorder and those with a non-psychotic disorder. The aim of the first comparison was to investigate whether any methylation disruption occurred as a in the context of the t(1;11) translocation, either locally within the t(1;11) breakpoint regions (e.g. as a direct result of an effect of the translocation on chromatin structure), or by means of a *trans*- effect at regions across the genome (e.g. due to methylation quantitative trait loci). The aim of the second comparison was to identify whether differences in methylation levels within affected t(1;11) carriers may be correlated with a psychiatric phenotype (i.e. psychosis) in these individuals.

4.2 Comparison of DNA methylation between t(1;11) carriers and non-carriers

4.2.1 Data preprocessing

Raw methylation data for the entire 450k array (n = 485,577 probes) were read into R. The initial filtering step involved the removal of probes with predicted cross-hybridising potential (Chen *et al.*, 2013; n = 30,969). A second filtering step was performed to remove probes with a variant at the target CpG, and, in the case of type I probes, the site of single base extension (the base before C), based on whole genome sequence data from the family (n = 10,548).

Based on the filtering criteria described in Chapter 2.10.2, the *pfilter()* function removed 799 probes and 0 samples, based on the proportion of probes with a detection p – value of > 0.05, and the proportion of these probes in a given sample, respectively. This resulted in a final sample set of 41 individuals for whom 443,196 sites were profiled for methylation.

4.2.2 Estimation of cellular proportions in whole blood

This study used whole blood-derived DNA: a heterogeneous tissue in terms of cellular composition. Estimated cell counts were compared between groups to determine whether cellular proportions differed as this may confound subsequent analyses. Estimated cell counts were calculated using the *estimateCellCounts()* function in *minfi*. A *t*-test was performed to compare these estimated cell proportions between t(1;11) carriers and non-carriers. No significant between-group differences were observed in cell count estimates for B-lymphocytes, granulocytes, monocytes, natural killer cells, $CD4^+$ T-lymphocytes and $CD8^+$ T-lymphocytes (Student's two-tailed independent samples *t*-test $p \geq 0.41$; Table 4.1). Based on this information, cellular estimates were not included as covariates in this differential methylation analysis.

	$CD8^+$ T-Cells	**$CD4^+$ T-Cells**	**Natural Killer Cells**	**B-Cells**	**Monocytes**	**Granulocytes**
Mean proportion in t(1;11) carriers	0.053	0.153	0.057	0.064	0.071	0.619
Mean proportion in t(1;11) non-carriers	0.048	0.161	0.065	0.051	0.069	0.615
p-value	0.704	0.578	0.502	0.408	0.846	0.866

Table 4.1: Estimated cellular proportions of blood in t(1;11) carriers and non-carriers.

Table summarises the minfi package's estimateCellCounts() function's estimated proportions of CD8+ and CD4+ T-cells, natural killer cells, B-cells, monocytes and granulocytes. Mean proportions for each group are presented for each cell type along with an unpaired t*-test* p*-value, assessing differences in cell subtypes between groups.*

4.2.3 Selection of normalisation method

The raw data were normalised by 14 methods: SWAN, Noob, beta mixture quantile normalisation (BMIQ), *dasen*, *nasen*, *nanet*, *naten*, *nanes*, *danes*, *danet*, *danen*, *daten1*, *daten2* and peak-based correction (PBC) (Table 4.2). Using the three metrics described in Chapter 2.10.3, each method's ability to reduce technical error was assessed and ranked with *dasen* achieving the highest overall rank. The data were then normalised by *dasen*. This method equalises type I and type II backgrounds, followed by quantile normalisation of methylated and unmethylated intensities separately each for type I and type II probes, then calculates normalised methylation β-values.

Normalisation method	DMRSE rank	GCOSE rank	Seabird rank	Rank mean	Rank of rank mean
dasen	1	11.5	2	4.833	1
nasen	2	11.5	3	5.5	2.5
danes	3	3.5	10	5.5	2.5
daten1	4	8	5	5.666	4
danen	11	7	1	6.333	5.5
daten2	6	9	4	6.333	5.5
nanes	7	3.5	11	7.1666	7
raw	12	5.5	6	7.8333	8
SWAN	10	2	12	8	9
danet	5	14	7	8.666	10
naten	9	10	8	9	11
BMIQ	14	1	14	9.666	12
nanet	8	13	9	10	13
Noob	13	5.5	13	10.5	14
PBC	15	15	15	15	15

Table 4.2: Assessment of the performance of 14 methods used to normalise the raw methylation data.

Table summarises the rankings of three metrics used to assess the performance of 14 normalisation strategies. From left to right, columns show the normalisation strategy, DMR-standard error (DMRSE) rank, genotype-combined standard error (GCOSE) rank, 1 - Seabird AUC rank, mean rank of the three metrics, and the rank of mean ranks.

4.2.4 Identification of differentially methylated positions

Linear regression was performed to identify differentially methylated positions (DMPs) between t(1;11) carriers and non-carriers. Surrogate variable analysis (SVA) was performed on the data to identify latent sources of variation. Seven surrogate variables (SVs) were identified and fitted as covariates, along with age and gender. The genomic inflation factor λ was calculated from the *p*-values for differential methylation to assess the goodness-of-fit of the model. This was within the acceptable range of 1-1.1 as defined by GenABLE (λ = 1.05; Wang and Leal, 2012; Figure 4.1)

Thirteen significant DMPs were identified in the comparison of t(1;11) carriers and non-carriers (FDR $q \leq 0.05$; Figure 4.2; Table 4.3). Four of these sites were in the *DISC1* gene itself: three mapping to the gene body and one to the 3' UTR. With the exception of one site on chromosome 10 (cg24508974), all were on either chromosome 1 or 11. Hypomethylation was observed in all but two of the DMPs in t(1;11) carriers. The two loci displaying hypermethylation were at an intergenic site on chromosome 10, and one site within the gene body of *EGLN1*, on chromosome 1. With regards to the translocation breakpoints, the most distal sites showing significant differential methylation on chromosomes 1 and 11 were approximately 10 Mb (cg26355502) and 31 Mb (cg02771260) centromeric, respectively.

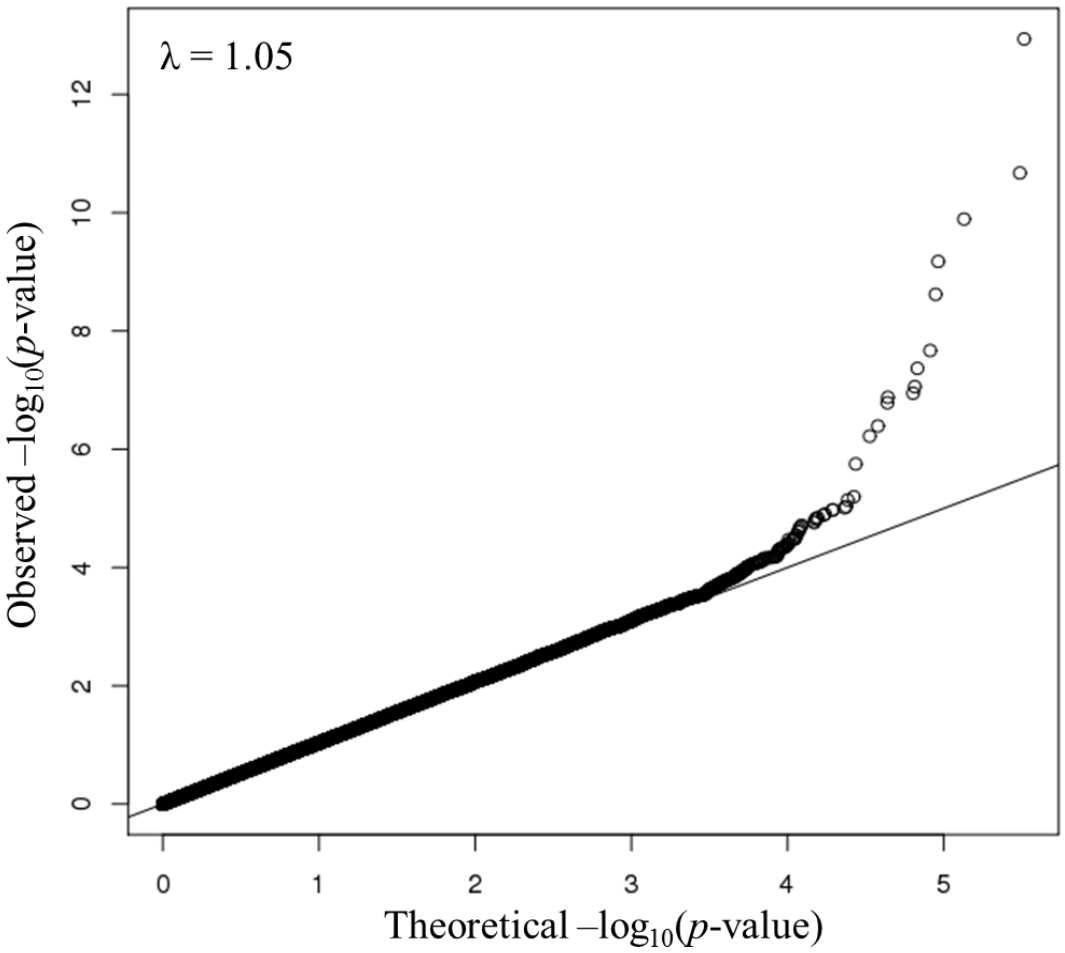


Figure 4.1: Quantile-quantile plot showing observed vs. expected *p*-values for differential methylation between t(1;11) carriers and non-carriers.

Shown are observed unadjusted –log10 p-values (y-axis; circular points) plotted against the expected distribution of –log10 p-values under the null hypothesis (x-axis; solid diagonal line). The genomic inflation factor λ is presented in the upper-left corner of the plot.

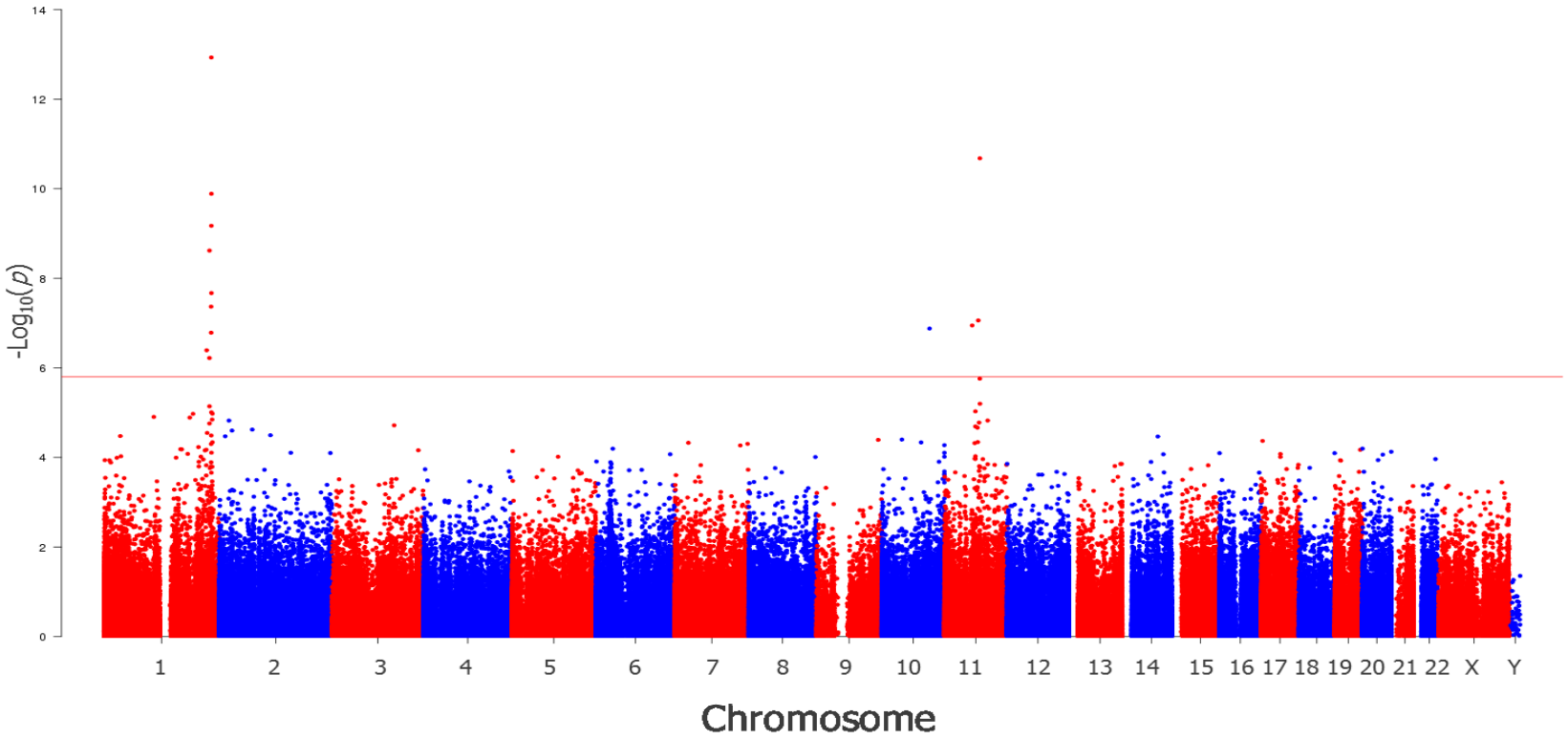


Figure 4.2: Manhattan plot for DNA methylation analysis of t(1;11) carriers and non-carriers.

Figure shows –log$_{10}$ p – values for differential methylation between t(1;11) carriers and non-carriers (y–axis) plotted against chromosomal position (x–axis). The horizontal red line represents the –log$_{10}$ p – value threshold for genome wide significance (FDR q *= 0.05)*

Probe ID	Hg19 Coordinates	Gene	Beta Difference	Fold-Change	t	*p-value*	*q* - value
cg09186051	Chr1:231981906	*DISC1*;T*SNAX-DISC1*	-0.07	-1.31	-11.81	1.17×10^{-13}	5.20×10^{-8}
cg26728851	Chr11:76430375	*GUCY2E*	-0.03	-1.63	-9.72	2.09×10^{-11}	4.64×10^{-6}
cg15157974	Chr1:232144702	*DISC1*;*TSNAX-DISC1*	-0.04	-1.27	-9.04	1.30×10^{-10}	1.92×10^{-5}
cg05656812	Chr1:232021560	*DISC1*;*TSNAX-DISC1*	-0.06	-1.33	-8.44	6.73×10^{-10}	7.46×10^{-5}
cg06928246	Chr1:227974645	NA	-0.07	-1.56	-7.99	2.41×10^{-09}	0.0002
cg16177633	Chr1:232172585	*DISC1*;*TSNAX-DISC1*	-0.03	-1.20	-7.23	2.13×10^{-08}	0.0016
cg18815120	Chr1:231512676	*EGLN1*	-0.12	-2.10	-6.99	4.29×10^{-08}	0.003
cg25899154	Chr11:72897143	NA	-0.07	-1.28	-6.75	8.73×10^{-08}	0.005
cg02771260	Chr11:59836817	*MS4A3*	-0.13	-1.79	-6.66	1.13×10^{-07}	0.006
cg24508974	Chr10:103330391	NA	0.01	1.17	6.61	1.32×10^{-07}	0.006
cg21875980	Chr1:231553510	*EGLN1*	0.06	1.40	6.54	1.64×10^{-07}	0.007
cg26355502	Chr1:221916303	*DUSP10*	-0.01	-1.30	-6.24	4.05×10^{-07}	0.01
cg00965168	Chr1:227974541	NA	-0.05	-1.42	-6.10	6.03×10^{-07}	0.02

Table 4.3: Significantly differentially methylated positions between t(1;11) carriers and non-carriers.

Table summarises significantly differentially methylated sites between t(1;11) carriers and non-carriers (FDR $q \leq 0.05$*). In order of column appearance are probe identifiers, Hg19 genomic coordinates, UCSC reference gene names ("NA" denotes intergenic regions), between-group difference in mean beta value, fold-change between groups, moderated t-statistic, p-value for differential methylation and FDR-adjusted p-value.*

4.2.5 Gene ontology analysis of differentially methylated genes in t(1;11) carriers

To identify whether any biological processes, molecular functions or cellular components were significantly over-represented amongst the most significant DMPs, a *p*–value ranked list of DMP-containing genes was submitted to GOrilla for ontology analysis (http://cbl-gorilla.cs.technion.ac.il/; Eden *et al.*, 2009; n = 20,752 genes submitted). In cases of multiple probes per gene, the probe with the lowest DMP *p*-value was selected.

Overall, 62 GO categories were significantly over-represented in these data ($q \leq 0.05$; Table 4.4). Categories are split into processes, components and functions by GOrilla. All significant terms were within the "process" and "component" categories. The most significantly over-represented term in this analysis was "neuron projection" (GO:0043005; $q = 3.72 \times 10^{-6}$). Additional categories relating to neuronal function were also significantly over-represented. These included "regulation of nervous system development", "regulation of synapse organisation", and "axon".

GO Term	Description	Class	Enrichment	*p*-value	*q*-value
GO:0043005	neuron projection	Component	1.73	2.29×10^{-9}	3.72×10^{-6}
GO:0042995	cell projection	Component	1.44	3.63×10^{-7}	0.0003
GO:0030425	dendrite	Component	1.43	2.10×10^{-6}	0.0009
GO:0009653	anatomical structure morphogenesis	Process	1.5	6.49×10^{-8}	0.0009
GO:0097458	neuron part	Component	1.96	1.93×10^{-6}	0.001
GO:0009987	cellular process	Process	1.33	5.98×10^{-7}	0.002
GO:0048562	embryonic organ morphogenesis	Process	4.36	4.93×10^{-7}	0.0022
GO:0048856	anatomical structure development	Process	1.07	4.14×10^{-7}	0.0028
GO:0048731	system development	Process	1.72	1.04×10^{-6}	0.0028
GO:0030424	axon	Component	1.12	3.05×10^{-5}	0.0083
GO:0044444	cytoplasmic part	Component	4.26	2.75×10^{-5}	0.0089
GO:0032502	developmental process	Process	1.3	6.12×10^{-6}	0.0093

Table 4.4: Summary of GO terms enriched amongst differentially methylated genes in t(1;11) carriers.

For each GO term, table summarises the GO identifier, the GO description, the GO class, enrichment, the enrichment p-values, and enrichment FDR q-values for genes showing the most differential methylation in t(1,11) carriers compared to non-carriers. Enrichment is defined as (b/n) / (B/N), where b = the number of genes at the top of the p-value ranked list that is associated with a given GO term, n = the number of genes at the top of the p-*value-ranked gene list, B = the total number of genes associated with a given GO term, and N = the total number of genes. Shown are the top 10 enrichments ranked by* p-*value. The total list of significant enrichments is presented in Appendix I (Table A2).*

4.2.6 Identification of differentially methylated regions

To identify genomic regions containing multiple nominally significant differentially sites between t(1;11) carriers and non-carriers ($p \leq 0.05$), differentially methylated region (DMR) analysis was performed. An advantage of the DMR analysis is the increased statistical power to detect small methylation differences in the context of multiple signals within a given region (Robinson *et al.*, 2014). The ChAMP package's DMR-calling algorithm, probe lasso, was used (Morris *et al.*, 2014; Butcher and Beck, 2015). This algorithm is agnostic to the direction of effect of differentially methylated probes, with DMRs containing both hyper- and hypomethylated probes.

Using the *ChAMP* package's *ChAMP.lasso()* function, 123 DMRs were identified (Table 4.5). The most significant DMR in t(1;11) carriers spanned a 1.6 kb region in the gene body of the chromosome 6 gene *TNXB* ($p = 2.46 \times 10^{-13}$). This DMR also comprises the most probes (n = 51). The largest region spanned 4.8 kb over nine probes and was located in an intergenic region on chromosome 1, approximately 84 Mb centromeric to the chromosome 1 breakpoint ($p = 0.0001$). The closest DMRs to the chromosome 1 breakpoint was approximately 4 Mb centromeric, in the gene body of *RHOU* ($p = 0.0007$). The closest DMR to the chromosome 11 breakpoint was located approximately 3 Mb telomeric, within 1500 bp of the TSS of *C11orf75* ($p = 1.67 \times 10^{-9}$).

Gene(s)	Feature(s)	Region	Probes	DMR *p*-value
TNXB	Body	Chr6:32063516-32065113	51	2.46×10^{-13}
NA	IGR	Chr3:196704439-196707088	5	7.89×10^{-10}
C11orf75	TSS1500	Chr11:93277097-93277255	3	1.67×10^{-9}
NA;*PRRT1*	IGR, 3'UTR	Chr6:32115866-32116728	14	2.60×10^{-9}
RNF5P1;AGPAT1	TSS1500, TSS200	Chr6:32145233-32145902	20	6.26×10^{-9}
GABRG1	TSS1500, TSS200, 5'UTR	Chr4:46125801-46126455	7	7.27×10^{-9}
KRTAP5-9	TSS1500, TSS200, 5'UTR	Chr11:71259142-71259846	5	3.17×10^{-8}
CYP2E1	Body	Chr10:135343047-135343426	3	5.93×10^{-8}
XRRA1	TSS200	Chr11:74660246-74660274	4	6.07×10^{-8}
RHOD	3'UTR	Chr11:66839183-66839543	3	8.99×10^{-8}

Table 4.5: Summary of t(1;11)-associated DMRs identified by the probe lasso algorithm.

Summary of the genes (if applicable), genomic features, hg19 genomic coordinates, number of probes and p-*value associated with each DMR. "NA" in the "Gene" column represents intergenic regions (i.e. regions not annotated to a RefSeq gene). Genomic features are coded "IGR" for intergenic regions, "TSS200" and "TSS1500" for probes occurring within 200 and 1500 of a gene's transcription start site, respectively; "5'UTR" and "3'UTR" for probes occurring within a gene's 5' and 3' untranslated region, respectively, "1stExon" for probes occurring within the first exon of a gene, and "Body" for probes occurring within the gene body. Shown are the top 10 DMRs ranked by* p-*value. The total list of significant DMRs is presented in Appendix I (Table A3).*

4.2.7 Identification of methylation quantitative trait loci

An analysis of methylation quantitative trait loci (meQTLs) was performed to determine whether the observed differences in methylation in t(1;11) carriers was associated with genotype at sites in linkage disequilibrium with the translocation. The thirteen loci displaying genome-wide significant methylation between translocation carriers and non-carriers were cross-referenced to a previously published list of *cis*- and *trans*- meQTLs identified in lymphocyte DNA by Lemire *et al.* (2015). Of the 13 DMPs, seven were present in this list, all of which were *cis*-acting (less than 1 Mb between SNP and CpG; Lemire *et al.*, 2015). Genotypes were obtained from whole-genome sequence data from the family at these previously-reported sites and their role in regulating DNA methylation was assessed. Six of the seven previously-reported meQTLs were significantly associated with translocation carrier status ($p \leq 0.05$; Table 4.6). Minor allele count at these meQTLs were significantly associated with DNA methylation at five of the seven sites assessed ($p \leq 0.05$; Figure 4.3; Table 4.6).

Reported meQTL	Probe ID	Distance between probe and SNP	Probe Gene	meQTL Gene	Translocation~meQTL+Sex	meQTL *p*-value (Age+Sex)
rs2486729*	cg18815120	23 kb	*EGLN1*	*EGLN1*	2.71×10^{-8}	2.02×10^{-16}
rs17154511*	cg02771260	11 kb	*MS4A3*	*MS4A3*	9.84×10^{-6}	8.14×10^{-11}
rs10899287*	cg26728851	84 kb	*GUCY2E*	Intergenic	7.26×10^{-10}	0.0002
rs545937*	cg21875980	10 kb	*EGLN1*	*EGLN1*	6.52×10^{-10}	3.53×10^{-5}
rs4366301*	cg16177633	366 kb	*DISC1*;*TSNAX-DISC1*	*DISC1*;*TSNAX-DISC1*	0.0003	0.0009
rs6541279*	cg15157974	391 kb	*DISC1*;*TSNAX-DISC1*	*TSNAX-DISC1*	0.045	0.737
rs9419922	cg24508974	50 kb	Intergenic	Intergenic	0.4518	0.548

Table 4.6: Summary of meQTLs reported to regulate DNA methylation at differentially methylated loci identified between t(1;11) carriers and non-carriers.

From left to right, columns show the probe identifier, the corresponding probe's associated gene, the meQTL reported by Lemire et al. (2015), the gene containing the meQTL, the p-value for the relationship between the translocation and meQTL genotype, adjusting for sex; and the p-value for the relationship between DNA methylation and meQTL genotype, adjusting for age and sex. meQTLs accompanied by an asterisk (*) denote those that are significantly associated with t(1,11) carrier status ($p \leq 0.05$)

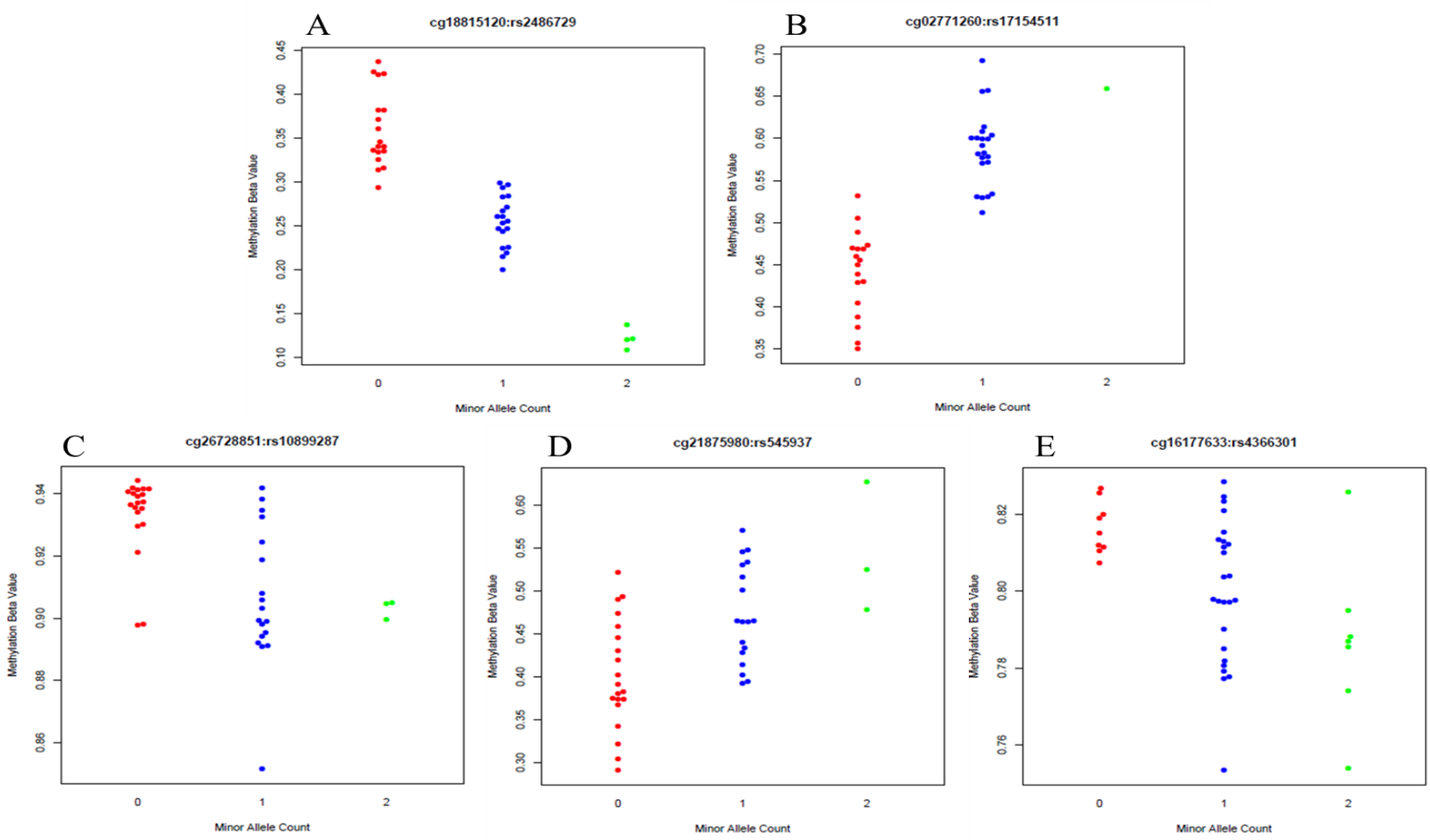

A
cg18815120:rs2486729
Methylation Beta Value
Minor Allele Count
B
cg02771260:rs17154511
Methylation Beta Value
Minor Allele Count
C
cg26728851:rs10899287
Methylation Beta Value
Minor Allele Count
D
cg21875980:rs545937
Methylation Beta Value
Minor Allele Count
E
cg16177633:rs4366301
Methylation Beta Value
Minor Allele Count

Figure 4.3: Significantly differentially methylated loci between t(1;11) carriers and non-carriers associated with genotype at previously-reported meQTLs ($p \leq 0.05$; Lemire et al., 2015).

Panels A-E display methylation β-values at five CpG sites (y-axes) plotted against minor allele counts of meQTLs ($p \leq 0.05$, x-axes). Titles of each panel identify the CpG probe and the SNP involved in each meQTL in the format "probe ID:SNP ID". Red points represent individuals homozygous for the major allele, blue points represent heterozygotes and green points represent homozygotes for the minor allele.

4.3 Analysis of DNA methylation in t(1;11) carriers with a psychotic disorder

4.3.1 Sample information

In order to determine whether DNA methylation could be associated with phenotypic outcome (i.e. diagnosis) in t(1;11) carriers, a second analysis was performed. DNA methylation was compared between eight individuals with a psychotic disorder (i.e. schizophrenia, bipolar disorder or schizoaffective disorder, all presenting hallucinations or delusions) and five individuals with a non-psychotic disorder (i.e. single episode MDD, cyclothymia, conduct disorder, generalised anxiety disorder). The aim of this study was to determine whether DNA methylation was associated with psychosis in affected t(1;11) carriers. Thirteen t(1;11) carriers were analysed: eight individuals with a non-psychotic disorder and five with a psychotic disorder. The remaining three t(1;11) carriers in this study had a diagnosis of recurrent MDD. These samples were omitted from the analysis in order to also permit a comparison between "severe" (i.e. psychotic) diagnoses and "minor" diagnoses, based on the assumption that the burden of illness in recurrent MDD is greater than that in single episode MDD but lesser than that in psychosis.

4.3.2 Estimation of cellular proportions in whole blood

The *estimateCellCounts()* function was performed to estimate proportions of B-lymphocytes, granulocytes, monocytes, natural killer cells, $CD4^+$ T-lymphocytes and $CD8^+$ T-lymphocytes in this subset of individuals. The mean estimated proportion of each cell type was compared between individuals with a psychotic disorder and those with a non-psychotic disorder using a two-tailed Student's unpaired *t*–test. No significant between-group differences were observed ($p \geq 0.16$; Table 4.7). Therefore, cellular estimates were not included as covariates in the differential methylation analysis.

4.3.3 Data normalisation

The sample subset was normalised using 14 methods described previously (Table 2.3), and each method was scored based on its ability to reduce technical variation,

as described in Chapter 2.10.3. As with the analysis of the total sample set, dasen was identified as the optimum normalisation method (Table 4.8).

	$CD8^+$ T-Cells	**$CD4^+$ T-Cells**	**Natural Killer Cells**	**B-Cells**	**Monocytes**	**Granulocytes**
Mean proportion in t(1;11) Carriers with Psychosis	0.079	0.149	0.063	0.116	0.063	0.563
Mean proportion in t(1;11) Carriers without psychosis	0.040	0.155	0.063	0.048	0.073	0.629
p- value	0.189	0.797	0.982	0.174	0.470	0.285

Table 4.7: Estimated cellular proportions of whole blood within t(1;11) carriers with psychotic and non-psychotic disorders.

From left to right, columns show the minfi package's estimateCellCounts() function's estimated proportions of CD8+ and CD4+ T-cells, natural killer cells, B-cells, monocytes and granulocytes. Mean proportions per experimental group are presented for each cell type along with two-tailed unpaired Student's t-*test* p-*values assessing differences in cell subtypes between groups.*

Normalisation Method	DMRSE Rank	GCOSE Rank	Seabird Rank	Rank Mean	Rank of Rank Mean
dasen	3	7.5	5	5.17	1
daten1	5	9	4	6	2.5
daten2	4	11	3	6	2.5
naten	9	10	2	7	4
nanes	8	7.5	6	7.166	5
danes	1	14.5	8	7.833	6
nanet	10	13	1	8	7
Raw	12	1.5	11	8.17	8
danet	6	12	7	8.33	9
nasen	2	14.5	9	8.50	10
SWAN	7	6	14	9	11.5
danen	11	3	13	9	11.5
BMIQ	15	4	10	9.66	13
Noob	13	1.5	15	9.83	14
PBC	14	5	12	10.33	15

Table 4.8: Performance of 14 normalisation methods tested in 13 samples.

Table summarises the rankings of three metrics used to assess the performance of 14 normalisation strategies. In order of appearance, columns show the normalisation strategy, DMR-standard error (DMRSE) rank, genotype-combined standard error (GCOSE) rank, 1 - Seabird AUC rank, mean rank of the three metrics, and the rank of mean ranks.

4.3.4 Identification of differentially methylated positions

Linear regression was performed to identify differentially methylated positions (DMPs) between t(1;11) carriers with psychosis and those without psychosis. SVA was performed on the data to identify latent sources of variation. Three SVs were identified and fitted as covariates, along with age and gender. The goodness-of-fit of the model was assessed by calculating the genomic inflation factor λ. This was within the acceptable range of 1-1.1 as defined by GenABLE (Wang and Leal, 2012; λ = 1.08; Figure 4.4)

Three sites were identified as being significantly differentially methylated in t(1;11) carriers with a psychotic disorder compared to carriers with a non-psychotic disorder ($q \leq 0.05$; Table 4.9; Figure 4.5). The most significantly differentially methylated site from this comparison was hypomethylated in individuals with a psychotic disorder, and was located within 1500 bp of the transcription start site of the chromosome 8 gene *STC1* ($q = 0.02$). A further two loci were also hypomethylated in t(1;11) carriers with psychosis, located within 10 kb of one another in major histocompatibility complex (MHC) region on chromosome 6: *DPCR1* ($q = 0.03$) and *SFTA2* ($q = 0.03$). Although not significant after correction for multiple testing, three sites within the Protocadherin-γ gene cluster (*PCDHGA*) on chromosome 5 were present within the top 20 DMPs when ranked by p – value. All three sites showed hypermethylation in individuals with psychosis ($q \leq 0.056$).

The bimodal distribution of the methylation signal at these three sites indicated a possible SNP effect. To investigate whether genetic variation was driving the differential methylation signal at these sites, their genomic coordinates were cross-referenced to a table of potential signal-affecting genetic variants provided by Chen *et al.* (2013). Common variants (European minor allele frequency ≥ 0.05) were reported to be present at all three of these sites. Targeted Sanger-based sequencing was performed at these loci by Susan Anderson, confirming the presence of genetic variation as predicted based on methylation levels (i.e. lower levels in an individual carrying a CpG-abolishing variant). Although not all samples were successfully

sequenced, confirmation that the expected genetic variation was present in a subset of individuals was sufficient evidence that these findings were driven by the presence of CpG-altering variants.

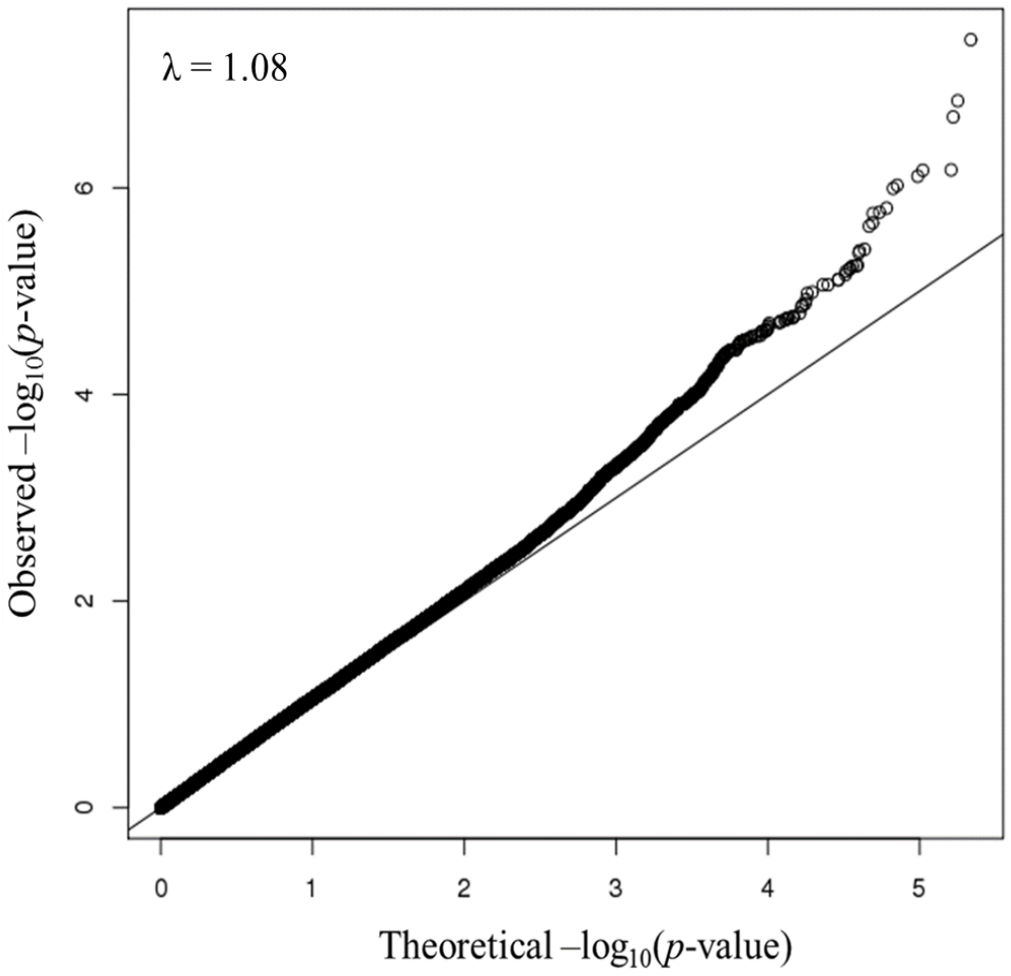


Figure 4.4: Quantile-quantile plot showing observed vs. expected *p*-values for differential methylation between t(1;11) carriers with a psychotic disorder and carriers with a non-psychotic disorder.

Shown are observed unadjusted $-\log_{10}$ p-values (y-axis; circular points) plotted against the expected distribution of $-\log_{10}$ p-values under the null hypothesis (x-axis; solid diagonal line). The genomic inflation factor λ is presented in the upper-left corner of the plot.

Probe ID	Hg19 Coordinates	Gene Symbol	Beta Difference	Fold-Change	t	*p-value*	*q* - value
cg16688533	Chr8:23713016	*STC1*	-0.19	-2.44	-12.98	3.69E-08	0.02
cg04559908	Chr6:30920123	*DPCR1*	0.30	4.28	11.41	1.44E-07	0.03
cg13561028	Chr6:30899649	*SFTA2*	0.18	2.50	11.03	2.06E-07	0.03

Table 4.9: Significantly differentially methylated positions between t(1;11) carriers with a psychotic disorder and t(1;11) carriers with a non-psychotic disorder.

Table summarises significantly differentially methylated sites between t(1;11) carriers with a psychotic disorder and carriers with a non-psychotic disorder (FDR q ≤ *0.05). Shown are: probe identifiers, Hg19 genomic coordinates, UCSC reference gene symbols, between-group differences in mean beta values, fold-change between groups, moderated* t-*statistic,* p-*value for differential methylation, and FDR-adjusted* p-*value.*

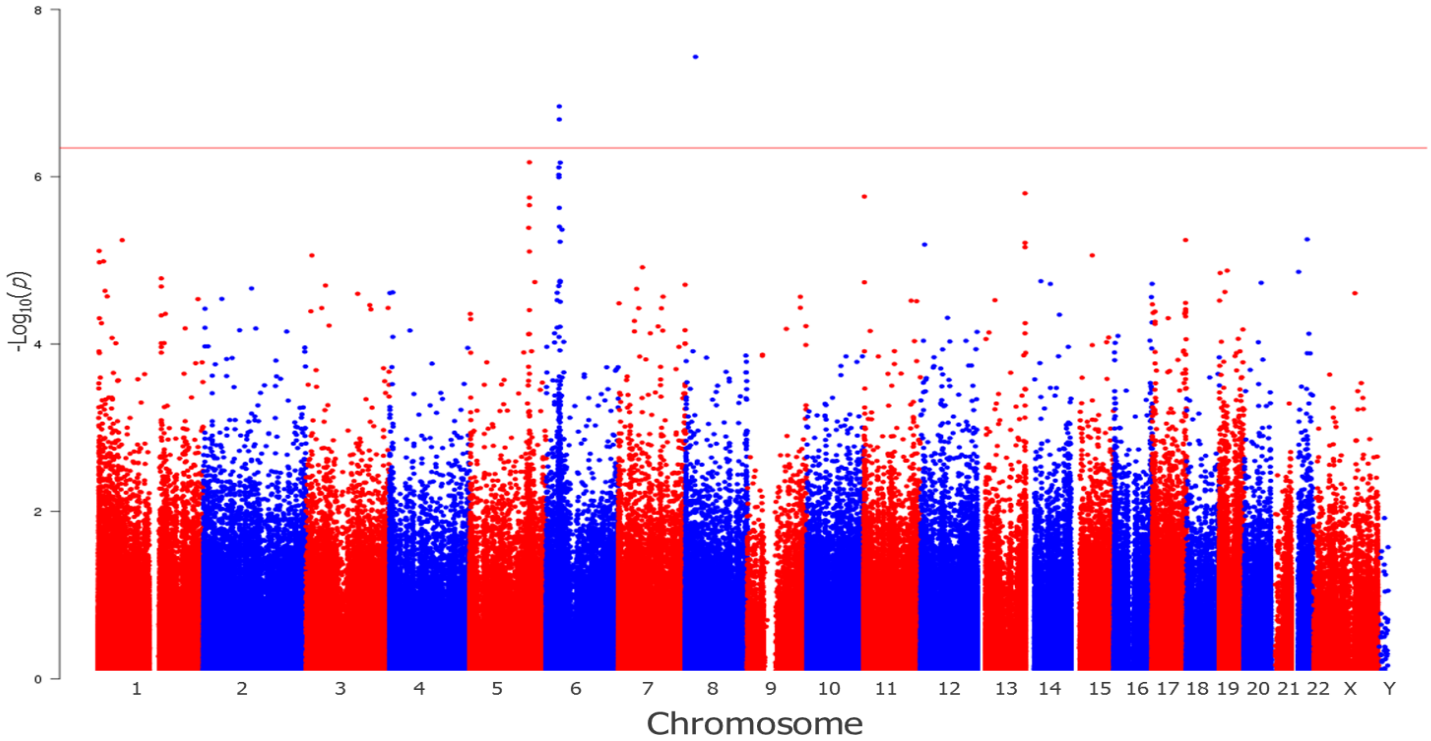


Figure 4.5: Manhattan plot for comparison of DNA methylation between t(1;11) carriers with a psychotic disorder and t(1;11) carriers with a non-psychotic disorder.

Shown are $-\log_{10} p$ – values for differential methylation between t(1;11) carriers with psychotic disorders and those with non-psychotic disorders (y – axis), plotted against chromosomal position (x – axis). The horizontal red line represents the $-\log_{10} p$ – value threshold for genome wide significance (FDR q *= 0.05).*

4.3.5 Gene ontology analysis of differentially methylated genes in psychosis

As with the analysis between carriers of the translocation non-carriers, a *p*-value-ranked list of genes containing DMPs was submitted to GORilla for ontology analysis (n = 20,752 genes).

A list of ten most significantly enriched GO terms is presented in Table 4.10. A complete list is available in Appendix I (Table A4). Of the 87 categories significantly over-represented amongst differentially methylated genes in t(1;11) carriers with psychosis ($q \leq 0.05$), several related to neuronal function. These included "Regulation of nervous system development", "Regulation of neurogenesis" and "postsynaptic density". The most significantly overrepresented term was "Integral component of lumenal side of endoplasmic reticulum membrane", in the "GO component" class (GO:0071556; $q = 5.89 \times 10^{-5}$).

GO Term	Description	Class	Enrichment	*p*-value	*q*-value
GO:0071556	integral component of lumenal side of endoplasmic reticulum membrane	Component	82.96	3.55×10^{-8}	5.89×10^{-5}
GO:0042611	MHC protein complex	Component	140.38	1.45×10^{-7}	0.0001
GO:0030658	transport vesicle membrane	Component	30.85	5.56×10^{-7}	0.0003
GO:0048518	positive regulation of biological process	Process	1.19	8.43×10^{-8}	0.0012
GO:0012507	ER to Golgi transport vesicle membrane	Component	68.76	3.41×10^{-6}	0.0014
GO:0048522	positive regulation of cellular process	Process	1.2	5.73×10^{-7}	0.0020
GO:0042605	peptide antigen binding	Function	4.97	2.29×10^{-6}	0.0024
GO:0042995	cell projection	Component	1.38	7.47×10^{-6}	0.0025
GO:0051960	regulation of nervous system development	Process	1.6	5.64×10^{-7}	0.0026
GO:0048856	anatomical structure development	Process	1.28	4.57×10^{-7}	0.0032

Table 4.10: Summary of GO terms enriched for differentially methylated genes in t(1;11) carriers with a psychotic disorder.

For each GO term, table summarises the GO identifier, the GO description, the GO class, enrichment, the enrichment p-values, and enrichment FDR q-values for genes showing the most differential methylation in t(1,11) carriers with psychotic disorders compared to carriers with a non-psychotic disorder. Enrichment is defined as (b/n) / (B/N), where b = the number of genes at the top of the p-value ranked list that is associated with a given GO term, n = the number of genes at the top of the p-*value-ranked gene list, B = the total number of genes associated with a given GO term, and N = the total number of genes. Shown are the top 10 enrichments ranked by* p-*value. The total list of significant enrichments is presented in Appendix I (Table A4).*

4.3.1 Identification of differentially methylated regions in t(1;11) carriers with a psychotic disorder

To identify genomic regions containing multiple nominally significant differentially sites between t(1;11) carriers with a psychotic illness and those with a non-psychotic illness ($p \leq 0.05$), DMR analysis was performed.

A total of 238 DMRs were identified using the *ChAMP* package (Table 4.11). The most significant DMR in t(1;11) carriers with psychosis occurred within the gene body of *RPTOR*, spanning eight probes ($p = 1.55 \times 10^{-21}$). The largest region was detected in an intergenic region on chromosome 13, consisting of 13 probes over 5.6 kb ($p = 0.0004$). The DMR with the most probes spanned 2.6 kb within *TNXB*, on chromosome 6 (n = 45 probes; $p = 0.001$).

Gene(s)	Feature	Region	Probe Count	DMR *p*-value
RPTOR	Body	Chr17:78865119-78867431	8	1.55×10^{-21}
NA, *MIR886*	IGR, TSS200, Body	Chr5:135415531-135416414	13	1.18×10^{-19}
PTPRN2	Body	Chr7:158045532-158046806	6	2.13×10^{-15}
TNXB	Body	Chr6:32064430-32064738	13	1.49×10^{-13}
NA	IGR	Chr6:29721548-29725160	4	1.6×10^{-13}
NA	IGR	Chr13:113295067-113297572	4	4.17×10^{-13}
SLC38A4	TSS200, 5'UTR, 1stExon	Chr12:47219385-47219867	6	4.95×10^{-12}
HCG27	TSS1500	Chr6:31164851-31165031	8	3.24×10^{-11}
NA	IGR	Chr6:29520536-29521310	15	4.37×10^{-11}
NA	IGR	Chr6:25882428-25882752	3	7.78×10^{-11}

Table 4.11: Psychosis-associated DMRs identified by the probe lasso algorithm.

Table summarises the genes (if applicable), genomic features, hg19 genomic coordinates, number of probes and p-value associated with each DMR. "NA" in the "Gene" column represents intergenic regions (i.e. regions not annotated to a RefSeq gene). Genomic features are coded "IGR" for intergenic regions, "TSS200" and "TSS1500" for probes occurring within 200 and 1500 bp of a gene's transcription start site, respectively; "5'UTR" and "3'UTR" for probes occurring within a gene's 5' and 3' untranslated region, respectively, "1stExon" for probes occurring within the first exon of a gene, and "Body" for probes occurring within the gene body. Shown are the top 10 DMRs ranked by p-*value. The total list of significant DMRs is presented in Appendix I (Table A5).*

4.4 Assessment of DNA methylation at sites correlated between blood and brain

A study by Walton *et al.* (2015) presented a list of 100 probes from the 450k methylation array where methylation levels were highly correlated between blood and brain (Spearman's rho $\geq$ 0.94). To determine whether any of these sites were differentially methylated between groups in blood from the t(1;11) family, methylation was examined in a subset of these probes that had passed quality control in the above analyses (n = 70). Within this set of a probes, a multiple testing correction was applied using a false discovery rate of 5% (Benjamini and Hochberg, 1995).

Of the 70 probes reported as correlated between blood and brain (Walton *et al.*, 2015), two sites were significantly differentially methylated between t(1;11) carriers and non-carriers ($q \leq 0.05$; Table 4.12). These probes were located within 350 bp of one another in the body of *CYP2E1*, on chromosome 10.

The same 70 probes were assessed for differential methylation between t(1;11) carriers with a psychotic disorder and those with a non-psychotic disorder. Two sites were significantly differentially methylated between groups ($q \leq 0.05$; Table 4.12). The most significantly differentially methylated site was located in the MHC region on chromosome 6, within the body of *HLA-DQB2*. The second site was within 1500 bp of the TSS of *MRI1*, on chromosome 19. Methylation at each of these loci was reported by Lemire *et al.* (2015) to be associated with a meQTL. To determine whether the difference in methylation was associated with this meQTL, whole-genome sequence data from the family were queried for this variant. None of the individuals sequenced showed variation at this site.

Probe ID	Hg19 Coordinates	Gene	Beta Difference	Fold-Change	t	*p-value*	*q* - value	Study
cg23400446	Chr10: 135342560	*CYP2E1*	0.13	2.25	4.41	9.60×10^{-5}	0.007	t(1;11) carrier vs. non-carrier
cg10862468	Chr10: 135342218	*CYP2E1*	0.12	1.57	3.67	0.0008	0.03	
cg07180897	Chr6: 32729130	*HLA-DQB2*	0.04	1.47	4.58	0.0007	0.035	Psychotic vs. non-psychotic disorder
cg16474696	Chr19: 13875014	*MRI1*	0.01	1.60	4.39	0.001	0.035	

Table 4.12: Significantly differentially methylated positions in probes reported to be correlated between blood and brain by Walton *et al.* (2015).

Table summarises sites showing significant differential methylation in a set of 70 probes reported by Walton et al. (2015) to be correlated between blood and brain. From left to right, columns show the probe identifiers, Hg19 genomic coordinates, UCSC reference gene IDs, between-group difference in mean beta value, fold-change between groups, average methylation M-value, moderated t-statistic, p-value for differential methylation, FDR-adjusted p-value, log odds for differential methylation and the study in which significant differential methylation was observed.

4.5 Estimation of DNA methylation age

Based on Illumina methylation array profiles from over 7,800 samples across 82 datasets, Horvath (2013) developed a DNA methylation age predictor applicable to the methylation profiles of multiple tissues, including whole blood. DNA methylation age has been proposed to reflect a cumulative effect of epigenetic maintenance, and is highly correlated with chronological age.

To assess whether carrying the translocation was associated with differences in DNA methylation age, age acceleration (i.e. the difference between DNA methylation age and chronological age) was compared between t(1;11) carriers and non-carriers. DNA methylation age and chronological age were highly correlated overall ($p < 2.2 \times 10^{-16}$, $R^2 = 0.94$; Figure 4.6), as well as within groups (translocation carriers $p = 1.23 \times 10^{-10}$, $R^2 = 0.94$; non-carriers $p = 1.16 \times 10^{-14}$, $r^2 = 0.94$). No significant differences in age acceleration (i.e the difference between chronological age and DNA methylation age) were observed between t(1;11) carriers and non-carriers (Student's two-tailed independent samples *t*-test $p = 0.4$; Figure 4.7).

To determine whether psychosis may be related to an effect on DNA methylation age in t(1;11) carriers, DNA methylation age was compared between t(1;11) carriers with a psychotic disorder and those with a non-psychotic disorder. No significant difference was observed ($p = 0.07$; Figure 4.8).

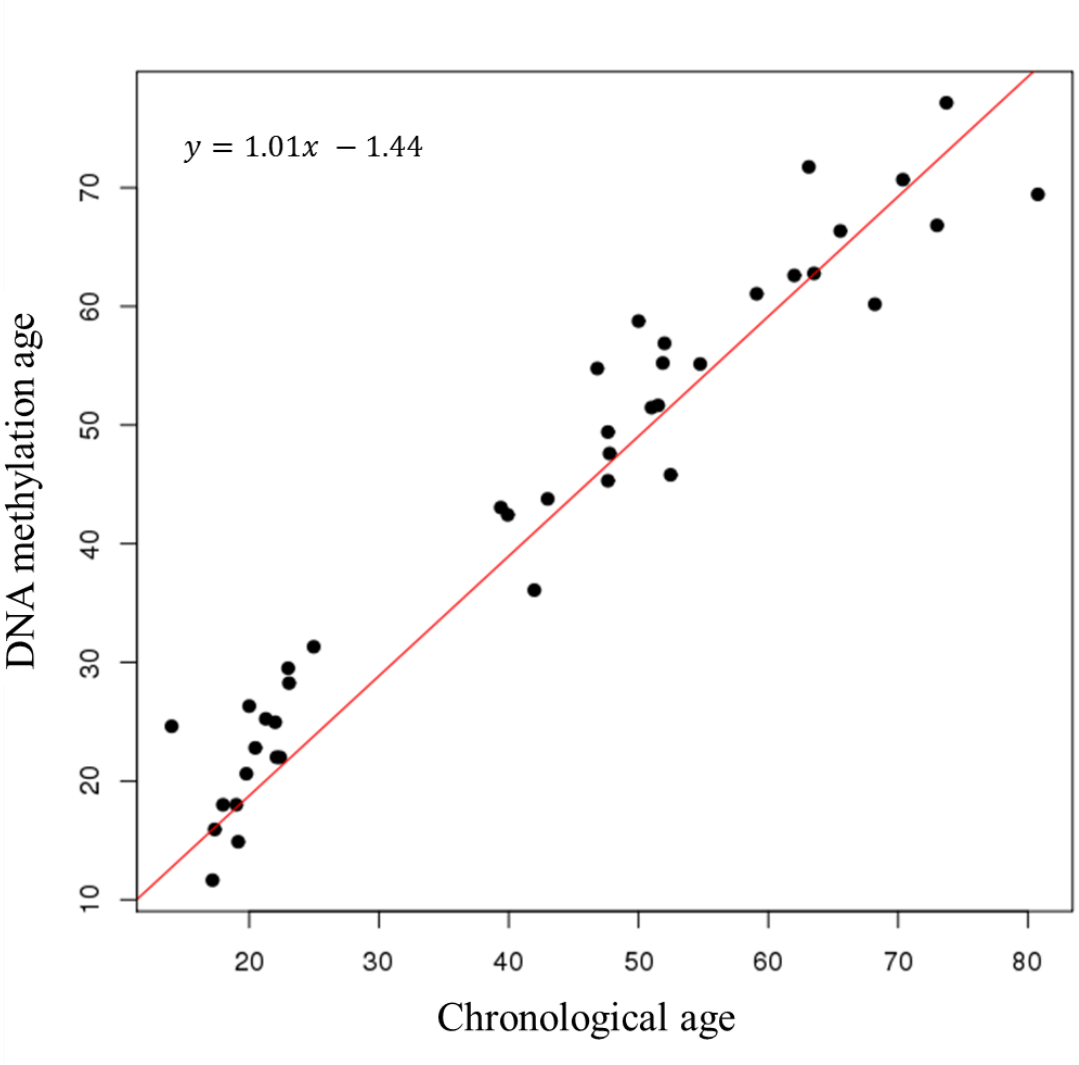


Figure 4.6: Relationship between DNA methylation age and chronological age in 41 blood-derived samples.

Chronological age (x-axis) is plotted against DNA methylation age (y-axis) derived from Horvath's DNA methylation age calculator (Horvath, 2013). The linear relationship between the two variables is summarised by the red line, the equation of which is presented in the top left corner.

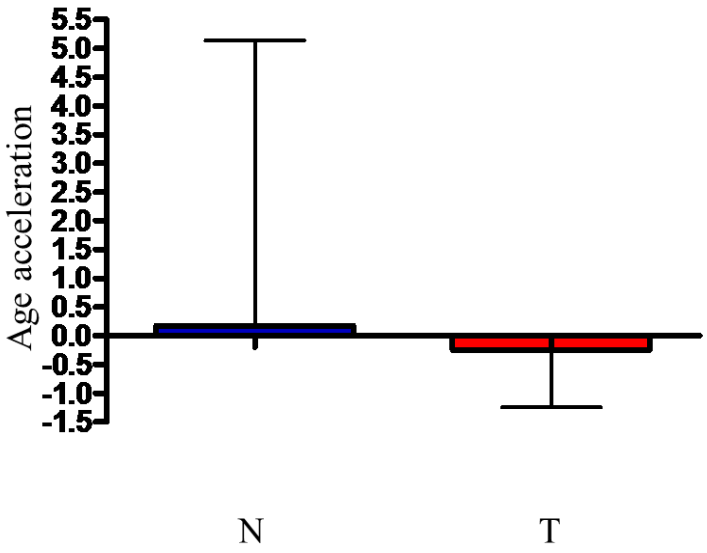


Figure 4.7: Age acceleration between t(1;11) carriers and non-carriers.

Mean age acceleration as calculated as the difference between methylation age and chronological age (y-axis) is plotted against translocation carrier status. "N" denotes t(1;11) non-carriers (blue bar), "T" denotes t(1;11) carriers (red bar). Error bars show the standard deviation from mean age acceleration in each group.

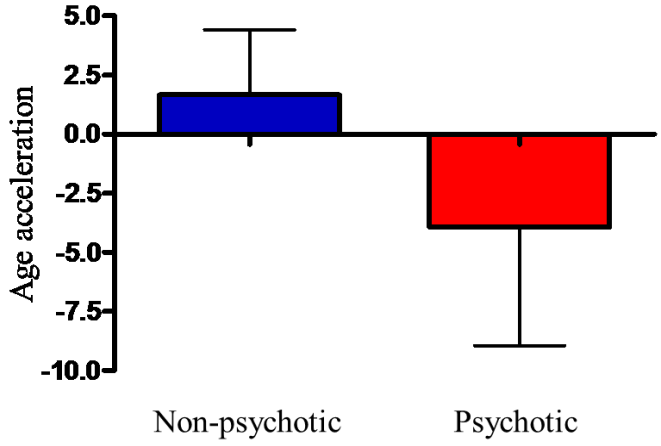


Figure 4.8: Age acceleration between t(1;11) carriers with a non-psychotic disorder and carriers with a psychotic disorder.

Mean age acceleration as calculated as the difference between methylation age and chronological age (y-axis) is plotted against diagnosis class. The blue bar shows the mean age acceleration age in t(1;11) carriers with a non-psychotic disorder while the red bar shows the mean age acceleration in t(1;11) carriers with a psychotic disorder. Error bars show the standard deviation from the mean age acceleration in each group.

.

4.6 DNA methylation at polymorphic sites

Prior to the analysis of the methylation data, probes affected by polymorphisms at their target sites were removed as a quality control measure. Differences in signal can be observed at such sites in the form of a tri-modal distribution of the three possible genotypes of the underlying variant (i.e. major allele homozygous, heterozygous, and minor allele homozygous). Such polymorphisms may result in disrupted DNA methylation and disruption of downstream regulatory functions through the abolition of CpG sites. To determine whether such polymorphisms were present in the t(1;11) family, methylation data were reanalysed, including these probes.

4.7 Selection of normalisation strategy for polymorphic probe set

The data were normalised by the 14 methods described in Table 2.3 and each method was scored based on its ability to minimise technical variation as described in Chapter 2.10.3. The top-ranking method was *dasen*, which was used to normalise the data (Table 4.13).

Normalisation Method	DMRSE Rank	GCOSE Rank	Seabird Rank	Rank Mean	Rank of Rank Mean
dasen	1	11.5	2	4.83	1
danes	3	3.5	9	5.16	2
nasen	2	11.5	3	5.5	3
daten1	4	8	5	5.66	4
daten2	5	9	4	6	5
danen	11	7	1	6.33	6
nanes	7	3.5	11	7.16	7
raw	12	5.5	6	7.83	8
SWAN	10	2	12	8	9
naten	8	10	8	8.66	10
danet	6	14	7	9	11
BMIQ	14	1	14	9.66	12
noob	13	5.5	13	10.5	13
nanet	9	13	10	10.66	14
PBC	15	15	15	15	15

Table 4.13: Performance of 14 normalisation methods tested in the dataset containing polymorphism-targeting probes

Table summarises the rankings of three metrics used to assess the performance of 14 normalisation strategies. From left to right, columns show the normalisation strategy, DMR-standard error (DMRSE) rank, genotype-combined standard error (GCOSE) rank, 1 - Seabird AUC rank, mean rank of the three metrics, and the rank of mean ranks.

4.8 Identification of differentially methylated positions in the polymorphic probe set

To identify latent sources of variation present in the data, surrogate variable analysis was performed. Nine significant surrogate variables were identified. Carriers of the translocation were compared to non-carriers by linear regression, fitting age, gender and the nine surrogate variables as covariates. In total, 20 sites were significantly differentially methylated between t(1;11) carriers and non-carriers ($q \leq 0.05$; Figure 4.9; Table 4.14). Among these sites were the 13 significantly differentially methylated loci identified in the SNP-filtered comparison of t(1;11) carriers and non-carriers. The two most significantly differentially methylated probes identified were on chromosome 11, within *CHRDL2* ($q = 3.16 \times 10^{-28}$; Figure 4.10) and *SHANK2* ($q = 1.12 \times 10^{-9}$; Figure 4.10). Cross-referencing of the genomic coordinates of these targets to 1000 genomes phase 3 data (1000 Genomes Project Consortium, 2015) confirmed the presence of variation at both sites. The *SHANK2* probe targets a common polymorphism (rs1000968; European MAF = 0.87) while the *CHRDL2* probe targets a rare polymorphism (rs75439151; European MAF = 0.005). Whole-genome sequence data from the family confirmed the presence of these variants. The rare *CHRDL2* variant was present exclusively among t(1;11) carriers while the *SHANK2* variant was observed in all t(1;11) carriers and two non-carriers. Methylation profiles were consistent with genotype: CpG homozygotes showed methylation levels approximately twice those seen in heterozygotes of these CpG-abolishing variants (Figure 4.10).

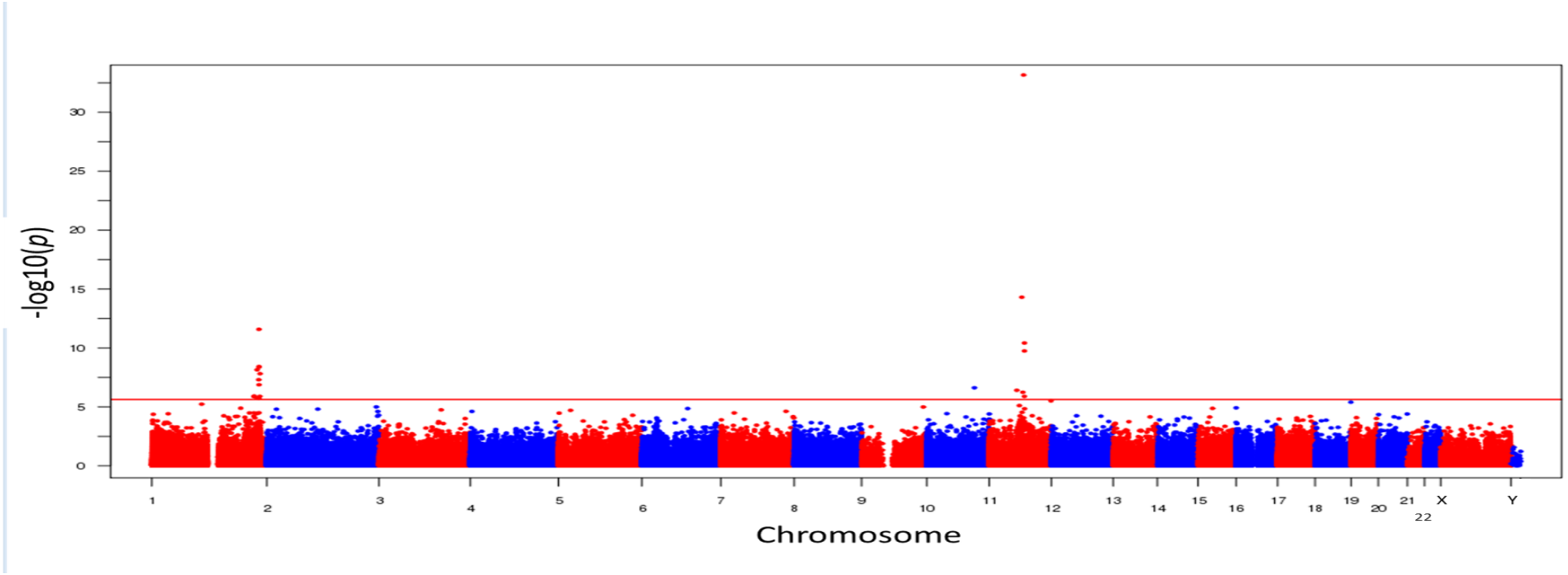


Figure 4.9: Manhattan plot of differentially methylated positions between t(1;11) carriers and non-carriers in the polymorphic probe set.

Shown are –log$_{10}$ p–values for differential methylation between t(1;11) carriers with psychotic disorders and those with non-psychotic disorders (y–axis), plotted against chromosomal position (x – axis). The horizontal red line represents the –log$_{10}$ p – value threshold for genome wide significance (FDR q *= 0.05).*

Probe ID	Hg19 Coordinates	Gene	Beta Difference	Fold-Change	t	*p-value*	*q* - value
cg16056964	Chr11:74430117	*CHRDL2*	-0.39	-7.49	-55.94	$6.96X\ 10^{-34}$	$3.16X\ 10^{-28}$
cg09157251	Chr11:70733251	*SHANK2*	0.44	7.86	13.68	$4.94X\ 10^{-15}$	$1.12X\ 10^{-09}$
cg09186051	Chr1:231981906	*DISC1,TSNAX-DISC1*	-0.07	-1.30	-10.82	$2.61X\ 10^{-12}$	$3.94X\ 10^{-07}$
cg10109470	Chr11:76430445	*GUCY2E*	-0.08	-1.77	-9.73	$3.79X\ 10^{-11}$	$4.29X\ 10^{-06}$
cg26728851	Chr11:76430375	*GUCY2E*	-0.03	-1.64	-9.12	$1.79X\ 10^{-10}$	$1.62X\ 10^{-05}$
cg15157974	Chr1:232144702	*DISC1,TSNAX-DISC1*	-0.04	-1.26	-7.97	$3.77X\ 10^{-09}$	0.0003
cg05656812	Chr1:232021560	*DISC1,TSNAX-DISC1*	-0.06	-1.31	-7.94	$4.17X\ 10^{-09}$	0.0003
cg06928246	Chr1:227974645	NA	-0.07	-1.58	-7.74	$7.09X\ 10^{-09}$	0.0004
cg09674468	Chr1:234300299	*SLC35F3*	-0.08	-1.44	-7.47	$1.51X\ 10^{-08}$	0.0008
cg18815120	Chr1:231512676	*EGLN1*	-0.12	-1.99	-7.05	$4.92X\ 10^{-08}$	0.002
cg16177633	Chr1:232172585	*DISC1,TSNAX-DISC1*	-0.03	-1.19	-6.71	$1.32X\ 10^{-07}$	0.005
cg24508974	Chr10:103330391	NA	-0.01	-1.19	6.50	$2.38X\ 10^{-07}$	0.009
cg02771260	Chr11:59836817	*MS4A3*	-0.13	-1.81	-6.33	$3.87X\ 10^{-07}$	0.013
cg25899154	Chr11:72897143	NA	-0.07	-1.30	-6.20	$5.66X\ 10^{-07}$	0.018
cg26355502	Chr1:221916303	*DUSP10*	-0.01	-1.30	-5.94	$1.20X\ 10^{-06}$	0.033
cg25191850	Chr1:233749633	*KCNK1*	0.03	-1.31	5.93	$1.26X\ 10^{-06}$	0.033
cg03807330	Chr11:76327232	NA	0.05	-1.33	5.92	$1.28X\ 10^{-06}$	0.033
cg00892096	Chr1:220987469	*MOSC1*	0.04	-1.36	5.91	$1.32X\ 10^{-06}$	0.033
cg00965168	Chr1:227974541	NA	-0.05	-1.44	-5.80	$1.83X\ 10^{-06}$	0.044
cg21875980	Chr1:231553510	*EGLN1*	0.06	-1.41	5.78	$1.93X\ 10^{-06}$	0.044

Table 4.14: Table of differentially methylated positions between t(1;11) carriers and non-carriers in the polymorphic dataset.

Table summarises significantly differentially methylated sites between t(1,11) carriers and non-carriers in a dataset including polymorphic probes. From left to right, columns show probe identifiers, Hg19 genomic coordinates, UCSC reference gene IDs ("NA" denotes intergenic), between-group difference in mean beta value, fold-change between groups, moderated t*-statistic,* p*-value for differential methylation and FDR-adjusted* p*-value (*q*-value).*

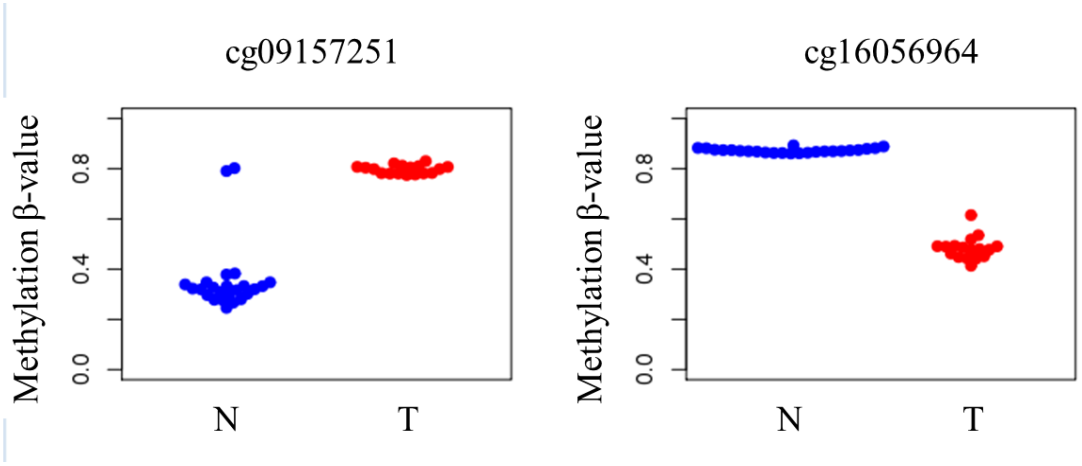


Figure 4.10: Top two significantly differentially methylated polymorphic CpGs in t(1;11) carriers.

Methylation β-values (y-axis) are plotted against t(1,11) carrier status at significantly differentially methylated polymorphic CpG sites in SHANK2 (cg09157251) and CHRDL2 (cg16056964). "N" denotes t(1,11) non-carriers (blue points) and "T" denotes t(1,11) carriers (red points).

4.9 Discussion

This chapter describes an analysis of methylation in whole blood-derived DNA from individuals from the t(1;11) family. Two comparisons were performed: first, DNA methylation was compared between t(1;11) carriers and non-carriers, and second, methylation in t(1;11) carriers with a psychotic disorder was compared to those with a non-psychotic disorder. The aim of the former study was to determine whether differential DNA methylation was associated with the translocation in carriers, while the aim of the latter study was to determine whether DNA methylation differences could be observed between psychotic and non-psychotic illness in t(1;11) carriers.

Comparison of DNA methylation between translocation carriers and non-carriers resulted in the identification of 13 DMPs: nine on chromosome 1, three on chromosome 11, and one on chromosome 10. With the exception of the chromosome 10 DMP, all probes mapped to the regions surrounding the translocation breakpoints. However, not all CpGs immediately adjacent to the translocation breakpoints showed differential methylation. These findings could be due to the co-inheritance of meQTLs with the translocation. Alternatively, this could be an issue of power, as the sample size was relatively small. To investigate this, a *post-hoc* power analysis revealed the study was 80 % powered to detect an effect size of a fold-change of at least 1.95 at an alpha of 1.13×10^{-7} (Bonferroni $p = 0.05$). This may indicate that smaller effects in the region of the translocation may not be detected, at the current sample size.

Four DMPs mapped within the *DISC1* gene, which is directly disrupted by the translocation on chromosome 1. DISC1 has been implicated in neurodevelopment, cognitive function and disease risk in psychiatric illness (Brandon and Sawa, 2011; Carless et al., 2011; Thomson et al., 2013; Rampino et al., 2014). Two DMPs mapped to the *EGLN1* gene, which encodes prolyl hydroxylase domain-containing protein 2 (PHD2), and is located within 500 kb of the chromosome 1 breakpoint. PHD2 regulates the transcription factor HIF-1α, the master transcriptional regulator of the cellular response to hypoxia (Berra et al., 2003). Hypoxia is a previously reported obstetric and

developmental risk factor for schizophrenia (Dalman et al., 2001; Byrne et al., 2007; Schmidt-Kastner et al., 2006). Differential methylation at this gene might indicate a mechanism whereby increased risk of illness is conferred by the translocation, through a disruption to the cellular response to hypoxia.

Comparison of t(1;11) carriers and non-carriers also identified a significantly differentially methylated site on chromosome 10 (cg24508974). This site is located in an intergenic region flanked by two RefSeq genes: *BTRC* centromeric, and *POLL* telomeric, as reported by RefSeq. ChIP-seq data from ENCODE data reports a transcription factor binding site at this locus in K562 cells (ZNF263; Gerstein *et al.*, 2012). This suggests that this locus is involved in regulating the expression of the nearby genes. Zariwala *et al.* (2004) reported a phenotype reminiscent of primary ciliary dyskinesia (PCD) in mice following deletion of *Poll.* DISC1 regulates the maintenance of dopamine receptor-expressing primary cilia in the brain (Marley and von Zastrow, 2010). However, there is currently no evidence to suggest that DISC1 and POLL act together in regulating primary cilia. Although this site has been previously reported to be regulated by a meQTL by Lemire et al. (2015), this was not replicated in the above study (Table 4.6).

In order to identify biological processes that may be affected by differential methylation associated with the translocation, gene ontology analysis was performed. This analysis identified several neurologically-relevant processes, including "regulation of nervous system development", "regulation of neurogenesis" and "positive regulation of nervous system development". These findings are in keeping with the theory of the neurodevelopmental origin of schizophrenia (reviewed by Rapoport et al., 2012), pointing additionally to the possibility of neurodevelopmental abnormalities in t(1;11) carriers. Recent work has identified structural differences in the brains of t(1;11) carriers compared to non-carriers. Doyle et al. (2015) reported reductions in cortical thickness measurements in t(1;11) carriers consistent with those observed in schizophrenia patients. Furthermore, deficits in white matter integrity have been reported in t(1;11) carriers by Whalley *et al.* (2015), while Thomson *et al.* (2016) reported t(1;11)-associated reductions in gyrification. It should be noted, however, that the enriched neurologically-relevant

terms were not the most significant findings in the two analyses when considering enrichment *p*-value in the psychotic vs. non-psychotic comparison. Here, the most enriched term was "integral component of lumenal side of endoplasmic reticulum membrane". Such enrichments may support a hypothesis of more general (i.e. non-brain-specific) effects of the translocation on basic cell functioning.

Other studies of DNA methylation in the context of chromosomal translocations and disease have generally been cancer-based. These studies have also reported significant differences in DNA methylation associated with translocations. Busche *et al.* (2013) examined DNA methylation in leukaemia patients with a t(12;21) translocation. They reported differential methylation associated with this translocation but, in contrast to the observations here, this was not in the regions surrounding the breakpoints. Walker *et al.* (2011) reported on methylation in 12 cancer-associated chromosomal abnormalities. They found the greatest differences in methylation to occur in the context of a t(4;14) translocation. Here, significant DMPs occurred throughout the genome with no apparent enrichment on the translocated chromosomes. However, the same study also reported on differential methylation associated with a t(11;14) translocation where the top six ranked DMPs were hypomethylated and located within *CCND1*: the chromosome 11 gene disrupted by this translocation.

Another possible mechanism for differential methylation at the t(1;11) breakpoint regions may relate to the passive transmission on the translocated chromosomes of meQTLs associated with methylation at sites around these breakpoints. Support has been provided for this mechanism in the results presented here. Of the 13 significant DMPs identified between t(1;11) carriers and non-carriers ($q \leq 0.05$), seven were previously reported to be influenced by a meQTL in a study of lymphocyte DNA methylation (Lemire et al., 2015). These findings were replicated for five of these seven probes. The failure to replicate the remaining loci may be attributable to limited statistical power due to the small number of homozygote carriers of the minor allele of the meQTL.

To determine whether DNA methylation differences were associated with a phenotypic outcome (i.e. diagnosis) in t(1;11) carriers, five individuals with a psychotic disorder were

compared to eight individuals with a non-psychotic disorder. Three DMPs were identified in this analysis: one on chromosome 8 and two on chromosome 6. These results were found to be due to genetic variation at the target locus for all three sites. Filtering of probes affected by polymorphisms at their target site was performed on the data prior to analysis, based on whole-genome sequence data from the family. However, it is clear from these results that a proportion of such probes still remained. Factors such as sequencing errors and variable read depth may limit the accuracy of variant detection and, therefore, pose limitations for this strategy. An alternative approach would be to filter based on population-based information, such as 1000 genomes data, as has been performed in other epigenome-wide association studies (1000 Genomes Consortium, 2015; Chen *et al.*, 2013). However, in the context of a family study, this strategy may result in the superfluous removal of probes targeting non-polymorphic sites or, alternatively, the failure to identify rare, family-specific variants. There may be some situations where retaining such probes may be desirable. For example, a difference in methylation due to a SNP that creates or destroys a CpG at a target site is informative if it represents the mechanism by which the SNP confers a change in disease risk. One site proximal to the chromosome 11 breakpoint was found to possess a rare CpG-abolishing variant, present exclusively in t(1;11) carriers (cg16056964; Figure 4.10). It is likely that the regions flanking the t(1;11) breakpoints will contain many similar sites simply due to being in linkage disequilibrium with the translocation. However, further experiments would be required to determine which, if any, of these variants have risk-conferring effects.

GO analysis of the DMPs identified from the within-carrier comparison of psychotic and non-psychotic disorders revealed significant enrichment for neurodevelopmental terms, including “negative regulation of nervous system development” and “negative regulation of neurogenesis” (Table 4.12). These findings may indicate that diagnostic differences in carriers of the translocation might, at least in part, be developmental in origin. Alternatively, it is possible that adult neurogenesis is disrupted by the translocation: a process in which *DISC1* has been implicated (Lee *et al.*, 2015).

DMP analysis assesses altered methylation on a site-by-site basis, thus ignoring potentially informative signal present at neighbouring sites. It can be hypothesised that

sub-genome-wide significant differences in methylation across a number of nearby sites might confer a biologically meaningful effect, but would remain undetected by DMP analysis. To address this, DMR analyses were carried out for both the translocation carrier vs. non-carrier and psychotic vs. non-psychotic disorder comparisons. These analyses identified one gene, *TNXB*, as being particularly noteworthy. This gene contained the most significant DMR, which comprised 51 probes spanning approximately 1.6 kb, identified in the comparison of t(1;11) carriers and non-carriers. This comparison also identified a second region within *TNXB*, 325 bp in length, which comprised eight probes. In the comparison of t(1;11) carriers with psychotic and non-psychotic disorders, four further *TNXB* DMRs were identified, one of which fell within the 51-probe DMR described above. *TNXB* encodes Tenascin X, an extracellular glycoprotein, which is predominantly expressed in connective tissues. It has been associated with schizophrenia in a study of 122 British trios (Wei and Hemmings, 2003) and is located within the extended MHC region, which was identified as the most significant locus in a recent GWAS of schizophrenia (Ripke *et al.*, 2014). There is both long-term and recent evidence for altered immune function in psychiatric illness (Smith, 1992; Corvin and Morris, 2014; Sekar *et al.*, 2016).

Along with the MHC, three additional DMRs were identified within genes from the same Psychiatric Genomics Consortium's most recent GWAS of schizophrenia (Ripke *et al.*, 2014). Two regions within the genes *IGSF9B* and *CNTN4* showed differential methylation in the comparison between t(1;11) carriers and non-carriers. The third, identified in the comparison of individuals with psychotic versus non-psychotic disorders, was within the cell cycle-regulating gene *MAD1L1*. This gene has also been found to be significantly differentially methylated in a large-scale DNA methylation study of schizophrenia (Montano *et al.*, 2016). Montano *et al.* (2016) also reported significant differential methylation of a locus within *RPTOR*: a gene in which three DMRs were identified in the present analyses: one in the comparison of t(1;11) carriers and non-carriers, and two in the comparison of t(1;11) carriers with a psychotic disorder compared to those with non-psychotic disorders. One of the *RPTOR* DMRs identified in the comparison of individuals with a psychotic disorder vs. non-psychotic disorder consisted of four probes, two of which (cg22091236 and cg27457201; both hypermethylated in

individuals with a psychotic disorder) have been reported to be hypermethylated in schizophrenia by Montano *et al*. (2016). *RPTOR* is a key component of mTOR signalling, which has been implicated in synaptic plasticity (Reviewed by Graber *et al.*, 2013). Another DMR identified in the comparison of t(1;11) carriers with psychotic and non-psychotic illness which might be relevant to disease is located within 1500 bp of the TSS of *HTR2A*, a serotonin receptor gene. This DMR was hypermethylated in t(1;11) carriers with psychosis ($p = 0.0003$). This region is within 1 kb of rs6311, which has been shown by Abdolmaleky *et al.* (2011) to be associated with hypermethylation of the promoter region of *HTR2A* in *post-mortem* brains of schizophrenia and bipolar disorder patients. The same study also reported that reduced expression of *HTR2A*, and age-of-onset of disease was associated with hypermethylation of the gene's promoter region. The serotonin pathway is a target of mood stabilising and antipsychotic drugs (Schloss *et al.*, 1998; Meltzer *et al.*, 2012). It is possible that this difference is due to an effect of medication in these individuals. However, the relationship between methylation and medication status could not be assessed as information regarding medication status of these individuals was unavailable. This was also the case for other potential confounders, such as smoking status and alcohol intake: both of which have been shown to influence DNA methylation (Philibert *et al.*, 2012; Ambatipudi *et al.*, 2016). Should they be independent of t(1;11) carrier status or diagnosis, it is hoped that surrogate variable analysis would account for these factors. In order to attain information regarding these factors, follow-up of these family members is warranted.

A further DMR identified in the comparison of t(1;11) carriers and non-carriers worth noting was within a gene coding for the DISC1 interactor TRAK1. TRAK1 and DISC1 have been shown to interact in mitochondrial trafficking complexes in neurons (Ogawa *et al.*, 2014). Mitochondrial trafficking is a critical process in neurons in order to meet high-energy requirements at specific regions, such as sites of neurotransmission (reviewed by Vos *et al.*, 2010).

It is important to note that none of the DMRs identified contained any of the 13 significant DMPs identified when comparing t(1;11) carriers and non-carriers. Of these thirteen

DMPs, seven could not have been found to contribute to DMRs due to the fact that they are located more than 2000 bp away from the nearest nominally significant DMP – outside the bounds of the DMR-calling lasso radius. This highlights a limitation of DMR analysis: the identification of DMRs requires several parameters, such as the minimum number of probes required to form a DMR and the distance permissible between these probes, to be set. As there is a dearth of experimental evidence linking the selection of DMR parameters to the identification of biologically meaningful DMRs, the setting of these parameters is arbitrary. Bearing this in mind, it may be of interest that three adjacent probes immediately telomeric to the chromosome 1 breakpoint were nominally significantly hypomethylated in t(1;11) carriers, all mapping to *DISC1* and *DISC2*, spanning a 1088 bp region (cg12751277, cg05812666 and cg10483534; unadjusted $p \leq 0.05$). However, these did not meet the criteria for DMR calling as they were spaced slightly too far apart.

It is not possible to distinguish between DNA methylation and hydroxymethylation by the methods used in this study, posing a limitation to the work presented in this chapter. Hydroxymethylation may act as an intermediate between methylation and demethylation of DNA. It is possible that the results presented in this chapter are due to mixed effects of DNA methylation and hydroxymethylation. However, the reported levels of global DNA hydroxymethylation are low in blood (Nestor *et al.*, 2012; Stewart *et al.*, 2015). Should both processes be occurring in these samples, further work would be required such as oxidative bisulphite treatment in order to determine their individual effects (Booth *et al.*, 2013)

It was not possible to determine whether the observed differential methylation was associated with a change in gene expression, due to lack of sample availability. This presents a challenge in identifying downstream effects of differential DNA methylation in these individuals. The recent availability of DNA, RNA and protein from t(1;11) family iPSC-derived neurons should go some way to address this. Methylation analysis has been performed in these samples, and is the subject of Chapter 5. This work was performed to address the hypothesis that the differences in methylation observed at the t(1;11)

breakpoints might be observed in other cell types – specifically neurons. Analysis of gene expression data from these samples is ongoing which will also allow for the assessment of the relationship between DNA methylation and gene expression in t(1;11) individuals.

It has yet to be determined whether the above blood-based methylation differences are also present in the brains of these individuals. Walton et al. (2015) reported that only 4.1% of 450k methylation array probes were strongly correlated between blood and neocortex biopsied at the point of surgical intervention for pharmaco-resistant epilepsy from the same individual. In contrast, others have reported meQTLs to be consistently detected across tissue types (Smith et al., 2014). By virtue of the nature of recombination events, it is likely that the translocation is flanked by regions of DNA that are shared more frequently between translocation carriers than between non-carriers. As such, it is possible that certain translocation-associated effects on methylation might be conveyed indirectly by co-inherited genetic sequence. Should the differences in methylation observed on chromosomes 1 and 11 in the t(1;11) carrier vs. non-carrier comparison reflect the effects of variants in linkage disequilibrium with the translocation, these methylation differences might consistently be observed across tissue types.

The results presented in this chapter provide evidence for disrupted DNA methylation in the context of a translocation linked to psychiatric illness. However, further experiments will be required to determine whether the differences observed here are also present in the brains of the individuals profiled, and whether such differences might lead to altered gene expression and increased risk of illness.

Chapter 5

Analysis of DNA methylation in t(1;11) family iPSC-derived neurons

5 DNA methylation analysis of t(1;11) family iPSC-derived neurons

5.1 Background and motivation

Until recently, a major limitation in psychiatric research has been the lack of reliable cellular models for illness. One option is *post-mortem* tissue. However, several confounding factors can affect studies of *post-mortem* brain tissue including post-mortem interval, cause of death and age of death (McCullumsmith *et al.*, 2014). Furthermore, it is not possible to perform functional assays on post-mortem tissue. Animal models go some way to address these issues, but they cannot fully recapitulate the genetic architecture or the phenotypic complexity underlying these human-specific disorders. The last decade has seen the development of induced pluripotent stem cell (iPSC) technology (Takahashi *et al.*, 2007). These cells have the potential to differentiate into any cell type, which has permitted the development of specialised *in-vitro* human disease models using relevant cellular systems (reviewed by Ebert *et al.*, 2013).

Chapter 4 describes the assessment of genome-wide DNA methylation in whole-blood derived DNA from the t(1;11) family using the Illumina Infinium HumanMethylation450 BeadChip (450k array; Illumina, San Diego, CA; Bibikova *et al.,* 2011). The 450k array was ceased as of December 2015 and has since been superseded by the Illumina Infinium HumanMethylationEPIC BeadChip (EPIC array). The EPIC array uses the same chemistry as its predecessor to detect methylation using a combination of Type I probes, which detect methylated and unmethylated loci using individual assays; and Type II probes, which detect both methylated and unmethylated loci with a single assay. Both probe types rely on single base extension of a fluorescently labelled nucleotide to generate signal. Containing 866,836 probes, the EPIC platform interrogates methylation at approximately twice as many loci as the 450k array. The EPIC array includes approximately 90% of the 450k array's probes, thus allowing for high inter-platform reproducibility.

The data described in this chapter were obtained using the EPIC array, using a cellular model more physiologically relevant to psychiatric illness than whole blood: iPSC-

derived cortical neurons from the t(1;11) family. These neurons were differentiated from neuronal precursor cells (NPCs) as described in section 2.11. Briefly, NPCs from three t(1;11) carriers and three non-carriers were grown to confluency across two six-well plates. This was followed by a 5 week differentiation protocol after which DNA, RNA and protein were harvested. This was performed three times for each of the six individuals, resulting in three technical replicates per individual (n = 18 samples). This portion of the work was performed over approximately ten months by Susan Anderson, Helen Torrance, Dr. Kirsty Millar and myself. Expression of the neuronal marker βIII-tubulin was confirmed in these samples by immunocytochemistry experiments performed by Dr. Kirsty Millar. There were two aims to this work. Firstly, to investigate whether the translocation was associated with differential DNA methylation in these neurons, and secondly, to permit the comparison of methylation profiles observed in blood and iPSC-derived neurons. The work described below was based on the hypothesis that *DISC1* would display differential methylation in t(1;11) neurons as previously observed in whole blood (Chapter 4).

5.2 Identification of sub-optimal probes on the Infinium HumanMethylationEPIC BeadChip

At the time of writing this thesis, few resources were available for the analysis of data generated on the EPIC array. Quality control and analysis strategies performed on the 450k array are largely transferable to this platform, due to the similarity of the two arrays. However, while Chen *et al.* (2013) had previously described probes with the potential for non-specific binding and probes targeting polymorphic sites on the 450k array, such issues had yet to be addressed on the EPIC array.

Prior to the analysis of DNA methylation in iPSC-derived neurons, a list of EPIC array probes with non-specific binding potential was generated by following the protocol described by Chen *et al.* (2013), described in section 2.10.2. Additionally, to identify probes targeting polymorphic loci, the coordinates of signal-generating sites (i.e. target cytosines, guanines and sites of single base extension) were cross-referenced to the

current release of the 1000 Genomes Project by Stewart Morris (Phase 3; 1000 Genomes Project Consortium, 2015). This resulted in a list of probes with potentially polymorphic targets.

5.2.1 Infinium MethylationEPIC BeadChip probes with cross-hybridisation potential

Chen *et al.* (2013) defined cross-hybridising probes as those with ≥ 47 nucleotide off-target matches that included the end base. This was based on an observed enrichment of such autosome-targeting probes which cross-hybridised to sites on the sex chromosomes, resulting in spurious sex-associated signals (Chen *et al.*, 2013). To identify potentially cross-hybridising probes on the EPIC array, all probe sequences were aligned *in-silico* to eight bisulphite converted genomes (hg19 sequence), each representing forward methylated, reverse methylated, forward unmethylated, reverse unmethylated genomes, and their complements. In total, 44,210 potentially cross-hybridising probes, meeting the criteria of ≥ 47 nucleotide off-target matches including the end base were identified. These comprised 11,772 Type I probes and 32,438 Type II probes.

Consistent with findings on the 450k array (Chen *et al.*, 2013), a larger proportion of non-CpG-targeting (CpH) probes (Probe ID prefix = "ch") were identified as potentially cross-hybridising compared with CpG-targeting probes (Probe ID prefix = "cg"). Of 863,904 CpG-targeting probes present on the array, 42,558 (5% of the total CpG-targeting probes) were identified as potentially cross-hybridising. In contrast, of 2,932 CpH-targeting probes, 1,652 were found to be potentially cross-hybridising (56% of the total non-CpG-targeting probes).

5.2.2 Infinium MethylationEPIC BeadChip probes with polymorphic targets

Coordinates for the 866,836 probes on the EPIC array were obtained from the array's manifest downloaded on the 8th February 2016. Excluding control probes, the array contains 142,262 Type I probes (426,786 potential signal-affecting positions), and

724,574 Type II probes (1,449,148 potential signal-affecting positions), giving a total of 1,875,934 sites. This list of coordinates was cross-referenced to 1000 genomes data (phase 3; 1000 Genomes Project Consortium, 2015).

In total, 340,327 sites were identified with either single nucleotide polymorphisms (SNPs), or insertions or deletions (indels). These sites were targeted by 297,744 unique probes: 34% of the total probe content of the array. Of these, 23,399 probes (2.7% of the total probe content) targeted polymorphic sites with a minor allele frequency (MAF) of ≥ 5% in at least one population studied. A table of probes affected by polymorphisms, with minor allele frequencies corresponding to African, admixed American, European, South Asian, and East Asian populations (AFR, AMR, EUR, SAS, EAS; respectively) was compiled as a resource for others to perform population-specific probe filtering.

These lists of potentially cross-hybridising and polymorphism-targeting probes were submitted as supplementary information to a publication on these findings, intended to be used as a resource for others when analysing data derived from the EPIC array (McCartney *et al.*, 2016; Appendix 1).

5.3 Sample information

DNA from iPSC-derived cortical neurons of three t(1;11) carriers and three non-carriers was assessed for differential methylation on the EPIC array. This study consisted of four females and two males. The relationship between gender and t(1;11) status was assessed by a Fisher's exact test. A significant relationship was observed between sample carrier status and gender (Fisher's exact test $p = 0.001$). This was as expected due to sample availability: the non-carrier group consisted exclusively of females, with two males present among the three t(1;11) carriers. Eighteen samples (i.e. three independent neuronal differentiations from each individual) were distributed across three slides, with each slide containing one technical replicate per individual.

5.3.1 Assessment of Intra-Individual Correlation

In order to identify outliers among technical replicates which might confound the data, correlation of raw methylation data (i.e. pre-normalisation) was assessed for all pairwise combinations of each triplicate within each individual. Pearson's correlation coefficients > 0.98 were observed between technical replicates, indicative of high intra-individual reproducibility (Table 5.1A-F). Cluster analysis was performed across all samples on variable probes with a coefficient of variation (CV) > 0.2 (n = 327,588). Significant clusters were observed corresponding to technical replicates, gender and an unknown factor separating one sample (individual A) from the remaining individuals (Figure 5.1).

A

Sample	A1	A2	A3
A1	1	0.992	0.982
A2	0.992	1	0.986
A3	0.982	0.986	1

B

Sample	B1	B2	B3
B1	1	0.995	0.995
B2	0.995	1	0.995
B3	0.995	0.995	1

C

Sample	C1	C2	C3
C1	1	0.988	0.981
C2	0.988	1	0.987
C3	0.981	0.987	1

D

Sample	D1	D2	D3
D1	1	0.994	0.992
D2	0.994	1	0.995
D3	0.992	0.995	1

E

Sample	E1	E2	E3
E1	1	0.991	0.991
E2	0.991	1	0.995
E3	0.991	0.995	1

F

Sample	F1	F2	F3
F1	1	0.994	0.994
F2	0.994	1	0.992
F3	0.994	0.992	1

Table 5.1: Correlation coefficients between technical replicates for six t(1;11) family members profiled for DNA methylation in iPSC-derived neurons.

Each table presents Pearson's coefficients of determination (R^2) between pairs of DNA methylation profiles from an individual profiled for DNA methylation in triplicate. Tables correspond to individuals coded A-F, respectively. Sample IDs are suffixed with numbers 1-3 corresponding to neuronal differentiations performed in triplicate for each individual.

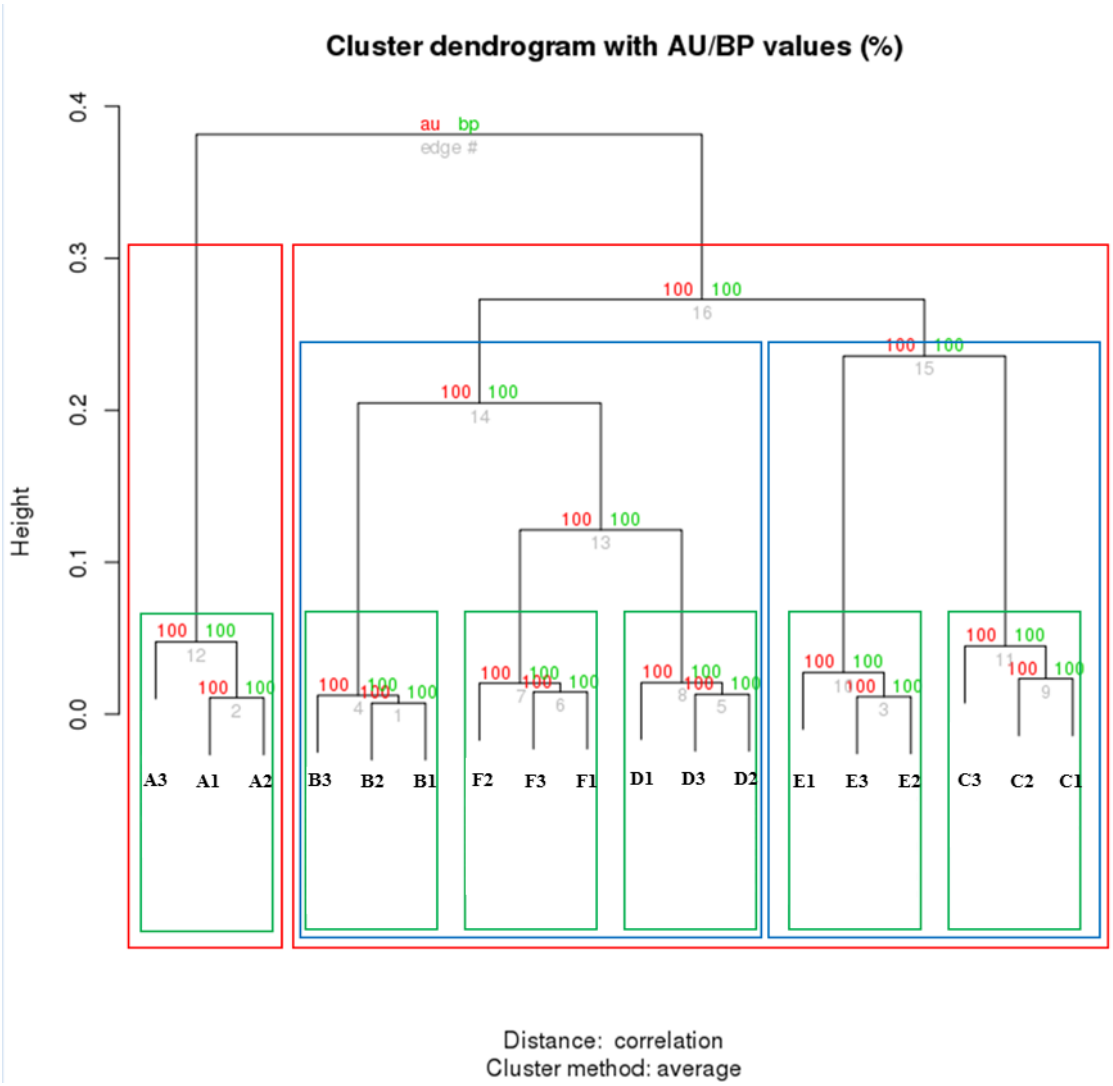


Figure 5.1: Hierarchical cluster analysis of raw (pre-normalised) methylation data from 18 iPSC-derived neuronal samples.

Shown is a dendrogram assessing the relationship between 18 iPSC-derived neuronal samples from six individuals (A-F) based on the raw methylation data. The height of the dendrogram (y-axis) represents sample variability. Red values at each node correspond to the approximately unbiased (au) p-*value and green values correspond to the bootstrap probability (bp) percentage* p*–values. au values > 95 indicate clusters supported by the data. Grey numbers at each node indicate the edge number. Green boxes highlight the technical replicate clusters (3 per individual), blue boxes highlight clustering by gender and red boxes highlight an unknown cluster separating individual A from the remaining samples.*

5.4 Preprocessing of the data generated by analysis of iPSC-derived neuronal samples

Prior to analysis of differential methylation, quality control procedures including probe filtering and normalisation were performed on the data. Probe filtering was performed to ensure only correctly functioning probes were assessed for differential methylation, thus reducing both the likelihood of type I error and the multiple testing burden. Normalisation was performed to remove systematic sources of variation such as array position and signal bias relating to probe type (Teschendorff *et al.*, 2013). Raw (i.e. non-normalised) methylation data was read into R with a starting probe count of 866,836. Probes with predicted cross-hybridisation potential identified in section 5.2 were removed (n = 44,210). Whole-genome sequence data from the family were cross-referenced to probe coordinates to identify sites containing genetic variation at target CpGs, and sites of single base extension, which were also removed (n = 10,107).

Using the filtering criteria described in section 2.10.3, 13,667 probes were removed due to having a beadcount of < 3 in 5% of samples, and 2026 probes based on a detection *p*–value > 0.05 in > 1% of samples. None of the samples met these exclusion criteria. Following this filtering step, 810,344 probes measuring methylation 18 samples remained.

5.5 Selection of normalisation strategy

In order to determine the optimum normalisation method for this dataset, fourteen normalisation methods were compared based on their performance at reducing technical variation using three metrics: DMRSE, GCOSE and Seabird (Pidsley *et al.*, 2013; Schalkwyk *et al*., 2013). The top-ranking method was danet (Table 5.2). This method equalises background from type I and type II probes, performs quantile normalisation of methylated and unmethylated intensities together, and then calculates normalised methylation β-values. As there was a significant relationship between t(1;11) carrier status and gender, probes targeting the sex chromosomes were removed prior to analysis

(n = 17,920). The 59 “rs”-prefixed control probes were also removed at this stage leaving 781,000 probes in the final dataset.

To determine whether these steps removed the clustering effects identified in Chapter 6.3.1, hierarchical cluster analysis was performed again on the normalised methylation data on probes with a CV ≥ 0.2 (n = 143,852). Six clusters were observed corresponding to each individual (Figure 5.2). Data normalisation, along with the removal of sex chromosome and control probes resulted in the elimination of the significant gender-associated cluster, and the cluster separating one individual (individual A) from the rest of the individuals.

Normalisation Method	DMRSE Rank	GCOSE Rank	Seabird Rank	Rank Mean	Rank of Rank Mean
danet	3	6	2	3.66	1
nanet	4	7	1	4	2.5
daten1	5	1	6	4	2.5
nanes	2	8.5	3	4.5	4.5
danes	1	8.5	4	4.5	4.5
naten	7	3	5	5	6.5
daten2	6	2	7	5	6.5
dasen	8	4.5	9	7.16	8.5
nasen	9	4.5	8	7.16	8.5
SWAN	10	10	12	10.66	10
BMIQ	11	14	11	12	11
Noob	14	12.5	10	12.16	12
PBC	12	11	15	12.66	13
Raw	13	12.5	13	12.83	14
danen	15	15	14	14.66	15

Table 5.2: Performance of fourteen normalisation strategies in neuronal DNA methylation samples.

Table summarises the rankings of three metrics used to assess the performance of 14 normalisation strategies applied to methylation data, along with non-normalised data (Raw) from 18 iPSC-derived neuronal samples (i.e. three replicates from six individuals). In order of appearance, columns show the normalisation strategy, DMR-standard error (DMRSE) rank, genotype-combined standard error (GCOSE) rank, 1 - Seabird AUC rank, mean rank of the three metrics, and the rank of mean ranks.

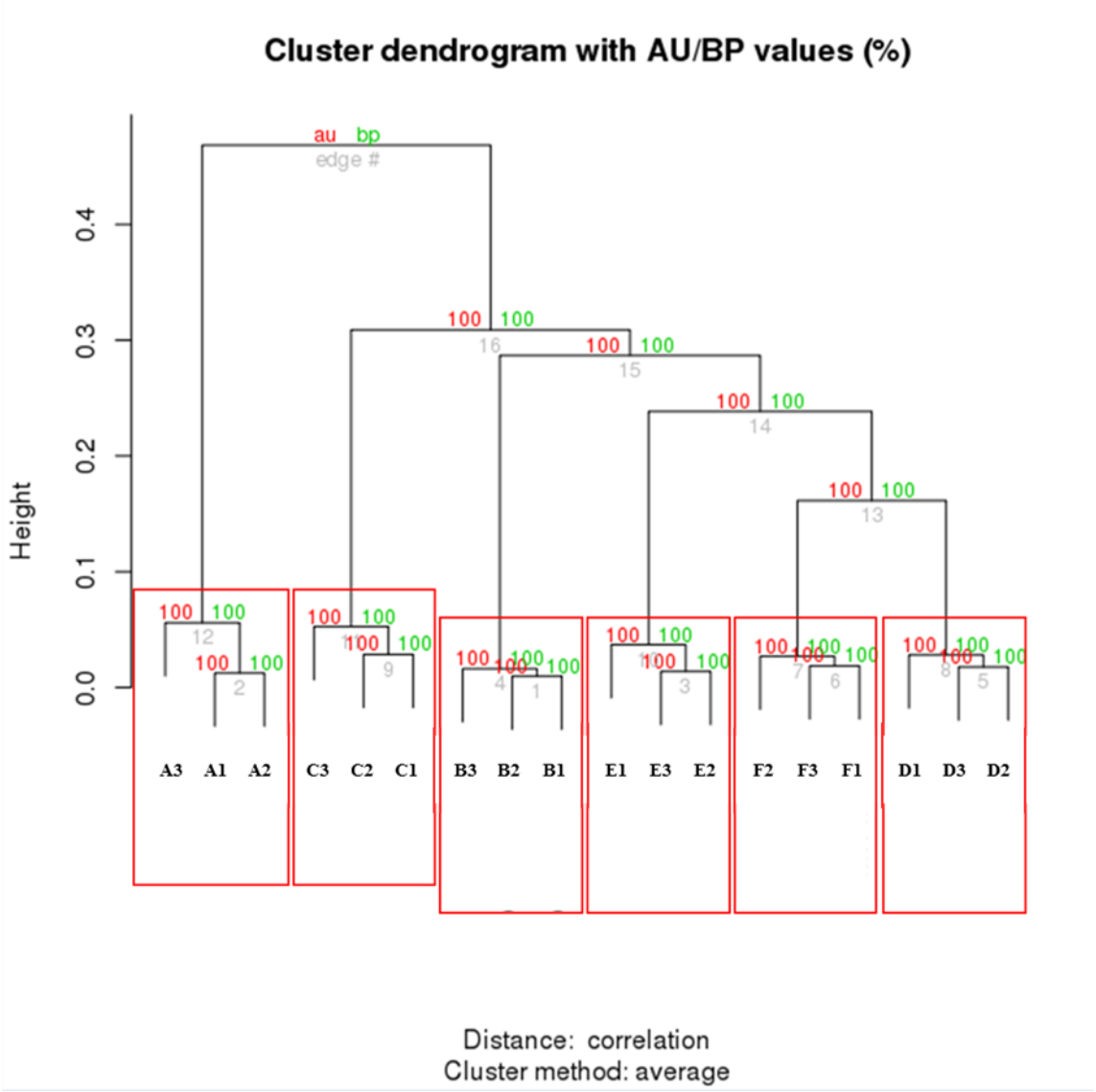


Figure 5.2: Hierarchical cluster analysis of danet-normalised methylation data from 18 iPSC-derived neuronal samples.

A dendrogram is presented assessing the relationship between 18 iPSC-derived neuronal samples based on the danet-normalised DNA methylation data after removing sex-chromosome probes and control probes. The height of the dendrogram (y-axis) represents sample variability. Red values at each node correspond to the approximately unbiased (au) p*-value and green values correspond to the bootstrap probability (bp) percentage* p*–values. au values > 95 indicate clusters supported by the data. Grey numbers at each node indicate the edge number. Red boxes highlight clustering by triplicates from each individual (A-F).*

5.6 Identification of differentially methylated positions in iPSC-derived neurons

In order to identify sites showing differential methylation associated with the t(1;11) translocation, linear regression was performed. For each set of technical replicates, the mean methylation β-value was assessed, comparing six data points. Carriers of the translocation (n = 3) were compared to non-carrying individuals (n = 3), covarying for gender and mean passage number of NPCs per individual at the point of neuronal differentiation. Surrogate variable analysis (SVA) was performed to identify potential latent sources of variation present in the data. No significant surrogate variables were identified. No DMPs were identified following correction for multiple testing (FDR $q > 0.05$ for all probes; Figure 5.3). A quantile-quantile plot and calculation of the genomic inflation factor (λ) indicated deflation of *p*-values ($\lambda = 0.84$; Figure 5.4).

The most significant differential methylation signal was observed within a 443 bp intergenic region of chromosome 5. This region contained three probes which were hypomethylated in t(1;11) carriers. The top 10 sites ranked by *p*-value for differential methylation are presented in Table 5.3. Large differences in average methylation were observed between groups at all of these sites. Beeswarm plots were drawn to visualise the distribution of methylation at these sites (Figure 5.5A-J). At nine of these sites, clustering was observed around a β-value of 0.2 for all individuals the exception of a t(1;11) carrier, who displayed consistently higher methylation levels ($\beta > 0.5$).

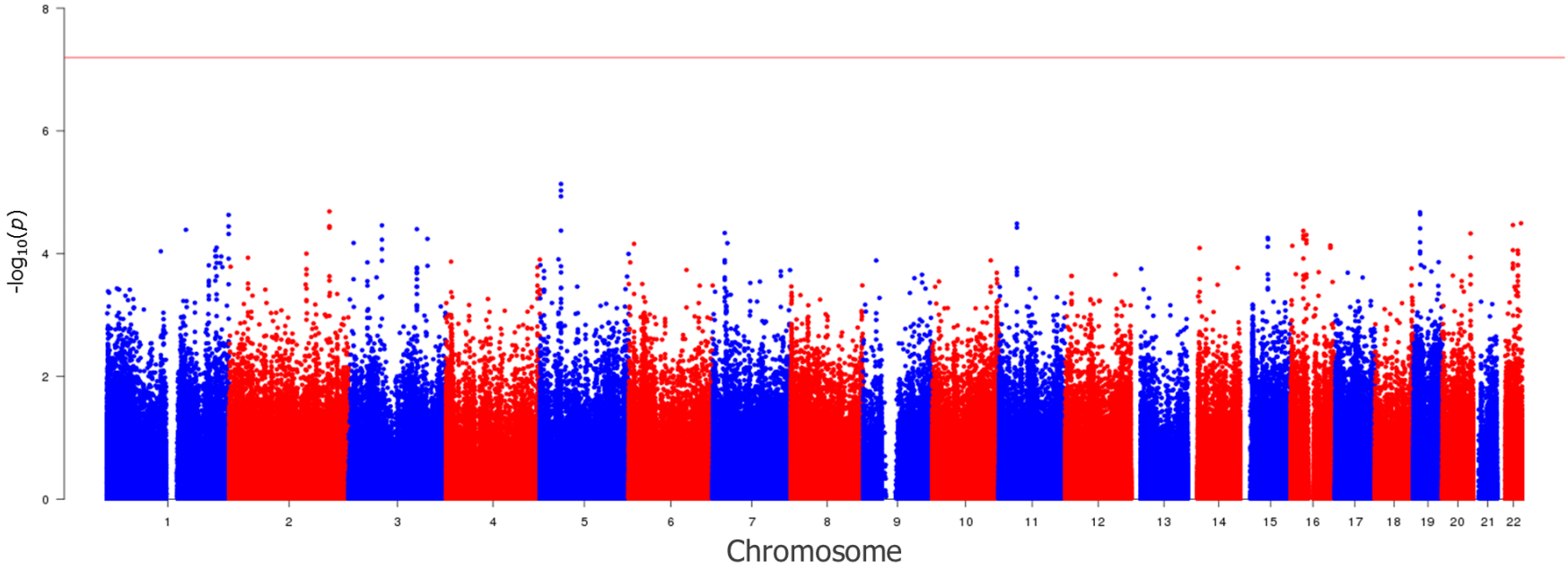


Figure 5.3: Manhattan plot of sites assessed for DNA methylation in iPSC-derived neurons from t(1;11) carriers and non-carriers.

Manhattan plot of the –log$_{10}$ p–value for differential methylation between t(1;11) carriers and non-carriers (y–axis) against chromosome and position (x–axis). The horizontal red line represents the –log$_{10}$ p*–value threshold for genome wide significance corresponding to a false discovery rate of 5%.*

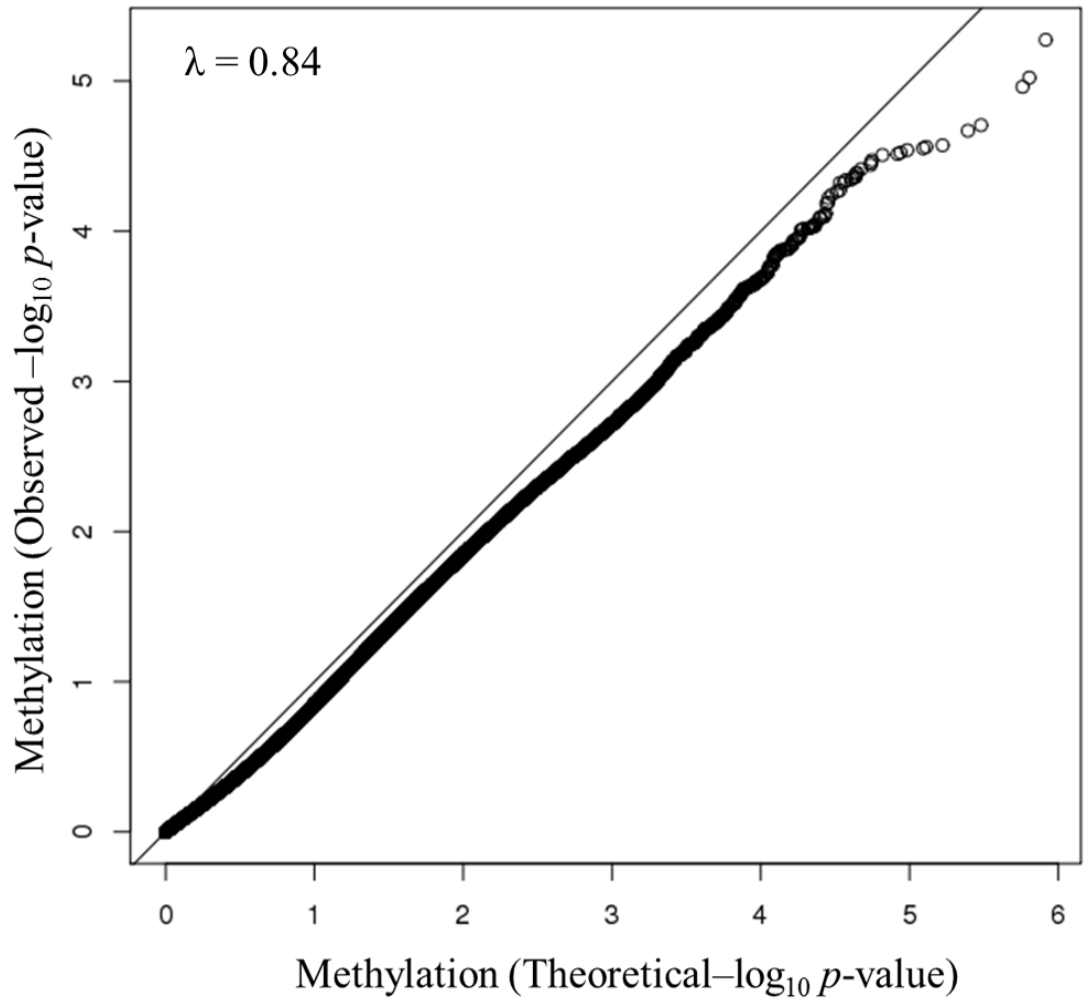


Figure 5.4: Quantile-quantile plot showing observed vs expected p-values for differential methylation between t(1;11) carriers and non-carriers.

Observed unadjusted $-\log_{10}$ p-values (y-axis; circular points) are plotted against the expected distribution of $-\log_{10}$ p-values under the null hypothesis (x-axis; solid diagonal line). The genomic inflation factor λ is presented in the upper-left corner of the plot.

Probe ID	Hg19 Coordinates	Gene	Beta Difference	Fold-Change	t	DMP *p-value*	DMP *q* - value
cg07137955	Chr5:42992774	NA	0.313916337	7.19	36.06	7.31×10^{-6}	1
cg21889472	Chr5:42992555	NA	0.259492632	6.78	33.74	9.35×10^{-6}	1
cg03894174	Chr5:42992998	NA	0.247769856	5.25	31.79	1.16×10^{-5}	1
cg12356111	Chr2:203879443	*NBEAL1*	0.185736306	5.15	27.30	2.05×10^{-5}	1
cg26171523	Chr19:13858605	*CCDC130*	0.178365793	6.17	27.06	2.11×10^{-5}	1
cg24153071	Chr19:13858573	*CCDC130*	0.19816271	4.62	26.5	2.29×10^{-5}	1
cg20146541	Chr1:248020697	*TRIM58*	0.583503247	12.92	26.35	2.33×10^{-5}	1
cg09209679	Chr22:49051077	*FAM19A5*;*FAM19A5*	0.227795233	2.35	24.21	3.19×10^{-5}	1
cg11631775	Chr11:36616251	*C11orf74*;*RAG2*	0.169545168	3.76	24.12	3.23×10^{-5}	1
cg26494441	Chr22:32601101	*RFPL2*	0.171565868	3.77	23.77	3.41×10^{-5}	1

Table 5.3: Top ten nominally significant differentially methylated positions identified (ranked by *p*-value) in the comparison of DNA methylation profiles from iPSC-derived neurons in t(1;11) carriers and non-carriers.

Shown are nominally significantly differentially methylated sites between t(1;11) carriers and non-carriers in iPSC-derived neuronal DNA ($p \leq 0.05$*). In order of column appearance are probe identifiers, Hg19 genomic coordinates, UCSC reference gene names ("NA" denotes intergenic regions), between-group difference in mean beta-value, fold-change in methylation between groups, moderated* t*-statistic,* p*-value for differential methylation and FDR-adjusted* p*-value.*

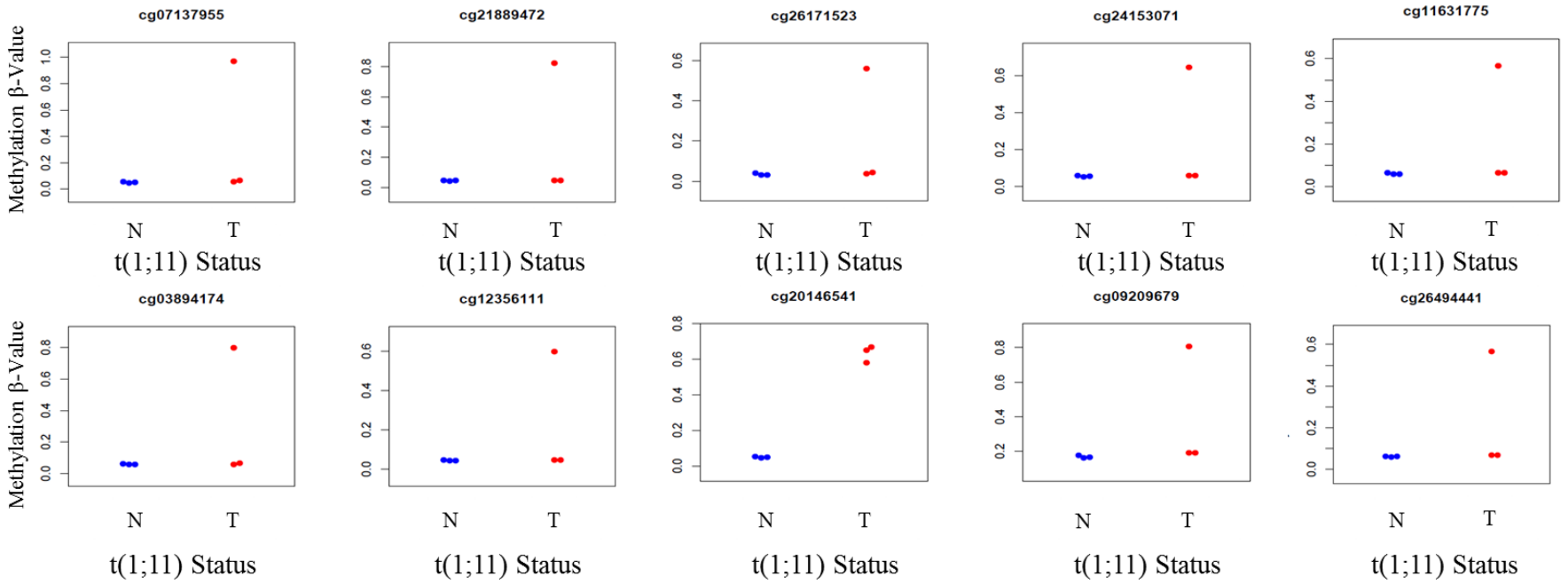


Figure 5.5: Beeswarm plots of top 10 differentially methylated sites in iPSC-derived t(1;11) samples.

Shown are methylation β-values for each sample (y-axis) plotted against t(1;11) status (x-axis) for the top 10 differentially methylated sites between t(1;11) carriers and non-carriers. Non-carriers of the translocation are denoted "N" while carriers are denoted "T". Mean methylation β-values from triplicates corresponding to each individual are represented by a point on each plot. Points are colour-coded blue for translocation non-carriers and red for t(1;11) carriers.

5.7 Gene ontology analysis of differentially methylated positions in iPSC-derived neurons

To identify any biological processes, molecular functions or cellular components that were significantly over-represented amongst the genes showing the most significant differences in methylation in t(1;11) carriers, gene ontology analysis was performed. A ranked list of *p*–values for differential methylation between t(1;11) carriers and non-carriers was submitted to GOrilla for analysis. A total of 194 terms were significantly over-represented amongst the most differentially methylated genes ($q \leq 0.05$; Table 5.4). The most significantly over-represented term was "plasma membrane part" (GO:0044459; $q = 3.57 \times 10^{-12}$), present in the GO components class. The most significantly over-represented GO process was "organ morphogenesis" (GO:0009887; $q = 4.72 \times 10^{-11}$) while the most significantly over-represented GO function was "sequence-specific DNA binding" (GO:0043565; $q = 9.02 \times 10^{-5}$). Multiple terms relating to neuronal function and neurodevelopment were enriched amongst the differentially methylated genes in t(1;11) carriers, including "glutamate receptor activity" (GO:0008066; $q = 0.0001$), "modulation of synaptic transmission" (GO:0050804; $q = 9.36 \times 10^{-7}$), "neuron part" (7.11×10^{-6}) and "central nervous system development" (GO:0007417; $q = 0.025$).

GO Term	Description	Class	Enrichment	P-value	FDR q-value
GO:0044459	Plasma membrane part	Component	433/2385	2.09×10^{-15}	3.57×10^{-12}
GO:0009887	Organ morphogenesis	Process	114/424	3.32×10^{-15}	4.72×10^{-11}
GO:0048856	Anatomical structure development	Process	516/2969	3.81×10^{-14}	2.70×10^{-10}
GO:0032502	Developmental process	Process	654/3896	3.86×10^{-13}	1.82×10^{-9}
GO:0005886	Plasma membrane	Component	596/3766	2.49×10^{-11}	2.12×10^{-8}
GO:0048731	System development	Process	141/629	7.86×10^{-12}	2.79×10^{-8}
GO:0009653	Anatomical structure morphogenesis	Process	245/1229	1.61×10^{-11}	4.58×10^{-8}
GO:0032501	Multicellular organismal process	Process	518/3067	1.37×10^{-10}	3.24×10^{-7}
GO:0044767	Single-organism developmental process	Process	567/3416	2.31×10^{-10}	4.69×10^{-7}
GO:0048562	Embryonic organ morphogenesis	Process	41/115	3.37×10^{-10}	5.99×10^{-7}

Table 5.4: Summary of GO terms found to be enriched amongst the most differentially methylated genes in iPSC-derived neurons of t(1;11) carriers.

For each GO term, table summarises the GO identifier, the GO description, the GO class, enrichment, the enrichment p-values, and enrichment FDR q-values for genes showing the most differential methylation in t(1,11) carriers compared to non-carriers. Enrichment is defined as (b/n) / (B/N), where b = the number of genes at the top of the p-value ranked list that is associated with a given GO term, n = the number of genes at the top of the p-*value-ranked gene list, B = the total number of genes associated with a given GO term, and N = the total number of genes. Shown are the top 10 enrichments ranked by* p-*value. The total list of significant enrichments is presented in Appendix I (Table A6).*

5.8 Identification of differentially methylated regions in iPSC-derived neurons

In order to identify discrete genomic regions containing multiple nominally significantly differentially methylated probes ($p \leq 0.05$) in t(1;11) carriers, differentially methylated region (DMR) analysis was performed. DMRs were assessed using the *ChAMP* package's probe lasso algorithm (Butcher and Beck, 2015). A total of 424 nominally significant DMRs ($p \leq 0.05$) were identified between t(1;11) carriers and non-carriers. The most significant DMR in t(1;11) carriers overlapped two genes, mapping to *IER3*, and *FLOT1*, which is located antisense to *IER3* ($p = 7.72 \times 10^{-29}$). This region is within the major histocompatibility complex (MHC). The nearest DMR to the chromosome 1 translocation breakpoint is located approximately 12 Mb telomeric in an intergenic region (chr1: 244395115-244395352; $p = 0.0002$). On chromosome 11, the nearest DMR to the

breakpoint is approximately 21 Mb centromeric, also within an intergenic region (Chr11:69286177-69286467; $p = 0.0002$).

The largest DMR identified fell within the chromosome 10 gene *MGMT*, spanning 8.9 kb and extending from the gene body to an intergenic region downstream of the gene. A further nine DMRs were identified within *MGMT*. Along with *MGMT*, 13 additional genes were identified containing multiple DMRs. Of these, *SORCS1*, a member of the Sortilin gene family, contained two DMRs ($p = 1.57 \times 10^{-5}$ and $p = 0.0004$). Four DMRs were identified within *DLGAP2* ($p = 4.26 \times 10^{-5} – 0.0118$). A single DMR was present in *SV2B*: a gene found to be differentially expressed in t(1;11) carrier lymphoblastoids in Chapter 3 (DMR $p = 0.0036$). Seven DMRs were within genes found to contain DMRs in the blood-based methylation analysis (Chapter 4).

Gene	Features	Region	No. Probes	DMR *p*
FLOT1, *IER3*	TSS1500, 3'UTR, Body	Chr6:30710912-30711968	26	7.72×10^{-29}
GNASAS, *GNAS*	3'UTR, TSS1500	Chr20:57425870-57427652	51	1.96×10^{-20}
CCDC130	TSS1500, TSS200	Chr19:13858480-13858585	5	1.46×10^{-15}
SNORD116-15 - SNORD116-19	TSS1500, TSS200, Body	Chr15:25325510-25330514	16	1.46×10^{-15}
C22orf32, *SMDT1*	TSS200, 1stExon	Chr22:42475680-42475844	6	4.10×10^{-15}
MIR4458HG, *LOC729506*	Body	Chr5:8457127-8457933	7	1.89×10^{-14}
DHRS4, *C14orf167*	Body, TSS200	Chr14:24422520-24423061	7	1.62×10^{-13}
C13orf38, *CCDC169*	5'UTR, 1stExon, TSS200	Chr13:36871465-36872189	10	6.16×10^{-13}
GDNF	Body, 5'UTR	Chr5:37834742-37835348	9	1.18×10^{-12}
RFPL2	TSS1500	Chr22:32601040-32601418	4	2.08×10^{-12}

Table 5.5: DMRs identified between t(1;11) carriers and non-carriers in iPSC-derived neurons.

From left to right, columns summarise the DMR-containing genes, the DMR's underlying genomic features, the Hg19 genomic coordinates of each DMR, the number of probes within each DMR, and the p*-value for differential methylation in t(1;11) carriers. Genes highlighted with a red asterisk (*) indicate those that also contained t(1;11)-associated DMRs identified in blood. Shown are the top 10 DMRs ranked by* p*-value. The total list of significant DMRs is presented in Appendix I (Table A7).*

5.9 Comparison with results obtained from the profiling of DNA methylation in blood samples from t(1;11) carriers and non-carriers

In order to compare the neuronal results to the t(1;11)-associated findings in blood (Chapter 4), a direct assessment of DNA methylation in neurons was carried out at the 13 differentially methylated probes identified in blood ($q \leq 0.05$; Table 4.3). Of these 13 probes, one showed nominally significant differential methylation between t(1;11) carriers and non-carriers in neurons (cg26728851; $p = 0.017$; Table 5.6). This probe mapped to the 3'UTR of *GUCY2E*, on chromosome 11 and was hypomethylated in t(1;11) carriers in both blood and iPSC-derived neurons.

Probe ID	Hg19 Coordinates	Gene	Blood			iPSC-derived Neurons						
			Average t(1;11) non-carrier beta value	Average t(1;11) carrier beta value	Beta Difference	Average t(1;11) non-carrier beta value	Average t(1;11) carrier beta value	Beta Difference	Fold-Change	t	*p-value*	*q* - value
cg26728851	Chr11:76430375	*GUCY2E*	0.93	0.90	-0.07	0.94	0.87	-0.07	-1.07	-4.05	0.02	1.00
cg26355502	Chr1:221916303	*DUSP10*	0.07	0.06	-0.01	0.08	0.06	-0.02	-1.27	-2.22	0.10	1.00
cg02771260	Chr11:59836817	*MS4A3*	0.58	0.45	-0.13	0.50	0.32	-0.18	-1.58	-1.60	0.19	1.00
cg24508974	Chr10:103330391	NA	0.12	0.14	0.02	0.25	0.22	-0.03	-1.14	-1.50	0.21	1.00
cg25899154	Chr11:72897143	NA	0.57	0.50	-0.07	0.75	0.74	-0.01	-1.01	-1.32	0.26	1.00
cg15157974	Chr1:232144702	*DISC1*; *TSNAX-DISC1*	0.82	0.78	-0.04	0.64	0.63	-0.02	-1.03	0.79	0.48	1.00
cg05656812	Chr1:232021560	*DISC1*; *TSNAX-DISC1*	0.79	0.73	-0.06	0.76	0.72	-0.03	-1.05	0.31	0.77	1.00
cg16177633	Chr1:232172585	*DISC1*; *TSNAX-DISC1*	0.81	0.78	-0.03	0.50	0.43	-0.07	-1.16	-0.28	0.79	1.00
cg18815120	Chr1:231512676	*EGLN1*	0.34	0.22	-0.12	0.81	0.81	0.00	1.01	-0.16	0.88	1.00
cg06928246	Chr1:227974645	NA	0.66	0.59	-0.07	0.76	0.67	-0.09	-1.13	0.15	0.89	1.00
cg21875980	Chr1:231553510	*EGLN1*	0.42	0.48	0.06	0.43	0.39	-0.04	-1.10	-0.13	0.91	1.00
cg00965168	Chr1:227974541	NA	0.68	0.63	-0.05	0.88	0.82	-0.06	-1.06	0.09	0.94	1.00
cg09186051	Chr1:231981906	*DISC1*; *TSNAX-DISC1*	0.74	0.67	-0.07	0.58	0.61	0.03	1.06	-0.05	0.96	1.00

Table 5.6: DNA methylation in iPSC-derived neurons at 13 probes found to show significant differential methylation in a comparison of blood DNA from t(1;11) carriers and non-carriers.

Summary of DNA methylation in iPSC-derivatives at the thirteen probes significantly differentially methylated in blood. Columns show the probe identifiers, Hg19 genomic coordinates, UCSC reference gene names ("NA" denotes intergenic regions). From the methylation analysis in blood, shown are mean methylation β-values in t(1;11) carriers, mean methylation β-values in t(1;11) non-carriers, and between-group difference in mean methylation β-values are presented. From the methylation analysis in iPSC-derived neurons, shown are mean methylation β-values in t(1;11) carriers, mean

methylation β-values in t(1;11) non-carriers, between-group difference in mean methylation β-values, fold-change between groups, moderated t-*statistic,* p-*value for differential methylation and FDR-adjusted* p-*value (*q-*value).*

5.10 DNA Methylation in six blood samples corresponding to iPSC-derived neurons

To determine whether the results in neurons might be more comparable to those in blood given the same set of individuals, the blood-based dataset described in Chapter 5 was subsetted to contain the same six individuals for whom neuronal DNA methylation was measured. As gender and t(1;11) status were significantly correlated in these individuals (section 5.3, Fisher's exact test $p = 0.001$), sex chromosome probes were removed. Surrogate variable analysis was performed on this dataset with no significant surrogate variables identified. Linear regression was performed to assess differential methylation in t(1;11) carriers, covarying for age and gender.

Six sites were significantly differentially methylated in the comparison of blood DNA from three t(1;11) carriers and three non-carriers ($q \leq 0.05$; Figure 5.6; Table 5.7). None of these six sites were among the 13 found to show differential methylation in the complete blood DNA sample set (n = 41 individuals). However, in the six blood samples, nine of these 13 sites showed nominally significant differential methylation between t(1;11) carriers and non-carriers, all of which showed the same direction of effect as in the total sample set ($p \leq 0.05$; Table 5.8).

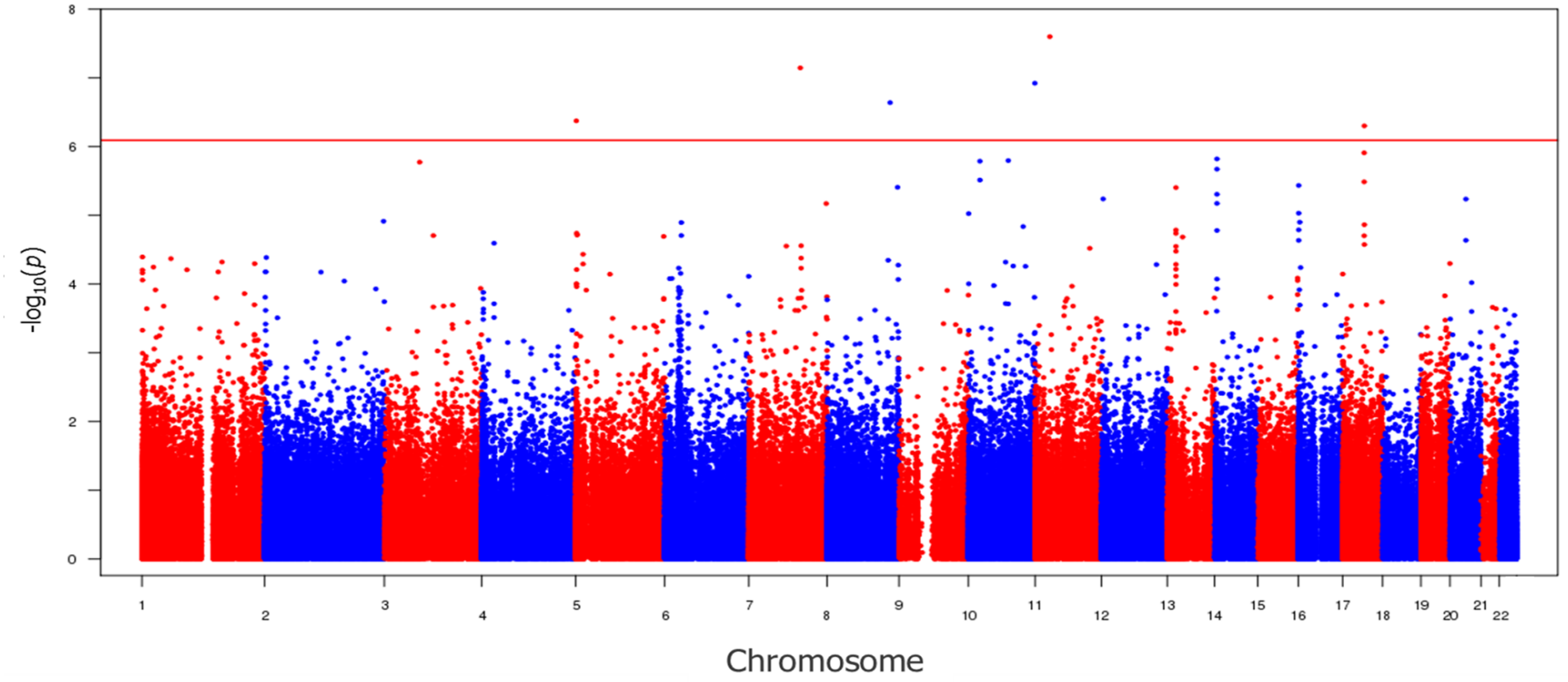


Figure 5.6: Manhattan plot of sites assessed for DNA methylation in six blood samples corresponding to iPS-derived neuronal samples.

Shown are $-\log_{10} p$ – values between t(1;11) carriers and non-carriers (y – axis) plotted against chromosomal position (x–axis). The horizontal red line represents the $-\log_{10} p$–value threshold for genome wide significance as defined by a false discovery rate of 5% (q = *0.05;* p = *8.08×10^{-7}*).

Probe ID	Hg19 Coordinates	Gene	Beta Difference	Fold-Change	t	DMP *p*-value	DMP *q*-value
cg27319123	Chr11:30038286	*KCNA4*	0.17	2.39	22.32	2.51×10^{-8}	0.011
cg06850186	Chr7:104580787	NA	-0.39	-1.72	-19.46	7.16×10^{-8}	0.015
cg24668570	Chr10:134973778	*KNDC1*	0.18	2.12	18.19	1.20×10^{-7}	0.017
cg23203918	Chr8:128235836	NA	-0.12	-1.17	-16.70	2.29×10^{-7}	0.025
cg15668967	Chr5:1180364	NA	-0.12	-1.18	-15.41	4.22×10^{-7}	0.036
cg24801230	Chr17:43978533	*MAPT*	-0.38	-1.96	-15.07	4.99×10^{-7}	0.036

Table 5.7: Table of significantly differentially methylated positions between t(1;11) carriers and non-carriers in the six blood samples corresponding to iPSC-derived neuronal samples.

Differentially methylated sites between t(1;11) carriers and non-carriers in six blood samples corresponding to iPSC-derived neuronal samples ($q \leq 0.05$*). From left to right, columns show the probe identifiers, Hg19 genomic coordinates, UCSC reference gene names ("NA" denotes intergenic regions), between-group difference in mean methylation β-values, fold-change between groups, moderated* t*-statistic,* p*-value for differential methylation, and FDR-adjusted* p*-value (*q*-value).*

Probe ID	Hg19 Coordinates	Gene	Beta Difference	Fold-Change	t	DMP *p-value*	DMP *q* - value
cg02771260	Chr11:59836817	*MS4A3*	-0.16	-1.42	-6.49	0.0002	0.60
cg21875980	Chr1:231553510	*EGLN1*	0.19	1.58	5.18	0.0009	0.97
cg26728851	Chr11:76430375	*GUCY2E*	-0.04	-1.05	-4.45	0.002	0.97
cg06928246	Chr1:227974645	NA	-0.06	-1.11	-3.08	0.016	0.97
cg15157974	Chr1:232144702	*DISC1*;*TSNAX-DISC1*	-0.05	-1.06	-2.93	0.02	0.97
cg05656812	Chr1:232021560	*DISC1;TSNAX-DISC1*	-0.07	-1.10	-2.85	0.02	0.97
cg18815120	Chr1:231512676	*EGLN1*	-0.05	-1.17	-2.83	0.02	0.97
cg26355502	Chr1:221916303	*DUSP10*	-0.02	-1.40	-2.80	0.02	0.97
cg00965168	Chr1:227974541	NA	-0.07	-1.10	-2.70	0.03	0.97
cg09186051	Chr1:231981906	*DISC1;TSNAX-DISC1*	-0.09	-1.13	-1.84	0.10	0.97
cg24508974	Chr10:103330391	NA	0.04	1.36	1.24	0.25	0.97
cg16177633	Chr1:232172585	*DISC1;TSNAX-DISC1*	-0.03	-1.04	-1.08	0.31	0.97
cg25899154	Chr11:72897143	NA	-0.008	-1.02	-0.25	0.81	0.99

Table 5.8: Summary of DNA methylation in the blood sample subset (n =6) at probes significantly differentially methylated in the blood sample superset (n =41).

Columns show probe identifiers, Hg19 genomic coordinates, UCSC reference gene names ("NA" denotes intergenic regions), between-group difference in mean methylation β-values, fold-change between groups, moderated t-*statistic,* p-*value for differential methylation and FDR-adjusted* p-*value (*q-*value) for six blood samples at 13 sites previously found to be significantly differentially methylated in the total blood sample (n = 41;* $q \leq 0.05$*).*

5.11 Correlation between blood and iPSC-derived neurons

As blood-based studies of DNA methylation in psychiatric disorders are unlikely to be entirely representative of DNA methylation in the brain, it is important to consider the correlation of methylation levels between tissues from the same individuals. Using cortical tissue biopsied from drug-resistant epilepsy patients, Walton *et al.*, (2015) reported significant correlation between DNA methylation in blood and brain at 4.1 % of the probes of the 450k array. They reported a ranked list of the top 100 probes correlated between blood and brain (Spearman's Rho $\geq$ 0.94). In order to determine whether the same correlation could be observed between whole blood and iPSC-derived neurons from the t(1;11) family, a pairwise analysis of methylation at the 397,244 probes common to both blood and neuronal analyses was performed. First, the 100 probes reported by Walton *et al.*, (2015) as correlated between blood and brain were assessed in t(1;11) individuals. Among the 100 blood-brain correlated probes reported, 65 probes were present in both the blood and neuronal datasets, due to the removal of 35 probes during quality control procedures in these analyses. For the six individuals with both blood and iPSC-derived neuronal methylation data, Spearman's correlation coefficients were calculated for methylation at these probes. These between-tissue correlation coefficients were compared with those by Walton *et al.* (2015) using a paired *t*-test. The mean difference between the correlation coefficients in the t(1;11) samples and the correlation coefficients presented by Walton *et al.*, (2015) was significantly different from zero (paired sample *t*-test $p < 2.2 \times 10^{-16}$; Figure 5.7).

In order to assess whether the 13 DMPs identified in the t(1;11) blood-based analysis were correlated with iPSC-derived neurons, correlation coefficients were assessed at these sites. Of the 13 DMPs, the highest correlation was observed at cg26728851, mapping to *GUCY2E* on chromosome 11 (Spearman's Rho = 0.91). This site showed nominally significant differential methylation ($p < 0.05$) in iPSC-derived neurons as well as the corresponding subset of blood samples.

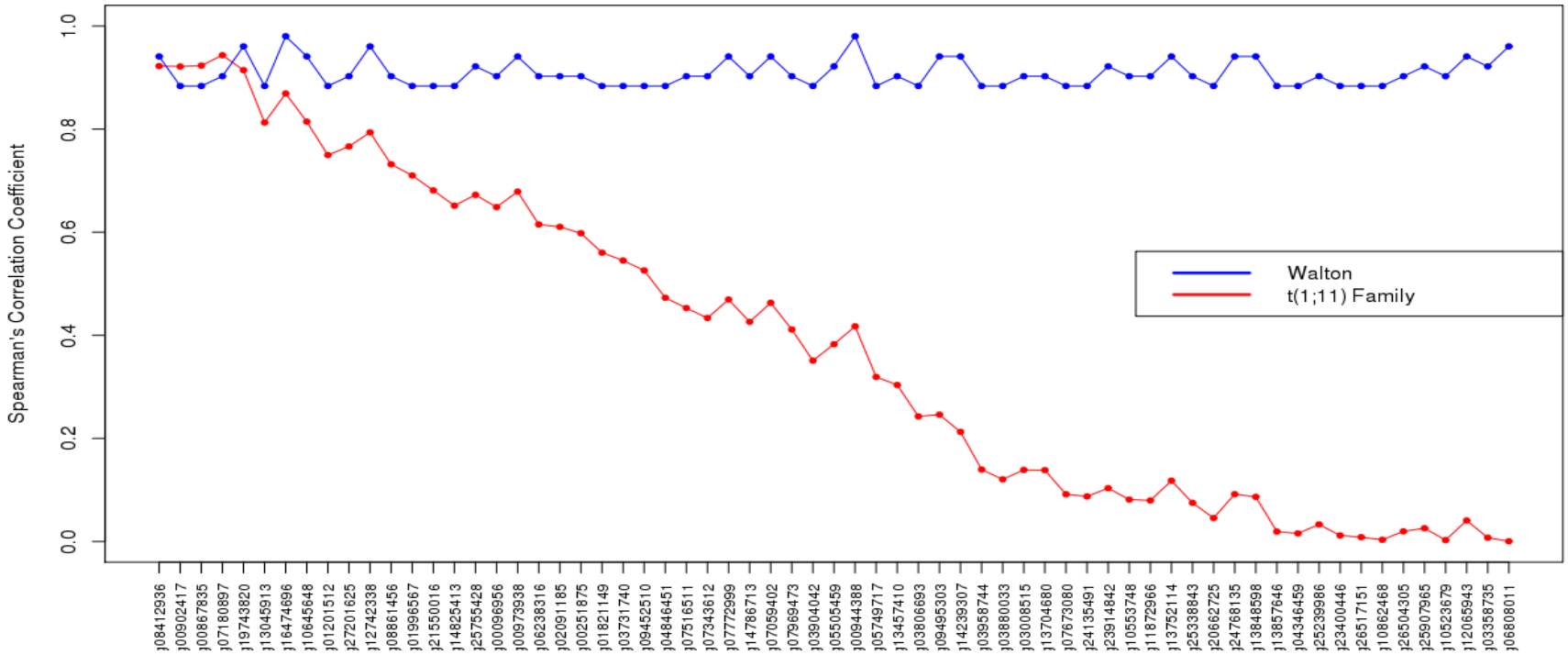


Figure 5.7: Comparison of correlation coefficients for 65 probes reported by Walton *et al.* (2015) as correlated between blood and brain.

Shown are Spearman's correlation coefficients (Rho; y-axis) for 65 probes reported to be correlated between blood and brain by Walton et al. (2015; x-axis). The red line represents the correlation coefficient between blood and iPSC-derivatives in six t(1;11) family samples. The blue line represents the correlation coefficients of same probes in blood and brain, as reported by Walton et al. (2015).

Probe ID	Gene	Blood		iPSC-derived neurons		Correlation (Spearman's Rho)
		p-value	*q* - value	*p-value*	*q* - value	
cg26728851	*GUCY2E*	0.002	0.97	0.02	1	0.91
cg26355502	*DUSP10*	0.02	0.97	0.1	1	0.75
cg09186051	*DISC1*;*TSNAX-DISC1*	0.1	0.97	0.96	1	0.41
cg02771260	*MS4A3*	0.0002	0.6	0.19	1	0.40
cg00965168	NA	0.03	0.97	0.94	1	0.34
cg05656812	*DISC1*;*TSNAX-DISC1*	0.02	0.97	0.77	1	0.26
cg18815120	*EGLN1*	0.02	0.97	0.88	1	0.19
cg06928246	NA	0.016	0.97	0.89	1	0.16
cg16177633	*DISC1*;*TSNAX-DISC1*	0.31	0.97	0.79	1	0.14
cg24508974	NA	0.25	0.97	0.21	1	0.02
cg15157974	*DISC1*;*TSNAX-DISC1*	0.02	0.97	0.48	1	0.003
cg21875980	*EGLN1*	0.0009	0.97	0.91	1	0.003
cg25899154	NA	0.81	0.99	0.26	1	0.003

Table 5.9: Blood/iPSC-derived neuron correlation coefficients between significantly differentially methylated sites in blood.

Table summarises DNA methylation data for 13 probes significantly differentially methylated between t(1;11) carriers and non-carriers in whole blood (n = 41 samples). From left to right, columns show the probe ID, the associated gene ("NA" denotes intergenic), the p-*value for differential methylation in six blood samples corresponding to iPSC-derived neurons, the FDR adjusted* p-*value for differential methylation in the same samples, the* p-*value for differential methylation in six iPSC-derived neurons, the adjusted* p-*value for differential methylation in the same samples, and the correlation coefficient for the probe between six blood samples and the corresponding iPSC-derived neuronal samples (Spearman's Rho).*

5.12 Analysis of methylation quantitative trait loci in iPSC-derived neurons

DNA methylation can be influenced by both environmental and genetic variation at linked and independent loci (Lemire *et al.*, 2015). A significant relationship was identified between DNA methylation and previously reported (Lemire *et al.*, 2015) methylation quantitative trait loci (meQTLs) at five of the 13 DMPs identified in t(1;11) blood samples (section 4.2.7; Table 4.6; Figure 4.3). To determine whether the same SNPs might act as meQTLs in iPSC-derived neurons, linear regression was performed to assess the relationship between minor allele count and DNA methylation at these five sites. No significant relationship was observed between genotype and DNA methylation for any of these SNPs ($p \geq 0.57$). To determine whether this could be due to the difference in the set of individuals tested for meQTLs, six blood samples corresponding to iPS neurons were also examined at these variants of interest. Here, a significant relationship was observed between genotype at rs4366301 and DNA methylation at cg16177633 ($p = 0.01$), with both the meQTL and CpG mapping to *DISC1*. For the remaining sites, no significant meQTL-CpG associations were observed ($p \geq 0.41$; Table 5.10).

meQTL/CpG Interaction	Neuronal *p*-value (n = 6)	Blood *p*-value (n = 6)	Blood *p*-value (n = 41)
rs2486729/cg18815120	0.95	0.75	1.26×10^{-16}
rs545937/cg21875980	0.99	0.41	4.26×10^{-5}
rs4366301/cg16177633	0.78	0.01	0.001
rs10899287/cg26728851	0.78	0.86	2.05×10^{-5}
rs17154511/cg02771260	0.57	0.90	4.68×10^{-12}

Table 5.10: Analysis of previously-reported meQTLs in blood and iPSC-derived neurons.

Relationships were assessed between meQTL genotype and CpG methylation levels at previously reported meQTLs present in t(1;11) family blood-derived samples (n = 41). Shown in order of column appearance are the meQTL SNP ID and associated CpG probe, p-values for association between meQTL genotype and DNA methylation at the reported site in iPSC-derived neurons (n = 6), the subset of blood samples corresponding to neurons (n = 6), and the total blood sample set in which significant meQTL/CpG associations were observed (n = 41).

5.13 Estimation of DNA methylation age of iPSC-derived Neurons

In order to assess whether the translocation had an effect on DNA methylation age in iPSC-derived neurons, DNA methylation age was calculated using Horvath's DNA methylation age calculator (Horvath, 2013; section 4.5). An additional aim of this analysis was to determine whether DNA methylation age corresponded with patient age at the time of fibroblast biopsy, or whether it showed correlation with NPC passage number at the point of neuronal differentiation. Linear regression was performed to assess the correlation between DNA methylation age and age at the time of fibroblast biopsy. No significant correlation was observed between DNA methylation age and age of fibroblast biopsy ($r^2 = 0.003$, $p = 0.82$). Additionally, correlation between DNA methylation age and cell passage number at the time of initiation of neuronal differentiation was assessed using linear regression. No significant correlation was observed between DNA methylation age and cell passage number ($r^2 = 0.13$, $p = 0.14$).

The average DNA methylation age of t(1;11) carriers was 9.9 years (SD = 6.11 years), while the average DNA methylation age of non-carriers was 7.01 (SD = 2.85 years). Age acceleration (i.e. DNA methylation age minus chronological age) was negative in all samples, with no significant between-group differences observed (student's *t*-test $p = 0.158$; Table 5.11).

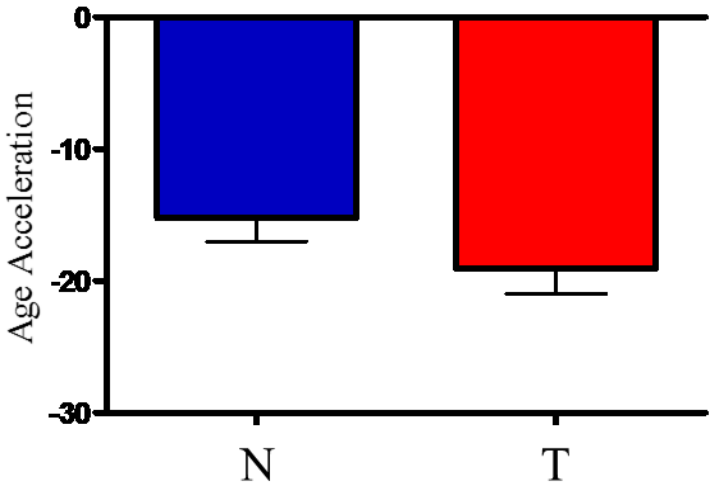


Figure 5.8: Age acceleration in iPSC-derived neurons.

Age acceleration (y-axis) is plotted against sample group (x-axis) for 18 iPSC-derived neuronal samples (n = 9 per group: 3 technical replicates from 3 individuals). "N" indicates t(1;11) non-carriers (blue bar) while "T" indicates t(1;11) carriers (red bar). Age acceleration is calculated as the difference between DNA methylation age and chronological age.

Sample	DNA methylation age	NPC passage
A1	7.17	18
A2	7.44	18
A3	14.08	25
B1	4.89	27
B2	4.99	27
B3	4.67	27
C1	22.35	27
C2	15.86	34
C3	14.27	36
D1	6.09	16
D2	6.66	20
D3	7.11	22
E1	6.16	26
E2	5.65	29
E3	7.52	30
F1	5.58	25
F2	6.45	28
F3	5.32	26

Table 5.11: Summary of DNA methylation age estimates in iPSC-derived neurons.

Columns show sample IDs corresponding to three differentiations each from individuals A-F, the DNA methylation age estimate, and the passage number of NPCs at initiation of neuronal differentiation.

5.14 Discussion

The aim of this study was to determine whether the t(1;11) translocation was associated with differential methylation in a cellular model relevant to psychiatric illness: iPSC-derived neurons. An additional aim of this work was to compare DNA methylation differences between t(1;11) carriers and non-carriers identified in blood with those identified in iPSC-derived neurons. Three lines of iPSC-derived neuronal precursor cells from each of six t(1;11) family members were differentiated into cortical neurons for DNA methylation analysis using the Infinium MethylationEPIC BeadChip (EPIC array). Of these six individuals, three were t(1;11) carriers and three were non-carriers.

Prior to analysing the data, sub-optimal probes present on the EPIC array were identified by *in-silico* methods, following a previously-published protocol (Chen *et al.*, 2013). A list of probes targeting known polymorphic sites was generated along with a list of potentially cross-hybridising probes. These lists were published as a resource for others analysing data derived from the EPIC array, who may wish to consider such probes for removal as a quality control measure (Appendix 1; McCartney *et al.*, 2016).

To determine whether the translocation was associated with differential methylation in iPSC-derived neurons, DNA methylation was compared between t(1;11) carriers to non-carriers. Methylation differences in nine of the top ten ranked loci were mostly driven by a single individual - a t(1;11) carrier - which consistently displayed hypermethylation relative to the remaining five individuals profiled. Among the top ten differentially methylated loci between t(1;11) carriers and non-carriers, one locus did not show this effect, with consistent levels of hypermethylation observed among translocation carriers (cg20146541; Figure 5.3). This site was located in *TRIM58*, approximately 16 Mb telomeric of the chromosome 1 breakpoint, and was among three probes constituting a significant DMR identified in t(1;11) carriers (Chr1:248020638-248020745, $p = 4.74 \times 10^{-10}$). This outlier sample was retained, having not been

flagged as a failure during the quality control check. It should therefore be noted that the results discussed here are in the context of this outlier sample. Due to the low number of samples in this study, it is possible that the methylation profile of this sample is not anomalous, but similar to the profiles of samples from other family members who have yet to be assessed for DNA methylation. This will likely be clarified should additional iPSC-derived neuronal samples undergo methylation profiling in the future.

No differences in DNA methylation were observed between t(1;11) carriers and non-carriers following correction for multiple testing ($q \leq 0.05$). Deflation of *p*-values for differential methylation was observed, suggesting the study is underpowered to detect genome-wide significant differences in methylation. The most significant differential methylation signal observed was on chromosome 5, in a 224 bp region containing the top three sites ranked by *p*-value. Although not annotated to a gene by the Illumina manifest, UCSC genome browser reports the presence of a validated RefSeq gene at this site: an uncharacterised long noncoding RNA (lncRNA; *FLJ32255*). Furthermore, the region containing these probes is a DNAse hypersensitivity site and contains multiple chromatin immunoprecipitation (ChIP)-identified transcription factor binding sites as reported by ENCODE: features that suggest a regulatory role. Further investigation of the expression and/or function of this gene would be useful to determine whether its disrupted methylation might confer downstream effects in t(1;11) carriers.

Gene ontology analysis was performed on the DMP data to identify functions, processes or components that are significantly enriched amongst the most differentially methylated genes. Multiple terms relating to synaptic activity, development and cell signalling were among those over-represented amongst the genes harbouring the most significant DMPs. This data might indicate that disrupted neurotransmission and neurodevelopment are associated with the translocation in neurons. This would be in keeping with the known functions of DISC1, which is interrupted by the translocation (Brandon and Sawa, 2011; section 1.4.3-4). The most

significantly over-represented function was "sequence-specific DNA binding" (GO0043565; $q = 9.05 \times 10^{-5}$). This may be indicative of downstream effects of the t(1;11) translocation on gene expression through disruptions to transcription factor binding. Soda *et al.* (2013) reported an interaction between the DISC1-ATF4 complex and the *PDE4D* locus, regulating its expression. Further experiments such as ChIP analysis of these samples might be useful to establish whether there are disruptions to such protein-DNA interactions in t(1;11) carriers, and whether such disruptions are linked to aberrant gene expression.

A total of 424 DMRs were identified in t(1;11) carriers. Multiple DMRs were identified in 14 genes. Four DMRs were identified within *DLGAP2*, a gene in which a psychosis-associated DMR was also identified in t(1;11) blood samples in Chapter 5. *DLGAP2* plays a role in synapse organisation and has been implicated in schizophrenia (Li *et al.*, 2014). Ten DMRs were identified in the DNA-repair gene *MGMT*. A DMR was also identified within 1500 bp of the TSS of *BRCA1*, another DNA repair gene (Moynahan *et al.*, 1999). Aberrant expression of DNA repair genes has been reported in psychotic disorders (Benes *et al.*, 2009). It is possible that disrupted methylation in such genes may affect their capacity for DNA repair in t(1;11) carriers, conferring an increased risk of illness in these individuals. Amongst the genes containing multiple DMRs was *SORCS1*, in which two hypermethylated regions were observed. *SORCS1* is a member of the brain-expressed Sortilin gene family described in Chapter 4. Expression of another member of this gene family, *SORL1*, was found to be decreased in t(1;11) carrier lymphoblastoid samples, described in Chapter 3. Moreover, others have reported dysregulation of Sortilin family members in *DISC1* mutant iPSC-derived neurons (Wen *et al.*, 2014; Srikanth *et al.*, 2015). Taken together, these findings may suggest interplay between DISC1 and the Sortilin genes. Based on this, a regulatory relationship between *DISC1* and Sortilin family members was further investigated. This work is reported in Chapters 6 and 7 of this thesis.

A limitation to the DMR analysis which must be considered is that larger genes will contain more probes to interrogate methylation, and therefore a greater likelihood of

obtaining a significant *p*-value for differential methylation due to chance. With this in mind, of the fourteen genes containing multiple DMRs, the median probe count per gene was 72. The median gene count across all genes on the EPIC array is 20 (Phipson *et al.*, 2016). Therefore, confirmation of findings from DMR analyses by targeted methods such as pyrosequencing of bisulphite-converted DNA is warranted, should meaningful conclusions be made from this analysis.

Seven DMR-containing genes identified in this study were also found to contain DMRs in the blood-based study. Disrupted methylation in one of these genes, *OR2L13*, has been implicated in autism by two independent studies using different tissue sources (Wong *et al.*, 2014 [peripheral blood]; Berko *et al.*, 2014 [buccal endothelium]). *OR2L13* is located at the telomeric end of chromosome 1q. Autism has not been reported in the t(1;11) family. However, overlap has been reported in the genetic architecture of several neuropsychiatric disorders, including autism, MDD, bipolar disorder and schizophrenia (Smoller *et al.*, 2013). Individuals carrying the t(1;11) translocation have an increased risk of recurrent MDD, bipolar disorder and schizophrenia (Thomson *et al.*, 2016). Therefore, expression and functional follow-up studies of *OR2L13* in the t(1;11) family may inform of possible roles for the gene in conferring increased risk of illness in t(1;11) carriers. Another DMR-containing gene worth noting is *COMT*, which, although not associated with schizophrenia through GWAS, has been the subject of numerous candidate gene association studies of schizophrenia described in section 1.3. This DMR spanned the TSS region of the gene, which may be indicative of altered *COMT* expression in t(1;11) carriers, as previously observed in cases of schizophrenia (Bray *et al.*, 2003).

The work presented in Chapter 4 reported significant differential methylation in t(1;11) carriers at 13 sites in an analysis of 41 blood-derived DNA samples. Twelve of these sites were within the t(1;11) breakpoint regions. Differential methylation was not observed at the t(1;11) breakpoint regions in the iPSC-derived neuronal samples. To determine whether the discrepancy between these findings might have been attributable to the difference in the individuals assessed for methylation, blood-derived

methylation data was examined in the six individuals corresponding to the iPSC-derived samples examined above. No genome-wide significant differences in methylation were observed at these 13 loci in the analysis of the subset of blood samples corresponding to the iPSC-derived neurons.

There are several possible reasons for the differing results found in blood compared to iPSC-derived neurons. It is possible that the differences are due to tissue-specific differences DNA methylation, with blood and brain having distinct DNA methylation signatures in these individuals. However, it is not possible to determine this from iPSC-derived neurons as they cannot faithfully recapitulate the human brain. The artificial nature in which these samples were grown is unlikely to reflect the three-dimensional system (i.e. the human brain) in which cortical neurons develop. Furthermore, the samples described in this analysis have undergone extensive reprogramming from the point of fibroblast biopsy, along with exposure to various cell culture conditions. Nazor *et al.* (2012) reported aberrant methylation levels at imprinted loci in iPSCs that were not repaired upon differentiation. It is possible that differential methylation is indeed present around the t(1;11) breakpoint regions in the brains of translocation carriers and this is not recapitulated by an iPSC-derived model due to disrupted DNA methylation during reprogramming. The small sample size may also have presented an issue with power to detect significant differential methylation. Horvath (2013) reported DNA methylation age in brain samples was significantly correlated with chronological age, as well as in whole blood. DNA methylation age and chronological age were significantly correlated in whole blood from these individuals (section 4.5, Figure 4.6) but not in the iPSC-derived neuronal samples. It is therefore likely that the methylomes of these samples are not reflective of those in the brains of t(1;11) family members. This is not unexpected, as others have reported significant correlation between transcriptomes of iPSC-derived neurons and foetal brains, but not adult brains (Handel *et al.*, 2016). Analysis of DNA methylation in *post-mortem* or biopsied brain tissue from t(1;11) family members would be required to generate conclusive results as to whether differential methylation occurs at the t(1;11) breakpoint regions in the brain. Analysis of other primary tissues from these

individuals might also be informative as to whether differential methylation at the t(1;11) breakpoint regions occurs across tissues in these individuals.

Correlation between DNA methylation in iPSC-derived neurons and blood from the t(1;11) family was assessed for a set of probes that were previously reported by Walton *et al.* (2015) to be significantly correlated between blood and brain. No significant agreement was observed between these studies. The samples used by Walton *et al*. (2015) consisted of temporal lobe biopsies (neocortex) while the iPSC-derivatives from this study were representative of frontal cortical neurons. Hannon *et al*. (2015) have reported high levels of correlation of DNA methylation between different cortical regions. This would suggest correlation should have been observed between methylation in the iPSC-derived neurons and blood at the 65 probes reported by Walton *et al.* (2015), if these iPSC-derivatives were faithfully recapitulating the DNA methylation in the brain. It should be noted, however, that the patients assessed by Walton *et al.* (2015) suffered from cortical dysplasia, which may have biased their results. The relatively small sample from the t(1;11) family from whom blood and iPSC-derivatives were available may also present an issue in terms of power to detect significant correlations.

The work described in Chapter 4 presented evidence in support of a role of meQTLs in driving the differential methylation signal at five of the 13 genome-wide significant DMPs associated with the translocation. To determine whether these variants were associated with methylation at the same sites in iPSC-derived neurons, the same SNP-CpG associations were assessed. No significant associations were observed between SNP genotype and methylation. No significant correlations were observed between genotype and methylation observed at five of the six sites when examined in the subset of six blood samples. This may have been due to the small sample size, or alternatively, between-individual variation.

A limitation of the blood-based DNA methylation analysis described in Chapter 5 was the absence of RNA to examine the effects of differential methylation on gene

expression. This shortcoming has been addressed in the iPSC-derived samples: RNA and protein were harvested concurrently with the DNA for transcriptomic and proteomic analyses, which are currently ongoing. This will provide the opportunity to cross-reference methylation data from these iPSC-derived neurons to data derived from gene and protein expression analyses in order to dissect the relationship between DNA methylation, gene expression and biological functions in these individuals. However, larger sample sizes will be required if meaningful conclusions are to be drawn from this work.

Chapter 6

Analysis of Sortilin family expression in a *Disc1* mutant mouse model

6 Analysis of Sortilin family expression in a *Disc1* mutant mouse model

6.1 Background and Motivation

The complex aetiology and clinical features associated with psychiatric illnesses have made it challenging to model such disorders in non-human animals. Nonetheless, mouse models can display behavioural endophenotypes which may reflect human-specific symptoms of psychiatric disorders. The use of animal models has also permitted insights into the underlying biology of several diseases, rendering them valuable resources for pharmacological and genetic studies both *in-vivo* and *in-vitro* (Salgado and Sandner, 2013). This chapter describes an analysis of gene expression in a mouse model of schizophrenia. This mouse contains a non-synonymous mutation in exon 2 of *Disc1*, whereby a leucine is substituted with a proline at amino acid position 100 (100P; Clapcote *et al.*, 2007). The aim of this chapter was to determine whether the *Disc1* L100P mutation played a regulatory role in Sortilin family gene expression during neurodevelopment. Support for a regulatory relationship between *DISC1* and Sortilin family members previously came from evidence of Sortilin dysregulation in *DISC1* mutant iPS-derived neuronal models (Wen *et al.*, 2014; Srikanth *et al.*, 2015), as well as evidence of disrupted expression and methylation of *SORL1* and *SORCS1*, respectively, in t(1;11) family samples described in Chapters 3 and 5 of this thesis.

The *Disc1* 100P mouse was identified from a screen of N-ethyl-N-nitrosourea (ENU)-mutagenised mice at the RIKEN Institute. The initial characterisation of this mouse by Clapcote *et al.* (2007) reported behavioural endophenotypes considered relevant to schizophrenia including hyperlocomotion, and deficits in pre-pulse and latent inhibition. Hyperlocomotion in mice is thought to reflect positive symptoms of schizophrenia observed in humans (Jones *et al.*, 2011), while pre-pulse and latent inhibition are thought to reflect the cognitive deficits observed in schizophrenia (Leumann *et al.*, 2002). These phenotypes were restored to wild-type levels in the 100P mice following antipsychotic treatment (Clapcote *et al.*, 2007). The presence of these endophenotypes coupled with the effects of antipsychotic medication rendered these

mice an attractive model for schizophrenia. In addition to behavioural characteristics, neuroanatomical deficits were also observed in the form of reduced brain volumes in these mice, providing additional support for *Disc1* in neurodevelopment. Furthermore, Lipina *et al*. (2012) reported dysregulated gene expression in the hippocampi and brain stems of these mice: a phenotype which was partially rescued following treatment with the mood stabiliser, valproic acid. This finding suggests that, through the 100P mutation, disruption of *Disc1* affects the expression of genes which may be involved in brain function.

Previously, a microarray-based analysis of gene expression was performed on brain-derived RNA from these mice by Sarah Brown (SB), a former PhD student. Here, genome-wide expression was compared between brains of 100P mice and their wild-type littermates at developmental stages from embryonic day 13 to adulthood using samples collected by Prof. John Roder's group in Toronto, Canada (referred to subsequently as the "SB samples"). SB reported significant downregulation of *Sort1* in 100P homozygotes at E18 and adult stages by qRT-PCR analysis. *Sort1* is a member of the Sortilin gene family, described in Chapter 1. A second member of the Sortilin gene family, *SorCS2*, was also found to be developmentally dysregulated in 100P homozygotes compared to wild-type mice from the SB samples by qRT-PCR. This work was performed by Franziska Sendfeld (FS), a former MSc. Student, who reported significant upregulation of *SORCS2* at E13 and E18; and significant downregulation at P1. *SORCS2* has previously been implicated in bipolar disorder by GWAS and follow-up studies (Baum *et al.*, 2008a; Baum *et al.*, 2008b; Ollila *et al.*, 2009; Christoforou *et al.*, 2011), although this has not been replicated at the genome-wide significant level in more recent, larger scale studies (Sklar *et al*., 2011; Muhleisen *et al.*, 2014; Hou *et al.*, 2016). The work described in Chapter 3 reported downregulation of *SORL1*, another Sortilin family member, in t(1;11) family-derived LCL samples. A study be Wen *et al.* (2014) examined gene expression in iPSC-derived neurons with a 4 bp *DISC1* frameshift mutation, in which synaptic deficits were observed. These cells showed dysregulation of *SORCS1*, *SORCS2* and *SORCS3*. Moreover, Srikanth *et al.* (2015) reported upregulation of *SORCS2* in iPSC-derived neurons containing a mutation predicted to induce nonsense-mediated decay in *DISC1*. Based on the above

findings it was hypothesised that a regulatory relationship might exist between *DISC1* and Sortilin family members.

SB normalised the gene expression data from her samples to two reference genes, one of which was *Gapdh*. The assay used (Taqman® assay ID: Mm99999915_g1, Thermo Fisher Scientific) has since been shown to be unreliable due to its potential to detect genomic DNA in addition to *Gapdh* transcripts. Moreover, others have reported high inter-individual and inter-tissue variability in *Gapdh*, which may render it sub-optimal as a reference gene (Barber *et al.*, 2005). Identifying an appropriate number of stably-expressed internal control genes for data normalisation is an important step in qRT-PCR-based analyses (Vandesompele *et al.*, 2002). However, such steps were not performed in the original analyses of the 100P mice. Taken together, the analysis strategy undertaken by SB and FS may have compromised the validity of the findings of differential expression of *Sort1* and *SorCS2*.

The work described in this chapter describes the reanalysis of Sortilin family gene expression in the SB samples using an optimised normalisation protocol. A replication analysis was then performed in an independent batch of samples (DM samples). The optimised protocol involved the use of geNorm software (Vandesompele et al., 2002) to identify the reference genes that are most stably-expressed in the sample from a panel of six frequently used reference genes. This step is followed by normalisation of the gene-of-interest expression data to the geometric mean of the expression data from the recommended reference genes. Based on findings of differential expression of *SORCS1-3* in the context of a *DISC1* frameshift mutation (Wen *et al.*, 2014), *SorCS1* and *SorCS3* were included for gene expression analysis. Furthermore, as *SORL1* was found to be differentially expressed in t(1;11) carriers (Chapter 3), this gene was also included for gene expression analysis in the 100P mice.

6.2 Results

6.2.1 Reference gene selection in the SB samples

As material from this batch of cDNA was limited, two samples containing sufficient volumes of cDNA were selected from each genotype (100P/100P and wild-type) for geNorm analysis at each developmental stage: embryonic days 13, 15 and 18; postnatal days 1,7 and 20; and adulthood (8 weeks). Six reference genes were tested in these samples: *Ppid, Ubc, Sdha*, *Hmbs*, *Rplp0* and *Hprt* and geNorm analysis was performed on their expression data. Values for geNorm M and geNorm V were obtained for each developmental stage to determine the relative stability of each gene and the recommended number of genes to use for normalisation, respectively. M-values correspond to the average pairwise variation for a given reference gene with all other reference genes considered, while V-values correspond to the stability of the normalisation factors for n reference genes compared to n+1 genes. Genes with a lower geNorm M-value are more stably expressed across all samples. A geNorm V-value of < 0.15 indicates that the normalisation factor obtained from n+1 genes is not significantly different to the normalisation factor obtained from n genes.

Using geNorm (Vandesompele *et al.*, 2002), it was possible to determine the optimum reference genes for all stages with the exception of the adult stage, for which expression between samples was too variable. In this instance, the number of genes in the set corresponding to the lowest geNorm V-value (4/5) were used: Hmbs, Ppid, Ubc and Rplp0, as this was deemed by geNorm as the most stably-expressed combination of reference genes tested in the adult samples. The geNorm results for each developmental stage are summarised in Table 6.1.

Stage	Optimum Number of Reference Genes (geNorm V < 0.15)	Recommended Reference Genes
E13	2	*Hprt, Ubc*
E15	2	*Hprt, Rplp0*
E18	2	*Sdha, Ubc*
P1	2	*Rplp0, Sdha*
P7	2	*Ubc, Hprt*
P20	2	*Hprt, Sdha*
Adult	4 (Lowest geNorm V = 0.18)	*Hmbs, Ppid, Ubc, Rplp0*

Table 6.1: Summary of the results of geNorm analysis of a panel of six reference genes in the SB samples.

The table lists developmental stage, the recommended number of reference genes and their identities.

6.2.2 Gene expression analysis in the SB samples

Differences in the expression of the five Sortilin family members between 100P homozygotes and wild-type littermates (n = 6 per group) was assessed at seven developmental time points. Samples were derived from whole brains from all mice with the exception of the adult stage. In the case of adult mice, expression was assessed in hippocampal samples. Significant differential expression was observed in at least one developmental stage for all genes with the exception of *Sort1* (Welch's t-test $p < 0.05$; Figure 6.1A-E; Table 6.2A-E).

A

Gene	Developmental Stage	*p*-value	Fold change 100P/100P vs wild type
Sort1	E13	0.463	1.27
Sort1	E15	0.148	-2.43
Sort1	E18	0.111	-2.21
Sort1	P1	0.344	1.69
Sort1	P7	*0.09*	4.67
Sort1	P20	*0.12*	5.10
Sort1	Adult	*0.38*	2.41

B

Gene	Developmental Stage	*p*-value	Fold change 100P/100P vs wild type
SorCS1	E13	0.795	1.06
SorCS1	E15	0.0003	2.17
SorCS1	E18	0.340	1.38
SorCS1	P1	0.0003	8.50
SorCS1	P7	0.003	2.24
SorCS1	P20	*0.476*	4.12
SorCS1	Adult	*1*	1.02

C

Gene	Developmental Stage	*p*-value	Fold change 100P/100P vs wild type
SorCS2	E13	0.073	-1.29
SorCS2	E15	0.021	1.86
SorCS2	E18	0.444	1.15
SorCS2	P1	0.088	2.36
SorCS2	P7	0.053	2.43
SorCS2	P20	0.029	5.02
SorCS2	Adult	0.144	9.29

D

Gene	Developmental Stage	*p*-value	Fold change 100P/100P vs wild type
SorCS3	E13	0.098	-1.31
SorCS3	E15	0.0002	1.62
SorCS3	E18	0.850	1.04
SorCS3	P1	0.010	7.17
SorCS3	P7	0.0004	1.87
SorCS3	P20	0.076	2.54
SorCS3	Adult	*0.143*	2.46

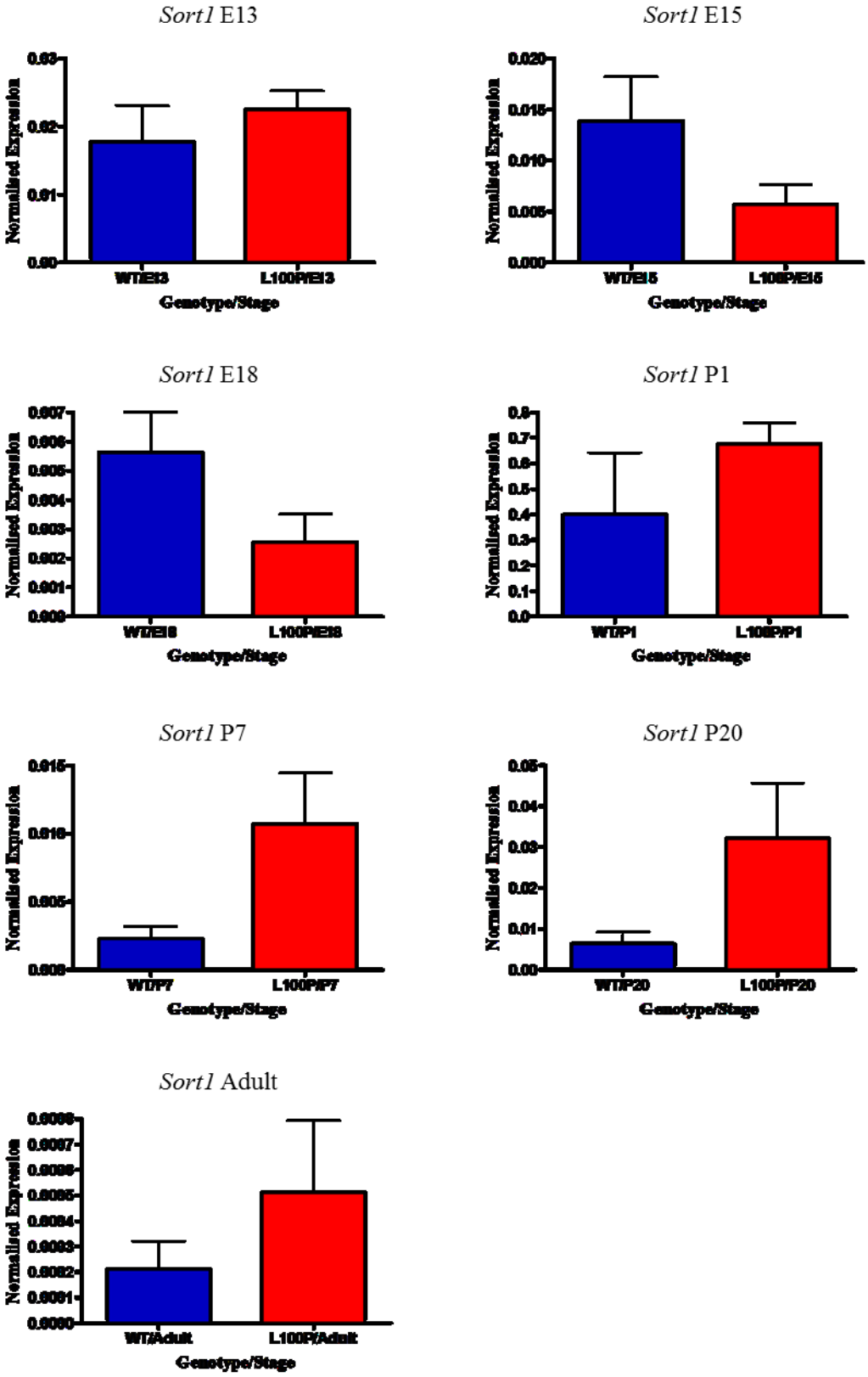
E

Gene	Developmental Stage	*p*-value	Fold change 100P/100P vs wild type
Sorl1	E13	0.007	2.01
Sorl1	E15	0.059	-2.89
Sorl1	E18	0.0004	-2.84
Sorl1	P1	0.002	12.05
Sorl1	P7	*0.004*	-3.17
Sorl1	P20	0.036	19.51
Sorl1	Adult	*0.1143*	8.09

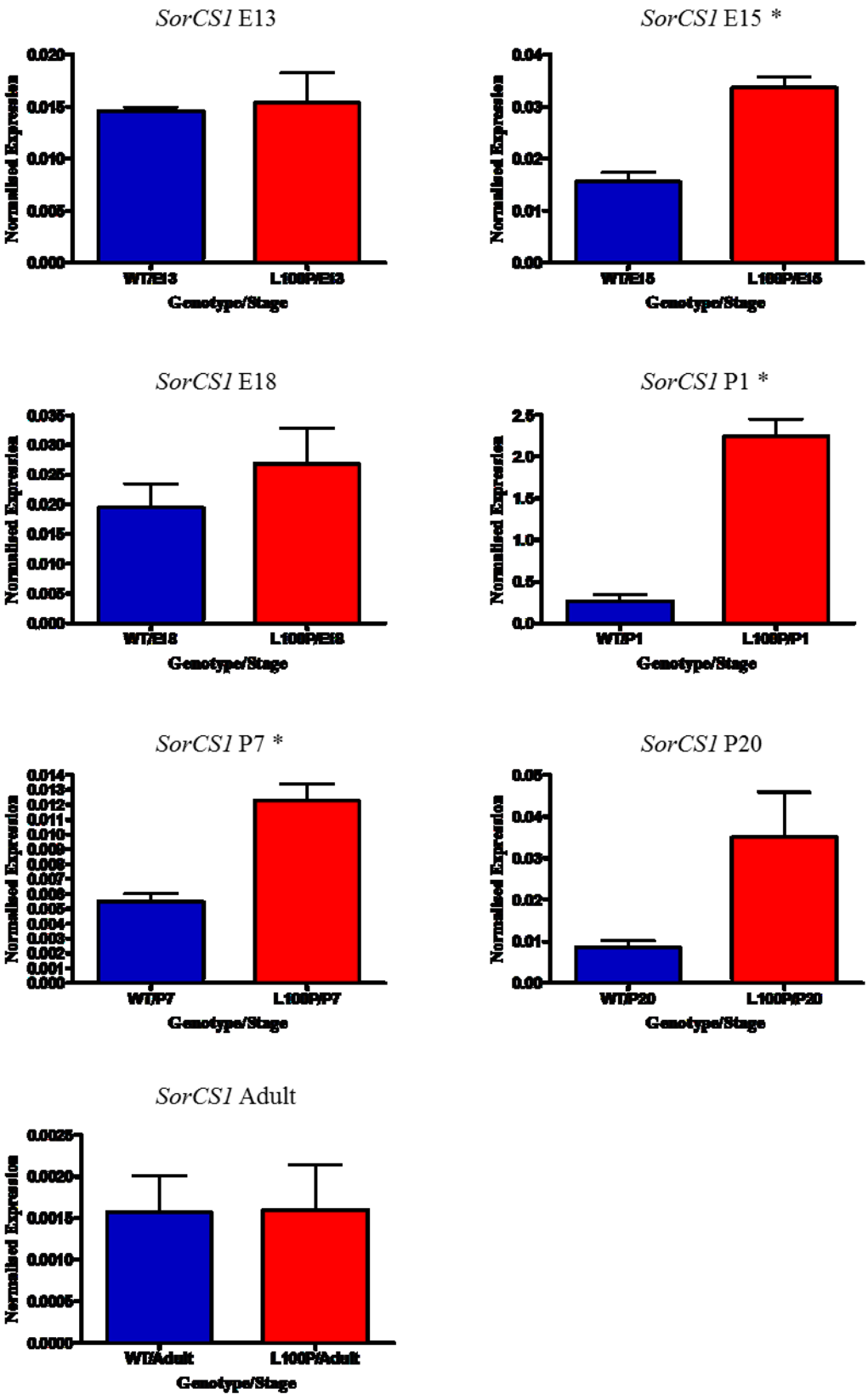
Table 6.2A-E: Summary of the analysis of gene expression of Sortilin family genes in the SB samples.

Shown are the genes tested for differential expression, the developmental stage of mice, the p-*value for differential expression and the fold-change in 100P homozygotes. Red font indicates significant differential expression ($p \leq 0.05$). Italicised* p-*values correspond to Mann-Whitney U test* p-*values whilst non-italicised* p-*values correspond to Welch's* t-*test* p-*values, depending on the distribution of the data in each group (section 2.8.4).*

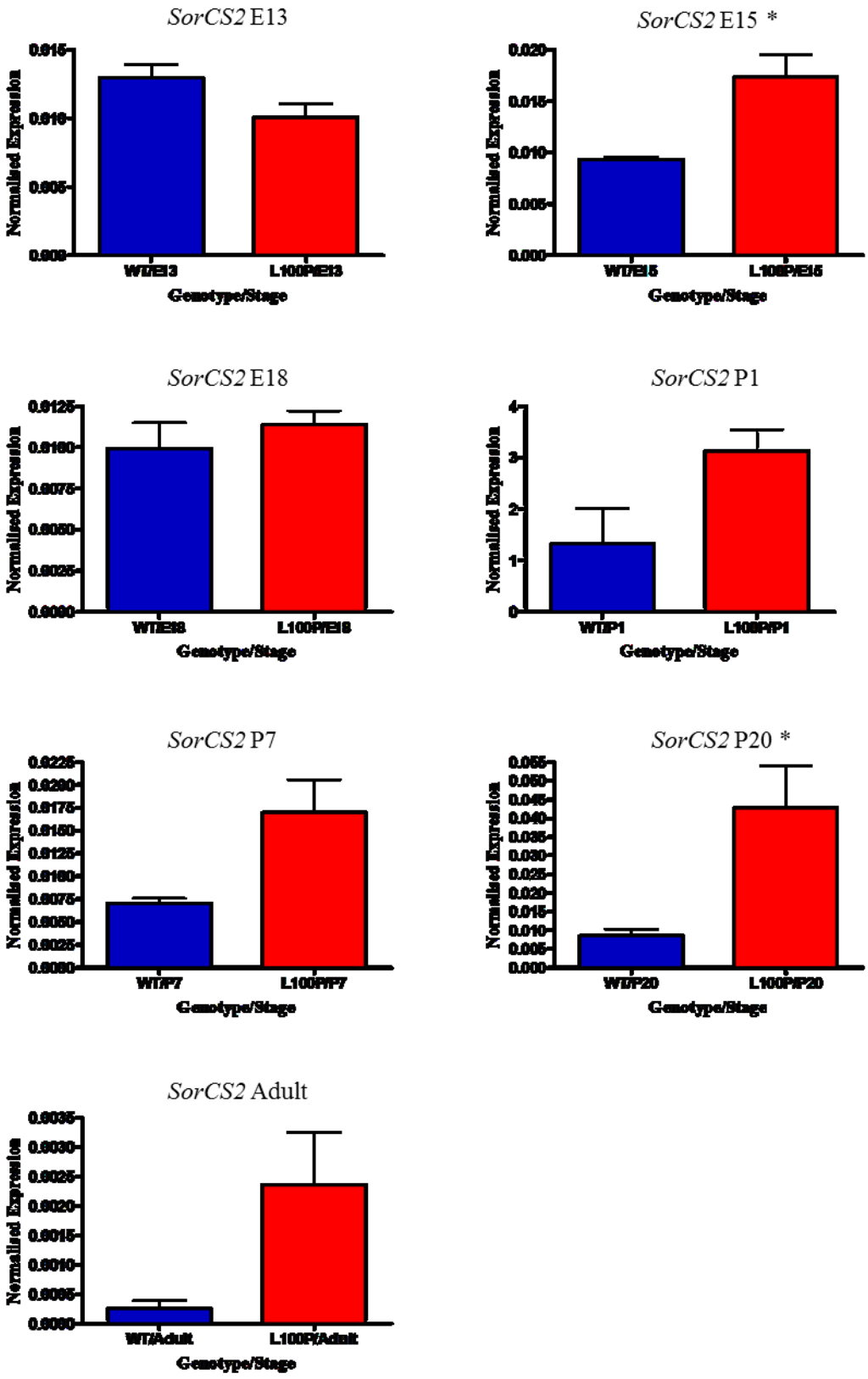
A


Sort1 E13
Normalised Expression
0.93
0.92
0.91
0.90
WTE13
L100PE13
Genotype/Stage
Sort1 E15
Normalised Expression
0.020
0.015
0.010
0.005
0.000
WTE15
L100PE15
Genotype/Stage
Sort1 E18
Normalised Expression
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
WTE18
L100PE18
Genotype/Stage
Sort1 P1
Normalised Expression
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
WTP1
L100PP1
Genotype/Stage
Sort1 P7
Normalised Expression
0.015
0.010
0.005
0.000
WTP7
L100PP7
Genotype/Stage
Sort1 P20
Normalised Expression
0.05
0.04
0.03
0.02
0.01
0.00
WTP20
L100PP20
Genotype/Stage
Sort1 Adult
Normalised Expression
0.0008
0.0007
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
0.0000
WTAdult
L100PAdult
Genotype/Stage

B

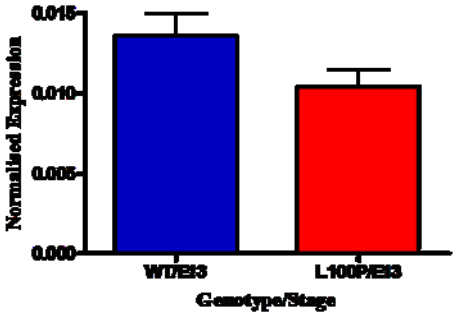

SorCS1 E13
Normalised Expression
0.020
0.015
0.010
0.005
0.000
WT/E13
L100P/E13
Genotype/Stage
SorCS1 E15 *
Normalised Expression
0.04
0.03
0.02
0.01
0.00
WT/E15
L100P/E15
Genotype/Stage
SorCS1 E18
Normalised Expression
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
WT/E18
L100P/E18
Genotype/Stage
SorCS1 P1 *
Normalised Expression
2.5
2.0
1.5
1.0
0.5
0.0
WT/P1
L100P/P1
Genotype/Stage
SorCS1 P7 *
Normalised Expression
0.014
0.013
0.012
0.011
0.010
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
WT/P7
L100P/P7
Genotype/Stage
SorCS1 P20
Normalised Expression
0.05
0.04
0.03
0.02
0.01
0.00
WT/P20
L100P/P20
Genotype/Stage
SorCS1 Adult
Normalised Expression
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
WT/Adult
L100P/Adult
Genotype/Stage

C

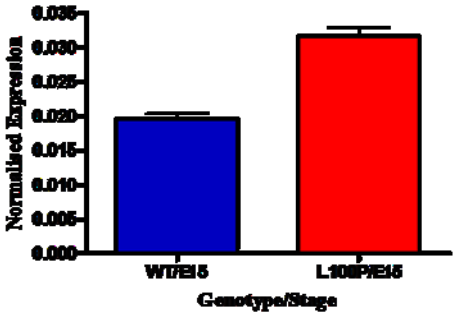

SorCS2 E13
Normalised Expression
0.015
0.010
0.005
0.000
WtE13
L100PE13
Genotype/Stage
SorCS2 E15 *
Normalised Expression
0.020
0.015
0.010
0.005
0.000
WtE15
L100PE15
Genotype/Stage
SorCS2 E18
Normalised Expression
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
WtE18
L100PE18
Genotype/Stage
SorCS2 P1
Normalised Expression
4
3
2
1
0
WtP1
L100PP1
Genotype/Stage
SorCS2 P7
Normalised Expression
0.0225
0.0200
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
WtP7
L100PP7
Genotype/Stage
SorCS2 P20 *
Normalised Expression
0.055
0.050
0.045
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
WtP20
L100PP20
Genotype/Stage
SorCS2 Adult
Normalised Expression
0.0035
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
WtAdult
L100PAdult
Genotype/Stage

D

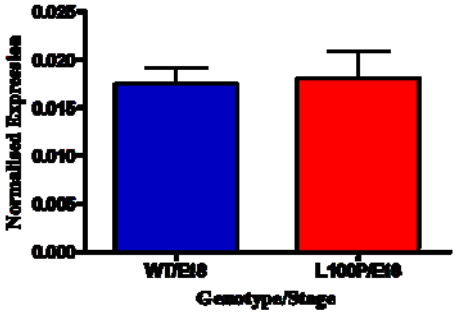
SorCS3 E13



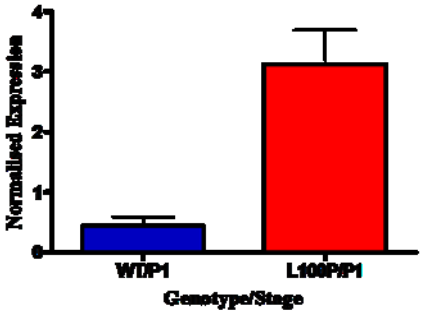
SorCS3 E15 *



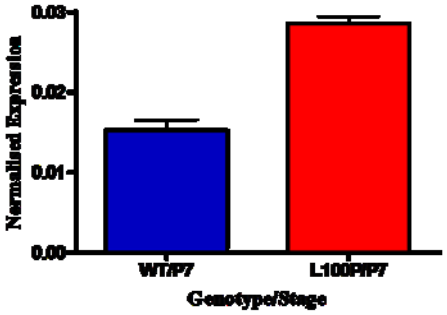
SorCS3 E18



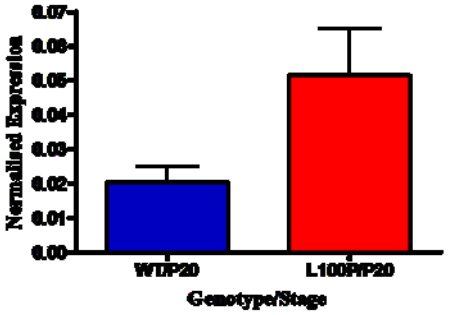
SorCS3 P1 *



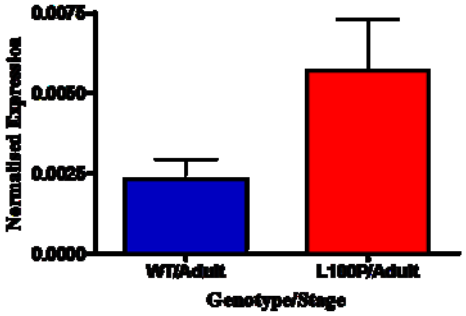
SorCS3 P7*



SorCS3 P20



SorCS3 Adult



E

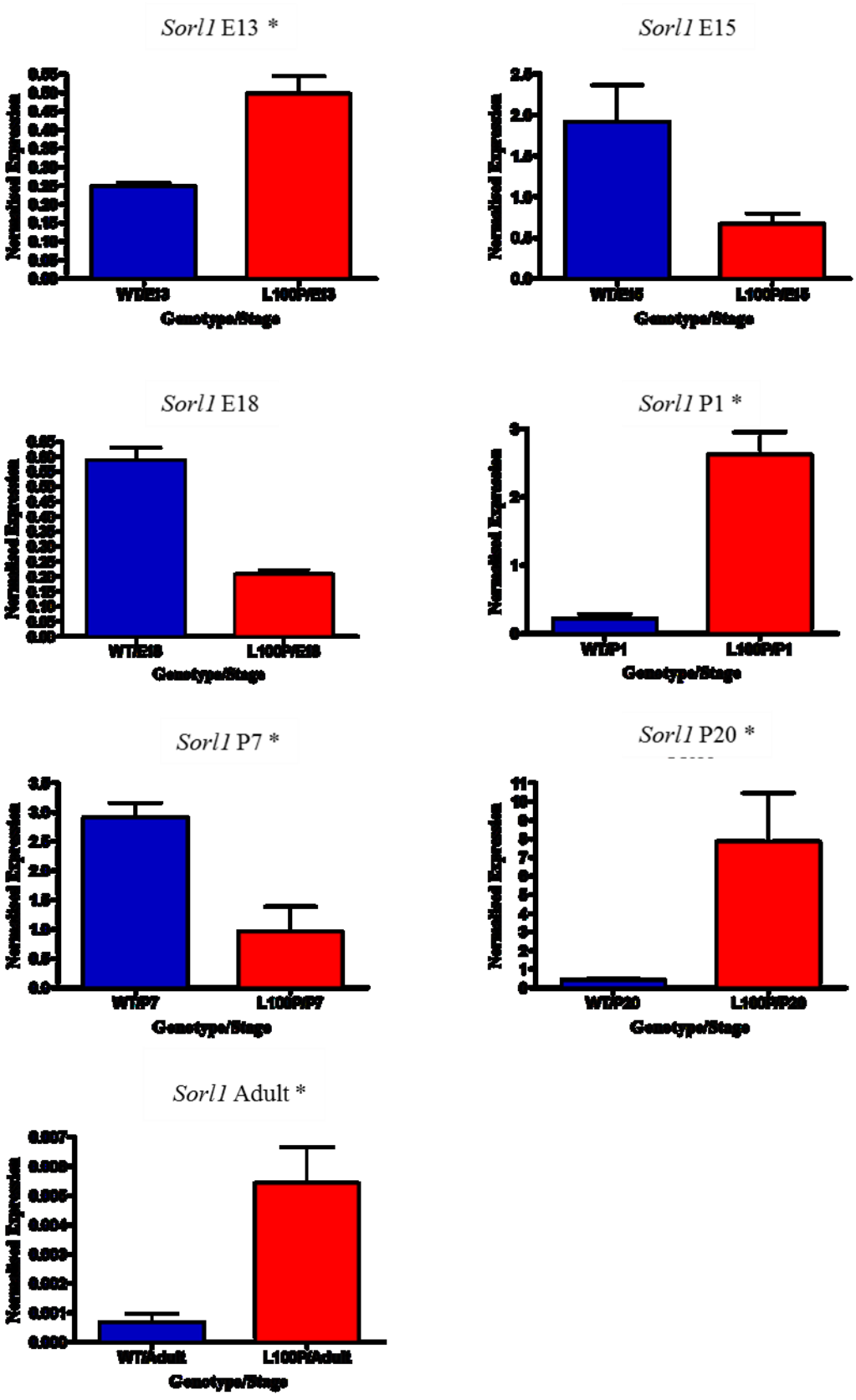

Sorl1 E13 *
Sorl1 E15
Sorl1 E18
Sorl1 P1 *
Sorl1 P7 *
Sorl1 P20 *
Sorl1 Adult *

Figure 6.1 A-E: Comparison of Sortilin family gene expression in 100P heterozygotes and wild-type littermates in the SB samples.

Shown are bar plots for normalised gene expression values (y-axes; arbitrary units) of Sortilin family genes (A-E = Sort1, SorCS1, SorCS2, SorCS3, Sorl1; respectively) in wild-type mice (blue bars) and their 100P homozygous littermates (red bars) from stages E13 to postnatal week 8 (Adult) in the SB samples. Genotype and developmental stage are presented on the x-axes. Error bars show the standard error of the mean. Significant differential expression is indicated by an asterisk (*) in the title of each plot.

In an attempt to replicate these findings, this experiment was repeated in samples collected from an independent set of 100P samples, from mice bred in Edinburgh (DM samples).

6.2.3 Analysis of differential gene expression in the DM samples

6.2.3.1 Sample information and quality control

100P heterozygote mice were obtained from Malgorzata Borkowska (Centre for Integrative Physiology, University of Edinburgh) for breeding. 100P homozygous mutants and their wild-type littermates were obtained from heterozygote crosses for gene expression analysis (N = 5-7 mice per group). As with the SB samples, gene expression was assessed at embryonic days 13, 15 and 18; postnatal days 1, 7 and 20; and adult mice.

Samples were assessed for RNA degradation and genomic DNA contamination as described in Chapter 3. RNA integrity was assessed as a measure of sample quality taking a minimum RIN cut-off of 7 (Thomson *et al.*, 2007). All samples were within the acceptable range of RIN scores (7 – 10), indicating intact RNA (Table 6.3). Genomic DNA contamination of cDNA was assessed using a primer pair spanning exons 24 to 25 of mouse *Sorcs2*, amplifying a 628 bp fragment from genomic DNA, and a 156 bp fragment from cDNA. Following reverse transcription of RNA, all cDNA samples were found to be free of genomic DNA.

Sample ID	Age	Genotype	Sex	RIN
175	Adult	100P	M	8.2
186	Adult	100P	M	8.6
188	Adult	100P	M	9.1
189	Adult	100P	F	9.8
198	Adult	100P	M	9
199	Adult	100P	F	8.8
177	Adult	WT	F	8.7
179	Adult	WT	F	9.2
183	Adult	WT	M	9.5
191	Adult	WT	F	9.1
193	Adult	WT	M	8.9
E13 15	E13	100P	M	9.8
E13 2	E13	100P	F	9.6
E13 20	E13	100P	M	10
E13 25	E13	100P	M	10
E13 8	E13	100P	M	9.7
E13 11	E13	WT	M	9.9
E13 17	E13	WT	F	9.9
E13 21	E13	WT	F	10
E13 22	E13	WT	F	10
E13 23	E13	WT	M	10
E13 3	E13	WT	M	9.7
E13 9	E13	WT	M	9.9
E15 1	E15	100P	F	9.9
E15 13	E15	100P	M	9.9
E15 14	E15	100P	F	9.9
E15 15	E15	100P	M	9.8
E15 16	E15	100P	M	9.9
E15 2	E15	100P	M	9.8
E15 3	E15	100P	M	9.8
E15 20	E15	WT	M	9.9
E15 21	E15	WT	F	9.9
E15 22	E15	WT	F	9.9

E15 23	E15	WT	M	9.9
E15 9	E15	WT	M	10
E18 12	E18	100P	F	9.7
E18 16	E18	100P	F	9.7
E18 23	E18	100P	F	9.6
E18 3	E18	100P	F	9.6
E18 4	E18	100P	F	9.6
E18 6	E18	100P	M	9.6
E18 8	E18	100P	M	9.6
E18 18	E18	WT	F	9.6
E18 21	E18	WT	M	9.6
E18 22	E18	WT	M	9.6
E18 5	E18	WT	F	9.7
E18 7	E18	WT	F	9.6
P1 11	P1	100P	M	9.5
P1 13	P1	100P	M	9.4
P1 21	P1	100P	F	7
P1 31	P1	100P	F	8.7
P1 33	P1	100P	M	8.6
P1 38	P1	100P	M	8.4
P1 1	P1	WT	F	9.6
P1 10	P1	WT	F	9.3
P1 15	P1	WT	M	7.4
P1 19	P1	WT	F	7.1
P1 29	P1	WT	M	8.7
P1 8	P1	WT	F	9.4
P20 1	P20	100P	M	8.1
P20 17	P20	100P	F	8.1
P20 24	P20	100P	M	7.2
P20 25	P20	100P	F	7.1
P20 3	P20	100P	M	7.7
P20 8	P20	100P	M	8.3
P20 16	P20	WT	M	8.3
P20 21	P20	WT	M	7.8
P20 35	P20	WT	F	7.2

P20 6	P20	WT	F	8.3
P20 7	P20	WT	F	7.9
P7 13	P7	100P	M	8.7
P7 17	P7	100P	M	7.8
P7 21	P7	100P	M	7.5
P7 27	P7	100P	F	8.7
P7 28	P7	100P	F	8.9
P7 7	P7	100P	F	8.1
P7 11	P7	WT	F	8.2
P7 15	P7	WT	M	7.4
P7 16	P7	WT	M	8.2
P7 2	P7	WT	F	8.8
P7 3	P7	WT	M	7.7
P7 5	P7	WT	F	8.3

Table 6.3: Summary of 100P RNA samples obtained from the DM samples.

Shown are sample IDs for each mouse, developmental stage (prefix E = embryonic, prefix P = postnatal), 100P genotype (WT= wild-type, 100P = 100P homozygote), sex (M = male, F = female), and RIN value.

6.2.3.2 Reference gene selection for the DM samples

As described in section 6.3, geNorm analysis was performed to identify the most stably-expressed reference gene set for each developmental stage. All samples from the DM batch were included in this geNorm analysis. It was possible to identify the most stably-expressed gene sets for all developmental stages with the exception of P7 due to high variability in these samples. In this instance, five reference genes with the lowest corresponding geNorm V score were used (V = 0.16) as they were deemed by geNorm to be the most stably-expressed combination of genes amongst those tested in the P7 samples. A summary of the geNorm analysis in the DM samples is presented in Table 6.4.

Stage	Optimum Number of Reference Genes (geNorm V < 0.15)	Recommended Reference Genes
E13	2	*Ppid, Rplp0*
E15	2	*Rplp0, Hmbs*
E18	2	*Ppid, Hprt*
P1	2	*Hprt, Sdha*
P7	Undetermined (Lowest geNorm V = 0.16)	*Rplp0*, *Hmbs*, *Ubc*, *Sdha*, *Ppid*
P20	2	*Rplp0, Hmbs*
Adult	2	*Ppid*, *Hprt*

Table 6.4: Summary of geNorm analysis using a panel of six reference genes in the DM samples.

The developmental stage is shown with the corresponding geNorm V and geNorm M scores determining the optimum number of reference genes and the recommended reference genes respectively.

6.2.3.3 qRT-PCR analysis of Sortilin family gene expression in the DM samples

Differential expression was assessed between wild-type and 100P homozygotes following normalisation of gene expression data to geometric mean of the appropriate reference genes. Three genes showed significant differential expression in 100P homozygotes, each at a single developmental time point (Table 6.5A-E; Figure 6.2A-E). At P20, 100P homozygotes showed a significant upregulation of *SorCS2* ($p = 0.04$, two-tailed Welch's *t*-test; FC = 1.47), and *Sort1* ($p = 0.03$, Mann-Whitney U test; FC = 1.47); while a significant upregulation of *Sorl1* was observed in 100P homozygotes at P7 ($p = 0.017$, two-tailed Welch's *t*-test; FC = 1.33). The *SorCS2* direction of effect in P20 100P homozygotes was consistent between the SB and DM samples, with both showing upregulation.

A

		DM Samples		SB Samples	
Gene	**Stage**	***p*-value**	**Fold change in 100P/100P vs wild type**	***p*-value**	**Fold change in 100P/100P vs wild type**
Sort1	E13	*0.930*	1.01	0.463	1.27
Sort1	E15	0.604	1.03	0.148	-2.43
Sort1	E18	0.256	1.12	0.111	-2.21
Sort1	P1	0.410	1.04	0.344	1.69
Sort1	P7	0.254	1.18	*0.09*	4.67
Sort1	P20	*0.038*	1.33	*0.12*	5.10
Sort1	Adult	0.759	-1.07	*0.381*	2.41

B

		DM Samples		SB Samples	
Gene	**Stage**	***p*-value**	**Fold change in 100P/100P vs wild type**	***p*-value**	**Fold change in 100P/100P vs wild type**
SorCS1	E13	0.296	1.18	0.795	1.06
SorCS1	E15	0.1990	-1.13	0.0003	2.17
SorCS1	E18	0.090	1.17	0.340	1.38
SorCS1	P1	0.6730	1.02	0.0003	8.50
SorCS1	P7	0.739	1.08	0.003	2.24
SorCS1	P20	*0.217*	-1.13	*0.476*	4.12
SorCS1	Adult	0.992	-1.00	*1*	1.02

C

		DM Samples		SB Samples	
Gene	**Stage**	***p*-value**	**Fold change in 100P/100P vs wild type**	***p*-value**	**Fold change in 100P/100P vs wild type**
SorCS2	E13	0.462	-1.09	0.073	-1.29
SorCS2	E15	0.874	-1.02	0.021	1.86
SorCS2	E18	0.602	1.09	0.444	1.15
SorCS2	P1	0.721	1.03	0.088	2.36
SorCS2	P7	0.060	1.58	0.053	2.43
SorCS2	P20	0.040	1.47	0.029	5.02
SorCS2	Adult	0.829	-1.03	0.144	9.29

D

		DM Samples		SB Samples	
Gene	**Stage**	***p*-value**	**Fold change in 100P/100P vs wild type**	***p*-value**	**Fold change in 100P/100P vs wild type**
SorCS3	E13	0.557	-1.19	0.098	-1.31
SorCS3	E15	*0.22*	-1.06	0.0002	1.62
SorCS3	E18	0.602	1.15	0.850	1.04
SorCS3	P1	0.496	1.09	0.010	7.17
SorCS3	P7	0.2120	1.30	0.0004	1.87
SorCS3	P20	0.323	1.11	0.076	2.54
SorCS3	Adult	0.847	-1.03	*0.143*	2.46

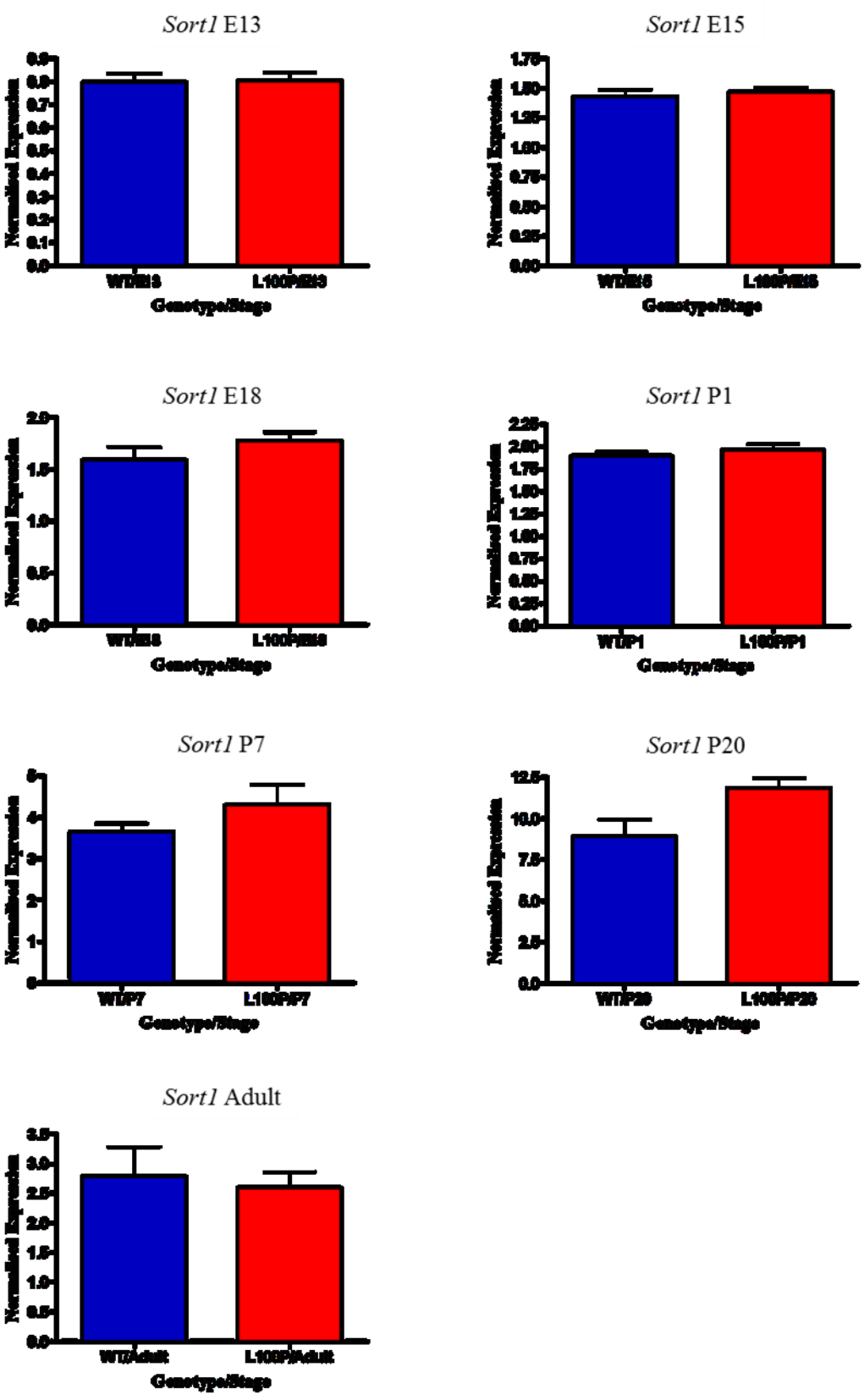
E

		DM Samples		SB Samples	
Gene	**Stage**	***p*-value**	**Fold change in 100P/100P vs wild type**	***p*-value**	**Fold change in 100P/100P vs wild type**
Sorl1	E13	*0.09*	-1.06	0.007	2.01
Sorl1	E15	0.715	1.02	0.059	-2.89
Sorl1	E18	0.0820	1.49	0.0004	-2.84
Sorl1	P1	*0.335*	1.01	0.002	12.05
Sorl1	P7	0.017	1.33	*0.004*	-3.17
Sorl1	P20	0.392	1.18	0.036	19.51
Sorl1	Adult	0.283	-1.13	*0.1143*	8.09

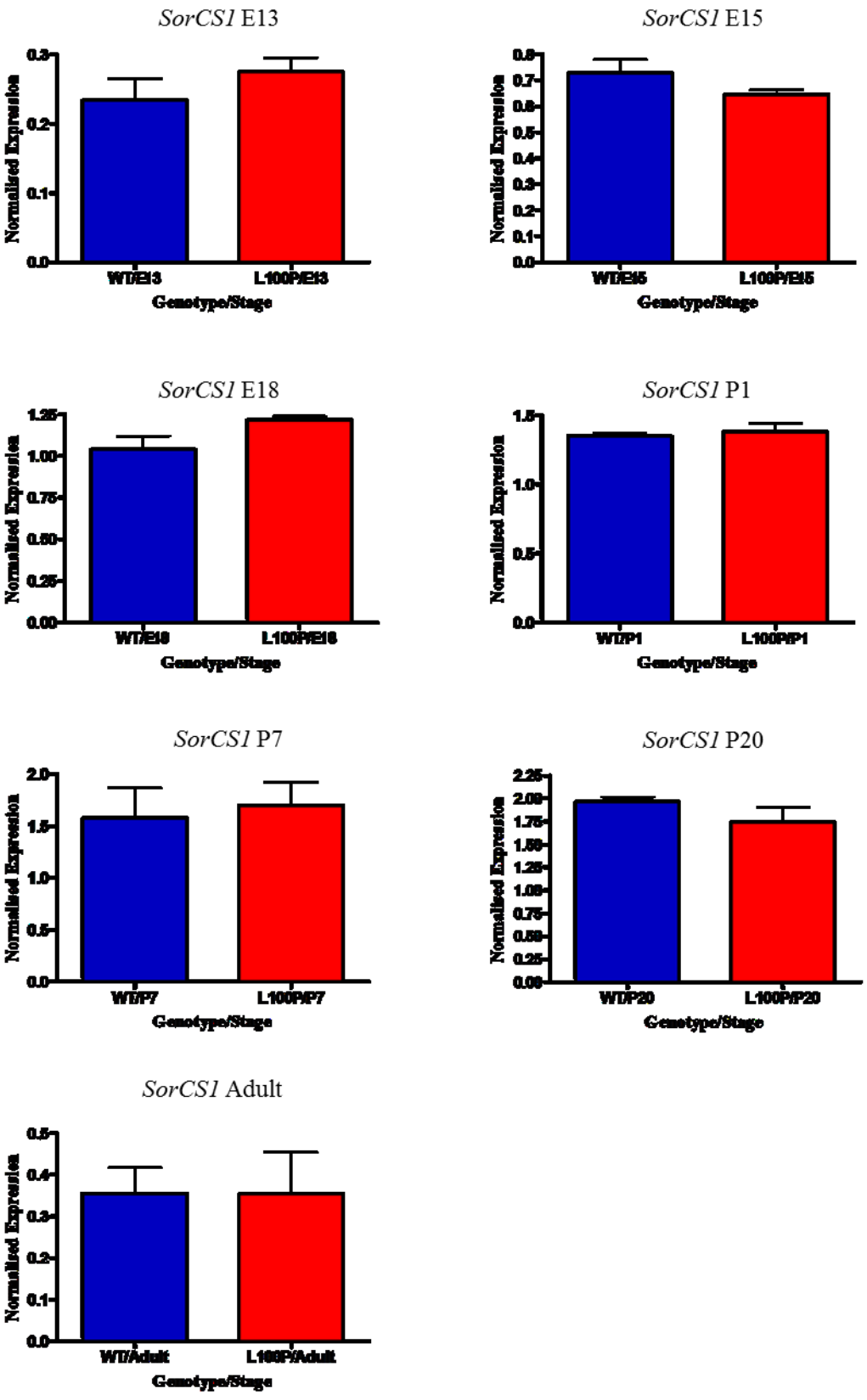
Table 6.5: Summary of gene expression analysis of Sortilin family genes in the DM samples compared with the SB samples.

Shown are the genes tested for differential expression and the developmental stage. For both the DM samples and SB samples, the p*-values for differential expression and the fold-changes in 100P homozygotes are shown. Red font indicates significantly differentially expressed genes (*p ≤ *0.05). Italicised* p*-values correspond to Mann-Whitney U test* p*-values whilst non-italicised* p*-values correspond to Welch's* t*-test* p*-values, depending on the distribution of the data in each group (section 2.8.4).*

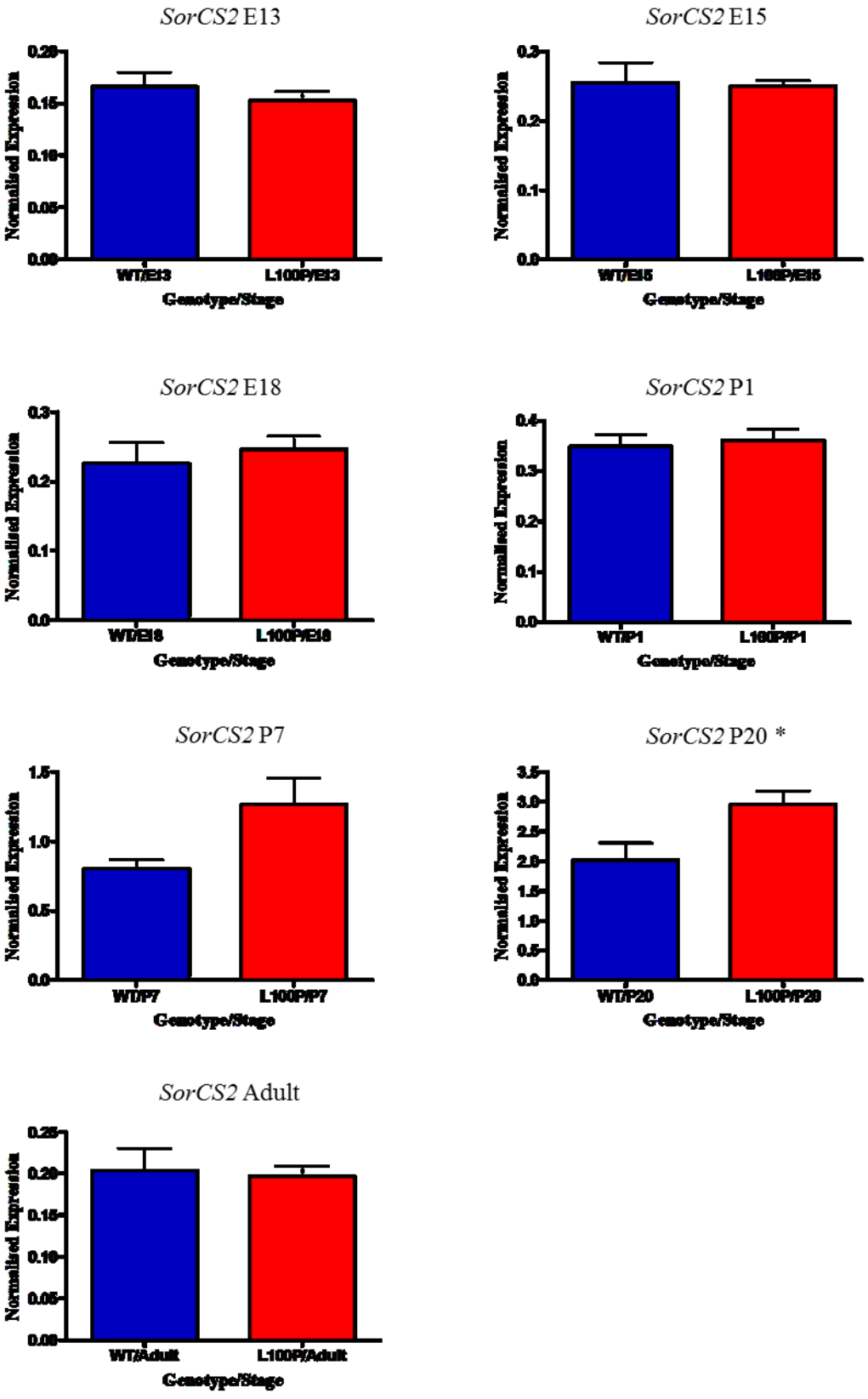
A


Sort1 E13
Sort1 E15
Sort1 E18
Sort1 P1
Sort1 P7
Sort1 P20
Sort1 Adult

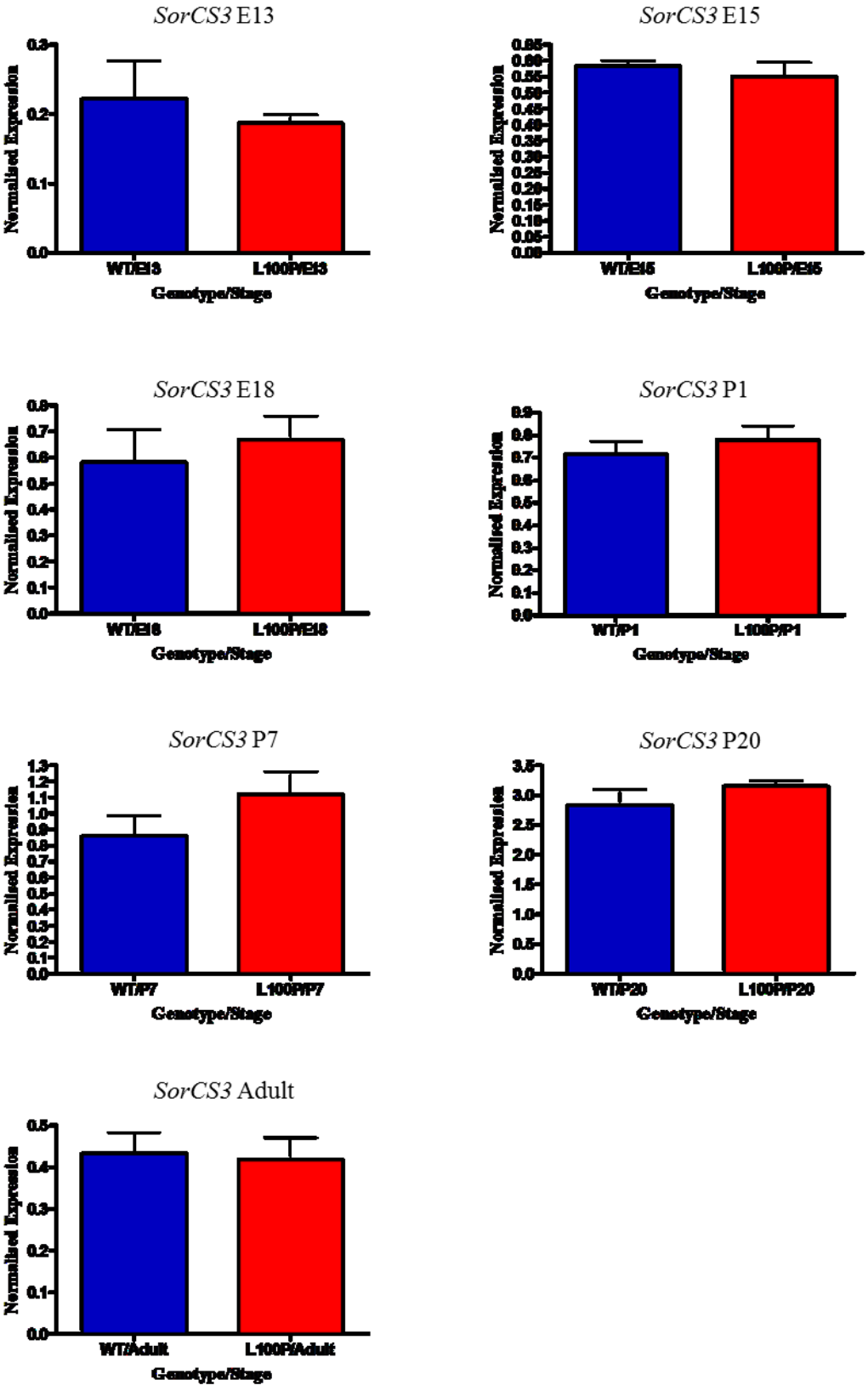
B


SorCS1 E13
Normalised Expression
0.0
0.1
0.2
0.3
WT/E13
L100P/E13
Genotype/Stage
SorCS1 E15
Normalised Expression
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
WT/E15
L100P/E15
Genotype/Stage
SorCS1 E18
Normalised Expression
0.00
0.25
0.50
0.75
1.00
1.25
WT/E18
L100P/E18
Genotype/Stage
SorCS1 P1
Normalised Expression
0.0
0.5
1.0
1.5
WT/P1
L100P/P1
Genotype/Stage
SorCS1 P7
Normalised Expression
0.0
0.5
1.0
1.5
2.0
WT/P7
L100P/P7
Genotype/Stage
SorCS1 P20
Normalised Expression
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
WT/P20
L100P/P20
Genotype/Stage
SorCS1 Adult
Normalised Expression
0.0
0.1
0.2
0.3
0.4
0.5
WT/Adult
L100P/Adult
Genotype/Stage

C


SorCS2 E13
Normalised Expression
0.20
0.15
0.10
0.05
0.00
WT/E13
L100P/E13
Genotype/Stage
SorCS2 E15
Normalised Expression
0.3
0.2
0.1
0.0
WT/E15
L100P/E15
Genotype/Stage
SorCS2 E18
Normalised Expression
0.3
0.2
0.1
0.0
WT/E18
L100P/E18
Genotype/Stage
SorCS2 P1
Normalised Expression
0.4
0.3
0.2
0.1
0.0
WT/P1
L100P/P1
Genotype/Stage
SorCS2 P7
Normalised Expression
1.5
1.0
0.5
0.0
WT/P7
L100P/P7
Genotype/Stage
SorCS2 P20 *
Normalised Expression
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
WT/P20
L100P/P20
Genotype/Stage
SorCS2 Adult
Normalised Expression
0.25
0.20
0.15
0.10
0.05
0.00
WT/Adult
L100P/Adult
Genotype/Stage

D


SorCS3 E13
Normalised Expression
WT/E13
L100P/E13
Genotype/Stage
SorCS3 E15
Normalised Expression
WT/E15
L100P/E15
Genotype/Stage
SorCS3 E18
Normalised Expression
WT/E18
L100P/E18
Genotype/Stage
SorCS3 P1
Normalised Expression
WT/P1
L100P/P1
Genotype/Stage
SorCS3 P7
Normalised Expression
WT/P7
L100P/P7
Genotype/Stage
SorCS3 P20
Normalised Expression
WT/P20
L100P/P20
Genotype/Stage
SorCS3 Adult
Normalised Expression
WT/Adult
L100P/Adult
Genotype/Stage

E

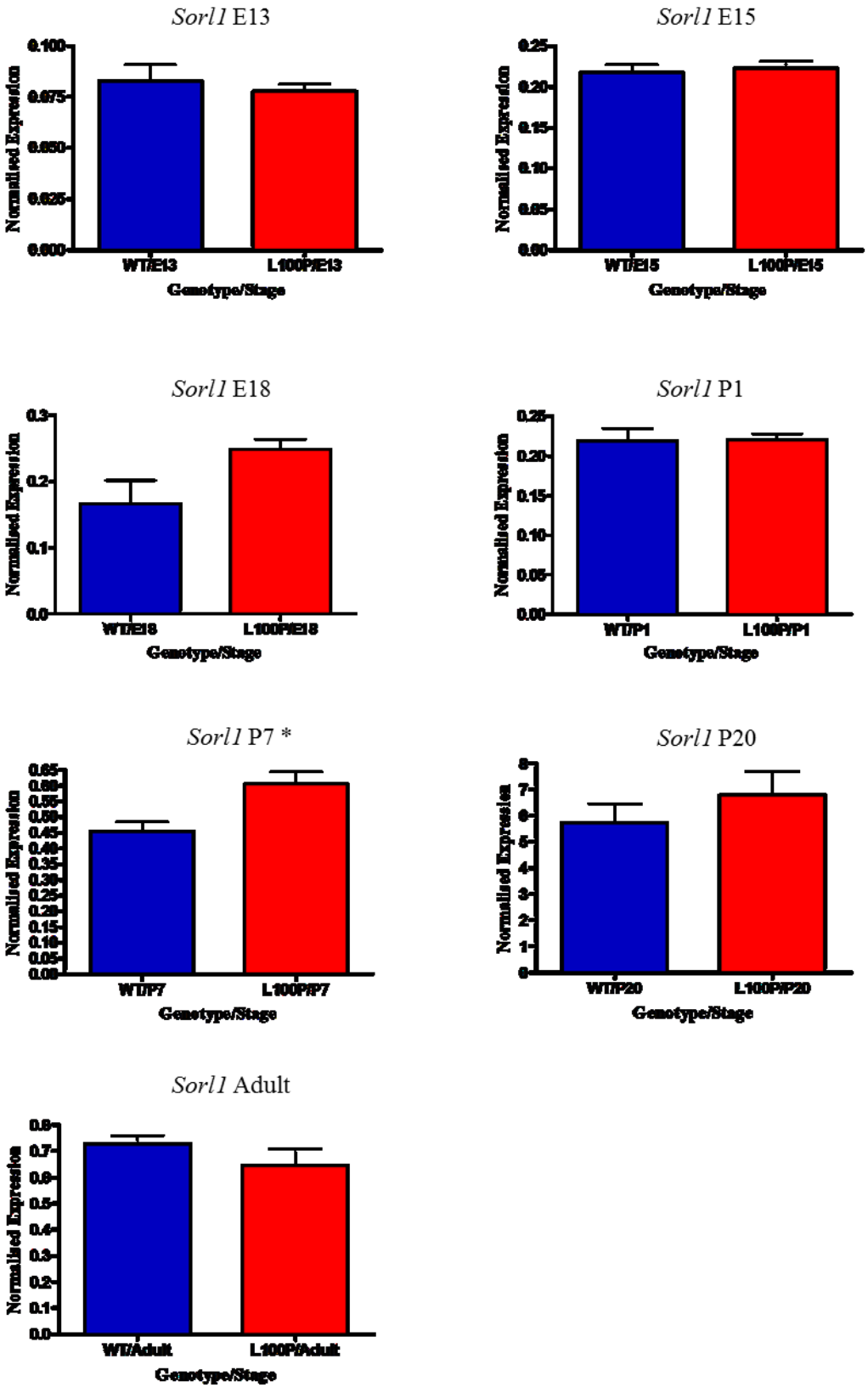

Sorl1 E13
Normalised Expression
WT/E13
L100P/E13
Genotype/Stage
Sorl1 E15
Normalised Expression
WT/E15
L100P/E15
Genotype/Stage
Sorl1 E18
Normalised Expression
WT/E18
L100P/E18
Genotype/Stage
Sorl1 P1
Normalised Expression
WT/P1
L100P/P1
Genotype/Stage
Sorl1 P7 *
Normalised Expression
WT/P7
L100P/P7
Genotype/Stage
Sorl1 P20
Normalised Expression
WT/P20
L100P/P20
Genotype/Stage
Sorl1 Adult
Normalised Expression
WT/Adult
L100P/Adult
Genotype/Stage

Figure 6.2A-E: Comparison of Sortilin family gene expression in 100P heterozygotes and wild-type littermates in the DM samples.

Shown are bar plots for normalised gene expression values (y-axes; arbitrary units) of Sortilin family genes (A-E = Sort1, SorCS1, SorCS2, SorCS3, Sorl1; respectively) in wild-type mice (blue bars) and their 100P homozygous littermates (red bars) from stages E13 to postnatal week 8 (Adult) in the DM samples. Genotype and developmental stage are presented on the x-axes. Error bars show the standard error of the mean. Significant differential expression is indicated by an asterisk (*) in the title of each plot.

6.3 Discussion

The aim of this chapter was to determine whether expression of members of the Sortilin gene family are dysregulated in a mutant mouse model that carries a homozygous nonsynonymous mutation in *Disc1*. This work aimed to address the hypothesis that *Disc1* is involved in regulating Sortilin family gene expression. This hypothesis was based on findings by others in the *Disc1* 100P mouse, along with findings of differential expression and methylation of *SORL1* and *SORCS1*, respectively, in carriers of t(1;11), which disrupts the *DISC1* gene, reported in this thesis (Chapters 3 and 5).

Previous work performed by SB and FS found *Sort1* and *SorCS2* expression to be altered during brain development in the 100P mouse. However, it was possible that these results may have been compromised by the use of a sub-optimal reference gene. These qRT-PCR experiments were repeated using an optimised protocol for reference gene selection. In addition to assessing the expression of *Sort1* and *SorCS2*, the other Sortilin family members, *SorCS1*, *SorCS3* and *Sorl1*, were also assessed for differential expression based on evidence for their dysregulation in the context of other *DISC1* mutations (Chapter 3, Chapter 5, Wen *et al.*, 2014, Srikanth *et al.*, 2015). These experiments were carried out using the SB samples, from which the initial findings by SB and FS were reported. In contrast to the data obtained by SB, *Sort1* showed no differential expression following normalisation to geNorm-recommended reference genes. This may illustrate the importance of a data-driven approach when selecting reference genes for qRT-PCR analyses. However, it is important to note that due to limitations in cDNA availability of the SB samples, only a subset of samples were used

for the geNorm analysis. The geNorm software identifies the most stably-expressed genes across an entire sample set and for each developmental stage, gene expression in the subset of samples was unlikely to be representative of the entire sample set. It is possible that, had another subset of samples been used for geNorm analysis, the recommended reference genes might have differed. Nonetheless, it is likely that the findings obtained following normalisation to geNorm-recommended genes are more valid than those in the analysis performed by SB and FS in these samples, due to their use of a sub-optimal reference assay.

As there was significant differential expression of Sortilin family members in the SB sample, an independent sample set was collected with an aim of replicating these findings. Little overlap was observed with regards to differential expression of Sortilin family members between the two sample sets. Only one result was consistent between these sample sets in terms of significance and direction of effect. It would not be unreasonable to expect this finding to occur by chance, however, as no multiple testing correction was implemented in this study.

There are several possible reasons for the discordance between the findings from the two sets of samples. As mentioned before, the normalisation strategy taken during reanalysis of the SB samples may still have been sub-optimal. Genetic and environmental factors may also have played a role. The mice from which the SB RNA batch was prepared were housed in a facility in Toronto, while the mice from the second batch of samples were housed in Edinburgh. While there is no evidence available for a systematic environmental difference between the two facilities, it is unlikely that pregnant dams and postnatally-collected mice from the SB samples shared identical environments with those from the DM samples prior to culling. Although animal research facilities generally take measures to maintain a controlled environment, variables such as handlers and ambient noise may result in stress, and potential stress-related effects on gene expression (Reinhardt, 2004; Murata *et al.*, 2005). In addition, factors such as season and maternal stress are known to impact

upon gene expression in the developing brain, both in humans and mice (Talge *et al.*, 2007).

Differences in genetic background may have also played a role in the different results obtained from the two sample sets. To generate the 100P mutation, C57/Bl6/J males were injected with ENU and crossed with females from a DBA/2J background. Males heterozygous for the 100P mutation were further backcrossed with C57/Bl6/J females for four generations to obtain a C57/Bl6/J background (Clapcote *et al.*, 2007). There have been conflicting reports of behavioural phenotypes in the 100P mouse. Clapcote *et al.*, (2007) previously reported schizophrenic-like behaviours in these mice including deficits in prepulse inhibition. However, Shoji *et al.* (2012) reported normal prepulse inhibition in 100P mice compared to wild-type. Furthermore, the mice in which Clapcote *et al.* (2007) initially reported the schizophrenic-like behaviours were assessed by Arime *et al.* (2014), who found residual genetic variation from the DBA/2J background. The authors of this study also raised the concern of confounding effects arising from any remaining ENU-induced mutations elsewhere in the genome. To this end, it is of possible note that exome sequencing of the male mouse from which the 100P strain was derived identified 116 variants, including a missense mutation in *SorCS3* (Arime *et al.*. 2014). The mice from the DM samples had undergone additional backcrosses to a C57/Bl6/J background (performed by Malgorzata Borkowska). This should have rendered these mice less likely to be confounded by residual ENU mutations and mixed genetic backgrounds than those from the SB samples. The SB samples were obtained from the same colony as the mice originally reported by Clapcote *et al.* (2007), and were bred approximately five years prior to the DM samples. This analyses of the DM samples might therefore be considered more valid than the analyses using the SB samples.

Evidence of altered *SORL1* expression and *SORCS1* methylation was reported in human t(1;11) samples in Chapters 3 and 5, respectively. It is possible that, in addition to *DISC1*, other consequences of the t(1;11) translocation are responsible for the disrupted regulation of Sortilin gene expression and/or methylation in the t(1;11)

family. Factors such as *DISC1FP1* on chromosome 11, fusion transcripts, or passive transmission of regulatory variants on the derived chromosomes might be responsible for the abnormalities observed in *SORCS1* and *SORL1*. However, should a regulatory relationship exist between *Disc1* and the Sortilin gene family, it may not be affected by the 100P mutation, but rather by disruptions to other regions of *Disc1*. The 100P mutation occurs at the N-terminal region of *Disc1* within exon 2 (Clapcote *et al.*, 2007), whereas the *DISC1* frameshift mutation associated with differential expression of *SORCS1-3* is present in the gene's C-terminal region (Sachs *et al.*, 2005; Wen *et al.*, 2014; Srikanth *et al*., 2015). The 100P substitution has not been reported to affect *Disc1* expression levels. It does, however, result in reduced binding of PDE4B to DISC1 (Clapcote *et al.*, 2007). Perturbed PDE4B activity may result in altered cAMP signalling, presenting a mechanism through which the 100P mutation may lead to dysregulated gene expression in these mice. Although the L100P site is not conserved between humans and mice (Soares *et al.*, 2011), similar functional consequences have may occur through reduced DISC1-PDE4B binding in the context of DISC1 haploinsufficiency (Millar *et al.*, 2005). Moreover, the frameshift mutation reported by Srikanth *et al.* (2015) results in nonsense-mediated decay of *DISC1*. Heterozygotes for this mutation showed decreased levels of *SORCS2*. Investigation of *SORCS2* expression in t(1;11) carrier iPSC-derived neurons is therefore warranted, as translocation carriers display half the normal levels of *DISC1* (Millar *et al.*, 2005)

As described in Chapter 5, neuronal precursors generated from t(1;11) family iPSCs have recently become available. These are likely to be useful resources to further investigate the possible relationship between Sortilin genes and the impact of the translocation event upon *DISC1* in a human neuronal model. This could be achieved through the analysis of gene and protein expression in these samples, as well as analyses of protein interactions. The 100P mouse model could also be improved upon using technologies developed since its original characterisation by Clapcote *et al.* (2007). Genome-editing methods such as CRISPR/Cas9 (Ran *et al.*, 2014) may prove useful in introducing the 100P mutation whilst avoiding the generation of confounding background mutations. Furthermore, this method could be utilised for the generation

of additional mutations in *Disc1*. Such experiments should inform as to which regions of *Disc1*, if any, are responsible for regulating Sortilin family expression.

Using an *in-silico* approach, the relationship between *DISC1* and Sortilin family members was further investigated, with an aim to identify gene-gene interactions which may be related to depressive and/or cognitive phenotypes in humans. This work was performed based on previous associations identified between Sortilin family members and *DISC1* in both cognitive and depressive phenotypes. This work is described in Chapter 7.

Chapter 7

Investigating epistatic interactions between *DISC1* and Sortilin family genes in cognition and depression

7 Investigating epistatic interactions between *DISC1* and Sortilin family genes in cognition and depression

7.1 Introduction

Genome-wide association studies of complex traits have had some success in identifying associations at the level of individual SNPs, taking, for example, the most recent GWASs of MDD (Cai *et al.*, 2015; Hyde *et al.*, 2016). However, such associations go a limited way to explaining the heritability of psychiatric disorders (Crow, 2011), which has been estimated at approximately 37% for MDD (Sullivan *et al.*, 2000). It has been argued that a proportion of this missing heritability could be accounted for by the interactions between already-identified risk loci (Zuk *et al.*, 2012). A statistical challenge presented by large-scale genomic datasets is a large number of observations among a comparatively small number of samples (otherwise known as the "large *p* small *n*" problem; Johnstone and Titterington, 2009). This has rendered the identification of epistatic interactions challenging in with regards to both the computational and statistical burdens involved: assuming a GWAS cohort in which 500,000 SNPs have been genotyped, the number of tests required to assess all pairwise genetic interactions would be in the order of billions (Wei *et al.*, 2014).

Machine learning algorithms (MLAs) may go some way to address this challenge. One such algorithm, random forest analysis, is based on classification and regression trees (CARTs) first introduced by Breiman *et al*. (1984). CART analysis is a method whereby a decision tree is drawn, recursively partitioning data based on known variables (e.g. genotype), in order to predict an outcome (e.g. disease/trait status). This method has been shown to achieve high prediction accuracy, with the ability to provide measures relating to the impact of individual effects (e.g. a single SNP) as well as interactions between variables (e.g. epistatic interactions). Ensemble methods can improve upon CART analysis through the construction of multiple trees (i.e. random forests). Random forest analysis is an ensemble technique involving the construction a forest of multiple, uncorrelated trees, achieved by random sampling of population and predictors for each tree, thereby increasing classification accuracy. After

constructing a forest, each predictor within a given tree is permuted and run through that tree among the unsampled observations to determine that predictor's classification accuracy. The average difference in classification accuracy between a permuted and real predictor is calculated across all trees, creating a variable importance measure for each SNP (VIM; Breiman, 2001). This VIM can then be used as a measure of a given SNP's association with the phenotype of interest in the context of all other SNPs assessed.

Random forest analysis has been suggested as a useful tool for epistatic analyses, by using a variant's associated VIM to rank and select candidate SNPs to address the computational and statistical burden associated with a typical genome-wide approach (Schwarz *et al.*, 2007). The underlying hypothesis of the work performed in this chapter was that *DISC1* interacts genetically with members of the Sortilin gene family. The work below describes the use of random forests to select SNPs to investigate genetic interactions within members the Sortilin gene family and *DISC1*, which may be associated with cognitive phenotypes and/or MDD. These genes were selected based on evidence of a relationship between *DISC1* and Sortilin family gene expression in iPSC-derived neurons containing *DISC1* disruptions (Wen *et al.*, 2014; Srikanth *et al.*, 2015; section 1.7.1), as well as from findings described in this thesis. In Chapter 3, an association was observed between *SORL1* gene expression and the t(1;11) translocation in lymphoblastoid cell lines while in Chapter 5, two differentially methylated regions were observed in *SORCS1* in iPSC-derived neurons from t(1;11) carriers. *DISC1* and Sortilin family genes have been implicated in psychiatric disorders including bipolar disorder and depression (Thomson *et al.*, 2014; Hyde *et al.*, 2016; Baum *et al.*, 2008). Furthermore, *DISC1* has been associated with working memory (Carless *et al.*, 2011) while Sortilin family genes have been implicated in Alzheimer's disease, both independently and through epistatic interactions (Reitz *et al.*, 2013). Analyses were performed using genotype data from the Generation Scotland cohort (Smith *et al.*, 2013).

7.2 Results

7.2.1 Generation of an unrelated sample set prioritising depressed individuals

The Generation Scotland sample is a family-based cohort. As sample relatedness can confound genetic data and increase a genetic bias through the over-representation of genotypes segregating within the families sampled, it is necessary to limit the number of closely related individuals in association studies. A genetic relationship matrix (GRM), was obtained from Dr. Mark Adams (Division of Psychiatry, Edinburgh). The GRM contained genetic relationship estimates based on genotype data from the Generation Scotland cohort (N = 19,994 genotyped for 561,125 SNPs). A total of 40,871 pairs of individuals with a genetic relationship coefficient of ≥ 0.025 (equivalent of an expected level of relatedness in between second and third cousins) were present in the sample. These 40,871 pairs comprised 18,659 different individuals.

As MDD was one of the phenotypes to be tested for association, a method was devised to obtain the maximum number of unrelated individuals while prioritising the retention of depressed individuals. Starting with individuals without a diagnosis of MDD, the individual with the most relatives in the sample set, based on the GRM, was progressively discarded and relationships were recalculated until there were zero such individuals with a relative in the total sample. This was repeated for individuals diagnosed with MDD, resulting in a final sample containing of 7235 individuals. Of these, 2017 were diagnosed with MDD.

7.2.2 Selection of markers for association analysis

The five Sortilin family genes: *SORT1*, *SORCS1*, *SORCS2*, *SORCS3* and *SORL1* were selected, along with *DISC1*, to investigate epistatic interactions associated with cognition and depression. Independent haplotype blocks (D' < 0.8 between blocks), identified from 100 non-depressed unrelated individuals from the Generation Scotland cohort were generated across each gene using Haploview (50 males and 50 females; section 2.13.2.2; Barrett *et al.*, 2005). SNPs were considered for association

if they occurred within the haplotype blocks spanning the start and end coordinates of the each gene (Hg19), or any intermediate blocks. To reduce the multiple testing burden, Haploview's tagger program was used to identify tagging SNPs that capture correlated SNPs ($R^2 \geq 0.8$; de Bakker *et al.*, 2009). Table 7.1 summarises the number of SNPs in each haplotype block, and the number of tag SNPs for each gene.

Gene	Total SNPs in Haplotype Block(s)	Tag SNPs
SORT1	54	28
SORL1	54	37
SORCS1	203	100
SORCS2	326	267
SORCS3	132	66
DISC1	129	87

Table 7.1: Summary of markers selected for assessment of association with cognitive phenotypes and depression.

Shown are the genes, the number of SNPs in each gene's corresponding haplotype blocks defined by Haploview, and the number of tag SNPs identified by Tagger ($R^2 < 0.8$ between tag SNPs).

7.2.3 Investigating genetic interactions associated with cognition

7.2.3.1 *Cognitive phenotypes in Generation Scotland*

The Sortilin gene family, and *DISC1* were assessed for association with cognition. The cognitive phenotypes available from the Generation Scotland cohort are vocabulary (Mill Hill test), verbal fluency, logical memory and digit symbol test performance. To reduce multiple testing, the latter three phenotypes were condensed to a single factor to represent general fluid intelligence, while the vocabulary phenotype was used as a measure of crystallised intelligence (Cattell, 1963). Of the 7235 unrelated individuals identified, 7122 had measurements for crystallised intelligence and 7192 had measurements for general fluid intelligence (i.e. non-missing values for all three of the verbal fluency, logical memory and digit symbol test scores).

7.2.3.2 *Assessment of genetic interactions associated with general fluid intelligence*

Variants in the Sortilin gene family and *DISC1* were assessed for association with cognition. This involved a two-stage process: random forest analysis was performed on 80% of the sample set, firstly using a real phenotype and secondly, on a dataset where the phenotype had been permuted. This created a VIM for each SNP in both the permuted and the real datasets. For each SNP, an empirical *p*-value was generated by comparing its corresponding VIM in the real data to the null distribution of VIMs in the permuted data. This permitted the prioritisation of SNPs for the second step: to test for association with general fluid intelligence by standard regression methods in the remaining 20% of the sample. In the case of general fluid intelligence, 15 SNPs had an empirical *p*–value of ≤ 0.05. These SNPs were within *SORCS1*, *SORCS2* and *DISC1* (Table 7.2). In order to consider *trans*-interactions only to reduce the number of interactions under analysis, only combinations of SNPs within at least two genes were considered. Following the removal of within-gene combinations, 48 two-SNP combinations were assessed, and 291 three-SNP combinations. No significant two-way interactions were observed following correction for multiple testing (Benjamini-Hochberg FDR; $q > 0.32$). However, four interactions were nominally significantly associated with general fluid intelligence ($p \leq 0.05$; Table 7.3). Twenty-eight three-SNP interactions were nominally significantly associated with general fluid intelligence ($p \leq 0.05$; Table 7.4).

SNP	Gene	Empirical *p*-value
rs4689835	*SORCS2*	0.001
rs6855453	*SORCS2*	0.001
rs7695937	*SORCS2*	0.001
rs821634	*DISC1*	0.001
rs4689845	*SORCS2*	0.002
rs4918274	*SORCS1*	0.007
rs7440772	*SORCS2*	0.007
rs821631	*DISC1*	0.021
rs6816649	*SORCS2*	0.029
rs2057723	*SORCS2*	0.037
rs17828052	*SORCS2*	0.042
rs7098468	*SORCS1*	0.043
rs13126941	*SORCS2*	0.044
rs17381732	*SORCS2*	0.044
rs4689838	*SORCS2*	0.045

Table 7.2: SNPs assessed for epistatic interactions associated with general fluid intelligence.

Shown are SNP identifiers, their corresponding genes and empirical p-*values calculated from the random forest analysis.*

SNP1	SNP2	Gene1	Gene2	LRT *p-value*	R^2	*q*-value
rs4918274	rs7695937	*SORCS1*	*SORCS2*	0.006	0.111	0.323
rs4918274	rs6816649	*SORCS1*	*SORCS2*	0.025	0.111	1
rs17828052	rs4918274	*SORCS2*	*SORCS1*	0.034	0.111	1
rs4918274	rs6855453	*SORCS1*	*SORCS2*	0.042	0.111	1

Table 7.3: Two-SNP interactions nominally associated with general fluid intelligence.

Shown is a summary of two-SNP interactions nominally significantly associated with general fluid intelligence ($p < 0.05$*). From left to right, columns state the SNP identifiers and their corresponding genes, the likelihood ratio test* p-*value for the difference in goodness-of-fit of the interaction model versus the null model, omitting the interaction term; the proportion of variance explained by the model (R^2), and the Benjamini-Hochberg FDR* q-*value for the SNP interaction following correction for 48 tests.*

SNP1	SNP2	SNP3	Gene1	Gene2	Gene3	LRT *p*-value	R^2	*q*-value
rs4689835	rs4918274	rs6855453	*SORCS2*	*SORCS1*	*SORCS2*	0.002	0.115	0.675
rs2057723	rs4689838	rs821634	*SORCS2*	*SORCS2*	*DISC1*	0.002	0.110	0.749
rs2057723	rs7098468	rs821634	*SORCS2*	*SORCS1*	*DISC1*	0.004	0.110	1
rs17828052	rs4689845	rs821631	*SORCS2*	*SORCS2*	*DISC1*	0.004	0.112	1
rs17381732	rs4689838	rs821634	*SORCS2*	*SORCS2*	*DISC1*	0.006	0.114	1
rs4689835	rs4918274	rs7695937	*SORCS2*	*SORCS1*	*SORCS2*	0.009	0.110	1
rs17381732	rs7098468	rs7440772	*SORCS2*	*SORCS1*	*SORCS2*	0.009	0.112	1
rs13126941	rs4689835	rs4918274	*SORCS2*	*SORCS2*	*SORCS1*	0.010	0.111	1
rs4689845	rs6855453	rs821631	*SORCS2*	*SORCS2*	*DISC1*	0.010	0.112	1
rs2057723	rs4689845	rs821634	*SORCS2*	*SORCS2*	*DISC1*	0.011	0.110	1
rs2057723	rs7098468	rs821631	*SORCS2*	*SORCS1*	*DISC1*	0.013	0.112	1
rs2057723	rs7695937	rs821631	*SORCS2*	*SORCS2*	*DISC1*	0.020	0.112	1
rs17381732	rs4689838	rs821631	*SORCS2*	*SORCS2*	*DISC1*	0.020	0.114	1
rs17828052	rs4689845	rs821634	*SORCS2*	*SORCS2*	*DISC1*	0.022	0.112	1
rs13126941	rs4689838	rs821634	*SORCS2*	*SORCS2*	*DISC1*	0.024	0.111	1
rs2057723	rs4689838	rs821631	*SORCS2*	*SORCS2*	*DISC1*	0.024	0.110	1
rs4689835	rs4689845	rs7098468	*SORCS2*	*SORCS2*	*SORCS1*	0.025	0.115	1
rs4918274	rs7440772	rs821631	*SORCS1*	*SORCS2*	*DISC1*	0.027	0.111	1
rs4918274	rs7440772	rs821634	*SORCS1*	*SORCS2*	*DISC1*	0.028	0.112	1
rs4689835	rs4689838	rs821631	*SORCS2*	*SORCS2*	*DISC1*	0.028	0.112	1
rs2057723	rs6855453	rs821631	*SORCS2*	*SORCS2*	*DISC1*	0.037	0.111	1

SNP1	SNP2	SNP3	Gene1	Gene2	Gene3	LRT *p*-value	R^2	*q*-value
rs4689838	rs6816649	rs821631	*SORCS2*	*SORCS2*	*DISC1*	0.039	0.112	1
rs17828052	rs4689838	rs821631	*SORCS2*	*SORCS2*	*DISC1*	0.039	0.113	1
rs6816649	rs6855453	rs821631	*SORCS2*	*SORCS2*	*DISC1*	0.039	0.112	1
rs13126941	rs4918274	rs7695937	*SORCS2*	*SORCS1*	*SORCS2*	0.040	0.112	1
rs4689845	rs7440772	rs821634	*SORCS2*	*SORCS2*	*DISC1*	0.041	0.111	1
rs13126941	rs4689838	rs821631	*SORCS2*	*SORCS2*	*DISC1*	0.045	0.113	1
rs4689845	rs821631	rs821634	*SORCS2*	*DISC1*	*DISC1*	0.047	0.115	1

Table 7.4: Three-SNP interactions nominally associated with general fluid intelligence.

Shown is a summary of nominally significant genetic interactions associated with general fluid intelligence ($p \leq 0.05$*). From left to right, columns show the SNP identifiers, their corresponding genes, the likelihood ratio test* p-*value for the difference in goodness-of-fit of the interaction model versus the null model, omitting the interaction term; the proportion of variance explained by the model (R^2), and the FDR* q-*value for the SNP interaction following correction for 291 tests.*

7.2.3.3 *Genetic interactions associated with crystallised intelligence*

Vocabulary, measured using the Mill Hill Vocabulary scale (Raven, 1941), was used as a measure of crystallised intelligence.

Random forest analysis identified 29 SNPs with an empirical *p*–value of ≤ 0.05 for association with crystallised intelligence (Table 7.5). These SNPs, which were located within *DISC1*, *SORCS1*, *SORCS2* and *SORCS3*, were selected for analysis of epistatic interactions in the remaining 20% of the sample set. Following removal of within-gene combinations, 273 between-gene two-way combinations were assessed, along with 3168 three-SNP combinations. No significant two-way interactions were observed following correction for multiple testing for either the two-SNP or three-SNP interactions (FDR $q > 0.32$). However, ten interactions were nominally significantly associated with crystallised intelligence ($p < 0.05$; Table 7.6). One hundred eighty-seven three-SNP interactions were nominally significantly associated with crystallised intelligence ($p < 0.05$; Table 7.7).

SNP	Gene	Empirical *p*-value
rs10032900	*SORCS2*	0.001
rs4350297	*SORCS3*	0.001
rs4613570	*SORCS2*	0.001
rs610785	*SORCS1*	0.001
rs6541281	*DISC1*	0.001
rs665679	*SORCS1*	0.001
rs756255	*SORCS2*	0.001
rs7667970	*SORCS2*	0.002
rs6835799	*SORCS2*	0.003
rs12730369	*DISC1*	0.004
rs2269850	*SORCS2*	0.004
rs10937826	*SORCS2*	0.005
rs1557816	*SORCS2*	0.005
rs2269852	*SORCS2*	0.005
rs1336979	*SORCS1*	0.007
rs2295959	*DISC1*	0.008
rs10884100	*SORCS3*	0.014
rs1565415	*SORCS3*	0.014
rs2107182	*SORCS2*	0.017
rs11932646	*SORCS2*	0.02
rs4637403	*SORCS2*	0.021
rs17466832	*SORCS2*	0.023
rs7897974	*SORCS1*	0.024
rs4689869	*SORCS2*	0.028
rs7440772	*SORCS2*	0.03
rs823162	*DISC1*	0.031
rs9432040	*DISC1*	0.032
rs4918288	*SORCS1*	0.04
rs1251753	*SORCS1*	0.043

Table 7.5: SNPs assessed for epistatic interactions associated with crystallised intelligence.

Shown are SNP identifiers, their corresponding genes and empirical p-*values calculated from the random forest analysis.*

SNP1	SNP2	Gene1	Gene2	LRT *p*- value	R^2	*q*-value
rs10032900	rs1565415	*SORCS2*	*SORCS3*	0.003	0.080	0.753
rs10937826	rs1565415	*SORCS2*	*SORCS3*	0.015	0.079	1
rs2269850	rs6541281	*SORCS2*	*DISC1*	0.016	0.077	1
rs11932646	rs1565415	*SORCS2*	*SORCS3*	0.016	0.078	1
rs11932646	rs4350297	*SORCS2*	*SORCS3*	0.017	0.078	1
rs10884100	rs11932646	*SORCS3*	*SORCS2*	0.018	0.077	1
rs1565415	rs4613570	*SORCS3*	*SORCS2*	0.029	0.078	1
rs4637403	rs6541281	*SORCS2*	*DISC1*	0.031	0.078	1
rs2269852	rs6541281	*SORCS2*	*DISC1*	0.034	0.079	1
rs610785	rs7667970	*SORCS1*	*SORCS2*	0.043	0.077	1

Table 7.6: Two-SNP interactions nominally associated with crystallised intelligence.

Shown is a summary of two-SNP interactions nominally significantly associated with crystallised intelligence ($p < 0.05$*). From left to right, columns state the SNP identifiers and their corresponding genes, the likelihood ratio test* p-*value for the difference in goodness-of-fit of the interaction model versus the null model, omitting the interaction term; the proportion of variance explained by the model (R^2), and the Benjamini-Hochberg FDR* q-*value for the SNP interaction following correction for 273 tests.*

SNP1	SNP2	SNP3	Gene1	Gene2	Gene3	LRT *p*-value	R^2	*q*-value
rs12730369	rs1336979	rs6541281	*DISC1*	*SORCS1*	*DISC1*	0.0001	0.08	0.36
rs1565415	rs2107182	rs823162	*SORCS3*	*SORCS2*	*DISC1*	0.0002	0.08	0.79
rs17466832	rs6541281	rs823162	*SORCS2*	*DISC1*	*DISC1*	0.0003	0.08	0.98
rs6835799	rs7440772	rs823162	*SORCS2*	*SORCS2*	*DISC1*	0.0003	0.08	1
rs1565415	rs756255	rs823162	*SORCS3*	*SORCS2*	*DISC1*	0.0004	0.08	1
rs1336979	rs2269852	rs4918288	*SORCS1*	*SORCS2*	*SORCS1*	0.0004	0.08	1
rs1565415	rs2269852	rs823162	*SORCS3*	*SORCS2*	*DISC1*	0.0005	0.08	1
rs2107182	rs6541281	rs823162	*SORCS2*	*DISC1*	*DISC1*	0.0006	0.08	1
rs10937826	rs1557816	rs6541281	*SORCS2*	*SORCS2*	*DISC1*	0.0007	0.08	1
rs6541281	rs756255	rs823162	*DISC1*	*SORCS2*	*DISC1*	0.0009	0.08	1

Table 7.7: Three-SNP interactions nominally significantly associated with crystallised intelligence.

Shown is a summary of nominally significant genetic interactions associated with crystallised intelligence ($p \leq 0.05$*). From left to right, columns show the SNP identifiers and their corresponding genes the likelihood ratio test* p-*value for the difference in goodness-of-fit of the interaction model versus the null model, omitting the interaction term; the proportion of variance explained by the model (R^2), and the Benjamini-Hochberg FDR* q-*value for the SNP interaction following correction for 3168 tests. Shown are the top ten interactions ranked by* p-*value. A full list of three-SNP interactions and their associations with crystallised intelligence is presented in Appendix I (Table A8).*

7.3 Genetic interactions associated with MDD

Two- and three-SNP interactions between DISC1 and Sortilin family genes were tested for association with MDD. Following random forest analysis of the permuted and real data, a total of 11 SNPs had an empirical *p* –value of ≤ 0.05 for association with MDD (Table 7.8). These SNPs were located in *SORCS2*, *SORCS3* and *DISC1*. Thirty-six two-SNP interactions were assessed for association with MDD. No significant interactions were identified following correction for multiple testing (FDR $q \geq 0.297$). However, one nominally significant interaction was observed ($p = 0.007$; Table 7.9). In the analysis of three-SNP interactions, 144 SNP combinations were assessed. No interactions were significantly associated with MDD following correction for multiple testing (FDR $q = 1$). However, 11 interactions were nominally significantly associated with MDD ($p < 0.05$; Table 7.10).

SNP	Gene	Empirical *p*-value
rs7692314	*SORCS2*	0.001
rs11122324	*DISC1*	0.002
rs12256390	*SORCS3*	0.002
rs4689789	*SORCS2*	0.01
rs10012347	*SORCS2*	0.012
rs970054	*SORCS3*	0.013
rs7679804	*SORCS2*	0.016
rs12040259	*DISC1*	0.021
rs4689682	*SORCS2*	0.034
rs1174741	*SORCS3*	0.04
rs10021084	*SORCS2*	0.049

Table 7.8: SNPs assessed for epistatic interactions associated with MDD.

Shown are SNP identifiers, their corresponding genes and empirical p-*values calculated from the random forest analysis.*

SNP1	SNP2	Gene1	Gene2	LRT *p*- value	R^2	*q*-value
rs12256390	rs10012347	*SORCS3*	*SORCS2*	0.007	0.12	0.258

Table 7.9: Two-SNP interactions nominally associated with MDD.

*Shown is a summary of two-SNP interactions nominally significantly associated with MDD (*p *< 0.05). From left to right, columns state the SNP identifiers and their corresponding genes the likelihood ratio test* p-*value for the difference in goodness-of-fit of the interaction model versus the null model, omitting the interaction term; the proportion of variance explained by the model (R^2), and the Benjamini-Hochberg FDR* q-*value for the SNP interaction following correction for 36 tests.*

SNP1	SNP2	SNP3	Gene1	Gene2	Gene3	LRT *p*-value	R^2	*q*-value
rs4689789	rs12040259	rs4689682	*SORCS2*	*DISC1*	*SORCS2*	0.009	0.11	1
rs11122324	rs4689789	rs4689682	*DISC1*	*SORCS2*	*SORCS2*	0.009	0.11	1
rs970054	rs12040259	rs4689682	*SORCS3*	*DISC1*	*SORCS2*	0.009	0.11	1
rs12256390	rs970054	rs7679804	*SORCS3*	*SORCS3*	*SORCS2*	0.011	0.11	1
rs11122324	rs970054	rs4689682	*DISC1*	*SORCS3*	*SORCS2*	0.012	0.11	1
rs11122324	rs970054	rs7679804	*DISC1*	*SORCS3*	*SORCS2*	0.028	0.12	1
rs10012347	rs970054	rs10021084	*SORCS2*	*SORCS3*	*SORCS2*	0.033	0.11	1
rs7692314	rs1174741	rs10021084	*SORCS2*	*SORCS3*	*SORCS2*	0.030	0.12	1
rs12256390	rs12040259	rs4689682	*SORCS3*	*DISC1*	*SORCS2*	0.039	0.11	1
rs11122324	rs10012347	rs4689682	*DISC1*	*SORCS2*	*SORCS2*	0.046	0.11	1
rs970054	rs7679804	rs12040259	*SORCS3*	*SORCS2*	*DISC1*	0.045	0.11	1

Table 7.10: Three-SNP interactions nominally associated with MDD.

*Shown are the three-SNP interactions nominally significantly associated with MDD (*p ≤ *0.05).From left to right, columns show the SNP identifiers and their corresponding genes the likelihood ratio test* p-*value for the difference in goodness-of-fit of the interaction model versus the null model, omitting the interaction term; the proportion of variance explained by the model (R^2), and the Benjamini-Hochberg FDR* q-*value for the SNP interaction following correction for 144 tests.*

7.4 Discussion

The aim of this chapter was to investigate whether epistatic interactions between SNPs in Sortilin family genes and *DISC1* contribute to variation in cognitive and depressive phenotypes. Variation in *DISC1* has been associated with working memory and depression (Carless *et al.*, 2011; Thomson *et al*., 2015) while others have reported both independent and epistatic associations between Sortilin family genes and Alzheimer's disease (Reitz *et al.*, 2013). Moreover, multiple lines of evidence suggest roles for *DISC1* and the Sortilin genes in psychiatric illness, as described in Chapters 1.4 and 1.5 while the work described in Chapter 3 and Chapter 5 reported nominally significant expression and methylation differences in *SORL1* and *SORCS1*, respectively, in t(1;11) carriers.

Prior to performing the analysis, a set of genetically unrelated individuals was generated from the total sample set, prioritising individuals with a diagnosis of MDD over those without. This was performed with an aim to maximise the number of cases, and therefore power to detect any significant genetic associations with MDD. This list of unrelated individuals was circulated to those studying MDD in the Generation Scotland cohort as a resource for subsequent analyses.

Random forest analysis was performed to identify SNPs most likely to interact as determined by their permutation-based *p*-values for association with depressive and cognitive phenotypes. For each phenotype, two- and three-SNP interactions were assessed. No significant interactions were observed after correction for multiple testing. Several nominally significant two-way and three-way interactions were observed in all three of the phenotypes examined ($p < 0.05$). Testing the nominally significant interactions identified here in an independent and/or larger population may, however, provide support for these findings.

In the case of depression, it is possible that disease heterogeneity and/or sample size had a role in the failure to detect significant epistatic interactions. Early large-scale

association studies of MDD had failed to detect variants significantly associated with MDD (Ripke *et al.*, 2013). However, Cai *et al.* (2015) reported significant genetic associations with depression through reducing heterogeneity amongst cases by limiting their study to Chinese females with severe (hospitalised) depression. More recently, Hyde *et al.* (2016) reported 15 significant associations with MDD, including a SNP in *SORCS3*, in the largest GWAS of the disorder to date consisting of over 300,000 individuals of European ancestry. The success of Cai *et al.* (2015) reflects the importance of phenotype homogeneity for the detection of genetic associations in MDD. It is likely, however, that due to the limited sample size of the Generation Scotland cohort, subdividing the depressed individuals to improve phenotypic homogeneity would render the current analysis underpowered. Conversely, Hyde *et al.* (2016) demonstrated the efficiency of using a large study cohort with less intensive phenotyping. It is possible that significant interactions may be observed in a larger cohort, such as UK Biobank.

It is possible that significant interactions between variants in *DISC1* and/or Sortilin family members might be detected when assessed in the presence of variants from other genes that are involved in common functions or pathways. An example of one such function common to DISC1, SORL1, SORT1 and SORCS1 is the proteolytic processing of APP (Reitz *et al.*, 2011; Gustafsen *et al.*, 2012; Shahani *et al.*, 2015). APP processing can occur through the action of either α-secretases or β-secretases, both of which are followed by γ-secretase processing (Zhang *et al.*, 2011). Dysregulated expression of *SORL1* and an α-secretase gene, *ADAM10*, was observed in t(1;11) carrier lymphoblastoids (described in Chapter 3). It is possible that interactions exist between *DISC1*, Sortilin family members and key players in the APP processing pathway, such as *APP* and/or the α-, β-, and γ-secretase-encoding genes. Processing of APP by action of β-secretases results in the generation of neurotoxic amyloid-β plaques: a hallmark of Alzheimer's disease pathology associated with cognitive deficits (Olsson *et al.*, 2003). Conversely, processing via the α-secretases results in the release of the neurotrophic soluble APP-α, levels of which have been positively correlated with cognitive function in rats (Anderson *et al.*, 1999). Assessment of interactions between *DISC1*, Sortilin family members and genes in the

APP processing pathway; and their association with cognition and Alzheimer's disease may be worthy of further investigation.

Interactions may also exist between *DISC1* and the Sortilin family in other psychiatric phenotypes. In addition to depression, there is some evidence for involvement of these genes in susceptibility to bipolar disorder and schizophrenia (Hennah *et al.*, 2003; Hennah *et al.*, 2009; Baum *et al.*, 2008; Montano *et al.*, 2016). It was not possible to assess this hypothesis as Generation Scotland did not record information on schizophrenia diagnosis and only 70 individuals in the cohort have a diagnosis of bipolar disorder. Further investigation in alternative datasets, such as those used by the psychiatric genomics consortium (PGC), may inform as to whether the interactions observed in this study are associated with other psychiatric disorders.

In summary, no genetic interactions were observed between variants in the Sortilin genes and *DISC1* that were significantly associated with cognition or depression in the Generation Scotland cohort. However, nominally significant associations were observed for all phenotypes assessed. Investigation of these interactions in larger independent cohorts and/or other psychiatric phenotypes is warranted. Ultimately, the functional effects of any variants of interest should be assessed in a laboratory setting to determine whether there exists a biological basis for the observed interactions.

Chapter 8

Discussion

8 Discussion

There were two aims to this thesis. The primary aim was to investigate the genome-wide effects of a psychiatric illness-associated translocation, t(1;11), on gene expression and DNA methylation in patient-derived samples. A secondary aim was to investigate a potential relationship between *DISC1*, the protein-coding gene disrupted by the t(1;11) translocation, and the Sortilin family of genes. This aim was predicted on both evidence of a regulatory relationship between *Disc1* and Sortilin gene family members observed in a *Disc1* mutant mouse (Brown *et al.*, unpublished data), and evidence for dysregulation of Sortilin family members in iPSC-derived neurons containing mutations disrupting *DISC1* (Wen *et al.*, 2014; Srikanth *et al.*, 2015).

To address the primary aim, transcriptomic and methylomic analyses were performed on three patient-derived sources: lymphoblastoid RNA, whole blood DNA, and iPSC-derived neuronal DNA. The secondary aim was addressed using both *in-vitro* and *in-silico* approaches, using brain tissue from the 100P *Disc1* mutant mouse, and genetic data from the Generation Scotland cohort, respectively.

This chapter will provide a summary of findings in this thesis, discuss how these might relate to the field of psychiatric genetics, highlight limitations to the work presented in this thesis, and recommend future work.

8.1 Overall findings

8.1.1 Analysis of gene expression in t(1;11) individuals

Chapter 3 described a genome-wide analysis of gene expression in lymphoblastoid – derived RNA, comparing individuals with the t(1;11) translocation (n = 8) to their karyotypically normal relatives (n = 5). This work was based on the hypothesis that the t(1;11) translocation might affect gene expression, mediated via several possible routes. As discussed in section 3.1, potential causal mechanisms include *DISC1* haploinsufficiency (Millar *et al.*, 2005), *DISC2* disruption, *DISC1FP1/Boymaw*

disruption, generation of deleterious fusion transcripts (Eykelenboom *et al.*, 2010), or through the passive transmission, due to linkage disequilibrium, of regulatory variants on the derived chromosomes. Genome-wide transcription was assessed in these samples using the Illumina HT-12 platform. No between-group differences were observed at a genome-wide significant level (FDR $q \leq 0.05$). However, imposing a relaxed *p*-value cut-off ($p \leq 0.05$) coupled with an absolute fold-change threshold of ≥ 1.25 revealed 303 differentially expressed genes as candidates for further follow-up. Among these genes was *SORL1*, a member of the Sortilin gene family. This finding was successfully validated in a targeted follow-up using qRT-PCR.

Previously, an analysis of gene expression was performed by Xu Tang on a separate growth of lymphoblastoids from the same individuals included in this study. Here, both microarray and qRT-PCR data supported dysregulation of *DLGAP1*, *HIPK2* and *SV2B*, among others (Table 3.3). *HIPK2* was among the genes meeting the criteria for nominally significant differential expression in the current study: a finding which was successfully validated by qRT-PCR. Furthermore, although *SV2B* and *DLGAP1* were not amongst the differentially expressed genes on the microarray, they were found to be differentially expressed by qRT-PCR analysis, with the same direction of effect as observed by Xu Tang. This study also supported a relationship between *DISC1* and Sortilin genes: carriers of the translocation showed downregulation of *SORL1*. This was observed both on the microarray and by qRT-PCR.

Other studies of gene expression in the context of *DISC1*-disrupting mutations have been performed on iPSC-derived neuronal samples (Wen *et al.*, 2014; Srikanth *et al.*, 2015). These studies have reported increased Wnt signaling, altered neuronal fate and altered expression of synaptic genes. Dysregulation of the synaptic genes *DLGAP1* and *SV2B* was confirmed by qRT-PCR in two separate growths of t(1;11) lymphoblastoid samples, further supporting the role of *DISC1* in synaptic function. These findings would suggest that normal DISC1 function is required to regulate the neuronal expression of genes involved in processes such as neurodevelopment and

synaptic function. This is supported by neuroanatomical findings in individuals carrying certain *DISC1* variants/genomic rearrangements, including the t(1;11) translocation. Doyle *et al.* (2015) reported reduced cortical thickness in t(1;11) carriers compared to their non-carrying relatives, while Whalley *et al.* (2015) reported an association between the t(1;11) translocation and reduced white matter integrity. Moreover, others have reported reduced grey matter volumes were associated with the *DISC1* variant rs6675281, suggestive of synaptic and neurodevelopmental deficits (Cannon *et al.*, 2005; Trost *et al.*, 2013).

Assessment of lymphoblastoid cell lines has revealed that individuals with the t(1;11) translocation display half the levels of *DISC1* seen in their karyotypically normal relatives, both at the mRNA and protein level (Millar *et al.*, 2005), presenting a possible mechanism through which its disruption might impact upon gene expression. Overlap was observed between dysregulated genes observed in t(1;11) carriers and those dysregulated in iPSC-derived neurons containing *DISC1* mutations (Wen *et al.*, 2014; Srikanth *et al.*, 2015). Of these genes, *HIPK2* may be of note as upregulation was observed in t(1;11) carriers using two separate growths of lymphoblastoid-derived samples: a finding confirmed by microarray and qRT-PCR analyses in both sets of samples. Although not implicated in psychiatric illness by GWAS, its function in promoting neuronal survival may be relevant to psychiatric illness (Zhang *et al.*, 2007; Jarskog *et al.*, 2005). Srikanth *et al.* (2015) reported downregulation of *HIPK2* in neurons with a *DISC1* frameshift mutation in exon 2: a mutation which targets all known coding isoforms of *DISC1*. The difference observed with regards to *HIPK2* expression could be attributed to independent effects of different mutations in *DISC1*. For example, upregulation of *HIPK2* may be associated with *DISC1* haploinsufficiency in t(1;11) carriers while downregulation may be associated with the effects of the *DISC1* frameshift mutation. Srikanth *et al.* (2015) reported normal *DISC1* expression levels in these cells. Alternatively, other consequences of the translocation (described in section 3.1) might impact upon *HIPK2* expression in a *DISC1*-independent manner. Multiple PDE4B binding sites are present in exon 2 of *DISC1* (Soares *et al.*, 2011). Should these be affected by the mutation reported by Srikanth *et al.* (2015), it may present a mechanism through

which dysregulation of gene expression occurs through the cAMP signaling pathway (Millar *et al.*, 2005; section 1.4.2).

The work carried out in Chapter 3 also supported a previously-reported role for DISC1 in the APP processing pathway (Shahani *et al.*, 2015). SORL1 has been shown by others to modulate levels of APP processing in which it inhibits both amyloidogenic and non-amyloidogenic processing (Gustafsen *et al.*, 2013; section 7.4). Shahani *et al*. (2015) reported a decrease in non-amyloidogenic processing and an increase in amyloidogenic processing following *DISC1* knockdown. Downregulation of *SORL1* was observed in t(1;11) carriers, which, along with *DISC1* haploinsufficiency, would be predicted to affect APP processing in t(1;11) carriers. This disruption to the APP processing pathway might occur independently, or through the joint effects of *DISC1* and *SORL1* misexpression. *SORL1* is located on chromosome 11, approximately 30 Mb telomeric to the translocation breakpoint. A possible mechanism for its observed downregulation in t(1;11) carriers is the presence of linkage disequilibrium between alleles associated with *SORL1* expression, and the translocation. The same might also apply to other differentially expressed genes on chromosomes 1 and 11. In Chapter 4 of this thesis, a potential mechanism for differential methylation at the t(1;11) breakpoint was observed. This pertained to the transmission of meQTLs in LD with the translocation causing genome-wide significant difference in methylation at regions surrounding the t(1;11) breakpoints. However, no differences in methylation were observed in *SORL1* in either blood or iPSC-derived neuronal DNA from the family. Experimental follow-up to investigate the effects of the t(1;11) translocation on APP processing was not possible as not all of the constituent genes of the APP processing pathway are expressed (according to the expression array) in these cells. Should these genes be expressed in the t(1;11) family iPS-derived neurons, further investigation using these samples might determine whether the translocation is associated with altered APP processing. Such experiments could assess the levels of soluble products of APP processing in conditioned media (i.e. sAPP-α and sAPP-β), to determine whether the translocation is associated with abnormal levels of APP processing by either α- or β-

secretase activity. Further experiments could assess whether DISC1 and SORL1 act together or independently in the APP processing pathway.

8.1.2 DNA methylation in t(1;11) individuals

Chapter 4 described the assessment of methylation in whole blood DNA samples from the t(1;11) family. There were two aims to this chapter. Firstly, to determine whether the t(1;11) translocation was associated with differential methylation in blood, and secondly, to assess whether t(1;11) carriers with a psychotic disorder show differences in methylation compared to t(1;11) carriers with a non-psychotic disorder. Thirteen sites were found to be significantly differentially methylated between translocation carriers and non-carriers (FDR $q < 0.05$). Of these, four sites were within *DISC1*. A comparison between translocation carriers with psychotic and non-psychotic diagnoses revealed three significantly differentially-methylated sites. Distribution of the methylation signal at these loci suggested that the findings at these three sites may be due to a technical artefact arising from impaired probe binding due to genetic variation. Analysis of differentially methylated regions revealed several loci with multiple nominally significant differentially methylated sites in both the carrier vs. non-carrier comparison, and the comparison of individuals with psychotic and non-psychotic diagnoses.

The observation of differential methylation around the t(1;11) breakpoint regions may have been due to the passive transmission of meQTLs in linkage disequilibrium with the translocation. In the comparison of t(1;11) carriers with a psychotic disorder to those with a non-psychotic disorder, three sites were significantly differentially methylated. At all three loci, genetic variation was discovered at the probe binding site, resulting in genotype-specific signal. In both comparisons, differentially methylated regions (DMRs) were identified in genes previously implicated in psychiatric illness by GWAS and candidate gene studies.

All but one of the significantly differentially methylated loci between t(1;11) carriers and non-carriers occurred in the regions of the translocation breakpoints. These

findings are in agreement with a previous study of differential methylation in a leukaemia-associated t(11;14) translocation (Walker *et al.* 2011). Here, the top six ranked DMPs were hypomethylated and located within one of the genes directly disrupted by the translocation.

A potential mechanism whereby differential methylation occurs at translocation breakpoint regions was identified in this study. Seven of the 13 t(1;11) associated DMPs were previously reported to be associated with variation at independent loci (meQTLs; Lemire *et al.*, 2015). Of these seven DMPs, five were found to be influenced by *cis*-acting meQTLs, all on either chromosome 1 or 11. Furthermore, these meQTLs were significantly associated with t(1;11) carrier status. Linkage disequilibrium between the translocation and regulatory variants may be a pathogenic mechanism in t(1;11) carriers. It is likely that other variants in LD with the translocation affect gene expression in addition to methylation by acting as expression QTLs. In addition, differential methylation mediated by meQTLs may result in altered gene expression (Shin *et al.*, 2015). It is possible that such consequences could affect downstream pathways relating to pathogenesis.

Overlap was observed between findings from this analysis and a recent blood-based epigenome-wide study of schizophrenia performed by Montano *et al*. (2016). Among other genes, Montano *et al.* (2016) reported differential methylation in *MAD1L1, RPTOR* and *DDR1*. Differentially methylated regions were observed in each of these genes when comparing t(1;11) carriers with a psychotic disorder to those with a non-psychotic disorder. *MAD1L1* has also been implicated in schizophrenia and bipolar disorder by GWAS (Ruderfer *et al.*, 2014; Ripke *et al.*, 2014; Hou *et al*., 2016). *RPTOR* has previously been implicated in bipolar disorder, having been identified at the breakpoint of a chromosomal translocation found in a case with bipolar disorder (Rajkumar et al., 2015). Roig *et al.* (2007) proposed *DDR1* as a susceptibility gene for schizophrenia based on evidence of its reduced expression in schizophrenic patients carrying a variant showing significant association with the disorder.

However, *RPTOR* and DDR1 have not been implicated in psychiatric illness by GWAS.

Differential methylation of genes relating to immune function was common to both the comparison of t(1;11) carriers and non-carriers, and the comparison of psychotic and non-psychotic illness in t(1;11) carriers. This adds support to a long-standing observation of immune system dysfunction in psychiatric illness (reviewed in Upthegroves and Barnes, 2014). Recent evidence further implicating the immune system in psychiatric illness has come from GWAS and studies of gene expression and function (Ripke *et al*., 2014; Sekar *et al.*, 2016).

Two significantly differentially methylated sites were observed in t(1;11) carriers within *EGLN1*. *EGLN1* plays a role in the transcriptional response to hypoxia through regulation of the hypoxia-inducible factor HIF-1α. Hypoxia is a known obstetric risk factor for schizophrenia. Methylation levels at both of these sites in *EGLN1* were found to be associated with genotype at nearby meQTLs in linkage disequilibrium with the translocation. Should differential methylation of this gene affect transcription in response to oxygen levels, this finding might indicate a mechanism for increased risk of illness through a t(1;11)-associated gene-environment interaction.

Chapter 5 described the assessment of DNA methylation in iPSC-derived neurons from six members of the t(1;11) family. There were two aims to this chapter. The first was to determine whether the t(1;11) translocation was associated with differential methylation in a cellular model more physiologically relevant to psychiatric illness. The second aim was to permit the comparison of DNA methylation levels in blood to those in neuron-like cells from t(1;11) family members. This work was performed using the Infinium HumanMethylationEPIC BeadChip. As this array has only recently been developed, prior to addressing the first two aims of this Chapter 5, it was necessary to identify probes on this array that (i) are affected by polymorphisms and/or (ii) have the potential to cross-hybridise.

Lists of these probes have been published in order to provide a resource for the research community (McCartney *et al.*, 2016; Appendix 1). No significant differences in methylation were observed between translocation carriers and non-carriers. Nine of the top ten loci ranked by *p*-value for differential methylation were found to be driven by a single individual, with consistently higher levels of DNA methylation compared to the remaining five samples. Differentially-methylated regions were identified containing multiple probes nominally significantly associated with t(1;11) carrier status. Of the regions identified, two were within *SORCS1*, a member of the Sortilin gene family. This, along with the finding of differential expression of *SORL1* in Chapter 3, provided justification to further investigate a relationship between *DISC1* and Sortilin family members.

Analysis of DMRs identified a number of genes previously implicated in psychiatric illness including *COMT*, a long time candidate gene for schizophrenia described in section 1.3.2.1. This DMR was hypomethylated in iPSC-derived neurons from t(1;11) carriers. Others have reported hypomethylation in *COMT* in the frontal lobes of *post-mortem* brains of schizophrenia and bipolar disorder patients, accompanied with increased *COMT* expression (Abdolmaleky *et al.*, 2006). Hypomethylation of *COMT* has also been observed in saliva-derived DNA of schizophrenia and bipolar disorder patients (Nohesara *et al.*, 2011). Dempster *et al.* (2006) did not find a difference in *COMT* methylation in the cerebellum of *post-mortem* brains of patients with schizophrenia, bipolar disorder or depression. This is perhaps unsurprising, as Hannon *et al.* (2015) have reported distinct methylation profiles between the cerebellum and cortical regions. Egan *et al.* (2001) proposed a mechanism whereby COMT enzyme activity regulates synaptic dopamine levels (section 1.3.2.1). This could point to a *COMT*-mediated effect on the dopaminergic system in t(1;11) carriers, which may be associated with increased risk of illness. As this finding is one of 424 DMRs, validation by targeted methods such as pyrosequencing and qRT-PCR in the t(1;11) iPSC-derived neurons is recommended to first confirm the presence of differential methylation at *COMT* in these individuals, and secondly, determine whether this has an effect on gene expression.

Two DMRs were discovered in *SORCS1*, supporting a regulatory relationship between *DISC1* and the Sortilin family. Both regions were hypomethylated in t(1;11) carriers. Although it was not possible to assess gene expression differences in these samples, others have reported downregulation of *SORCS1* in iPSC-derived neurons containing a 4 bp *DISC1* frameshift mutation (Wen *et al.*, 2014). RNA-seq analysis is currently ongoing which will permit assessment of the relationship between *SORCS1* expression and methylation, and whether differential expression of *SORL1* can be observed, as seen in t(1;11) family LCLs.

Numerous GO terms relating to synaptic function and transmission were over-represented by genes harbouring the most significantly differentially methylated loci. *DISC1* localises to synapses where it regulates dendritic spine morphology and function, as well as levels of post-synaptic density proteins (Hayashi-Takagi *et al.*, 2010; Wang *et al.*, 2011). Moreover, Wen *et al.* (2014) reported dysregulation of synaptic gene expression in iPSC-derived neurons with a 4 bp frameshift mutation in *DISC1*. Dysregulation of expression and methylation of synaptic genes was reported in the work performed in this thesis, including *SV2B* and *DLGAP2*. SV2B is a synaptic vesicle protein which functions in regulating presynaptic calcium levels (Wan *et al.*, 2010). Calcium signaling is a key regulatory pathway in the brain and has been implicated in both schizophrenia and bipolar disorder (Berridge, 2014). *SV2B* had previously been found to be upregulated in t(1;11) lymphoblastoid samples by microarray expression and qRT-PCR. This finding was replicated in RNA from a separate growth of lymphoblastoid samples, as reported in Chapter 3. DLGAP2 is a postsynaptic density protein which has been implicated in neuropsychiatric disorders including autism and schizophrenia (Marshall *et al.*, 2008; Li *et al.*, 2014). DMRs were reported in *DLGAP2* in iPSC-derived neurons and blood from the t(1;11) family in Chapter 4. Taken together, these findings might support a pathogenic mechanism via synaptic dysfunction in carriers of the translocation. Electrophysiological assessment of the iPSC-derived neurons from the t(1;11) family would aid in addressing this hypothesis.

In the blood-based analysis of DNA methylation in the t(1;11) family, significant differential methylation was reported at the translocation breakpoints. This was possibly due to the influence of meQTLs in LD with the translocation. There were no significant differences in DNA methylation at the breakpoint regions in iPSC-derived neurons. Investigation of the blood-based meQTLs in iPSC-derived neurons showed no significant associations between genotype and methylation. This may have been due to limitations in sample size. Only one of the five meQTLs identified in 41 blood samples was also detected when subsetting the sample to the same six individuals profiled for DNA methylation in iPSC-derived neurons. This suggests that assessment of a larger sample would be required in order to identify meQTLs in the iPSC-derived neurons.

8.1.3 Investigating a potential relationship between *DISC1* and the Sortilin gene family

Chapter 6 described an analysis of developmental gene expression of Sortilin family members in a mouse model containing a missense L100P mutation in exon 2 of *Disc1*. These mice had previously been reported to display schizophrenic-like behaviours (Clapcote *et al.*, 2007) and transcriptional dysregulation (Lipina *et al.*, 2012), which were ameliorated following antipsychotic and anticonvulsant treatment, respectively. This work aimed to investigate a regulatory relationship between *Disc1* and Sortilin family genes based on findings of dysregulated expression of *SorCS1* and *SorCS2* in the brains of 100P mice by others (Brown *et al.*, unpublished). The samples assessed by Brown *et al.* were reanalysed using an optimised normalisation strategy revealing dysregulated Sortilin genes at several developmental stages in the mutant mice. However, attempts to validate these findings in an independently-collected batch of RNA were unsuccessful.

There are several possible causes for the discordance of findings between the two batches of samples. One possible reason pertains to differences in the genetic backgrounds of the mice studied. Arime *et al*. (2014) reported residual ENU-generated mutations in the mice assessed by Clapcote *et al.* (2007), including a missense mutation in *SorCS3*. The SB samples were more closely related to these

mice than the DM samples, as they were obtained from the same colony as those reported by Clapcote *et al.* (2007). This suggests they were more likely to be affected by a mixed genetic background. Although the analysis performed on the DM samples is likely to be more valid due to reduced genetic heterogeneity, the degree of residual ENU mutations in these samples has not been ascertained. It is possible, should a regulatory relationship exist between *Disc1* and the Sortilin genes in these mice, the potential confounder of a mixed genetic background would mask any effects attributable to this relationship.

Additional analyses were performed to investigate a relationship between these genes in a human system, using an *in*-silico approach. Chapter 7 described an exploratory investigation of whether variants in *DISC1* and Sortilin family genes interact epistatically to exert an effect on cognitive phenotypes and/or depression. Epistatic interactions have been hypothesised to partially account for the missing heritability of complex traits (Eichler *et al.*, 2010). Machine learning algorithms are an attractive option to assess epistatic interactions in genomics data as they can efficiently address the computational and statistical burden associated with traditional association methods (Wei *et al.*, 2014). Random forest analysis was selected because of its ability to score each variable on their predictive abilities, whilst considering the effects of other variables (Lunetta *et al.*, 2004). Others have successfully identified genetic interactions associated with psychiatric illness using this method. Nicodemus *et al.*, (2010) reported genetic interactions associated with schizophrenia risk using multiple machine learning algorithms, including random forests. *DISC1* and the five members of the Sortilin gene family were assessed on the basis of evidence of dysregulation of Sortilin family genes in the context of *DISC1* mutations, which might suggest a relationship between these genes (Wen *et al.*, 2014; Srikanth *et al.*, 2015). Depression and cognition were investigated as *DISC1* and Sortilin family genes have been implicated in these phenotypes (Carless *et al.*, 2011; Thomson *et al.*, 2014; Hyde *et al.*, 2016; Reitz *et al.*, 2011). Random forest analysis was used to identify SNPs likely to interact in each phenotype. Interactions between these SNPs were then assessed for association with cognition and depression using likelihood ratio tests to compare the goodness-of-fit of a regression model containing an

interaction term; to a null model, omitting the interaction term. Although no significant interactions were observed following correction for multiple testing (all FDR $q > 0.05$). However, several two- and three- way interactions were nominally significantly associated with the phenotypes assessed ($p \leq 0.05$).

Although the absence of significant findings may suggest there are no epistatic interactions to be detected between *DISC1* and Sortilin family members in the phenotypes tested, it is possible that phenotypic heterogeneity contributed to the failure to detect significant interactions. This is likely to be the reason for the failure to detect significant associations with MDD in early GWASs (Ripke *et al.*, 2013). More recent studies have identified genome-wide significant associations with MDD through reducing phenotypic heterogeneity by considering only females with severe illness (Cai *et al.*, 2015), or increasing sample size (Hyde *et al.*, 2016). However, due to the limited sample size involved in the current analysis, stratification of the phenotype would likely impede the power of this study to detect significant interactions associated with MDD. Moreover, in order to stratify by phenotype, detailed measures of illness, such as information regarding hospitalisation, would be required. Repeating the analysis in a larger sample of individuals would likely be a more accurate assessment of epistasis between these genes, and may support the nominally significant findings reported here.

8.2 Limitations

8.2.1 Potential confounders

An important point to consider when interpreting the findings of the comparisons between t(1;11) carriers and non-carriers is the presence of factors which may be correlated with translocation carrier status. These factors (e.g. smoking status, alcohol intake) may, at least in part, be responsible for any differences observed in the above analyses. Although surrogate variable analysis was performed on the data to identify and remove latent sources of variation, this would not be sufficient to eliminate factors which are highly correlated with translocation carrier status. Further follow-up of t(1;11) family members is recommended, if possible, in order to attain

information pertaining to factors known to contribute to gene expression, DNA methylation, and increased risk of major mental illness.

8.2.2 Multiple study tissues

There are several limitations to consider when interpreting findings from the work presented in this thesis. The use of three different study tissues (i.e. LCLs, whole blood and iPSC-derived neurons) in the analysis of the t(1;11) family presents a challenge when drawing parallels between the results of Chapters 3-5. Moreover, each study tissue has its own drawbacks which must be acknowledged when considering the findings from the three t(1;11)-based studies.

Whether the differences in DNA methylation translate to effects on gene expression remains to be ascertained. It was not possible to determine whether changes in methylation had an effect on gene expression in the individuals assessed in this chapter, as RNA was not available. Although RNA is available from LCLs and iPSC-derived neurons from the t(1;11) family members, the overlap with the individuals for whom whole blood DNA was available is limited. Furthermore, inherent gene expression differences between blood, LCLs and neuronal cells would render findings from inter-tissue comparisons of DNA methylation and gene expression difficult to interpret.

With regards to Chapter 3, it is unlikely that LCLs will faithfully recapitulate all gene expression differences present in the brains of t(1;11) carriers. In part, this is because several brain-expressed genes are not expressed in blood-derived cell lines (Sullivan *et al.*, 2006). Moreover, cell passage numbers were not available for these samples. Long-term culture has been shown to result in the accumulation of random mutations, which, if present in these samples, might have confounded the results (Mohyuddin *et al.*, 2004). Use of whole blood in this analysis would have been a more attractive option, as it would have allowed for a more direct comparison with the findings from Chapter 4, with the additional benefit of being free of possible confounders resulting from long-term culture. Whole-blood derived samples, however, are not without their

disadvantages – particularly when studying disorders of the brain, as in the methylation analysis described in Chapter 4. Peripheral blood is an attractive tissue for the analysis of DNA methylation, due to its accessibility. However, differential methylation in blood may not always correlate with differential methylation in the brain (Walton *et al.*, 2015; Hannon *et al.*, 2015). Equally importantly, whether these differences in methylation are correlated with DNA methylation and gene expression in the brain cannot be determined with certainty. Smith *et al.* (2014) have reported a significant overlap of meQTLs between blood and brain. Should differences in methylation observed in this study be influenced by such meQTLs, this may suggest similar differences in methylation occur in the brains of these individuals. This represents a limitation of blood-based studies of psychiatric illness. Should meaningful conclusions be drawn from such studies, supporting evidence from a more physiologically-relevant tissue would be necessary. An attempt to address this limitation is presented in Chapter 5, in which DNA methylation was profiled in iPSC-derived neurons from six t(1;11) family members who were profiled for DNA methylation in the blood-based analysis. However, no overlap was evident between the findings of Chapter 4 and Chapter 5. This may be due to limitations in the ability of iPSC-derived neurons to model the brain. Although more physiologically relevant to psychiatric illness than whole blood, iPSC-derived neurons cannot faithfully model a human brain, in which different cell subtypes interact in a three-dimensional environment. Furthermore, iPSC-derived neurons are not representative of adult neurons: Mariani *et al.* (2012) reported significant similarities between the transcriptomes of human iPSC-derived neurons and foetal neurons between 8 and 10 weeks post-conception while no correlation was observed between patient age and DNA methylation age in the iPSC-derived neurons in this analysis.

8.2.3 Sample size

Although the use of a closely-related pedigree should go some way to reduce genetic variation between samples, thereby increasing power; the work presented in Chapters 3 and 5 was further hindered by the small sample size, which may have rendered these studies underpowered to detect genome-wide significant differences in gene expression and methylation, should any exist in these samples. Moreover, within the

six individuals profiled for DNA methylation in iPSC-derived neurons, gender was confounded with translocation carrier status. This may have resulted in sex-specific effects on DNA methylation being misinterpreted as t(1;11)-associated effects. It is also possible that t(1;11) effects were inadvertently removed after covarying for gender in the regression model. Although sex chromosome-targeting probes were removed from this analysis in an attempt to address the relationship between t(1;11) status and gender, it is known that autosomal sites also display sex-associated differences in DNA methylation (Liu *et al.*, 2010). Furthermore, a single sample appeared to drive the most differentially-methylated sites in t(1;11) samples in Chapter 5. It is difficult to conclude whether this outlier sample was artefactual, however, considering such a small sample. Generation of iPSC-derived neurons from additional members, followed by further methylation profiling and appropriate batch effect corrections may go some way to clarify the distribution of DNA methylation levels in the family. Optimally, these samples should be matched to those profiled in blood, to permit a direct comparison between the two tissues.

With regards to the work perfomed in Chapter 7, a cohort of approximately 20,000 individuals was used. However, to remove potentially confounding effects of shared genetic sequences amongst related individuals, only 7000 unrelated individuals were included in this analysis. The sample size was further reduced when splitting the cohort, with 80% included in the random forest analysis and the remaining 20% included for the validation stage. It is possible that, should interactions between *DISC1* and Sortilin family genes contribute to small effects in cognition and/or depression, such effects may not be detected due to a low sample number.

8.2.4 DNA hydroxymethylation

An additional limitation to this study concerns DNA hydroxymethylation. It is not possible to distinguish between DNA methylation and hydroxymethylation using the bisulphite-based method of detection used by the Infinium HumanMethylation450 BeadChip (Nestor *et al.*, 2013). Any methylation differences reported in this thesis therefore might also reflect changes in hydroxymethylation.

DNA hydroxymethylation is abundant in neurons, where it is thought to play a regulatory role (Mellén *et al.*, 2012). Moreover, Hahn *et al.* (2013) reported increased DNA hydroxymethylation during neurogenesis in embryonic mouse brains. Should the t(1;11) family iPSC-derived samples be representative of embryonic neurons, it is possible that increased DNA hydroxymethylation levels are present. A further possibility is that changes DNA methylation and hydroxymethylation might have cancelled each other out at some loci, resulting in a failure to identify any differences in DNA methylation. Dissection of these individual effects by methods such as oxidative bisulphite conversion is therefore warranted.

8.2.5 The L100P mouse

The 100P mouse was an attractive option to investigate a relationship between *DISC1* and Sortilin family members in Chapter 6, due to previous findings of differential expression by Brown *et al.* (unpublished), and reports of behaviours reminiscent of psychiatric disorders and their amelioration by antipsychotic and antidepressant medications (Clapcote *et al.*, 2007). As it has since emerged that these mice were genetically heterogeneous due to residual ENU-derived mutations, the original findings of differentially-expressed Sortilin family members may be compromised. Differences in expression of Sortilin family genes have been reported in the context of *DISC1* mutations in human iPSC-derived neurons (Wen *et al.*, 2014; Srikanth *et al.*, 2015). This might suggest that a regulatory relationship does indeed exist between *DISC1* and these genes, but is either human-specific or not modeled by the 100P mutation. Due to the differences in complexity between the brains of mice and humans, the former is a strong possibility.

Tissue heterogeneity from the use of whole brains is a confounding factor as murine *SorCS1-3* have shown region-specific expression patterns in the brain (Hermey *et al.*, 2001; Hermey *et al.*, 2004). It is therefore possible that the 100P mutation exerts different effects on Sortilin family expression in different brain regions. This may

render 100P-associated expression differences difficult to detect in whole-brain samples, through the dilution of small effects in individual brain regions.

8.3 Future work

8.3.1 Recommended future work on t(1;11) samples

It is likely that the limited sample sizes in Chapters 3 and 5 have contributed to increase in both type I and type II errors. Therefore, follow-up of findings in additional individuals is warranted. Moreover, differences in study tissue are also likely to be obstructive in permitting a direct assessment in the relationship between gene expression and DNA methylation in the t(1;11) family. To this end, the iPSC-derived neurons may prove a valuable resource for future analyses. Fibroblasts from additional family members are available which will provide an increased sample size upon differentiation to iPSC-derived neurons. Others are currently carrying out RNA-seq analysis in the iPSC-derived neurons from the t(1;11) family described in Chapter 5 of this thesis. Protein is also available from these cells which will permit investigation of the relationship between gene expression, DNA methylation and protein expression in t(1;11) individuals. Furthermore, genetic data is available from all individuals profiled for DNA methylation in the iPSC-derived neurons. This data could be applied to the gene expression data, when available, to investigate whether illness- or translocation-associated eQTLs are present in these samples. This should also be utilised in combination with the methylation data to identify iPSC-derived neuronal meQTLs for comparison with those present in blood. However, sample size may pose an issue in the identification of such sites, and power analysis would be recommended before performing an association analysis in order to generate robust results. Such studies might be informative as to which sites in blood are likely to be affected by differential methylation in the brain, permitting the selection of candidate loci to examine their possible roles in psychiatric illness.

As the findings from Chapter 4 and Chapter 5 may, at least in part, be due to differences in DNA hydroxymethylation, oxidative bisulphite treatment of the DNA

is recommended to dissect any effects of DNA hydroxymethylation from those of DNA methylation (Booth *et al.*, 2013).

Brain imaging data from the family is also now available from which t(1;11)-associated deficits in white matter integrity and cortical thinning have been reported (Whalley *et al.*, 2015; Doyle *et al.*, 2015). Future experiments could also assess the relationship between DNA methylation, gene expression and imaging data in order to identify possible effects of differential DNA methylation and gene expression on neurodevelopment in the family, which might be associated with illness or the translocation.

8.3.2 Recommended future work to investigate the DISC1-Sortilin relationship

Since the generation of the 100P mouse, more efficient means of mutagenesis have been developed, such as CRISPR-Cas9 gene editing (Ran *et al.*, 2013). Mutant animal models derived using this method are less likely to be confounded by residual mutations (Arime *et al.*, 2014). A possible future experiment could involve the introduction of the 100P mutation by means of gene editing, coupled with screening for additional, off-target mutations using high-throughput sequencing methods. Any findings of differential gene expression observed in such a model would be more robustly associated with the 100P mutation than the findings in the ENU model, due to the reduced heterogeneity of its genetic background. Furthermore, investigation of other *Disc1* mouse models (Johnstone *et al.*, 2011) might also be warranted to determine whether expression differences in Sortilin genes vary according to the type of disruption to *Disc1*. The 100P mutation involves a single amino acid change in exon 2 *Disc1* (Clapcote *et al.*, 2007). It is possible, however, that other disruptions to *Disc1* might affect Sortilin family expression. To this end, a CRISPR-Cas9 mutagenised model might be of use to induce mutations in candidate regions of *Disc1* to dissect this potential regulatory relationship. Additionally, *in-vitro* knock-out and knock-down studies might inform of such a relationship.

Assessing the relationship between *DISC1* and Sortilin family genes in a human system should also be informative. In Chapter 3, t(1;11)-associated downregulation of *SORL1* was reported in LCLs; while in Chapter 5, two hypomethylated *SORCS1* DMRs were identified in t(1;11) iPSC-derived neurons. Investigation of the expression and/or functions of these genes in the iPSC-derived samples from the t(1;11) family, or iPSC-derived neurons carrying CRISPR-Cas9-induced *DISC1* variants; may go some way to support or refute the hypothesis of a regulatory relationship between *DISC1* and Sortilin family genes.

Although no genome-wide significant interactions were identified in the analysis in Chapter 7, future analysis in a larger cohort might be informative as to whether significant interactions are associated with depression or cognition when increasing the sample size and/or reducing phenotypic heterogeneity. This could be addressed through the inclusion of additional cohorts, such as UK Biobank: a cohort over 500,000 genotyped individuals (Sudlow *et al.*, 2015). Furthermore, inclusion of additional genes which may interact with *DISC1* and Sortilin family genes may be necessary to detect genetic interactions. Should any interactions be observed by these methods, investigation of their role in a biological system would be warranted in order to identify mechanisms through which they contribute to the phenotype of interest. This could be through gene expression analyses to determine the co-expression patterns of interacting genes. Further investigation involving co-immunoprecipitation or mass spectrometry analyses could also be performed to assess whether such genetic interactions translate to a relationship at the protein level. The availability of t(1;11) neuronal precursors and iPSC-derived neurons may be a useful resource for such experiments.

Restricting the genes studied to *DISC1* and Sortilin family genes may have posed a limitation in identifying significant interactions. The study was limited to these genes to reduce the statistical and computational burden involved with investigating combinations of a large set of genes. However, it is possible that *DISC1* and Sortilin family members interact in the context of other genes that were not included in this

analysis. Selection of additional candidates that might occur in an interaction complex with *DISC1* and Sortilin genes may result in the identification of significant interactions.

8.4 Conclusion

Although the work presented here has provided suggestive evidence for t(1;11) associated differential methylation and gene expression, caution should be exerted when drawing links studies described in this thesis, due to the use of three different study tissues and low sample sizes, particularly in Chapters 3 and 5. The recent generation of iPSC derivatives from the family may address this due to the availability of concurrently-harvested DNA, RNA and protein. At the time of writing this thesis, gene expression, proteomic and morphological analyses are ongoing for t(1;11) family iPSC-derived neurons. There is also scope to generate iPSC-derived neurons from additional family members, permitting improvements in sample size. In addition, whole-genome sequence data is now available from the family. This is likely to be a key resource for the identification of additional risk factors pertaining to illness in the t(1;11) family. Future work should aim to integrate findings from gene expression, protein expression and DNA methylation studies of these samples, whilst also considering the available genetic data, in order to identify mechanisms through which increased risk of illness is conferred in these individuals. Such mechanisms may relate both functional and regulatory variants associated with the translocation, and their effects on gene expression, DNA methylation and protein function in t(1;11) carriers.

This work presented in this thesis has involved the use of a pedigree with a highly penetrant disease-associated mutation to identify genes and processes which might relate to a psychiatric phenotype. It is clear that the t(1;11) translocation has complex and far-reaching effects beyond the disruption of the breakpoint genes as illustrated by the expression and methylation analyses presented in this thesis. Furthering our knowledge of the genes and pathways disrupted by the translocation should lead to identification of mechanisms conferring increased risk of illness in the family, which

may extend to the general population. Ultimately, such findings should improve our understanding of psychiatric disorders, permitting the development of more effective treatments.

References

References

1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., … Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. doi:10.1038/nature15393

Abdolmaleky, H. M., Cheng, K.-H., Faraone, S. V, Wilcox, M., Glatt, S. J., Gao, F., … Thiagalingam, S. (2006). Hypomethylation of MB-COMT promoter is a major risk factor for schizophrenia and bipolar disorder. *Human Molecular Genetics*, *15*(21), 3132–45. doi:10.1093/hmg/ddl253

Abdolmaleky, H. M., Yaqubi, S., Papageorgis, P., Lambert, A. W., Ozturk, S., Sivaraman, V., & Thiagalingam, S. (2011). Epigenetic dysregulation of HTR2A in the brain of patients with schizophrenia and bipolar disorder. *Schizophrenia Research*, *129*(2-3), 183–90. doi:10.1016/j.schres.2011.04.007

Agostini, M., Tucci, P., Steinert, J. R., Shalom-Feuerstein, R., Rouleau, M., Aberdam, D., … Melino, G. (2011). microRNA-34a regulates neurite outgrowth, spinal morphology, and function. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(17), 1–6. doi:10.1073/pnas.1112063108

Albert, P. R., Benkelfat, C., Beaudet, A., Descarries, L., Descarries, L., Beaudet, A., … Cowen, P. (2013). The neurobiology of depression--revisiting the serotonin hypothesis. II. Genetic, epigenetic and clinical studies. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *368*(1615), 20120535. doi:10.1098/rstb.2012.0535

Allen, N. C., Bagade, S., McQueen, M. B., Ioannidis, J. P. A., Kavvoura, F. K., Khoury, M. J., … Bertram, L. (2008). Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nature Genetics*, *40*(7), 827–834. doi:10.1038/ng.171

American Psychiatric Association. (2013). *American Psychiatric Association, 2013. Diagnostic and statistical manual of mental disorders (5th ed.)*. *American Journal of Psychiatry*. doi:10.1176/appi.books.9780890425596.744053

Anderson, J. J., Holtz, G., Baskin, P. P., Wang, R., Mazzarelli, L., Wagner, S. L., & Menzaghi, F. (1999). Reduced cerebrospinal fluid levels of alpha-secretase-cleaved amyloid precursor protein in aged rats: correlation with spatial memory deficits. *Neuroscience*, *93*(4), 1409–20. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10501466

Andreasen, J. T., Gynther, M., Rygaard, A., Bøgelund, T., Nielsen, S. D., Clausen, R. P., … Pickering, D. S. (2013). *Does increasing the ratio of AMPA-to-NMDA receptor mediated neurotransmission engender antidepressant action? Studies*

in the mouse forced swim and tail suspension tests. *Neuroscience Letters* (Vol. 546). doi:10.1016/j.neulet.2013.04.045

Arime, Y., Fukumura, R., Miura, I., Mekada, K., Yoshiki, A., Wakana, S., … Akiyama, K. (2014). Effects of background mutations and single nucleotide polymorphisms (SNPs) on the Disc1 L100P behavioral phenotype associated with schizophrenia in mice. *Behavioral and Brain Functions : BBF*, *10*, 45. doi:10.1186/1744-9081-10-45

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., & Irizarry, R. a. (2014). Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*, *30*(10), 1363–1369. doi:10.1093/bioinformatics/btu049

Badner, J. A., & Gershon, E. S. (2002). Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Molecular Psychiatry*, *7*(4), 405–411. doi:10.1038/sj.mp.4001012

Barber, R. D., Harmer, D. W., Coleman, R. A., & Clark, B. J. (2005). GAPDH as a housekeeping gene: analysis of GAPDH mRNA expression in a panel of 72 human tissues. *Physiological Genomics*, *21*(3), 389–95. doi:10.1152/physiolgenomics.00025.2005

Barch, D. M., Bustillo, J., Gaebel, W., Gur, R., Heckers, S., Malaspina, D., … Carpenter, W. Logic and justification for dimensional assessment of symptoms and related clinical phenomena in psychosis: Relevance to DSM-5. , 150 Schizophrenia Research 15–20 (2013). doi:10.1016/j.schres.2013.04.027

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics*, *21*(2), 263–265. doi:10.1093/bioinformatics/bth457

Barsh, G. S., Copenhaver, G. P., Gibson, G., & Williams, S. M. (2012). Guidelines for Genome-Wide Association Studies. *PLoS Genetics*, *8*(7), e1002812. doi:10.1371/journal.pgen.1002812

Bassett, A. S., Chow, E. W. C., Husted, J., Weksberg, R., Caluseriu, O., Webb, G. D., & Gatzoulis, M. A. (2005). Clinical features of 78 adults with 22q11 Deletion Syndrome. *American Journal of Medical Genetics. Part A*, *138*(4), 307–13. doi:10.1002/ajmg.a.30984

Baum, A. E., Akula, N., Cabanero, M., Cardona, I., Corona, W., Klemens, B., … McMahon, F. J. (2008). A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. *Molecular Psychiatry*, *13*(2), 197–207. doi:10.1038/sj.mp.4002012

Baum, A. E., Hamshere, M., Green, E., Cichon, S., Rietschel, M., Noethen, M. M., … McMahon, F. J. (2008). Meta-analysis of two genome-wide association studies of bipolar disorder reveals important points of agreement. *Molecular Psychiatry*, *13*(5), 466–7. doi:10.1038/mp.2008.16

Bauman, A. L., Goehring, A. S., & Scott, J. D. (2004). Orchestration of synaptic plasticity through AKAP signaling complexes. *Neuropharmacology*, *46*(3), 299–310. doi:10.1016/j.neuropharm.2003.09.016

Bavamian, S., Mellios, N., Lalonde, J., Fass, D. M., Wang, J., Sheridan, S. D., … Haggarty, S. J. (2015). Dysregulation of miR-34a links neuronal development to genetic risk factors for bipolar disorder. *Molecular Psychiatry*, *20*(5), 573–84. doi:10.1038/mp.2014.176

Bell, J. T., & Spector, T. D. (2011). A twin approach to unraveling epigenetics. *Trends in Genetics*. doi:10.1016/j.tig.2010.12.005

Bellivier, F., Leboyer, M., Courtet, P., Buresi, C., Beaufils, B., Samolyk, D., … Malafosse, A. (1998). Association between the tryptophan hydroxylase gene and manic-depressive illness. *Archives of General Psychiatry*, *55*(1), 33–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9435758

Benes, F. M. (2006). Strategies for improving sensitivity of gene expression profiling: regulation of apoptosis in the limbic lobe of schizophrenics and bipolars. *Progress in Brain Research*, *158*, 153–72. doi:10.1016/S0079-6123(06)58008-2

Benes, F. M., Lim, B., & Subburaju, S. (2009). Site-specific regulation of cell cycle and DNA repair in post-mitotic GABA cells in schizophrenic versus bipolars. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(28), 11731–6. doi:10.1073/pnas.0903066106

Benjamini, Y., & Hochberg, Y. (1995). Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, *57*(1), 289–300. doi:10.2307/2346101

Berko, E. R., Suzuki, M., Beren, F., Lemetre, C., Alaimo, C. M., Calder, R. B., … Greally, J. M. (2014). Mosaic epigenetic dysregulation of ectodermal cells in autism spectrum disorder. *PLoS Genetics*, *10*(5), e1004402. doi:10.1371/journal.pgen.1004402

Berra, E., Benizri, E., Ginouvès, A., Volmat, V., Roux, D., & Pouysségur, J. (2003). HIF prolyl-hydroxylase 2 is the key oxygen sensor setting low steady-state levels of HIF-1alpha in normoxia. *The EMBO Journal*, *22*(16), 4082–90. doi:10.1093/emboj/cdg392

Berridge, M. J. (2014). Calcium signalling and psychiatric disease: bipolar disorder and schizophrenia. *Cell and Tissue Research*, *357*(2), 477–92. doi:10.1007/s00441-014-1806-z

Bhutani, N., Burns, D. M., & Blau, H. M. (2011). DNA demethylation dynamics. *Cell*. doi:10.1016/j.cell.2011.08.042

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J. M., … Shen, R. (2011). High density DNA methylation array with single CpG site resolution. *Genomics*, *98*(4), 288–95. doi:10.1016/j.ygeno.2011.07.007

Blackwood, D. H., He, L., Morris, S. W., McLean, a, Whitton, C., Thomson, M., … Muir, W. J. (1996). A locus for bipolar affective disorder on chromosome 4p. *Nature Genetics*, *12*, 427–430. doi:10.1038/ng0496-427

Blackwood, D. H. R., Fordyce, A., Walker, M. T., St. Clair, D. M., Porteous, D. J., & Muir, W. J. (2001). Schizophrenia and Affective Disorders—Cosegregation with a Translocation at Chromosome 1q42 That Directly Disrupts Brain-Expressed Genes: Clinical and P300 Findings in a Family. *The American Journal of Human Genetics*, *69*(2), 428–433. doi:10.1086/321969

Boks, M. P. M., Rietkerk, T., van de Beek, M. H., Sommer, I. E., de Koning, T. J., & Kahn, R. S. (2007). Reviewing the role of the genes G72 and DAAO in glutamate neurotransmission in schizophrenia. *European Neuropsychopharmacology*, *17*(9), 567–572. doi:10.1016/j.euroneuro.2006.12.003

Bolton, P. F., Veltman, M. W. M., Weisblatt, E., Holmes, J. R., Thomas, N. S., Youings, S. A., … Brown, J. (2004). Chromosome 15q11-13 abnormalities and other medical conditions in individuals with autism spectrum disorders. *Psychiatric Genetics*, *14*(3), 131–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15318025

Bonoldi, I., & Howes, O. D. (2014). Presynaptic dopaminergic function: implications for understanding treatment response in psychosis. *CNS Drugs*, *28*(7), 649–63. doi:10.1007/s40263-014-0177-z

Booth, M. J., Ost, T. W. B., Beraldi, D., Bell, N. M., Branco, M. R., Reik, W., & Balasubramanian, S. (2013). Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nature Protocols*, *8*(10), 1841–1851. doi:10.1038/nprot.2013.115

Bracaglia, G., Conca, B., Bergo, A., Rusconi, L., Zhou, Z., Greenberg, M. E., … Kilstrup-Nielsen, C. (2009). Methyl-CpG-binding protein 2 is phosphorylated by homeodomain-interacting protein kinase 2 and contributes to apoptosis. *EMBO Reports*, *10*(12), 1327–33. doi:10.1038/embor.2009.217

Bradshaw, N. J., Soares, D. C., Carlyle, B. C., Ogawa, F., Davidson-Smith, H., Christie, S., … Millar, J. K. (2011). PKA phosphorylation of NDE1 is DISC1/PDE4 dependent and modulates its interaction with LIS1 and NDEL1. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(24), 9043–54. doi:10.1523/JNEUROSCI.5410-10.2011

Brand, H., Pillalamarri, V., Collins, R. L., Eggert, S., O'Dushlaine, C., Braaten, E. B., … Doyle, A. E. (2014). Cryptic and complex chromosomal aberrations in early-onset neuropsychiatric disorders. *American Journal of Human Genetics*, *95*(4), 454–61. doi:10.1016/j.ajhg.2014.09.005

Brandon, N. J., & Sawa, A. (2011). Linking neurodevelopmental and synaptic theories of mental illness through DISC1. *Nature Reviews. Neuroscience*, *12*(12), 707–22. doi:10.1038/nrn3120

Bray, N. J., Buckland, P. R., Williams, N. M., Williams, H. J., Norton, N., Owen, M. J., & O'Donovan, M. C. (2003). A haplotype implicated in schizophrenia susceptibility is associated with reduced COMT expression in human brain. *American Journal of Human Genetics*, *73*(1), 152–61. doi:10.1086/376578

Breiderhoff, T., Christiansen, G. B., Pallesen, L. T., Vaegter, C., Nykjaer, A., Holm, M. M., … Rubio, M. (2013). Sortilin-Related Receptor SORCS3 Is a Postsynaptic Modulator of Synaptic Depression and Fear Extinction. *PLoS ONE*, *8*(9), e75006. doi:10.1371/journal.pone.0075006

Breiman, L. (2001). Random forests. *Machine Learning*, 5–32. doi:10.1023/A:1010933404324

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. *The Wadsworth statisticsprobability series* (Vol. 19).

Brennand, K. J., Simone, A., Jou, J., Gelboin-Burkhart, C., Tran, N., Sangar, S., … Gage, F. H. (2011). Modelling schizophrenia using human induced pluripotent stem cells. *Nature*, *473*(7346), 221. doi:10.1038/nature09915

Brennand, K., Savas, J. N., Kim, Y., Tran, N., Simone, A., Hashimoto-Torii, K., … Gage, F. H. (2015). Phenotypic differences in hiPSC NPCs derived from patients with schizophrenia. *Molecular Psychiatry*, *20*(3), 361–368. doi:10.1038/mp.2014.22

Brown, S. M., Clapcote, S. J., Millar, J. K., Torrance, H. S., Anderson, S. M., Walker, R., … Evans, K. L. (2011). Synaptic modulators Nrxn1 and Nrxn3 are disregulated in a Disc1 mouse model of schizophrenia. *Molecular Psychiatry*, *16*(6), 585–587. doi:10.1038/mp.2010.134

Burton, P. R., Clayton, D. G., Cardon, L. R., Craddock, N., Deloukas, P., Duncanson, A., … Worthington, J. (2007). Genome-wide association study of

14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, *447*(7145), 661–678. doi:10.1038/nature05911

Busche, S., Ge, B., Vidal, R., Spinella, J.-F., Saillour, V., Richer, C., … Pastinen, T. (n.d.). Integration of High-Resolution Methylome and Transcriptome Analyses to Dissect Epigenomic Changes in Childhood Acute Lymphoblastic Leukemia. *Cancer Res*, *73*(14), 4323–36. doi:10.1158/0008-5472.CAN-12-4367

Butcher, L. M., & Beck, S. (2015). Probe Lasso: a novel method to rope in differentially methylated regions with 450K DNA methylation data. *Methods*, *72*, 21–28. doi:10.1016/j.ymeth.2014.10.036

Buttenschøn, H. N., Demontis, D., Kaas, M., Elfving, B., Mølgaard, S., Gustafsen, C., … Glerup, S. (2015). Increased serum levels of Sortilin are associated with depression and correlated with BDNF and VEGF. *Translational Psychiatry*, *5*, e677. doi:10.1038/tp.2015.167

Byrne, M., Agerbo, E., Bennedsen, B., Eaton, W. W., & Mortensen, P. B. (2007). Obstetric conditions and risk of first admission with schizophrenia: a Danish national register based study. *Schizophrenia Research*, *97*(1-3), 51–9. doi:10.1016/j.schres.2007.07.018

Cai, N., Bigdeli, T. B., Kretzschmar, W., Li, Y., Liang, J., Song, L., … Flint, J. (2015). Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature*, *523*(7562), 588–591. doi:10.1038/nature14659

Callicott, J. H., Straub, R. E., Pezawas, L., Egan, M. F., Mattay, V. S., Hariri, A. R., … Weinberger, D. R. (2005). Variation in DISC1 affects hippocampal structure and function and increases risk for schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(24), 8627–32. doi:10.1073/pnas.0500515102

Camargo, L. M., Collura, V., Rain, J.-C., Mizuguchi, K., Hermjakob, H., Kerrien, S., … Brandon, N. J. (2007). Disrupted in Schizophrenia 1 Interactome: evidence for the close connectivity of risk genes and a potential synaptic basis for schizophrenia. *Molecular Psychiatry*, *12*(1), 74–86. doi:10.1038/sj.mp.4001880

Cannon, T. D., Hennah, W., van Erp, T. G. M., Thompson, P. M., Lonnqvist, J., Huttunen, M., … Peltonen, L. (2005). Association of DISC1/TRAX haplotypes with schizophrenia, reduced prefrontal gray matter, and impaired short- and long-term memory. *Archives of General Psychiatry*, *62*(11), 1205–13. doi:10.1001/archpsyc.62.11.1205

Cannon, T. D., Thompson, P. M., Van Erp, T. G. M., Toga, A. W., Poutanen, V.-P., Huttunen, M., … Kaprio, J. (n.d.). Cortex mapping reveals regionally specific patterns of genetic and disease-specific gray-matter deficits in twins discordant for schizophrenia.

Carless, M. A., Glahn, D. C., Johnson, M. P., Curran, J. E., Bozaoglu, K., Dyer, T. D., … Blangero, J. (2011). Impact of DISC1 variation on neuroanatomical and neurocognitive phenotypes. *Molecular Psychiatry*, *16*(11), 1096–104, 1063. doi:10.1038/mp.2011.37

Carlo, A.-S., Gustafsen, C., Mastrobuoni, G., Nielsen, M. S., Burgert, T., Hartl, D., … Willnow, T. E. (2013). The pro-neurotrophin receptor Sortilin is a major neuronal apolipoprotein E receptor for catabolism of amyloid-β peptide in the brain. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(1), 358–70. doi:10.1523/JNEUROSCI.2425-12.2013

Carlsson, A., & Lindqvist, M. (2009). Effect of Chlorpromazine or Haloperidol on Formation of 3-Methoxytyramine and Normetanephrine in Mouse Brain. *Acta Pharmacologica et Toxicologica*, *20*(2), 140–144. doi:10.1111/j.1600-0773.1963.tb01730.x

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22. doi:10.1037/h0046743

Chakos, M. H., Lieberman, J. A., Bilder, R. M., Borenstein, M., Lerner, G., Bogerts, B., … Ashtari, M. (1994). Increase in caudate nuclei volumes of first-episode schizophrenic patients taking antipsychotic drugs. *American Journal of Psychiatry*, *151*(10), 1430–1436. doi:10.1017/CBO9781107415324.004

Chen, D., Liu, F., Yang, C., Liang, X., Shang, Q., He, W., & Wang, Z. (2012). Association between the TPH1 A218C polymorphism and risk of mood disorders and alcohol dependence: Evidence from the current studies. *Journal of Affective Disorders*, *138*(1), 27–33. doi:10.1016/j.jad.2011.04.018

Chen, D. T., Jiang, X., Akula, N., Shugart, Y. Y., Wendland, J. R., Steele, C. J. M., … Strauss, J. (2013). Genome-wide association study meta-analysis of European and Asian-ancestry samples identifies three novel loci associated with bipolar disorder. *Molecular Psychiatry*, *18*(2), 195–205. doi:10.1038/mp.2011.157

Chen, H. M., DeLong, C. J., Bame, M., Rajapakse, I., Herron, T. J., McInnis, M. G., & O'Shea, K. S. (2014). Transcripts involved in calcium signaling and telencephalic neuronal fate are altered in induced pluripotent stem cells from bipolar disorder patients. *Translational Psychiatry*, *4*(3), e375. doi:10.1038/tp.2014.12

Chen, J., Lipska, B. K., Halim, N., Ma, Q. D., Matsumoto, M., Melhem, S., … Weinberger, D. R. (2004). Functional Analysis of Genetic Variation in Catechol-O-Methyltransferase (COMT): Effects on mRNA, Protein, and Enzyme Activity in Postmortem Human Brain. *The American Journal of Human Genetics*, *75*(5), 807–821. doi:10.1086/425589

Chen, Y. A., Lemire, M., Choufani, S., Butcher, D. T., Grafodatskaya, D., Zanke, B. W., … Weksberg, R. (2013). Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics : Official Journal of the DNA Methylation Society*, *8*(2), 203–209. doi:10.4161/epi.23470

Christoforou, A., Le Hellard, S., Thomson, P. A., Morris, S. W., Tenesa, A., Pickard, B. S., … Evans, K. L. (2007). Association analysis of the chromosome 4p15-p16 candidate region for bipolar disorder and schizophrenia. *Molecular Psychiatry*, *12*(11), 1011–25. doi:10.1038/sj.mp.4002003

Christoforou, A., McGhee, K. A., Morris, S. W., Thomson, P. A., Anderson, S., McLean, A., … Evans, K. L. (2011). Convergence of linkage, association and GWAS findings for a candidate region for bipolar disorder and schizophrenia on chromosome 4p. *Molecular Psychiatry*, *16*(3), 240–2. doi:10.1038/mp.2010.25

Chubb, J. E., Bradshaw, N. J., Soares, D. C., Porteous, D. J., & Millar, J. K. (2008). The DISC locus in psychiatric illness. *Molecular Psychiatry*, *13*(1), 36–64. doi:10.1038/sj.mp.4002106

Cichon, S., Mühleisen, T. W., Degenhardt, F. A., Mattheisen, M., Miró, X., Strohmaier, J., … Nöthen, M. M. (2011). Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. *American Journal of Human Genetics*, *88*(3), 372–81. doi:10.1016/j.ajhg.2011.01.017

Cipriani, A., Barbui, C., Salanti, G., Rendell, J., Brown, R., Stockton, S., … Geddes, J. R. (2011). Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple-treatments meta-analysis. *The Lancet*, *378*(9799), 1306–1315. doi:10.1016/S0140-6736(11)60873-8

Clapcote, S. J., Lipina, T. V., Millar, J. K., Mackie, S., Christie, S., Ogawa, F., … Roder, J. C. (2007). Behavioral Phenotypes of Disc1 Missense Mutations in Mice. *Neuron*, *54*(3), 387–402. doi:10.1016/j.neuron.2007.04.015

Clapcote, S. J., & Roder, J. C. (2005). Simplex PCR assay for sex determination in mice. *702 BioTechniques*, *38*(5).

Connor, C. M., Crawford, B. C., & Akbarian, S. (2011). White matter neuron alterations in schizophrenia and related disorders. *International Journal of Developmental Neuroscience*, *29*(3), 325–334. doi:10.1016/j.ijdevneu.2010.07.236

Connor, C. M., Guo, Y., & Akbarian, S. (2009). Cingulate white matter neurons in schizophrenia and bipolar disorder. *Biological Psychiatry*, *66*(5), 486–93. doi:10.1016/j.biopsych.2009.04.032

Coppen, A., Shaw, D. M., Malleson, A., Eccleston, E., & Gundy, G. (1965). Tryptamine Metabolism in Depression, 993–998.

Corvin, A., & Morris, D. W. (2014). Genome-wide association studies: findings at the major histocompatibility complex locus in psychosis. *Biological Psychiatry*, *75*(4), 276–83. doi:10.1016/j.biopsych.2013.09.018

Craddock, N., & Sklar, P. (2009). Genetics of bipolar disorder: successful start to a long journey. *Trends in Genetics : TIG*, *25*(2), 99–105. doi:10.1016/j.tig.2008.12.002

Creese, I., Burt, D. R., & Snyder, S. H. (1976). Dopamine receptor binding predicts clinical and pharmacological potencies of antischizophrenic drugs. *Science (New York, N.Y.)*, *192*(4238), 481–3. doi:10.1126/science.3854

Crow, T. J. (2011). "The missing genes: what happened to the heritability of psychiatric disorders?" *Molecular Psychiatry*, *16*(4), 362–364. doi:mp201092 [pii]\r10.1038/mp.2010.92

Dalman, C., Thomas, H. V, David, A. S., Gentz, J., Lewis, G., & Allebeck, P. (2001). Signs of asphyxia at birth and risk of schizophrenia. Population-based case-control study. *The British Journal of Psychiatry : The Journal of Mental Science*, *179*, 403–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11689395

Davis, K. L., Kahn, R. S., Ko, G., & Davidson, M. (1991). Dopamine in schizophrenia: a review and reconceptualization. *The American Journal of Psychiatry*, *148*(11), 1474–86. doi:10.1176/ajp.148.11.1474

de Bakker, P. I. W. (2009). Selection and evaluation of Tag-SNPs using Tagger and HapMap. *Cold Spring Harbor Protocols*, *2009*(6), pdb.ip67. doi:10.1101/pdb.ip67

Dedeurwaerder, S., Defrance, M., Calonne, E., Denis, H., Sotiriou, C., & Fuks, F. (2011). Evaluation of the Infinium Methylation 450K technology. *Epigenomics*, *3*(6), 771–784. doi:10.2217/epi.11.105

Dempster, E. L., Pidsley, R., Schalkwyk, L. C., Owens, S., Georgiades, A., Kane, F., … Mill, J. (2011). Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. *Human Molecular Genetics*, *20*(24), 4786–4796. doi:10.1093/hmg/ddr416

Dempster, E. L., Wong, C. C. Y., Lester, K. J., Burrage, J., Gregory, A. M., Mill, J., & Eley, T. C. (2014). Genome-wide methylomic analysis of monozygotic twins discordant for adolescent depression. *Biological Psychiatry*, *76*(12), 977–983. doi:10.1016/j.biopsych.2014.04.013

Detera-Wadleigh, S. D., Badner, J. A., Berrettini, W. H., Yoshikawa, T., Goldin, L. R., Turner, G., … Gershon, E. S. (1999). A high-density genome scan detects evidence for a bipolar-disorder susceptibility locus on 13q32 and other potential loci on 1q32 and 18p11.2. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(10), 5604–5609. doi:10.1073/pnas.96.10.5604

Doyle, O. M., Bois, C., Thomson, P., Romaniuk, L., Whitcher, B., Williams, S. C. R., … Lawrie, S. M. (2015). The cortical thickness phenotype of individuals with DISC1 translocation resembles schizophrenia. *The Journal of Clinical Investigation*, *125*(9), 3714–22. doi:10.1172/JCI82636

Du, P., Kibbe, W. A., & Lin, S. M. (2008). lumi: A pipeline for processing Illumina microarray. *Bioinformatics*, *24*(13), 1547–1548. doi:10.1093/bioinformatics/btn224

Du, P., Zhang, X., Huang, C.-C., Jafari, N., Kibbe, W. A., Hou, L., & Lin, S. M. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*, *11*(1), 587. doi:10.1186/1471-2105-11-587

Duan, X., Chang, J. H., Ge, S., Faulkner, R. L., Kim, J. Y., Kitabatake, Y., … Song, H. (2007). Disrupted-In-Schizophrenia 1 regulates integration of newly generated neurons in the adult brain. *Cell*, *130*(6), 1146–58. doi:10.1016/j.cell.2007.07.010

Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., … ENCODE Project Consortium. (2012). An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature*, *489*(7414), 57–74. doi:10.1038/nature11247

Ebert, A. D., Liang, P., & Wu, J. C. (2012). Induced pluripotent stem cells as a disease modeling and drug screening platform. *Journal of Cardiovascular Pharmacology*, *60*(4), 408–16. doi:10.1097/FJC.0b013e318247f642

Eden, E., Lipson, D., Yogev, S., & Yakhini, Z. (2007). Discovering motifs in ranked lists of DNA sequences. *PLoS Computational Biology*, *3*(3), e39. doi:10.1371/journal.pcbi.0030039

Eden, E., Navon, R., Steinfeld, I., Lipson, D., & Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, *10*(1), 48. doi:10.1186/1471-2105-10-48

Egan, M. F., Goldberg, T. E., Kolachana, B. S., Callicott, J. H., Mazzanti, C. M., Straub, R. E., … Weinberger, D. R. (2001). Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(12), 6917–22. doi:10.1073/pnas.111134598

Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H., & Nadeau, J. H. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Reviews. Genetics*, *11*(6), 446–50. doi:10.1038/nrg2809

Ekholm, J. M., Kieseppä, T., Hiekkalinna, T., Partonen, T., Paunio, T., Perola, M., … Peltonen, L. (2003). Evidence of susceptibility loci on 4q32 and 16p12 for bipolar disorder. *Human Molecular Genetics*, *12*(15), 1907–15. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12874110

Emsley, R., Chiliza, B., Asmal, L., Harvey, B. H., Robinson, D., Woerner, M., … Sampson, A. (2013). The nature of relapse in schizophrenia. *BMC Psychiatry*, *13*(1), 50. doi:10.1186/1471-244X-13-50

Ewald, H., Flint, T., Kruse, T. A., & Mors, O. (2002). A genome-wide scan shows significant linkage between bipolar disorder and chromosome 12q24.3 and suggestive linkage to chromosomes 1p22-21, 4p16, 6q14-22, 10q26 and 16p13.3. *Molecular Psychiatry*, *7*(7), 734–44. doi:10.1038/sj.mp.4001074

Eykelenboom, J. E., Briggs, G. J., Bradshaw, N. J., Soares, D. C., Ogawa, F., Christie, S., … Millar, J. K. (2012). A t(1;11) translocation linked to schizophrenia and affective disorders gives rise to aberrant chimeric DISC1 transcripts that encode structurally altered, deleterious mitochondrial proteins. *Human Molecular Genetics*, *21*(15), 3374–86. doi:10.1093/hmg/dds169

Faghihi, M. A., & Wahlestedt, C. (2009). Regulatory roles of natural antisense transcripts. *Nature Reviews. Molecular Cell Biology*, *10*(9), 637–643. doi:10.1038/nrm2738

Farrell, M. S., Werge, T., Sklar, P., Owen, M. J., Ophoff, R. A., O'Donovan, M. C., … Sullivan, P. F. (2015). Evaluating historical candidate genes for schizophrenia. *Molecular Psychiatry*, *20*(5), 555–62. doi:10.1038/mp.2015.16

Ferrari, A. J., Charlson, F. J., Norman, R. E., Flaxman, A. D., Patten, S. B., Vos, T., & Whiteford, H. A. (2013). The epidemiological modelling of major depressive disorder: application for the Global Burden of Disease Study 2010. *PloS One*, *8*(7), e69637. doi:10.1371/journal.pone.0069637

Ferreira, M. A. R., O'Donovan, M. C., Meng, Y. A., Jones, I. R., Ruderfer, D. M., Jones, L., … Wellcome Trust Case Control Consortium. (2008). Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nature Genetics*, *40*(9), 1056–8. doi:10.1038/ng.209

Fitzpatrick, P. F. (2003). Tetrahydropterin-Dependent Amino Acid Hydroxylases. *http://dx.doi.org/10.1146/annurev.biochem.68.1.355*.

Fukumoto, T., Morinobu, S., Okamoto, Y., Kagaya, A., & Yamawaki, S. (2001). Chronic lithium treatment increases the expression of brain-derived

neurotrophic factor in the rat brain. *Psychopharmacology*, *158*(1), 100–6. doi:10.1007/s002130100871

Fuller, R., Nopoulos, P., Arndt, S., O'Leary, D., Ho, B.-C., & Andreasen, N. C. (2002). Longitudinal assessment of premorbid cognitive functioning in patients with schizophrenia through examination of standardized scholastic test performance. *The American Journal of Psychiatry*, *159*(7), 1183–9. doi:10.1176/appi.ajp.159.7.1183

Galehdari, H., Pooryasin, A., Foroughmand, A., Daneshmand, S., & Saadat, M. (2009). Association between the G1001C polymorphism in the GRIN1 gene promoter and schizophrenia in the Iranian population. *Journal of Molecular Neuroscience : MN*, *38*(2), 178–81. doi:10.1007/s12031-008-9148-5

Garey, L. J., Ong, W. Y., Patel, T. S., Kanani, M., Davis, A., Mortimer, A. M., … Hirsch, S. R. (1998). Reduced dendritic spine density on cerebral cortical pyramidal neurons in schizophrenia. *Journal of Neurology, Neurosurgery, and Psychiatry*, *65*(4), 446–53. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9771764

Geddes, J. R., & Miklowitz, D. J. (2013). Treatment of bipolar disorder. *Lancet (London, England)*, *381*(9878), 1672–82. doi:10.1016/S0140-6736(13)60857-0

Georgi, A., Jamra, R. A., Schumacher, J., Becker, T., Schmael, C., Deschner, M., … Schulze, T. G. (2006). No association between genetic variants at the GRIN1 gene and bipolar disorder in a German sample. *Psychiatric Genetics*, *16*(5), 183–4. doi:10.1097/01.ypg.0000242194.36150.2b

Gerstein, M. B., Kundaje, A., Hariharan, M., Landt, S. G., Yan, K.-K., Cheng, C., … Snyder, M. (2012). Architecture of the human regulatory network derived from ENCODE data. *Nature*, *489*(7414), 91–100. doi:10.1038/nature11245

Gilmour, G., Dix, S., Fellini, L., Gastambide, F., Plath, N., Steckler, T., … Tricklebank, M. (2012). NMDA receptors, cognition and schizophrenia – Testing the validity of the NMDA receptor hypofunction hypothesis. *Neuropharmacology*, *62*(3), 1401–1412. doi:10.1016/j.neuropharm.2011.03.015

Glahn, D. C., Laird, A. R., Ellison-Wright, I., Thelen, S. M., Robinson, J. L., Lancaster, J. L., … Fox, P. T. (2008). Meta-Analysis of Gray Matter Anomalies in Schizophrenia: Application of Anatomic Likelihood Estimation and Network Analysis. *Biological Psychiatry*, *64*(9), 774–781. doi:10.1016/j.biopsych.2008.03.031

Glerup, S., Bolcho, U., Mølgaard, S., Bøggild, S., Vaegter, C. B., Smith, A. H., … Nykjaer, A. (2016). SorCS2 is required for BDNF-dependent plasticity in the hippocampus. *Molecular Psychiatry*. doi:10.1038/mp.2016.108

Glerup, S., Lume, M., Olsen, D., Nyengaard, J. R., Vaegter, C. B., Gustafsen, C., … Petersen, C. M. (2013). SorLA Controls Neurotrophic Activity by Sorting of GDNF and Its Receptors GFRα1 and RET. *Cell Reports*, *3*(1), 186–199. doi:10.1016/j.celrep.2012.12.011

Glerup, S., Olsen, D., Vaegter, C. B., Gustafsen, C., Sjoegaard, S. S., Hermey, G., … Teng, K. K. (2014). SorCS2 Regulates Dopaminergic Wiring and Is Processed into an Apoptotic Two-Chain Receptor in Peripheral Glia. *Neuron*, *82*(5), 1074–1087. doi:10.1016/j.neuron.2014.04.022

González-Castro, T. B., Nicolini, H., Lanzagorta, N., López-Narváez, L., Genis, A., Pool García, S., & Tovilla-Zárate, C. A. (2015). The role of brain-derived neurotrophic factor (BDNF) Val66Met genetic polymorphism in bipolar disorder: a case-control study, comorbidities, and meta-analysis of 16,786 subjects. *Bipolar Disorders*, *17*(1), 27–38. doi:10.1111/bdi.12227

Goodwin, F. K., Jamison, K. R., & Ghaemi, S. N. (2007). *Manic-depressive illness : bipolar disorders and recurrent depression*. Oxford University Press.

Gopalakrishnan, S., Van Emburgh, B. O., & Robertson, K. D. (2008). DNA methylation in development and human disease. *Mutat Res*, *647*(1-2), 30–38. doi:10.1016/j.mrfmmm.2008.08.006

Gottesman, I. I., & Shields, J. (1967). A polygenic theory of schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*, *58*(1), 199–205. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5231600

Graber, T. E., McCamphill, P. K., & Sossin, W. S. (2013). A recollection of mTOR signaling in learning and memory. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *20*(10), 518–30. doi:10.1101/lm.027664.112

Grafodatskaya, D., Choufani, S., Ferreira, J. C., Butcher, D. T., Lou, Y., Zhao, C., … Weksberg, R. (2010). EBV transformation and cell culturing destabilizes DNA methylation in human lymphoblastoid cell lines. *Genomics*, *95*(2), 73–83. doi:10.1016/j.ygeno.2009.12.001

Grahn, J. A., Parkinson, J. A., & Owen, A. M. (2008). The cognitive functions of the caudate nucleus. *Progress in Neurobiology*, *86*(3), 141–155. doi:10.1016/j.pneurobio.2008.09.004

Grayson, D. R., & Guidotti, A. (2012). The Dynamics of DNA Methylation in Schizophrenia and Related Psychiatric Disorders. *Neuropsychopharmacology*, *38*(1), 138–166. doi:10.1038/npp.2012.125

Green, E. K., Grozeva, D., Jones, I., Jones, L., Kirov, G., Caesar, S., … Craddock, N. (2010). The bipolar disorder risk allele at CACNA1C also confers risk of

recurrent major depression and of schizophrenia. *Molecular Psychiatry*, *15*(10), 1016–1022. doi:10.1038/mp.2009.49

Green, E. K., Norton, N., Peirce, T., Grozeva, D., Kirov, G., Owen, M. J., … Craddock, N. (2006). Evidence that a DISC1 frame-shift deletion associated with psychosis in a single family may not be a pathogenic mutation. *Molecular Psychiatry*, *11*(9), 798–799. doi:10.1038/sj.mp.4001853

Green, E. K., Rees, E., Walters, J. T. R., Smith, K.-G., Forty, L., Grozeva, D., … Kirov, G. (2016). Copy number variation in bipolar disorder. *Molecular Psychiatry*, *21*(1), 89–93. doi:10.1038/mp.2014.174

Gregg, J. P., Lit, L., Baron, C. A., Hertz-Picciotto, I., Walker, W., Davis, R. A., … Sharp, F. R. (2008). Gene expression changes in children with autism. *Genomics*, *91*(1), 22–9. doi:10.1016/j.ygeno.2007.09.003

Grozeva, D., Kirov, G., Ivanov, D., Jones, I. R., Jones, L., Green, E. K., … Wellcome Trust Case Control Consortium. (2010). Rare copy number variants: a point of rarity in genetic risk for bipolar disorder and schizophrenia. *Archives of General Psychiatry*, *67*(4), 318–27. doi:10.1001/archgenpsychiatry.2010.25

Gurling, H. M. D., Kalsi, G., Brynjolfson, J., Sigmundsson, T., Sherrington, R., Mankoo, B. S., … Curtis, D. (2001). Genomewide Genetic Linkage Analysis Confirms the Presence of Susceptibility Loci for Schizophrenia, on Chromosomes 1q32.2, 5q33.2, and 8p21-22 and Provides Support for Linkage to Schizophrenia, on Chromosomes 11q23.3-24 and 20q12.1-11.23. *The American Journal of Human Genetics*, *68*(3), 661–673. doi:10.1086/318788

Gurung, R., & Prata, D. P. (2015). What is the impact of genome-wide supported risk variants for schizophrenia and bipolar disorder on brain structure and function? A systematic review. *Psychological Medicine*, 1–20. doi:10.1017/S0033291715000537

Gusella, J. F., Wexler, N. S., Conneally, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E., … Martin, J. B. (1983). A polymorphic DNA marker genetically linked to Huntington's disease. *Nature*, *306*(5940), 234–238. doi:10.1038/306234a0

Gustafsen, C., Glerup, S., Pallesen, L. T., Olsen, D., Andersen, O. M., Nykjær, A., … Petersen, C. M. (2013). Sortilin and SorLA display distinct roles in processing and trafficking of amyloid precursor protein. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(1), 64–71. doi:10.1523/JNEUROSCI.2371-12.2013

Hafeman, D. M., Chang, K. D., Garrett, A. S., Sanders, E. M., & Phillips, M. L. (2012). Effects of medication on neuroimaging findings in bipolar disorder: an updated review. *Bipolar Disorders*, *14*(4), 375–410. doi:10.1111/j.1399-5618.2012.01023.x

Hahn, M. A., Qiu, R., Wu, X., Li, A. X., Zhang, H., Wang, J., … Lu, Q. (2013). Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. *Cell Reports*, *3*(2), 291–300. doi:10.1016/j.celrep.2013.01.011

Hall, E., Volkov, P., Dayeh, T., Esguerra, J. L. S., Salö, S., Eliasson, L., … Newgard, C. (2014). Sex differences in the genome-wide DNA methylation pattern and impact on gene expression, microRNA levels and insulin secretion in human pancreatic islets. *Genome Biology*, *15*(12), 522. doi:10.1186/s13059-014-0522-z

Handel, A. E., Chintawar, S., Lalic, T., Whiteley, E., Vowles, J., Giustacchini, A., … Cader, M. Z. (2016). Assessing similarity to primary tissue and cortical layer identity in induced pluripotent stem cell-derived cortical neurons through single-cell transcriptomics. *Human Molecular Genetics*, *25*(5), 989–1000. doi:10.1093/hmg/ddv637

Hannon, E., Spiers, H., Viana, J., Pidsley, R., Burrage, J., Murphy, T. M., … Mill, J. (2015). Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nature Neuroscience*, *19*(November), 1–9. doi:10.1038/nn.4182

Hargreaves, A., Anney, R., O'Dushlaine, C., Nicodemus, K. K., Schizophrenia Psychiatric Genome-Wide Association Study Consortium (PGC-SCZ), Wellcome Trust Case Control Consortium 2, … Donohoe, G. (2014). The one and the many: effects of the cell adhesion molecule pathway on neuropsychological function in psychosis. *Psychological Medicine*, *44*(10), 2177–87. doi:10.1017/S0033291713002663

Harrison, P. J. (2000). Postmortem studies in schizophrenia. *Dialogues in Clinical Neuroscience*, *2*(4), 349–57. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22033474

Harrison, P. J., & Weinberger, D. R. (2005). Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. *Molecular Psychiatry*, *10*(1), 40–68; image 5. doi:10.1038/sj.mp.4001558

Hartl, D., Klatt, S., Roch, M., Konthur, Z., Klose, J., Willnow, T. E., … Pons, T. (2013). Soluble Alpha-APP (sAPPalpha) Regulates CDK5 Expression and Activity in Neurons. *PLoS ONE*, *8*(6), e65920. doi:10.1371/journal.pone.0065920

Hasan, A., Falkai, P., Wobrock, T., Lieberman, J., Glenthoj, B., Gattaz, W. F., & Thibaut, F. (n.d.). World Federation of Societies of Biological Psychiatry (WFSBP) Guidelines for Biological Treatment of Schizophrenia, Part 1: Update 2012 on the acute treatment of schizophrenia and the management of treatment resistance. doi:10.3109/15622975.2012.696143

Hashimoto, R., Numakawa, T., Ohnishi, T., Kumamaru, E., Yagasaki, Y., Ishimoto, T., … Kunugi, H. (2006). Impact of the DISC1 Ser704Cys polymorphism on

risk for major depression, brain morphology and ERK signaling. *Human Molecular Genetics*, *15*(20), 3024–33. doi:10.1093/hmg/ddl244

Hashimshony, T., Zhang, J., Keshet, I., Bustin, M., & Cedar, H. (2003). The role of DNA methylation in setting up chromatin structure during development. *Nature Genetics*, *34*(2), 187–192. doi:10.1038/ng1158

Haslam, N. (2006). Bias in psychopathology research. *Current Opinion in Psychiatry*, *19*(6), 625–30. doi:10.1097/01.yco.0000245745.68255.db

Hattori, E., Liu, C., Badner, J. A., Bonner, T. I., Christian, S. L., Maheshwari, M., … Gershon, E. S. (2003). Polymorphisms at the G72/G30 gene locus, on 13q33, are associated with bipolar disorder in two independent pedigree series. *American Journal of Human Genetics*, *72*(5), 1131–40. doi:10.1086/374822

Health, C. for M. (2009). The economic and social costs of mental health problems in 2009 / 10. *Health San Francisco*, 4. Retrieved from http://www.centreformentalhealth.org.uk/pdfs/Economic_and_social_costs_2010.pdf

Hennah, W., Thomson, P., McQuillin, A., Bass, N., Loukola, A., Anjorin, A., … Porteous, D. (2009). DISC1 association, heterogeneity and interplay in schizophrenia and bipolar disorder., *14*(9). doi:10.1038/mp.2008.22

Hennah, W., Varilo, T., Kestilä, M., Paunio, T., Arajärvi, R., Haukka, J., … Ekelund, J. (2003). Haplotype transmission analysis provides evidence of association for DISC1 to schizophrenia and suggests sex-dependent effects. *Human Molecular Genetics*, *12*(23), 3151–9. doi:10.1093/hmg/ddg341

Hermey, G., Plath, N., Hübner, C. A., Kuhl, D., Schaller, H. C., & Hermans-Borgmeyer, I. (2004). The three sorCS genes are differentially expressed and regulated by synaptic activity. *Journal of Neurochemistry*, *88*(6), 1470–1476. doi:10.1046/j.1471-4159.2004.02286.x

Hermey, G., Riedel, I. B., Rezgaoui, M., Westergaard, U. B., Schaller, C., & Hermans-Borgmeyer, I. (2001). SorCS1, a member of the novel sorting receptor family, is localized in somata and dendrites of neurons throughout the murine brain. *Neuroscience Letters*, *313*(1-2), 83–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11684345

Herzing, L. B., Kim, S. J., Cook, E. H., & Ledbetter, D. H. (2001). The human aminophospholipid-transporting ATPase gene ATP10C maps adjacent to UBE3A and exhibits similar imprinted expression. *American Journal of Human Genetics*, *68*(6), 1501–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11353404

Hikida, T., Jaaro-Peled, H., Seshadri, S., Oishi, K., Hookway, C., Kong, S., … Sawa, A. (2007). Dominant-negative DISC1 transgenic mice display schizophrenia-

associated phenotypes detected by measures translatable to humans. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(36), 14501–6. doi:10.1073/pnas.0704774104

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, *14*(10), R115. doi:10.1186/gb-2013-14-10-r115

Hou, L., Bergen, S. E., Akula, N., Song, J., Hultman, C. M., Landén, M., … McMahon, F. J. (2016). Genome-wide association study of 40,000 individuals identifies two novel loci associated with bipolar disorder. *Human Molecular Genetics*. doi:10.1093/hmg/ddw181

Houseman, E. A., Accomando, W. P., Koestler, D. C., Christensen, B. C., Marsit, C. J., Nelson, H. H., … Kelsey, K. T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*, *13*(1), 86. doi:10.1186/1471-2105-13-86

Howes, O. D., Kambeitz, J., Kim, E., Stahl, D., Slifstein, M., Abi-Dargham, A., & Kapur, S. (2012). The nature of dopamine dysfunction in schizophrenia and what this means for treatment. *Archives of General Psychiatry*, *69*(8), 776–86. doi:10.1001/archgenpsychiatry.2012.169

Howes, O. D., & Kapur, S. (2009). The dopamine hypothesis of schizophrenia: version III--the final common pathway. *Schizophrenia Bulletin*, *35*(3), 549–62. doi:10.1093/schbul/sbp006

Huang, E. J., & Reichardt, L. F. (2001). Neurotrophins: roles in neuronal development and function. *Annual Review of Neuroscience*, *24*, 677–736. doi:10.1146/annurev.neuro.24.1.677

Hunsberger, J. G., Fessler, E. B., Chibane, F. L., Leng, Y., Maric, D., Elkahloun, A. G., & Chuang, D. M. (2013). Mood stabilizer-regulated miRNAs in neuropsychiatric and neurodegenerative diseases: Identifying associations and functions. *American Journal of Translational Research*, *5*(4), 450–464.

Hyde, C. L., Nagle, M. W., Tian, C., Chen, X., Paciga, S. A., Wendland, J. R., … Winslow, A. R. (2016). Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature Genetics*, *48*(9), 1031–1036. doi:10.1038/ng.3623

Inayama, Y., Yoneda, H., Fukushima, K., Sakai, J., Asaba, H., & Sakai, T. (2008). Paracentric inversion of chromosome 9 with schizoaffective disorder. *Clinical Genetics*, *51*(1), 69–70. doi:10.1111/j.1399-0004.1997.tb02419.x

Ingason, A., Rujescu, D., Cichon, S., Sigurdsson, E., Sigmundsson, T., Pietiläinen, O. P. H., … St Clair, D. M. (2011). Copy number variations of chromosome 16p13.1 region associated with schizophrenia. *Molecular Psychiatry*, *16*(1), 17–25. doi:10.1038/mp.2009.101

International Schizophrenia Consortium, Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., … Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, *460*(7256), 748–52. doi:10.1038/nature08185

International, T., & Consortium, S. (2008). Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*, *455*(7210), 237–241. doi:10.1038/nature07239

Ishizuka, K., Kamiya, A., Oh, E. C., Kanki, H., Seshadri, S., Robinson, J. F., … Sawa, A. (2011). DISC1-dependent switch from progenitor proliferation to migration in the developing cortex. *Nature*, *473*(7345), 92–96. doi:10.1038/nature09859

Jaaro-Peled, H., Niwa, M., Foss, C. A., Murai, R., de Los Reyes, S., Kamiya, A., … Sawa, A. (2013). Subcortical dopaminergic deficits in a DISC1 mutant model: a study in direct reference to human molecular brain imaging. *Human Molecular Genetics*, *22*(8), 1574–80. doi:10.1093/hmg/ddt007

Jacobs, P. a, Brunton, M., Frackiewicz, A., Newton, M., Cook, P. J. L., & Robson, E. B. (1970). Studies on a family with three cytogenetic markers. *Annals of Human Genetics*, *33*(4), 325–336. doi:10.1111/j.1469-1809.1970.tb01658.x

Jaffe, A. E., & Irizarry, R. a. (2014). Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biology*, *15*(2), R31. doi:10.1186/gb-2014-15-2-r31

Jarskog, L. F., Glantz, L. A., Gilmore, J. H., & Lieberman, J. A. (2005). Apoptotic mechanisms in the pathophysiology of schizophrenia. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, *29*(5), 846–58. doi:10.1016/j.pnpbp.2005.03.010

Jernigan, T. L., Trauner, D. a, Hesselink, J. R., & Tallal, P. a. (1991). Maturation of human cerebrum observed in vivo during adolescence. *Brain : A Journal of Neurology*, *114 (Pt 5*(5), 2037–49. doi:10.1093/brain/114.5.2037

Ji, B., Higa, K. K., Kim, M., Zhou, L., Young, J. W., Geyer, M. A., & Zhou, X. (2014). Inhibition of protein translation by the DISC1-Boymaw fusion gene from a Scottish family with major psychiatric disorders. *Human Molecular Genetics*, *23*(21), 5683–705. doi:10.1093/hmg/ddu285

Johnstone, E., Frith, C. D., Crow, T. J., Husband, J., & Kreel, L. (1976). Cerebral ventricular size and cognitive impairment in chronic schizophrenia. *The Lancet*, *308*(7992), 924–926. Doi:10.1016/S0140-6736(76)90890-4

Johnstone, I. M., & Titterington, D. M. (2009). Statistical challenges of high-dimensional data. *Philosophical Transactions of the Royal Society A:*

Mathematical, Physical and Engineering Sciences, *367*(1906), 4237–4253. doi:10.1098/rsta.2009.0159

Johnstone, M., Thomson, P. A., Hall, J., McIntosh, A. M., Lawrie, S. M., & Porteous, D. J. (2011). DISC1 in schizophrenia: genetic mouse models and human genomic imaging. *Schizophrenia Bulletin*, *37*(1), 14–20. doi:10.1093/schbul/sbq135

Jones, C. A., Watson, D. J. G., & Fone, K. C. F. Animal models of schizophrenia. , 164 British Journal of Pharmacology 1162–1194 (2011). Wiley-Blackwell. doi:10.1111/j.1476-5381.2011.01386.x

Kamiya, A., Kubo, K., Tomoda, T., Takaki, M., Youn, R., Ozeki, Y., … Sawa, A. (2005). A schizophrenia-associated mutation of DISC1 perturbs cerebral cortex development. *Nature Cell Biology*, *7*(12), 1167–1178. doi:10.1038/ncb1328

Kamiya, A., Tan, P. L., Kubo, K., Engelhard, C., Ishizuka, K., Kubo, A., … SR, M. (2008). Recruitment of PCM1 to the Centrosome by the Cooperative Action of DISC1 and BBS4. *Archives of General Psychiatry*, *65*(9), 996. doi:10.1001/archpsyc.65.9.996

Kent, W. J. (2002). BLAT - The BLAST-like alignment tool. *Genome Research*, *12*(4), 656–664. doi:10.1101/gr.229202. Article published online before March 2002

Kieseppä, T., Eerola, M., Mäntylä, R., Neuvonen, T., Poutanen, V.-P., Luoma, K., … Isometsä, E. (2010). Major depressive disorder and white matter abnormalities: a diffusion tensor imaging study with tract-based spatial statistics. *Journal of Affective Disorders*, *120*(1-3), 240–4. doi:10.1016/j.jad.2009.04.023

Kim, E., Naisbitt, S., Hsueh, Y.-P., Rao, A., Rothschild, A., Craig, A. M., & Sheng, M. (1997). GKAP, a Novel Synaptic Protein That Interacts with the Guanylate Kinase-like Domain of the PSD-95/SAP90 Family of Channel Clustering Molecules. *The Journal of Cell Biology*, *136*(3), 669–678.

Kim, J. S., Kornhuber, H. H., Schmid-Burgk, W., & Holzmüller, B. (1980). Low cerebrospinal fluid glutamate in schizophrenic patients and a new hypothesis on schizophrenia. *Neuroscience Letters*, *20*(3), 379–382. doi:10.1016/0304-3940(80)90178-0

Kim, S., Choi, K.-H., Baykiz, A., Gershenfeld, H. K., McKeown, R., Cuffe, S., … Lempicki, R. (2007). Suicide candidate genes associated with bipolar disorder and schizophrenia: An exploratory gene expression profiling analysis of post-mortem prefrontal cortex. *BMC Genomics*, *8*(1), 413. doi:10.1186/1471-2164-8-413

Kinoshita, M., Numata, S., Tajima, A., Shimodera, S., Ono, S., Imamura, A., … Ohmori, T. (2013). DNA methylation signatures of peripheral leukocytes in

schizophrenia. *NeuroMolecular Medicine*, *15*(1), 95–101. doi:10.1007/s12017-012-8198-6

Kirkpatrick, B., Xu, L., Cascella, N., Ozeki, Y., Sawa, A., & Roberts, R. C. (2006). DISC1 immunoreactivity at the light and ultrastructural level in the human neocortex. *The Journal of Comparative Neurology*, *497*(3), 436–450. doi:10.1002/cne.21007

Kirov, G., Grozeva, D., Norton, N., Ivanov, D., Mantripragada, K. K., Holmans, P., … O'Donovan. (2009). Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Human Molecular Genetics*, *18*(8), 1497–503. doi:10.1093/hmg/ddp043

Klengel, T., Pape, J., Binder, E. B., & Mehta, D. (2014). The role of DNA methylation in stress-related psychiatric disorders. *Neuropharmacology*. doi:10.1016/j.neuropharm.2014.01.013

Klose, R. J., & Bird, A. P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends in Biochemical Sciences*, *31*(2), 89–97. doi:10.1016/j.tibs.2005.12.008

Korver-Nieberg, N., Quee, P. J., Boos, H. B., Simons, C. J., & GROUP. (2011). The validity of the DSM-IV diagnostic classification system of non-affective psychoses. *The Australian and New Zealand Journal of Psychiatry*, *45*(12), 1061–8. doi:10.3109/00048674.2011.620562

Kriaucionis, S., & Heintz, N. (2009). The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science*, *324*, 929–931. doi:10.1126/science.1169786

Krücken, J., Schroetel, R. M. U., Müller, I. U., Saïdani, N., Marinovski, P., Benten, W. P. M., … Wunderlich, F. (2004). Comparative analysis of the human gimap gene cluster encoding a novel GTPase family. *Gene*, *341*, 291–304. doi:10.1016/j.gene.2004.07.005

Krystal, J. H., Karper, L. P., Seibyl, J. P., Freeman, G. K., Delaney, R., Bremner, J. D., … Charney, D. S. (1994). Subanesthetic effects of the noncompetitive NMDA antagonist, ketamine, in humans. Psychotomimetic, perceptual, cognitive, and neuroendocrine responses. *Archives of General Psychiatry*, *51*(3), 199–214. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8122957

Kunugi, H., Iijima, Y., Tatsumi, M., Yoshida, M., Hashimoto, R., Kato, T., … Yoshikawa, T. (2004). *No association between the Val66Met polymorphism of the brain-derived neurotrophic factor gene and bipolar disorder in a Japanese population: A multicenter study*. *Biological Psychiatry* (Vol. 56). doi:10.1016/j.biopsych.2004.06.017

Lachman, H. M., Papolos, D. F., Saito, T., Yu, Y. M., Szumlanski, C. L., & Weinshilboum, R. M. (1996). Human catechol-O-methyltransferase pharmacogenetics: description of a functional polymorphism and its potential application to neuropsychiatric disorders. *Pharmacogenetics*, *6*(3), 243–50. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8807664

Lai, C.-H. (2013). Gray matter volume in major depressive disorder: A meta-analysis of voxel-based morphometry studies. doi:10.1016/j.pscychresns.2012.06.006

Lakhan, S. E., Caro, M., & Hadzimichalis, N. (2013). NMDA Receptor Activity in Neuropsychiatric Disorders. *Frontiers in Psychiatry*, *4*, 52. doi:10.3389/fpsyt.2013.00052

Le Hellard, S., Lee, A. J., Underwood, S., Thomson, P. A., Morris, S. W., Torrance, H. S., … Evans, K. L. (2007). Haplotype Analysis and a Novel Allele-Sharing Method Refines a Chromosome 4p Locus Linked to Bipolar Affective Disorder. *Biological Psychiatry*, *61*(6), 797–805. doi:10.1016/j.biopsych.2006.06.029

Leal, S. M. (2012). Genome-wide Association Study using GenABEL, 1–22.

Lee, H., Kang, E., GoodSmith, D., Yoon, D. Y., Song, H., Knierim, J. J., … Christian, K. M. (2015). DISC1-mediated dysregulation of adult hippocampal neurogenesis in rats. *Frontiers in Systems Neuroscience*, *9*, 93. doi:10.3389/fnsys.2015.00093

Leibenluft, E. (1996). Women with bipolar illness: clinical and research issues. *The American Journal of Psychiatry*, *153*(2), 163–73. doi:10.1176/ajp.153.2.163

Leliveld, S. R., Bader, V., Hendriks, P., Prikulis, I., Sajnani, G., Requena, J. R., & Korth, C. (n.d.). Insolubility of Disrupted-in-Schizophrenia 1 Disrupts Oligomer-Dependent Interactions with Nuclear Distribution Element 1 and Is Associated with Sporadic Mental Disease. doi:10.1523/JNEUROSCI.5389-07.2008

Lemire, M., Zaidi, S. H. E., Ban, M., Ge, B., Aïssi, D., Germain, M., … Hudson, T. J. (2015). Long-range epigenetic regulation is conferred by genetic variation located at thousands of independent loci. *Nature Communications*, *6*, 6326. doi:10.1038/ncomms7326

Lencz, T., Lambert, C., DeRosse, P., Burdick, K. E., Morgan, T. V., Kane, J. M., … Malhotra, A. K. (2007). Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proceedings of the National Academy of Sciences*, *104*(50), 19942–19947. doi:10.1073/pnas.0710021104

Leumann, L., Feldon, J., Vollenweider, F. X., & Ludewig, K. (2002). Effects of typical and atypical antipsychotics on prepulse inhibition and latent inhibition in chronic schizophrenia. *Biological Psychiatry*, *52*(7), 729–739. doi:10.1016/S0006-3223(02)01344-6

Levinson, D. F., Holmans, P. A., Laurent, C., Riley, B., Pulver, A. E., Gejman, P. V, … Walters, M. (2002). No major schizophrenia locus detected on chromosome 1q in a large multicenter sample. *Science (New York, N.Y.)*, *296*(5568), 739–41. doi:10.1126/science.1069914

Lewis, C. M., Levinson, D. F., Wise, L. H., DeLisi, L. E., Straub, R. E., Hovatta, I., … Helgason, T. (2003). Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *American Journal of Human Genetics*, *73*(1), 34–48. doi:10.1086/376549

Lewis, J., & Bird, a. (1991). DNA methylation and chromatin structure. *FEBS Letters*, *285*(2), 155–159. doi:10.1016/0014-5793(91)80795-5

Li, J. J.-M., Lu, C. C.-L., Cheng, M. M.-C., Luu, S. S.-U., Hsu, S.-H., Hu, T.-M., … Kindler, S. (2014). Role of the DLGAP2 gene encoding the SAP90/PSD-95-associated protein 2 in schizophrenia., *9*(1), e85373. doi:10.1371/journal.pone.0085373

Li, N., Lee, B., Liu, R.-J., Banasr, M., Dwyer, J. M., Iwata, M., … Duman, R. S. (2010). mTOR-dependent synapse formation underlies the rapid antidepressant effects of NMDA antagonists. *Science (New York, N.Y.)*, *329*(5994), 959–64. doi:10.1126/science.1190287

Lichtenstein, P., Yip, B. H., Björk, C., Pawitan, Y., Cannon, T. D., Sullivan, P. F., & Hultman, C. M. (2009). Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet*, *373*(9659), 234–239. doi:10.1016/S0140-6736(09)60072-6

Lim, K. ., Rosenbloom, M. ., Faustman, W. ., Sullivan, E. ., & Pfefferbaum, A. (1999). Cortical gray matter deficit in patients with bipolar disorder. *Schizophrenia Research*, *40*(3), 219–227. doi:10.1016/S0920-9964(99)00063-8

Lim, U., & Song, M. (2012). Dietary and lifestyle factors of DNA methylation. *Methods in Molecular Biology (Clifton, N.J.)*, *863*, 359–76. doi:10.1007/978-1-61779-612-8_23

Lin, S. M., Du, P., Huber, W., & Kibbe, W. A. (2008). Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Research*, *36*(2), e11. doi:10.1093/nar/gkm1075

Lipina, T. V, Haque, F. N., McGirr, A., Boutros, P. C., Berger, T., Mak, T. W., … Wong, A. H. C. (2012). Prophylactic valproic acid treatment prevents schizophrenia-related behaviour in Disc1-L100P mutant mice. *PloS One*, *7*(12), e51562. doi:10.1371/journal.pone.0051562

Lipina, T. V, Niwa, M., Jaaro-Peled, H., Fletcher, P. J., Seeman, P., Sawa, A., & Roder, J. C. (2010). Enhanced dopamine function in DISC1-L100P mutant

mice: implications for schizophrenia. *Genes, Brain, and Behavior*, *9*(7), 777–89. doi:10.1111/j.1601-183X.2010.00615.x

Liu, J., Morgan, M., Hutchison, K., & Calhoun, V. D. (2010). A study of the influence of sex on genome wide methylation. *PloS One*, *5*(4), e10028. doi:10.1371/journal.pone.0010028

Lohmueller, K. E., Pearce, C. L., Pike, M., Lander, E. S., & Hirschhorn, J. N. (2003). Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nature Genetics*, *33*(2), 177–82. doi:10.1038/ng1071

Lohoff, F. W., Sander, T., Ferraro, T. N., Dahl, J. P., Gallinat, J., & Berrettini, W. H. (2005). Confirmation of association between the Val66Met polymorphism in the brain-derived neurotrophic factor (BDNF) gene and bipolar I disorder. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics : The Official Publication of the International Society of Psychiatric Genetics*, *139B*(1), 51–3. doi:10.1002/ajmg.b.30215

López León, S., Croes, E. A., Sayed-Tabatabaei, F. A., Claes, S., Broeckhoven, C. Van, & van Duijn, C. M. (2005). The dopamine D4 receptor gene 48-base-pair-repeat polymorphism and mood disorders: A meta-analysis. *Biological Psychiatry*, *57*(9), 999–1003. doi:10.1016/j.biopsych.2005.01.030

Lunetta, K. L., Hayward, L. B., Segal, J., & Van Eerdewegh, P. (2004). Screening large-scale association study data: exploiting interactions using random forests. *BMC Genetics*, *5*, 32. doi:10.1186/1471-2156-5-32

Madison, J. M., Zhou, F., Nigam, a, Hussain, a, Barker, D. D., Nehme, R., … Haggarty, S. J. (2015). Characterization of bipolar disorder patient-specific induced pluripotent stem cells from a family reveals neurodevelopmental and mRNA expression abnormalities. *Molecular Psychiatry*, *20*(November 2013), 703–17. doi:10.1038/mp.2015.7

Mah, S., Nelson, M. R., Delisi, L. E., Reneland, R. H., Markward, N., James, M. R., … Braun, A. (2006). Identification of the semaphorin receptor PLXNA2 as a candidate for susceptibility to schizophrenia. *Molecular Psychiatry*, *11*(5), 471–8. doi:10.1038/sj.mp.4001785

Maksimovic, J., Gordon, L., & Oshlack, A. (2012). SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biology*, *13*(6), R44. doi:10.1186/gb-2012-13-6-r44

Malavasi, E. L. V, Ogawa, F., Porteous, D. J., & Millar, J. K. (2012). DISC1 variants 37W and 607F disrupt its nuclear targeting and regulatory role in ATF4-mediated transcription. *Human Molecular Genetics*, *21*(12), 2779–92. doi:10.1093/hmg/dds106

Malhotra, D., McCarthy, S., Michaelson, J. J., Vacic, V., Burdick, K. E., Yoon, S., … Sebat, J. (2011). High frequencies of de novo CNVs in bipolar disorder and schizophrenia. *Neuron*, *72*(6), 951–63. doi:10.1016/j.neuron.2011.11.007

Malhotra, D., & Sebat, J. (2012). CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell*, *148*(6), 1223–41. doi:10.1016/j.cell.2012.02.039

Maller, J. J., Thaveenthiran, P., Thomson, R. H., Mcqueen, S., Fitzgerald, P. B., Barysheva, M., … Goodwin, G. M. (2014). Volumetric, cortical thickness and white matter integrity alterations in bipolar disorder type I and II. *Journal of Affective Disorders*, *169*, 118–127. doi:10.1016/j.jad.2014.08.016

Mandela, P., Ma, X.-M., Mandela, P., & Ma, X.-M. (2012). Kalirin, a Key Player in Synapse Formation, Is Implicated in Human Diseases. *Neural Plasticity*, *2012*, 1–9. doi:10.1155/2012/728161

Mandelli, L., Mazza, M., Di Nicola, M., Martinotti, G., Tavian, D., Colombo, E., … Serretti, A. (2011). SEROTONINERGIC AND DOPAMINERGIC GENES IN BIPOLAR DISORDER AND RESPONSE TO TREATMENTS IN BIPOLAR DEPRESSION. INVESTIGATION ON A WELL-CHARACTERIZED NATURALISTIC SAMPLE. *Clinical Neuropsychiatry*, *8*(5).

Manji, H., Kato, T., Di Prospero, N. A., Ness, S., Beal, M. F., Krams, M., & Chen, G. (2012). Impaired mitochondrial function in psychiatric disorders. *Nature Reviews. Neuroscience*, *13*(5), 293–307. doi:10.1038/nrn3229

Mao, Y., Ge, X., Frank, C. L., Madison, J. M., Koehler, A. N., Doud, M. K., … Tsai, L.-H. (2009). Disrupted in schizophrenia 1 regulates neuronal progenitor proliferation via modulation of GSK3beta/beta-catenin signaling. *Cell*, *136*(6), 1017–31. doi:10.1016/j.cell.2008.12.044

Mariani, J., Simonini, M. V., Palejev, D., Tomasini, L., Coppola, G., Szekely, A. M., … Vaccarino, F. M. (2012). Modeling human cortical development in vitro using induced pluripotent stem cells. *Proceedings of the National Academy of Sciences*, *109*(31), 12770–12775. doi:10.1073/pnas.1202944109

Marshall, C. R., Noor, A., Vincent, J. B., Lionel, A. C., Feuk, L., Skaug, J., … Scherer, S. W. (2008). Structural variation of chromosomes in autism spectrum disorder. *American Journal of Human Genetics*, *82*(2), 477–88. doi:10.1016/j.ajhg.2007.12.009

Matthysse, S. (1973). Antipsychotic drug actions: a clue to the neuropathology of schizophrenia? *Federation Proceedings*, *32*(2), 200–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4348519

Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., … Stamatoyannopoulos, J. A. (2012). Systematic localization of common disease-

associated variation in regulatory DNA. *Science (New York, N.Y.)*, *337*(6099), 1190–5. doi:10.1126/science.1222794

McCartney, D. L., Walker, R. M., Morris, S. W., McIntosh, A. M., Porteous, D. J., & Evans, K. L. (2016). Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genomics Data*, *9*, 22–4. doi:10.1016/j.gdata.2016.05.012

McCullumsmith, R. E., Hammond, J. H., Shan, D., & Meador-Woodruff, J. H. (2014). Postmortem brain: an underutilized substrate for studying severe mental illness. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, *39*(1), 65–87. doi:10.1038/npp.2013.239

McGrath, J., Saha, S., Chant, D., & Welham, J. Schizophrenia: A concise overview of incidence, prevalence, and mortality. , 30 Epidemiologic Reviews 67–76 (2008). doi:10.1093/epirev/mxn001

McIntosh, A. M., Job, D. E., Moorhead, T. W. J., Harrison, L. K., Lawrie, S. M., & Johnstone, E. C. (2005). *White Matter Density in Patients with Schizophrenia, Bipolar Disorder and Their Unaffected Relatives*. *Biological Psychiatry* (Vol. 58). doi:10.1016/j.biopsych.2005.03.044

McQueen, M. B., Devlin, B., Faraone, S. V., Nimgaonkar, V. L., Sklar, P., Smoller, J. W., … Laird, N. M. (2005). Combined analysis from eleven linkage studies of bipolar disorder provides strong evidence of susceptibility loci on chromosomes 6q and 8q. *American Journal of Human Genetics*, *77*(4), 582–595. doi:10.1086/491603

Meaney, M. J., & Szyf, M. (2005). Environmental programming of stress responses through DNA methylation: Life at the interface between a dynamic environment and a fixed genome. *Dialogues in Clinical Neuroscience*, *7*(2), 103–123.

Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S., & Heintz, N. (2012). MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell*, *151*(7), 1417–30. doi:10.1016/j.cell.2012.11.022

Michael, N., Erfurth, A., Ohrmann, P., Gössling, M., Arolt, V., Heindel, W., & Pfleiderer, B. (2003). Acute mania is accompanied by elevated glutamate/glutamine levels within the left dorsolateral prefrontal cortex. *Psychopharmacology*, *168*(3), 344–6. doi:10.1007/s00213-003-1440-z

Milaneschi, Y., Lamers, F., Peyrot, W. J., Abdellaoui, A., Willemsen, G., Hottenga, J.-J., … Penninx, B. W. J. H. (2016). Polygenic dissection of major depression clinical heterogeneity. *Molecular Psychiatry*, *21*(4), 516–22. doi:10.1038/mp.2015.86

Mill, J., Tang, T., Kaminsky, Z., Khare, T., Yazdanpanah, S., Bouchard, L., … Petronis, A. (2008). Epigenomic Profiling Reveals DNA-Methylation Changes

Associated with Major Psychosis. *American Journal of Human Genetics*, *82*(3), 696–711. doi:10.1016/j.ajhg.2008.01.008

Millar, J. K., Christie, S., & Porteous, D. J. (2003). Yeast two-hybrid screens implicate DISC1 in brain development and function. *Biochemical and Biophysical Research Communications*, *311*(4), 1019–1025. doi:10.1016/j.bbrc.2003.10.101

Millar, J. K., Pickard, B. S., Mackie, S., James, R., Christie, S., Buchanan, S. R., … Porteous, D. J. (2005). DISC1 and PDE4B are interacting genetic factors in schizophrenia that regulate cAMP signaling. *Science (New York, N.Y.)*, *310*(5751), 1187–91. doi:10.1126/science.1112915

Millar, J. K., Wilson-Annan, J. C., Anderson, S., Christie, S., Taylor, M. S., Semple, C. a, … Porteous, D. J. (2000). Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Human Molecular Genetics*, *9*(9), 1415–1423. doi:10.1093/hmg/9.9.1415

Miller, D. B., & O'Callaghan, J. P. (2013). Personalized medicine in major depressive disorder -- opportunities and pitfalls. *Metabolism: Clinical and Experimental*, *62 Suppl 1*, S34–9. doi:10.1016/j.metabol.2012.08.021

Miyai, T., Maruyama, Y., Osakabe, Y., Nejima, R., Miyata, K., & Amano, S. (2008). Karyotype changes in cultured human corneal endothelial cells. *Molecular Vision*, *14*, 942–50. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18509550

Mohyuddin, A., Ayub, Q., Siddiqi, S., Carvalho-Silva, D. R., Mazhar, K., Rehman, S., … Mehdi, S. Q. (2004). Genetic instability in EBV-transformed lymphoblastoid cell lines. *Biochimica et Biophysica Acta*, *1670*(1), 81–3. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14729144

Montano, C., Taub, M. A., Jaffe, A., Briem, E., Feinberg, J. I., Trygvadottir, R., … Feinberg, A. P. (2016). Association of DNA Methylation Differences With Schizophrenia in an Epigenome-Wide Association Study. *JAMA Psychiatry*, *73*(5), 506–14. doi:10.1001/jamapsychiatry.2016.0144

Moore, G. J., Bebchuk, J. M., Wilds, I. B., Chen, G., & Menji, H. K. (2000). Lithium-induced increase in human brain grey matter. *The Lancet*, *356*(9237), 1241–1242. doi:10.1016/S0140-6736(00)02793-8

Moreau, M. P., Bruse, S. E., David-Rus, R., Buyske, S., & Brzustowicz, L. M. (2011). Altered MicroRNA expression profiles in postmortem brain samples from individuals with schizophrenia and bipolar disorder. *Biological Psychiatry*, *69*(2), 188–193. doi:10.1016/j.biopsych.2010.09.039

Morris, J. A., Kandpal, G., Ma, L., & Austin, C. P. (2003). DISC1 (Disrupted-In-Schizophrenia 1) is a centrosome-associated protein that interacts with MAP1A,

MIPT3, ATF4/5 and NUDEL: regulation and loss of interaction with mutation. *Human Molecular Genetics*, *12*(13), 1591–608. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12812986

Moynahan, M. E., Chiu, J. W., Koller, B. H., & Jasin, M. (1999). Brca1 Controls Homology-Directed DNA Repair. *Molecular Cell*, *4*(4), 511–518. doi:10.1016/S1097-2765(00)80202-6

Muglia, P., Petronis, A., Mundo, E., Lander, S., Cate, T., & Kennedy, J. L. (2002). Dopamine D4 receptor and tyrosine hydroxylase genes in bipolar disorder: evidence for a role of DRD4. *Molecular Psychiatry*, *7*(8), 860–6. doi:10.1038/sj.mp.4001098

Mühleisen, T. W., Leber, M., Schulze, T. G., Strohmaier, J., Degenhardt, F., Treutlein, J., … Cichon, S. (2014). Genome-wide association study reveals two new risk loci for bipolar disorder. *Nature Communications*, *5*, 3339. doi:10.1038/ncomms4339

Muir, W. J., Pickard, B. S., & Blackwood, D. H. R. (2006). Chromosomal abnormalities and psychosis. *British Journal of Psychiatry*. doi:10.1192/bjp.bp.106.023895

Müller, D. J., Zai, C. C., Shinkai, T., Strauss, J., & Kennedy, J. L. (2011). Association between the DAOA/G72 gene and bipolar disorder and meta-analyses in bipolar disorder and schizophrenia. *Bipolar Disorders*, *13*(2), 198–207. doi:10.1111/j.1399-5618.2011.00905.x

Mundo, E., Tharmalingham, S., Neves-Pereira, M., Dalton, E. J., Macciardi, F., Parikh, S. V, … Kennedy, J. L. (2003). Evidence that the N-methyl-D-aspartate subunit 1 receptor gene (GRIN1) confers susceptibility to bipolar disorder. *Molecular Psychiatry*, *8*(2), 241–5. doi:10.1038/sj.mp.4001218

Murata, S., Yoshiara, T., Lim, C. R., Sugino, M., Kogure, M., Ohnuki, T., … Matsubara, K. (2005). Psychophysiological stress-regulated gene expression in mice. *FEBS Letters*, *579*(10), 2137–42. doi:10.1016/j.febslet.2005.02.069

Murdoch, H., Mackie, S., Collins, D. M., Hill, E. V, Bolger, G. B., Klussmann, E., … Houslay, M. D. (2007). Isoform-selective susceptibility of DISC1/phosphodiesterase-4 complexes to dissociation by elevated intracellular cAMP levels. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(35), 9513–24. doi:10.1523/JNEUROSCI.1493-07.2007

Murray, C.] L. (n.d.). BURDEN OF DISEASE A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020 EDITED BY.

Nazor, K. L., Altun, G., Lynch, C., Tran, H., Harness, J. V., Slavin, I., … Laurent, L. C. (2012). Recurrent Variations in DNA Methylation in Human Pluripotent

Stem Cells and Their Differentiated Derivatives. *Cell Stem Cell*, *10*(5), 620–634. doi:10.1016/j.stem.2012.02.013

Nestor, C., Ruzov, A., Meehan, R. R., & Dunican, D. S. (2010). Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. *BioTechniques*, *48*(4), 317–319. doi:10.2144/000113403

Neves-Pereira, M., Mundo, E., Muglia, P., King, N., Macciardi, F., & Kennedy, J. L. (2002). *The Brain-Derived Neurotrophic Factor Gene Confers Susceptibility to Bipolar Disorder: Evidence from a Family-Based Association Study*. *The American Journal of Human Genetics* (Vol. 71). doi:10.1086/342288

Ng, M. Y. M., Levinson, D. F., Faraone, S. V, Suarez, B. K., DeLisi, L. E., Arinami, T., … Lewis, C. M. (2009). Meta-analysis of 32 genome-wide linkage studies of schizophrenia. *Molecular Psychiatry*, *14*(8), 774–785. doi:10.1038/mp.2008.135

Nicodemus, K. K., Callicott, J. H., Higier, R. G., Luna, A., Nixon, D. C., Lipska, B. K., … Weinberger, D. R. (2010). Evidence of statistical epistasis between DISC1, CIT and NDEL1 impacting risk for schizophrenia: biological validation with functional neuroimaging. *Human Genetics*, *127*(4), 441–52. doi:10.1007/s00439-009-0782-y

Niculescu, A. B. (2014). DISCovery in psychiatric genetics. *Molecular Psychiatry*, *19*(2), 145–145. doi:10.1038/mp.2013.176

Nishioka, M., Bundo, M., Koike, S., Takizawa, R., Kakiuchi, C., Araki, T., … Iwamoto, K. (2013). Comprehensive DNA methylation analysis of peripheral blood cells derived from patients with first-episode schizophrenia. *Journal of Human Genetics*, *58*(2), 91–7. doi:10.1038/jhg.2012.140

Nivoli, A. M. A., Pacchiarotti, I., Rosa, A. R., Popovic, D., Murru, A., Valenti, M., … Colom, F. (2011). Gender differences in a cohort study of 604 bipolar patients: The role of predominant polarity. *Journal of Affective Disorders*, *133*(3), 443–449. doi:10.1016/j.jad.2011.04.055

Nohesara, S., Ghadirivasfi, M., Mostafavi, S., Eskandari, M.-R. R., Ahmadkhaniha, H., Thiagalingam, S., & Abdolmaleky, H. M. (2011). DNA hypomethylation of MB-COMT promoter in the DNA derived from saliva in schizophrenia and bipolar disorder. *Journal of Psychiatric Research*, *45*(11), 1432–1438. doi:10.1016/j.jpsychires.2011.06.013

Numata, S., Ishii, K., Tajima, A., Iga, J., Kinoshita, M., Watanabe, S., … Ohmori, T. (2015). Blood diagnostic biomarkers for major depressive disorder using multiplex DNA methylation profiles: discovery and validation. *Epigenetics : Official Journal of the DNA Methylation Society*, *10*(2), 135–41. doi:10.1080/15592294.2014.1003743

Numata, S., Ye, T., Herman, M., & Lipska, B. K. (2014). DNA methylation changes in the postmortem dorsolateral prefrontal cortex of patients with schizophrenia. *Frontiers in Genetics*, *5*, 280. doi:10.3389/fgene.2014.00280

Nykjaer, A., & Willnow, T. E. (2012). Sortilin: a receptor to regulate neuronal viability and function. *Trends in Neurosciences*, *35*(4), 261–70. doi:10.1016/j.tins.2012.01.003

O 'carroll, R. (2000). Cognitive impairment in schizophrenia. *APT*, *6*(6), 161–161.

O'Donnell, J. M., & Zhang, H.-T. (2004). Antidepressant effects of inhibitors of cAMP phosphodiesterase (PDE4). *Trends in Pharmacological Sciences*, *25*(3), 158–163. doi:10.1016/j.tips.2004.01.003

O'Donovan, M. C., Craddock, N., Norton, N., Williams, H., Peirce, T., Moskvina, V., … Owen, M. J. (2008). Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nature Genetics*, *40*(9), 1053–1055. doi:10.1038/ng.201

Ogawa, F., Malavasi, E. L. V, Crummie, D. K., Eykelenboom, J. E., Soares, D. C., Mackie, S., … Millar, J. K. (2014). DISC1 complexes with TRAK1 and Miro1 to modulate anterograde axonal mitochondrial trafficking. *Human Molecular Genetics*, *23*(4), 906–919. doi:10.1093/hmg/ddt485

Ogawa, F., Murphy, L. C., Malavasi, E. L. V, O'Sullivan, S. T., Torrance, H. S., Porteous, D. J., & Millar, J. K. (2016). NDE1 and GSK3β Associate with TRAK1 and Regulate Axonal Mitochondrial Motility: Identification of Cyclic AMP as a Novel Modulator of Axonal Mitochondrial Trafficking. *ACS Chemical Neuroscience*, *7*(5), 553–64. doi:10.1021/acschemneuro.5b00255

Olsson, A., Höglund, K., Sjögren, M., Andreasen, N., Minthon, L., Lannfelt, L., … Blennow, K. (2003). Measurement of α- and β-secretase cleaved amyloid precursor protein in cerebrospinal fluid from Alzheimer patients. *Experimental Neurology*, *183*(1), 74–80. doi:10.1016/S0014-4886(03)00027-X

Ong, D., Walterfang, M., Malhi, G. S., Styner, M., Velakoulis, D., & Pantelis, C. (2012). Size and shape of the caudate nucleus in individuals with bipolar affective disorder. *The Australian and New Zealand Journal of Psychiatry*, *46*(4), 340–51. doi:10.1177/0004867412440191

Ou, Z., Berg, J. S., Yonath, H., Enciso, V. B., Miller, D. T., Picker, J., … Patel, A. (2008). Microduplications of 22q11.2 are frequently inherited and are associated with variable phenotypes. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, *10*(4), 267–77. doi:10.1097/GIM.0b013e31816b64c2

Overhauser, J., Berrettini, W. H., & Rojas, K. (1998). Affective disorder associated with a balanced translocation involving chromosome 14 and 18. *Psychiatric*

Genetics, *8*(2), 53–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9686423

Owczarek, S., Bang, M. L., Berezin, V., Owczarek, S., Bang, M. L., & Berezin, V. (2015). Neurexin-Neuroligin Synaptic Complex Regulates Schizophrenia-Related DISC1/Kal-7/Rac1 "Signalosome". *Neural Plasticity*, *2015*, 167308. doi:10.1155/2015/167308

Owen, M. J., Craddock, N., & O'Donovan, M. C. (2005). Schizophrenia: genes at last? *Trends in Genetics*, *21*(9), 518–525. doi:10.1016/j.tig.2005.06.011

Ozeki, Y., Tomoda, T., Kleiderlein, J., Kamiya, A., Bord, L., Fujii, K., … Sawa, A. (2003). Disrupted-in-Schizophrenia-1 (DISC-1): Mutant truncation prevents binding to NudE-like (NUDEL) and inhibits neurite outgrowth. *Proceedings of the National Academy of Sciences*, *100*(1), 289–294. doi:10.1073/pnas.0136913100

Park, J. P., Moeschler, J. B., Berg, S. Z., & Wurster-Hill, D. H. (1991). Schizophrenia and mental retardation in an adult male with a de novo interstitial deletion 9(q32q34.1). *Journal of Medical Genetics*, *28*(4), 282–3. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1856838

Park, N., Juo, S. H., Cheng, R., Liu, J., Loth, J. E., Lilliston, B., … Baron, M. (2004). Linkage analysis of psychosis in bipolar pedigrees suggests novel putative loci for bipolar disorder and shared susceptibility with schizophrenia. *Molecular Psychiatry*, *9*(12), 1091–9. doi:10.1038/sj.mp.4001541

Pe'er, I., Yelensky, R., Altshuler, D., & Daly, M. J. (2008). Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genetic Epidemiology*, *32*(4), 381–5. doi:10.1002/gepi.20303

Piccinelli, M., & Wilkinson, G. (2000). Gender differences in depression. Critical review. *The British Journal of Psychiatry : The Journal of Mental Science*, *177*(6), 486–92. doi:10.1192/bjp.177.6.486

Pickard, B. S., Malloy, M. P., Clark, L., Lehellard, S., Ewald, H. L., Mors, O., … Muir, W. J. (2005). Candidate psychiatric illness genes identified in patients with pericentric inversions of chromosome 18. *Psychiatric Genetics*, *15*(1), 37–44. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15722956

Pidsley, R., & Mill, J. (2011). Epigenetic studies of psychosis: Current findings, methodological approaches, and implications for postmortem research. *Biological Psychiatry*. doi:10.1016/j.biopsych.2010.03.029

Pidsley, R., Viana, J., Hannon, E., Spiers, H., Troakes, C., Al-Saraj, S., … Mill, J. (2014). Methylomic profiling of human brain tissue supports a neurodevelopmental origin for schizophrenia. *Genome Biology*, *15*(10), 483. doi:10.1186/s13059-014-0483-2

Pidsley, R., Y Wong, C. C., Volta, M., Lunnon, K., Mill, J., & Schalkwyk, L. C. (2013). A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*, *14*, 293. doi:10.1186/1471-2164-14-293

Porteous, D. J., Thomson, P. A., Millar, J. K., Evans, K. L., Hennah, W., Soares, D. C., … Blackwood, D. H. (2014). DISC1 as a genetic risk factor for schizophrenia and related major mental illness: response to Sullivan. *Molecular Psychiatry*, *19*(2), 141–143. doi:10.1038/mp.2013.160

Prata, D., Breen, G., Osborne, S., Munro, J., St. Clair, D., & Collier, D. (2008). Association of DAO and G72(DAOA)/G30 genes with bipolar affective disorder. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, *147*(6), 914–917. doi:10.1002/ajmg.b.30682

Psychiatric GWAS Consortium Bipolar Disorder Working Group. (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genetics*, *43*(10), 977–83. doi:10.1038/ng.943

Pu, L., Igbavboa, U., Wood, W. G., Roths, J. B., Kier, A. B., Spener, F., & Schroeder, F. (1999). Expression of fatty acid binding proteins is altered in aged mouse brain. *Molecular and Cellular Biochemistry*, *198*(1-2), 69–78. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10497880

Purcell, S. M., Moran, J. L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., … Sklar, P. (2014). A polygenic burden of rare disruptive mutations in schizophrenia. *Nature*, *506*(7487), 185–90. doi:10.1038/nature12975

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. doi:10.1086/519795

R Core Team. (2013). R Core Team. *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.*, ISBN 3–900051–07–0, URL http://www.R–project.org/. Retrieved from http://www.mendeley.com/research/r-language-environment-statistical-computing-96/\npapers2://publication/uuid/A1207DAB-22D3-4A04-82FB-D4DD5AD57C28

Rajkumar, A. P., Christensen, J. H., Mattheisen, M., Jacobsen, I., Bache, I., Pallesen, J., … Børglum, A. D. (2015). Analysis of t(9;17)(q33.2;q25.3) chromosomal breakpoint regions and genetic association reveals novel candidate genes for bipolar disorder. *Bipolar Disorders*, *17*(2), 205–11. doi:10.1111/bdi.12239

Ramalingam, A., Zhou, X.-G., Fiedler, S. D., Brawner, S. J., Joyce, J. M., Liu, H.-Y., & Yu, S. (2011). 16p13.11 duplication is a risk factor for a wide spectrum of

neuropsychiatric disorders. *Journal of Human Genetics*, *56*(7), 541–544. doi:10.1038/jhg.2011.42

Rampino, A., Walker, R. M., Torrance, H. S., Anderson, S. M., Fazio, L., Di Giorgio, A., … Evans, K. L. (2014). Expression of DISC1-interactome members correlates with cognitive phenotypes related to schizophrenia. *PloS One*, *9*(6), e99892. doi:10.1371/journal.pone.0099892

Ran, F. A., Hsu, P. D. P. P. D., Wright, J., Agarwala, V., Scott, D. a, & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, *8*(11), 2281–308. doi:10.1038/nprot.2013.143

Rapoport, J. L., Giedd, J. N., & Gogtay, N. (2012). Neurodevelopmental model of schizophrenia: update 2012. *Molecular Psychiatry*, *17*(12), 1228–38. doi:10.1038/mp.2012.23

Rees, E., Walters, J. T. R., Georgieva, L., Isles, A. R., Chambert, K. D., Richards, A. L., … Kirov, G. (2014). Analysis of copy number variations at 15 schizophrenia-associated loci. *The British Journal of Psychiatry : The Journal of Mental Science*, *204*(2), 108–14. doi:10.1192/bjp.bp.113.131052

Reinhardt, V. (2004). Common husbandry-related variables in biomedical research with animals. *Laboratory Animals*, *38*(3), 213–235. doi:10.1258/002367704323133600

Reitz, C., Tokuhiro, S., Clark, L. N., Conrad, C., Vonsattel, J.-P., Hazrati, L.-N., … Mayeux, R. (2011). SORCS1 alters amyloid precursor protein processing and variants may increase Alzheimer's disease risk. *Annals of Neurology*, *69*(1), 47–64. doi:10.1002/ana.22308

Reitz, C., Tosto, G., Vardarajan, B., Rogaeva, E., Ghani, M., Rogers, R. S., … Alzheimer's Disease Genetics Consortium (ADGC). (2013). Independent and epistatic effects of variants in VPS10-d receptors on Alzheimer disease risk and processing of the amyloid precursor protein (APP). *Translational Psychiatry*, *3*, e256. doi:10.1038/tp.2013.13

Rice, S., Niu, N., Berman, D., Heston, L., & Sobell, J. (n.d.). Identification of single nucleotide polymorphisms (SNPs) and other sequence changes and estimation of nucleotide diversity in coding and flanking regions of the NMDAR1 receptor gene in schizophrenic patients. doi:10.1038/sj.mp.4000838

Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J. L., Kähler, A. K., Akterin, S., … Sullivan, P. F. (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genetics*, *45*(10), 1150–9. doi:10.1038/ng.2742

Ripke, S., Sanders, A. R., Kendler, K. S., Levinson, D. F., Sklar, P., Holmans, P. A., … Gejman, P. V. (2011). Genome-wide association study identifies five new schizophrenia loci. *Nature Genetics*, *43*(10), 969–976. doi:10.1038/ng.940

Ripke, S., Wray, N. R., Lewis, C. M., Hamilton, S. P., Weissman, M. M., Breen, G., … Sullivan, P. F. (2013). A mega-analysis of genome-wide association studies for major depressive disorder. *Molecular Psychiatry*, *18*(4), 497–511. doi:10.1038/mp.2012.21

Rocke, D. M., & Durbin, B. (2001). A model for measurement error for gene expression arrays. *Journal of Computational Biology : A Journal of Computational Molecular Cell Biology*, *8*(6), 557–569. doi:10.1089/106652701753307485

Rogaeva, E., Meng, Y., Lee, J. H., Gu, Y., Kawarai, T., Zou, F., … St George-Hyslop, P. (2007). The neuronal Sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nature Genetics*, *39*(2), 168–77. doi:10.1038/ng1943

Roig, B., Virgos, C., Franco, N., Martorell, L., Valero, J., Costas, J., … Vilella, E. (2007). The discoidin domain receptor 1 as a novel susceptibility gene for schizophrenia. *Molecular Psychiatry*, *12*(9), 833–41. doi:10.1038/sj.mp.4001995

Roth, R. M., Flashman, L. A., Saykin, A. J., Thomas McAllister, P. W., & Vidaver, R. (2004). Apathy in Schizophrenia: Reduced Frontal Lobe Volume and Neuropsychological Deficits. *Am J PsychiatryAm J Psychiatry*, *1611*(161), 157–159.

Ruan, C.-S., Yang, C.-R., Li, J.-Y., Luo, H.-Y., Bobrovskaya, L., & Zhou, X.-F. (2016). Mice with Sort1 deficiency display normal cognition but elevated anxiety-like behavior. *Experimental Neurology*, *281*, 99–108. doi:10.1016/j.expneurol.2016.04.015

Ruderfer, D. M., Fanous, A. H., Ripke, S., McQuillin, A., Amdur, R. L., Schizophrenia Working Group of Psychiatric Genomics Consortium, … Kendler. (2014). Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Molecular Psychiatry*, *19*(9), 1017–24. doi:10.1038/mp.2013.138

Sachs, G. S., Nierenberg, A. A., Calabrese, J. R., Marangell, L. B., Wisniewski, S. R., Gyulai, L., … Thase, M. E. (2009). Effectiveness of Adjunctive Antidepressant Treatment for Bipolar Depression. *http://dx.doi.org/10.1056/NEJMoa064135*.

Sachs, N. A., Sawa, A., Holmes, S. E., Ross, C. A., DeLisi, L. E., & Margolis, R. L. (2005). A frameshift mutation in Disrupted in Schizophrenia 1 in an American

family with schizophrenia and schizoaffective disorder. *Molecular Psychiatry*, *10*(8), 758–764. doi:10.1038/sj.mp.4001667

Schalkwyk, L. C., Pidsley, R., & CCY, W. (2013). wateRmelon: Illumina 450 methylation array normalization and metrics. *R Package Version 1.2.2*.

Schaub, M. A., Boyle, A. P., Kundaje, A., Batzoglou, S., & Snyder, M. (2012). Linking disease associations with regulatory information in the human genome. *Genome Research*, *22*(9), 1748–1759. doi:10.1101/gr.136127.111

Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, *511*(7510), 421–7. doi:10.1038/nature13595

Schloss, P., & Williams, D. C. (1998). The serotonin transporter: a primary target for antidepressant drugs. *Journal of Psychopharmacology (Oxford, England)*, *12*(2), 115–121. doi:10.1177/026988119801200201

Schmidt-Kastner, R., van Os, J., W M Steinbusch, H., & Schmitz, C. (2006). Gene regulation by hypoxia and the neurodevelopmental origin of schizophrenia. *Schizophrenia Research*, *84*(2-3), 253–71. doi:10.1016/j.schres.2006.02.022

Schwarz, D. F., König, I. R., & Ziegler, A. (2010). On safari to random Jungle: A fast implementation of random forests for high-dimensional data. *Bioinformatics*, *26*(14), 1752–1758. doi:10.1093/bioinformatics/btq257

Schwarz, D. F., Szymczak, S., Ziegler, A., & König, I. R. (2007). Picking single-nucleotide polymorphisms in forests. *BMC Proceedings*, *1 Suppl 1*, S59. doi:10.1186/1753-6561-1-s1-s59

Seeman, P., & Lee, T. (1975). Antipsychotic drugs: direct correlation between clinical potency and presynaptic action on dopamine neurons. *Science (New York, N.Y.)*, *188*(4194), 1217–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1145194

Segurado, R., Detera-Wadleigh, S. D., Levinson, D. F., Lewis, C. M., Gill, M., Nurnberger, J. I., … Seguardo Detera-Wadleigh, S.D., Levinson, D.F., et al., R. (2003). Genome scan meta-analysis of schizophrenia and bipolar disorder, part III: bipolar disorder. *American Journal of Human Genetics*, *73*(1), 49–62. doi:10.1086/376547

Seifuddin, F., Mahon, P. B., Judy, J., Pirooznia, M., Jancic, D., Taylor, J., … Zandi, P. P. (2012). Meta-analysis of genetic association studies on bipolar disorder. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics : The Official Publication of the International Society of Psychiatric Genetics*, *159B*(5), 508–18. doi:10.1002/ajmg.b.32057

Sekar, A., Bialas, A. R., de Rivera, H., Davis, A., Hammond, T. R., Kamitaki, N., … McCarroll, S. A. (2016). Schizophrenia risk from complex variation of complement component 4. *Nature*, *530*(7589), 177–83. doi:10.1038/nature16549

Serretti, A., Cristina, S., Lilli, R., Cusin, C., Lattuada, E., Lorenzi, C., … Smeraldi, E. (2002). Family-based association study of 5-HTTLPR, TPH, MAO-A, and DRD4 polymorphisms in mood disorders. *American Journal of Medical Genetics - Neuropsychiatric Genetics*, *114*(4), 361–369. doi:10.1002/ajmg.10356

Serretti, A., & Mandelli, L. (2008a). The genetics of bipolar disorder: genome 'hot regions,' genes, new potential candidates and future directions. *Molecular Psychiatry*, *13*(8), 742–71. doi:10.1038/mp.2008.29

Serretti, A., & Mandelli, L. (2008b). The genetics of bipolar disorder: genome "hot regions," genes, new potential candidates and future directions. *Molecular Psychiatry*, *13*(8), 742–771. doi:10.1038/mp.2008.29

Shahani, N., Seshadri, S., Jaaro-Peled, H., Ishizuka, K., Hirota-Tsuyada, Y., Wang, Q., … Sawa, A. (2015). DISC1 regulates trafficking and processing of APP and Aβ generation. *Molecular Psychiatry*, *20*(7), 874–9. doi:10.1038/mp.2014.100

Sher, L. (2001). Candidate gene studies in psychiatric disorders: promises and limitations. *Journal of Psychiatry & Neuroscience*, *26*(103).

Shi, J., Badner, J. A., Gershon, E. S., & Liu, C. (2008). Allelic association of G72/G30 with schizophrenia and bipolar disorder: A comprehensive meta-analysis. *Schizophrenia Research*, *98*(1-3), 89–97. doi:10.1016/j.schres.2007.10.004

Shi, J., Levinson, D. F., Duan, J., Sanders, A. R., Zheng, Y., Pe'er, I., … Gejman, P. V. (2009). Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature*, *460*(7256), 753–7. doi:10.1038/nature08192

Shih, R. A., Belmonte, P. L., & Zandi, P. P. (2004). A review of the evidence from family, twin and adoption studies for a genetic contribution to adult psychiatric disorders. *International Review of Psychiatry (Abingdon, England)*, *16*(4), 260–83. doi:10.1080/09540260400014401

Shimamoto, C., Ohnishi, T., Maekawa, M., Watanabe, A., Ohba, H., Arai, R., … Yoshikawa, T. (2014). Functional characterization of FABP3, 5 and 7 gene variants identified in schizophrenia and autism spectrum disorder and mouse behavioral studies. *Human Molecular Genetics*, *23*(24), 6495–511. doi:10.1093/hmg/ddu369

Shin, J., Bourdon, C., Bernard, M., Wilson, M. D., Reischl, E., Waldenberger, M., … Pausova, Z. (2015). Layered genetic control of DNA methylation and gene

expression: A locus of multiple sclerosis in healthy individuals. *Human Molecular Genetics*, *24*(20), 5733–5745. doi:10.1093/hmg/ddv294

Shirley, M. D., Baugher, J. D., Stevens, E. L., Tang, Z., Gerry, N., Beiswanger, C. M., … Pevsner, J. (2012). Chromosomal variation in lymphoblastoid cell lines. *Human Mutation*, *33*(7), 1075–86. doi:10.1002/humu.22062

Shoji, H., Toyama, K., Takamiya, Y., Wakana, S., Gondo, Y., & Miyakawa, T. (2012). Comprehensive behavioral analysis of ENU-induced Disc1-Q31L and -L100P mutant mice. *BMC Research Notes*, *5*, 108. doi:10.1186/1756-0500-5-108

Singh, K. K., De Rienzo, G., Drane, L., Mao, Y., Flood, Z., Madison, J., … Tsai, L. H. (2011). Common DISC1 polymorphisms disrupt Wnt/GSK3?? signaling and brain development. *Neuron*, *72*(4), 545–558. doi:10.1016/j.neuron.2011.09.030

Sklar, P., Smoller, J. W., Fan, J., Ferreira, M. A. R., Perlis, R. H., Chambert, K., … Purcell, S. M. (2008). Whole-genome association study of bipolar disorder. *Molecular Psychiatry*, *13*(6), 558–69. doi:10.1038/sj.mp.4002151

Smith, A. K., Kilaru, V., Kocak, M., Almli, L. M., Mercer, K. B., Ressler, K. J., … Conneely, K. N. (2014). Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type. *BMC Genomics*, *15*, 145. doi:10.1186/1471-2164-15-145

Smith, B. H., Campbell, A., Linksted, P., Fitzpatrick, B., Jackson, C., Kerr, S. M., … Morris, A. D. (2013). Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *International Journal of Epidemiology*, *42*(3), 689–700. doi:10.1093/ije/dys084

Smith, G. K. (2005). limma: Linear Models for Microarray Data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (pp. 397–420). doi:citeulike-article-id:5722720

Smith, R. S. (1992). A comprehensive macrophage-T-lymphocyte theory of schizophrenia. *Medical Hypotheses*, *39*(3), 248–257. doi:10.1016/0306-9877(92)90117-U

Smoller, J. W. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. *The Lancet*, *381*(9875), 1371–1379. doi:10.1016/S0140-6736(12)62129-1

Soares, D. C., Carlyle, B. C., Bradshaw, N. J., & Porteous, D. J. (2011). DISC1: Structure, function, and therapeutic potential for major mental illness. *ACS Chemical Neuroscience*. doi:10.1021/cn200062k

Soda, T., Frank, C., Ishizuka, K., Baccarella, A., Park, Y.-U., Flood, Z., … Tsai, L.-H. (2013). DISC1-ATF4 transcriptional repression complex: dual regulation of the cAMP-PDE4 cascade by DISC1. *Molecular Psychiatry*, *18*(8), 898–908. doi:10.1038/mp.2013.38

Sofer, T., Schifano, E. D., Hoppin, J. A., Hou, L., & Baccarelli, A. A. (2013). A-clustering: a novel method for the detection of co-regulated methylation regions, and regions associated with exposure. *Bioinformatics*, *29*(22), 2884–2891. doi:10.1093/bioinformatics/btt498

Srikanth, P., Han, K., Callahan, D. G., Makovkina, E., Muratore, C. R., Lalli, M. A., … Young-Pearse, T. L. (2015). Genomic DISC1 Disruption in hiPSCs Alters Wnt Signaling and Neural Cell Fate. *Cell Reports*, *12*(9), 1414–1429. doi:10.1016/j.celrep.2015.07.061

St Clair, D., Blackwood, D., Muir, W., Walker, M., St Clair, D., Muir, W., … Evans, H. J. (1990). Association within a family of a balanced autosomal translocation with major mental illness. *The Lancet*, *336*(8706), 13–16. doi:10.1016/0140-6736(90)91520-K

Stefansson, H., Ophoff, R. A., Steinberg, S., Andreassen, O. A., Cichon, S., Rujescu, D., … Collier, D. A. (2009). Common variants conferring risk of schizophrenia. *Nature*, *460*(7256), 744–7. doi:10.1038/nature08186

Steinberg, S., de Jong, S., Irish Schizophrenia Genomics Consortium, O. A., Andreassen, O. A., Werge, T., Børglum, A. D., … Stefansson, K. (2011). Common variants at VRK2 and TCF4 conferring risk of schizophrenia. *Human Molecular Genetics*, *20*(20), 4076–81. doi:10.1093/hmg/ddr325

Stelzhammer, V., Guest, P. C., Rothermundt, M., Sondermann, C., Michael, N., Schwarz, E., … Bahn, S. (2013). Electroconvulsive therapy exerts mainly acute molecular changes in serum of major depressive disorder patients. *European Neuropsychopharmacology : The Journal of the European College of Neuropsychopharmacology*, *23*(10), 1199–207. doi:10.1016/j.euroneuro.2012.10.012

Stewart, S. K., Morris, T. J., Guilhamon, P., Bulstrode, H., Bachman, M., Balasubramanian, S., & Beck, S. (2015). OxBS-450K: A method for analysing hydroxymethylation using 450K BeadChips. *Methods*, *72*(C), 9–15. doi:10.1016/j.ymeth.2014.08.009

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., … Collins, R. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine*, *12*(3), e1001779. doi:10.1371/journal.pmed.1001779

Sullivan, P. F. (2005). The Genetics of Schizophrenia. *PLoS Medicine*, *2*(7), e212. doi:10.1371/journal.pmed.0020212

Sullivan, P. F. (2013). Questions about DISC1 as a genetic risk factor for schizophrenia. *Molecular Psychiatry*, *18*(10), 1050–2. doi:10.1038/mp.2012.182

Sullivan, P. F., Fan, C., & Perou, C. M. (2006). Evaluating the comparability of gene expression in blood and brain. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *141B*(3), 261–268. doi:10.1002/ajmg.b.30272

Sullivan, P. F., Neale, M. C., Ph, D., & Kendler, K. S. (2000). Genetic epidemiology of major depression: review and meta-analysis. *The American Journal of Psychiatry*, *157*(10), 1552–62. doi:10.1176/appi.ajp.157.10.1552

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*, *131*(5), 861–872. doi:10.1016/j.cell.2007.11.019

Talge, N. M., Neal, C., & Glover, V. (2007). Antenatal maternal stress and long-term effects on child neurodevelopment: how and why? *Journal of Child Psychology and Psychiatry*, *48*(3-4), 245–261. doi:10.1111/j.1469-7610.2006.01714.x

Tang, Y., Wang, F., Xie, G., Liu, J., Li, L., Su, L., … Blumberg, H. P. Reduced ventral anterior cingulate and amygdala volumes in medication-naïve females with major depressive disorder: A voxel-based morphometric magnetic resonance imaging study. , 156 Psychiatry Research: Neuroimaging 83–86 (2007). doi:10.1016/j.pscychresns.2007.03.005

Teroganova, N., Girshkin, L., Suter, C. M., & Green, M. J. (2016). DNA methylation in peripheral tissue of schizophrenia and bipolar disorder: a systematic review. *BMC Genetics*, *17*(1), 27. doi:10.1186/s12863-016-0332-2

Teschendorff, A. E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., & Beck, S. (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics (Oxford, England)*, *29*(2), 189–96. doi:10.1093/bioinformatics/bts680

Thompson, K. L., Pine, P. S., Rosenzweig, B. a, Turpaz, Y., & Retief, J. (2007). Characterization of the effect of sample quality on high density oligonucleotide microarray data using progressively degraded rat liver RNA. *BMC Biotechnology*, *7*, 57. doi:10.1186/1472-6750-7-57

Thomson, P. A., Duff, B., Blackwood, D. H. R., Romaniuk, L., Watson, A., Whalley, H. C., … Almasy, L. (2016). Balanced translocation linked to psychiatric disorder, glutamate, and cortical structure/function. *Npj Schizophrenia*, *2*, 16024. doi:10.1038/npjschz.2016.24

Thomson, P. A., Malavasi, E. L. V, Grünewald, E., Soares, D. C., Borkowska, M., & Millar, J. K. (2013). DISC1 genetics, biology and psychiatric illness. *Frontiers in Biology*, *8*(1), 1–31. doi:10.1007/s11515-012-1254-7

Thomson, P. A., Parla, J. S., McRae, A. F., Kramer, M., Ramakrishnan, K., Yao, J., … Porteous, D. J. (2014). 708 Common and 2010 rare DISC1 locus variants identified in 1542 subjects: analysis for association with psychiatric disorder and cognitive traits. *Molecular Psychiatry*, *19*(6), 668–75. doi:10.1038/mp.2013.68

Triche, T. J., Weisenberger, D. J., Van Den Berg, D., Laird, P. W., & Siegmund, K. D. (2013). Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Research*, *41*(7), 1–11. doi:10.1093/nar/gkt090

Turetsky, B., Cowell, P. E., Gur, R. C., Grossman, R. I., Shtasel, D. L., & Gur, R. E. (1995). Frontal and temporal lobe brain volumes in schizophrenia. Relationship to symptoms and clinical subtype. *Archives of General Psychiatry*, *52*(12), 1061–70. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7492258

Upthegrove, R., & Barnes, N. (2014). The immune system and schizophrenia: \nan update for clinicians. *Advances in Psychiatric Treatment*, *20*(2), 83–91. doi:10.1192/apt.bp.113.011452

Van Cauwenberghe, C., Van Broeckhoven, C., & Sleegers, K. (2016). The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genetics in Medicine*, *18*(5), 421–430. doi:10.1038/gim.2015.117

van Spronsen, M., & Hoogenraad, C. C. (2010). Synapse pathology in psychiatric and neurologic disease. *Current Neurology and Neuroscience Reports*, *10*(3), 207–14. doi:10.1007/s11910-010-0104-8

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, *3*(7), RESEARCH0034. doi:10.1186/gb-2002-3-7-research0034

Viikki, M., Kampman, O., Illi, A., Setälä-Soikkeli, E., Anttila, S., Huuhka, M., … Leinonen, E. (2010). *TPH1 218A/C polymorphism is associated with major depressive disorder and its treatment response*. *Neuroscience Letters* (Vol. 468). doi:10.1016/j.neulet.2009.10.069

Vos, M., Lauwers, E., & Verstreken, P. (2010). Synaptic mitochondria in synaptic transmission and organization of vesicle pools in health and disease. *Frontiers in Synaptic Neuroscience*. doi:10.3389/fnsyn.2010.00139

Vos, T., Flaxman, A. D., Naghavi, M., Lozano, R., Michaud, C., Ezzati, M., … Memish, Z. A. (2012). Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global

Burden of Disease Study 2010. *Lancet*, *380*(9859), 2163–96. doi:10.1016/S0140-6736(12)61729-2

Walker, B. A., Wardell, C. P., Chiecchio, L., Smith, E. M., Boyd, K. D., Neri, A., … Vanwier, S. (2011). Aberrant global methylation patterns affect the molecular pathogenesis and prognosis of multiple myeloma. *Blood*, *117*(2), 553–62. doi:10.1182/blood-2010-04-279539

Walker, R. M., Rybka, J., Anderson, S. M., Torrance, H. S., Boxall, R., Sussmann, J. E., … Evans, K. L. (2015). Preliminary investigation of miRNA expression in individuals at high familial risk of bipolar disorder. *Journal of Psychiatric Research*, *62*, 48–55. doi:10.1016/j.jpsychires.2015.01.006

Walsh, T., McClellan, J. M., McCarthy, S. E., Addington, A. M., Pierce, S. B., Cooper, G. M., … Sebat, J. (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science (New York, N.Y.)*, *320*(5875), 539–43. doi:10.1126/science.1155174

Walters, J. T. R., Corvin, A., Owen, M. J., Williams, H., Dragovic, M., Quinn, E. M., … Donohoe, G. (2010). Psychosis susceptibility gene ZNF804A and cognitive performance in schizophrenia. *Archives of General Psychiatry*, *67*(7), 692–700. doi:10.1001/archgenpsychiatry.2010.81

Walton, E., Hass, J., Liu, J., Roffman, J. L., Bernardoni, F., Roessner, V., … Ehrlich, S. (2015). Correspondence of DNA Methylation Between Blood and Brain Tissue and its Application to Schizophrenia Research. *Schizophrenia Bulletin*, *42*(2), sbv074–. doi:10.1093/schbul/sbv074

Wan, Q.-F., Zhou, Z.-Y., Thakur, P., Vila, A., Sherry, D. M., Janz, R., & Heidelberger, R. (2010). SV2 acts via presynaptic calcium to regulate neurotransmitter release. *Neuron*, *66*(6), 884–95. doi:10.1016/j.neuron.2010.05.010

Wang, Q., Charych, E. I., Pulito, V. L., Lee, J. B., Graziane, N. M., Crozier, R. A., … Brandon, N. J. (2011). The psychiatric disease risk factors DISC1 and TNIK interact to regulate synapse composition and function. *Molecular Psychiatry*, *16*(10), 1006–1023. doi:10.1038/mp.2010.87

Wang, S., Liang, Q., Qiao, H., Li, H., Shen, T., Ji, F., & Jiao, J. (2016). DISC1 regulates astrogenesis in the embryonic brain via modulation of RAS/MEK/ERK signaling through RASSF7. *Development (Cambridge, England)*, *143*(15), 2732–40. doi:10.1242/dev.133066

Watanabe, K., Hashimoto, E., Ukai, W., Ishii, T., Yoshinaga, T., Ono, T., … Saito, T. (2010). Effect of antidepressants on brain-derived neurotrophic factor (BDNF) release from platelets in the rats. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *34*(8), 1450–1454. doi:10.1016/j.pnpbp.2010.07.036

Wei, J., & Hemmings, G. P. (2004). TNXB locus may be a candidate gene predisposing to schizophrenia. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics : The Official Publication of the International Society of Psychiatric Genetics*, *125B*(1), 43–49. doi:10.1002/ajmg.b.20093

Wei, W.-H., Hemani, G., & Haley, C. S. (2014). Detecting epistasis in human complex traits. *Nature Reviews Genetics*, *15*(11), 722–733. doi:10.1038/nrg3747

Weinberger, D. R., Egan, M. F., Bertolino, A., Callicott, J. H., Mattay, V. S., Lipska, B. K., … Goldberg, T. E. (2001). Prefrontal neurons and the genetics of schizophrenia. *Biological Psychiatry*, *50*(11), 825–844. doi:10.1016/S0006-3223(01)01252-5

Weinberger, D. R., Meltzer HY, S. S., Weinberger DR, K. J., DR, W., Adams RD, L. G., Bogerts B, M. E. S.-B. R., … Linn MQW, C. E. K. J. H. G. L. H. (1987). Implications of Normal Brain Development for the Pathogenesis of Schizophrenia. *Archives of General Psychiatry*, *44*(7), 660. doi:10.1001/archpsyc.1987.01800190080012

Wen, Z., Nguyen, H. N., Guo, Z., Lalli, M. A., Wang, X., Su, Y., … Ming, G. (2014). Synaptic dysregulation in a human iPS cell model of mental disorders. *Nature*, *515*(7527), 414–418. doi:10.1038/nature13716

Whalley HC, Dimitrova R, Sprooten E, Dauvermann MR, Romaniuk L, Duff B, et al. (2015). Effects of a Balanced Translocation between Chromosomes 1 and 11 Disrupting the DISC1 Locus on White Matter Integrity. *PLoS ONE 10*(6): e0130900. doi:10.1371/journal.pone.0130900

Williams, H. J., Norton, N., Dwyer, S., Moskvina, V., Nikolov, I., Carroll, L., … O'Donovan, M. C. (2011). Fine mapping of ZNF804A and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder. *Molecular Psychiatry*, *16*(4), 429–41. doi:10.1038/mp.2010.36

Williams, N. M., Green, E. K., MacGregor, S., Dwyer, S., Norton, N., Williams, H., … Craddock, N. (2006). Variation at the DAOA/G30 locus influences susceptibility to major mood episodes but not psychosis in schizophrenia and bipolar disorder. *Archives of General Psychiatry*, *63*(0003-990X (Print)), 366–373. doi:10.1001/archpsyc.63.4.366

Wockner, L. F., Noble, E. P., Lawford, B. R., Young, R. M., Morris, C. P., Whitehall, V. L. J., & Voisey, J. (2014). Genome-wide DNA methylation analysis of human brain tissue from schizophrenia patients. *Translational Psychiatry*, *4*(October 2013), e339. doi:10.1038/tp.2013.111

Wong, C. C. Y., Meaburn, E. L., Ronald, A., Price, T. S., Jeffries, A. R., Schalkwyk, L. C., … Mill, J. (2014). Methylomic analysis of monozygotic twins discordant

for autism spectrum disorder and related behavioural traits. *Molecular Psychiatry*, *19*(4), 495–503. doi:10.1038/mp.2013.41

Wu, H., D'Alessio, A. C., Ito, S., Wang, Z., Cui, K., Zhao, K., … Zhang, Y. (2011). Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes & Development*, *25*(7), 679–684. doi:10.1101/gad.2036011

Xu, B., Roos, J. L., Dexheimer, P., Boone, B., Plummer, B., Levy, S., … Karayiorgou, M. (2011). Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nature Genetics*, *43*(9), 864–8. doi:10.1038/ng.902

Xu, B., Roos, J. L., Levy, S., van Rensburg, E. J., Gogos, J. a, & Karayiorgou, M. (2008). Strong association of de novo copy number mutations with sporadic schizophrenia. *Nature Genetics*, *40*(7), 880–885. doi:10.1038/ng.162

Y. Meltzer, H., W. Massey, B., & Horiguchi, M. (2012). Serotonin Receptors as Targets for Drugs Useful to Treat Psychosis and Cognitive Impairment in Schizophrenia. *Current Pharmaceutical Biotechnology*, *13*(8), 1572–1586. doi:10.2174/138920112800784880

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, *88*(1), 76–82. doi:10.1016/j.ajhg.2010.11.011

Young, S. N., Smith, S. E., Pihl, R. O., & Ervin, F. R. (1985). Tryptophan depletion causes a rapid lowering of mood in normal males. *Psychopharmacology*, *87*(2), 173–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3931142

Young-Pearse, T. L., Suth, S., Luth, E. S., Sawa, A., & Selkoe, D. J. (2010). Biochemical and functional interaction of disrupted-in-schizophrenia 1 and amyloid precursor protein regulates neuronal migration during mammalian cortical development. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(31), 10431–40. doi:10.1523/JNEUROSCI.1445-10.2010

Yu, S., Levi, L., Casadesus, G., Kunos, G., & Noy, N. (2014). Fatty acid-binding protein 5 (FABP5) regulates cognitive function both by decreasing anandamide levels and by activating the nuclear receptor peroxisome proliferator-activated receptor β/δ (PPARβ/δ) in the brain. *The Journal of Biological Chemistry*, *289*(18), 12748–58. doi:10.1074/jbc.M114.559062

Zarate, C. A., Singh, J. B., Carlson, P. J., Brutsche, N. E., Ameli, R., Luckenbaugh, D. A., … Manji, H. K. (2006). A randomized trial of an N-methyl-D-aspartate antagonist in treatment-resistant major depression. *Archives of General Psychiatry*, *63*(8), 856–64. doi:10.1001/archpsyc.63.8.856

Zariwala, M., O'Neal, W. K., Noone, P. G., Leigh, M. W., Knowles, M. R., & Ostrowski, L. E. (2004). Investigation of the possible role of a novel gene, DPCD, in primary ciliary dyskinesia. *American Journal of Respiratory Cell and Molecular Biology*, *30*(4), 428–34. doi:10.1165/rcmb.2003-0338RC

Zhang, J., Pho, V., Bonasera, S. J., Holtzman, J., Tang, A. T., Hellmuth, J., … Huang, E. J. (2007). Essential function of HIPK2 in TGFbeta-dependent survival of midbrain dopamine neurons. *Nature Neuroscience*, *10*(1), 77–86. doi:10.1038/nn1816

Zhang, Y., Thompson, R., Zhang, H., & Xu, H. (2011). APP processing in Alzheimer's disease. *Molecular Brain*, *4*, 3. doi:10.1186/1756-6606-4-3

Zhou, X., Chen, Q., Schaukowitch, K., Kelsoe, J. R., & Geyer, M. A. (2010). Insoluble DISC1-Boymaw fusion proteins generated by DISC1 translocation. *Molecular Psychiatry*, *15*(7), 669–72. doi:10.1038/mp.2009.127

Ziff, E. B. (1997). Enlightening the postsynaptic density. *Neuron*, *19*(6), 1163–74. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9427241

Ziller, M. J., Hansen, K. D., Meissner, A., & Aryee, M. J. (2014). Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. *Nature Methods*, *12*(3), 230–2, 1 p following 232. doi:10.1038/nmeth.3152

Zimmerman, M., Ellison, W., Young, D., Chelminski, I., & Dalrymple, K. (2015). How many different ways do patients meet the diagnostic criteria for major depressive disorder? *Comprehensive Psychiatry*, *56*, 29–34. doi:10.1016/j.comppsych.2014.09.007

Zuk, O., Hechter, E., Sunyaev, S. R., & Lander, E. S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(4), 1193–8. doi:10.1073/pnas.1119675109

Appendix I

Supplementary Tables

ProbeID	Chromosome	Gene	Fold-Change	t	Differential Expression *p*-value
ILMN_1767523	3	*IL17RB*	-2.58	-9.89	3.52×10^{-6}
ILMN_2098446	18	*PMAIP1*	-1.46	-7.10	5.27×10^{-5}
ILMN_1675797	7	*EPDR1*	-2.63	-6.94	6.32×10^{-5}
ILMN_2046470	14	*DAAM1**	-1.43	-6.21	0.0001
ILMN_1659270	5	*OTP*	-1.72	-5.73	0.0003
ILMN_1709484	15	*BLM*	1.32	5.26	0.0005
ILMN_1708936	9	*EXOSC3*	1.23	5.22	0.0005
ILMN_1724181	4	*IL15*	-1.25	-5.13	0.0006
ILMN_1789106	1	*IPP*	-1.21	-5.10	0.0006
ILMN_3251423	3	*CHDH*	-1.50	-5.05	0.0007
ILMN_3178258	13	*FABP5L2**	1.34	5.04	0.0007
ILMN_1763129	16	*DCTPP1*	1.21	5.02	0.0007
ILMN_1659365	4	*LOC653071*	-1.47	-4.99	0.0007
ILMN_1733515	2	*LOXL3*	1.53	4.98	0.0007
ILMN_1668092	11	*ESAM*	-1.74	-4.80	0.0009
ILMN_2121068	2	*ADAM17*	-1.22	-4.77	0.0010
ILMN_1729749	4	*HERC5*	-1.63	-4.76	0.0010
ILMN_2089175	X	*SYAP1*	-1.24	-4.73	0.0010
ILMN_1803995	12	*TM7SF3*	-1.32	-4.73	0.0010
ILMN_1788283	16	*COTL1*	1.46	4.70	0.0011
ILMN_1665219	19	*LTBP4*	-1.22	-4.68	0.0011
ILMN_1725387	6	*TMEM200A*	1.47	4.67	0.0011
ILMN_1680037	16	*FAM65A*	-1.41	-4.66	0.0011
ILMN_1763198	12	*STAT6*	-1.22	-4.57	0.0013
ILMN_1708906	2	*C2orf29*	-1.21	-4.54	0.0014
ILMN_1758731	1	*CYP2J2*	-1.29	-4.51	0.0014
ILMN_1694711	6	*MAD2L1BP*	1.21	4.48	0.0015
ILMN_2289593	11	*FXYD2*	-1.47	-4.45	0.0015
ILMN_1708427	13	*KPNA3*	1.23	4.40	0.0017
ILMN_2364062	17	*THOC4**	1.26	4.38	0.0017
ILMN_2324056	3	*GNL3*	1.25	4.31	0.0019
ILMN_1678086	2	*CCDC74A*	1.33	4.31	0.0019
ILMN_1812721	12	*LOC728014*	1.29	4.30	0.0019
ILMN_2080080	X	*MAP7D2*	-1.36	-4.29	0.0019
ILMN_1761069	11	*UVRAG*	1.30	4.26	0.0020
ILMN_1803194	17	*GALK1*	1.21	4.25	0.0021
ILMN_1776181	11	*BIRC3**	-1.75	-4.24	0.0021
ILMN_1656274	1	*PRPF38A*	1.22	4.22	0.0022

ProbeID	Chromosome	Gene	Fold-Change	t	Differential Expression *p*-value
ILMN_2076463	12	*SLC15A4*	-1.25	-4.19	0.0023
ILMN_1799069	15	*LOC440280*	1.35	4.17	0.0023
ILMN_1702279	20	*KIF3B*	1.21	4.16	0.0024
ILMN_1713636	1	*S100A6*	-1.51	-4.14	0.0024
ILMN_1767320	22	*MAPK1**	1.21	4.13	0.0025
ILMN_2198878	4	*INPP4B*	-1.54	-4.12	0.0025
ILMN_1759991	22	*MGC3731*	1.28	4.12	0.0025
ILMN_1706959	17	*TIMM22*	1.25	4.04	0.0028
ILMN_1787251	14	*DAAM1**	-1.36	-4.02	0.0029
ILMN_2341254	13	*STARD13*	-1.30	-4.02	0.0030
ILMN_1717886	8	*PKHD1L1*	-1.21	-4.00	0.0030
ILMN_1787265	10	*ZNF503*	-1.37	-3.98	0.0031
ILMN_2372413	22	*BID*	1.25	3.98	0.0031
ILMN_3249216	13	*PDX1*	-1.34	-3.97	0.0032
ILMN_1733256	19	*PSMD8*	1.26	3.95	0.0033
ILMN_2235851	2	*LINCR*	-1.53	-3.92	0.0034
ILMN_2146761	8	*FABP5*	1.35	3.89	0.0036
ILMN_1741406	1	*HOOK1*	-1.50	-3.87	0.0037
ILMN_1759326	12	*P2RX7*	-1.30	-3.85	0.0038
ILMN_2301083	20	*UBE2C**	1.21	3.84	0.0039
ILMN_1714730	20	*UBE2C**	1.23	3.84	0.0039
ILMN_1748844	6	*CNKSR3*	-1.39	-3.82	0.0040
ILMN_1663042	20	*SDC4*	-1.22	-3.73	0.0046
ILMN_1696485	5	*HNRNPAB*	1.24	3.73	0.0046
ILMN_2073307	1	*IL10*	-1.32	-3.72	0.0046
ILMN_1682368	7	*LRWD1*	1.34	3.72	0.0046
ILMN_1745420	9	*PHF19*	1.21	3.71	0.0047
ILMN_3273229	16	*LOC100129781*	-1.22	-3.70	0.0048
ILMN_1808783	9	*STRBP*	1.29	3.68	0.0050
ILMN_2329429	7	*GIMAP6*	1.71	3.67	0.0050
ILMN_1729180	15	*GATM*	-1.24	-3.67	0.0050
ILMN_2231928	21	*MX2*	-1.37	-3.67	0.0051
ILMN_2089329	13	*SPRY2*	-1.37	-3.66	0.0051
ILMN_1765578	3	*TIPARP*	1.21	3.64	0.0053
ILMN_1756942	2	*SP3*	1.25	3.63	0.0054
ILMN_1720829	19	*ZFP36*	-1.34	-3.62	0.0054
ILMN_1732296	1	*ID3*	1.51	3.55	0.0061
ILMN_1721457	22	*RANBP1**	1.43	3.55	0.0061

ProbeID	Chromosome	Gene	Fold-Change	t	Differential Expression *p*-value
ILMN_1794825	17	*ALDH3A2*	1.23	3.54	0.0062
ILMN_3236036	15	*LOC283663*	-1.25	-3.54	0.0062
ILMN_1740466	6	*FAM46A*	-1.84	-3.49	0.0067
ILMN_1687440	7	*HIPK2*	1.25	3.48	0.0068
ILMN_2192072	11	*MMP7**	-2.08	-3.46	0.0070
ILMN_1800225	3	*PPARG*	-1.28	-3.46	0.0071
ILMN_1719696	3	*PLD1*	-1.66	-3.45	0.0071
ILMN_3212373	12	*LOC727803*	1.34	3.44	0.0073
ILMN_2357361	11	*THYN1*	1.20	3.43	0.0073
ILMN_3251383	2	*CCDC74B*	1.24	3.43	0.0074
ILMN_1764577	22	*MFNG*	1.22	3.42	0.0074
ILMN_2060115	11	*SORL1*	-1.29	-3.42	0.0074
ILMN_2357855	9	*NTRK2*	-1.24	-3.42	0.0075
ILMN_2173451	19	*GPI*	1.20	3.41	0.0076
ILMN_1685124	12	*TCTN1*	-1.21	-3.41	0.0076
ILMN_2347068	19	*MKNK2**	-1.41	-3.39	0.0078
ILMN_2157277	7	*AKR1D1*	-1.25	-3.37	0.0081
ILMN_1724497	2	*ABI2*	1.22	3.36	0.0083
ILMN_1697363	20	*C20orf27*	1.25	3.34	0.0085
ILMN_1772359	1	*LAPTM5*	-1.21	-3.33	0.0086
ILMN_1811433	8	*RPL8*	1.31	3.33	0.0087
ILMN_1814465	17	*UBE2G1*	1.30	3.32	0.0088
ILMN_1658464	13	*GTF3A*	1.23	3.30	0.0090
ILMN_1871233	14		1.20	3.29	0.0092
ILMN_1738491	9	*SNX30*	1.23	3.28	0.0094
ILMN_1752668	6	*DAAM2*	-1.26	-3.27	0.0094
ILMN_3276990	3	*LOC389141*	1.26	3.27	0.0095
ILMN_1685403	11	*MMP7**	-1.78	-3.26	0.0096
ILMN_1797332	11	*NARS2*	1.20	3.26	0.0097
ILMN_2070044	4	*PPM1K*	-1.32	-3.25	0.0098
ILMN_2320330	2	*MAL**	-1.45	-3.25	0.0099
ILMN_1748923	9	*SMC2*	1.26	3.24	0.0099
ILMN_1728710	19	*ZNF816A*	-1.21	-3.23	0.0101
ILMN_1757723	6	*C6orf106*	1.20	3.23	0.0101
ILMN_1737517	3	*RPL29*	1.30	3.23	0.0101
ILMN_1660462	1	*MCOLN2*	-1.37	-3.22	0.0103
ILMN_1815527	11	*HBD*	1.22	3.22	0.0104
ILMN_1709233	1	*F5*	-1.26	-3.21	0.0105

ProbeID	Chromosome	Gene	Fold-Change	t	Differential Expression *p*-value
ILMN_2072296	9	*CKS2**	1.32	3.19	0.0108
ILMN_1719695	3	*NFKBIZ*	-1.38	-3.17	0.0111
ILMN_1724422	1	*SELL*	-1.94	-3.17	0.0112
ILMN_1811933	17	*SHMT1*	1.39	3.15	0.0115
ILMN_1700182	19	*LOC400721*	-1.21	-3.15	0.0116
ILMN_1695157	17	*CA4*	-1.32	-3.13	0.0120
ILMN_1794707	11	*ATHL1*	-1.25	-3.11	0.0124
ILMN_1654060	19	*MKNK2**	-1.47	-3.11	0.0124
ILMN_1781536	15	*FAH*	1.42	3.09	0.0126
ILMN_1689908	12	*ANKRD13A*	1.31	3.09	0.0127
ILMN_2115862	2	*ESPNL*	-1.22	-3.09	0.0127
ILMN_1720513	18	*SETBP1*	1.37	3.08	0.0130
ILMN_3235709	17	*HNF1B*	-1.35	-3.07	0.0130
ILMN_1726064	6	*PAK1IP1*	1.21	3.07	0.0131
ILMN_2405684	11	*BIRC3**	-1.53	-3.07	0.0131
ILMN_2413015	2	*EMX1*	-1.24	-3.07	0.0132
ILMN_1721580	1	*TBX15*	-1.39	-3.06	0.0134
ILMN_2100209	17	*CCL4L1*	-1.73	-3.06	0.0134
ILMN_3266944	7	*LOC100129599*	-1.27	-3.04	0.0138
ILMN_1700967	3	*C3orf59*	-1.88	-3.04	0.0139
ILMN_2094360	15	*NR2F2*	-1.26	-3.01	0.0145
ILMN_1775734	9	*SH2D3C*	-1.26	-3.00	0.0146
ILMN_1808584	1	*FAM36A*	1.22	3.00	0.0147
ILMN_2109156	22	*RANBP1**	1.26	2.98	0.0152
ILMN_1781752	16	*CLEC16A*	-1.25	-2.97	0.0154
ILMN_2140799	10	*FAM24B*	-1.29	-2.97	0.0154
ILMN_1721713	4	*EXOSC9*	1.21	2.97	0.0155
ILMN_1763638	1	*BCAR3*	1.37	2.97	0.0156
ILMN_1662846	3	*GPR160*	1.32	2.95	0.0159
ILMN_1771987	19	*SLC44A2*	-1.51	-2.95	0.0161
ILMN_3300797	5	*LOC729090*	-1.35	-2.95	0.0161
ILMN_1760509	3	*EOMES*	-1.45	-2.93	0.0165
ILMN_1756326	9	*CKS2**	1.28	2.93	0.0166
ILMN_1672389	1	*CRYZ*	-1.49	-2.92	0.0167
ILMN_1773963	19	*GNA15*	-1.48	-2.91	0.0170
ILMN_1724139	11	*TMEM123*	1.20	2.91	0.0171
ILMN_2338785	5	*RPS14*	-1.21	-2.91	0.0171
ILMN_2318011	14	*PSMA3*	-1.22	-2.89	0.0175

ProbeID	Chromosome	Gene	Fold-Change	t	Differential Expression *p*-value
ILMN_2235283	22	*MAPK1**	1.26	2.89	0.0176
ILMN_1807825	6	*LY86*	-1.31	-2.89	0.0177
ILMN_1803254	14	*KIAA2010*	1.21	2.89	0.0178
ILMN_2307455	X	*UBE2A*	1.21	2.88	0.0178
ILMN_2220187	2	*GFPT1*	-1.20	-2.87	0.0183
ILMN_2344373	16	*MVP**	-1.22	-2.86	0.0185
ILMN_1742382	1	*RIMS3*	-1.57	-2.86	0.0185
ILMN_1654685	5	*MCTP1*	-1.50	-2.84	0.0190
ILMN_1661717	13	*TFDP1*	1.27	2.82	0.0198
ILMN_2048700	8	*ATAD2*	1.27	2.82	0.0198
ILMN_1795811	14	*ASB2*	1.39	2.81	0.0199
ILMN_1698213	X	*RBM3*	1.21	2.81	0.0201
ILMN_1787628	20	*NOP56**	1.24	2.81	0.0202
ILMN_1713156	X	*MSL3L1*	-1.21	-2.80	0.0203
ILMN_2120982	13	*ALG11*	-1.21	-2.80	0.0205
ILMN_2367113	4	*CASP6*	1.23	2.79	0.0206
ILMN_2124769	1	*YBX1*	1.26	2.79	0.0209
ILMN_1807372	22	*ADORA2A*	-1.31	-2.77	0.0214
ILMN_1842582	18		1.36	2.76	0.0218
ILMN_1750518	17	*THOC4**	1.24	2.76	0.0218
ILMN_1680738	5	*C5orf13*	-1.30	-2.76	0.0219
ILMN_2407824	1	*ATP1B1**	-1.56	-2.75	0.0220
ILMN_1658494	13	*C13orf15*	1.59	2.75	0.0220
ILMN_2373791	8	*ENPP2**	-1.44	-2.75	0.0223
ILMN_1800390	10	*ZNF511*	1.21	2.74	0.0226
ILMN_3266606	13	*FABP5L2**	1.27	2.74	0.0227
ILMN_1775708	12	*SLC2A3*	-1.27	-2.73	0.0228
ILMN_1776678	7	*GIMAP7*	1.51	2.73	0.0231
ILMN_1776653	X	*SCML1*	-1.22	-2.72	0.0232
ILMN_1750409	X	*RAB9A*	-1.28	-2.72	0.0235
ILMN_1658071	1	*ATP1B1**	-1.49	-2.71	0.0238
ILMN_1740170	22	*CHCHD10*	1.27	2.70	0.0239
ILMN_2374692	10	*WAC*	1.21	2.70	0.0240
ILMN_1761820	1	*EDARADD*	-1.25	-2.69	0.0245
ILMN_1731070	16	*ORC6L*	1.24	2.69	0.0246
ILMN_1752947	17	*C17orf79*	1.23	2.69	0.0247
ILMN_1685580	3	*CBLB*	1.31	2.68	0.0247
ILMN_1796417	7	*ASNS*	-1.24	-2.68	0.0249

ProbeID	Chromosome	Gene	Fold-Change	t	Differential Expression *p*-value
ILMN_2058251	10	*VIM**	-1.69	-2.68	0.0249
ILMN_1707763	7	*ST7*	1.25	2.67	0.0255
ILMN_1745112	9	*FAM102A*	1.30	2.66	0.0257
ILMN_3235593	19	*ZNF841*	-1.23	-2.65	0.0260
ILMN_1738578	3	*FILIP1L*	-1.21	-2.64	0.0267
ILMN_1663035	17	*SREBF1*	-1.21	-2.64	0.0267
ILMN_1760779	1	*ENSA*	1.22	2.64	0.0268
ILMN_2044471	6	*NCR3*	1.28	2.64	0.0268
ILMN_1734596	14	*TC2N*	-1.33	-2.63	0.0268
ILMN_1781373	2	*IFIH1*	-1.30	-2.63	0.0271
ILMN_1803277	16	*MVP**	-1.33	-2.62	0.0273
ILMN_1694780	15	*GCHFR*	1.20	2.62	0.0274
ILMN_1781472	1	*CDC42BPA*	-1.24	-2.62	0.0275
ILMN_1653466	1	*HES4*	-1.36	-2.62	0.0276
ILMN_1811426	12	*TMTC1*	-1.26	-2.61	0.0277
ILMN_1778337	17	*TCF2*	-1.27	-2.61	0.0278
ILMN_1730612	20	*DBNDD2*	-1.42	-2.60	0.0283
ILMN_1733937	17	*MMD*	-1.32	-2.60	0.0285
ILMN_1653871	7	*NAMPT*	-1.23	-2.60	0.0285
ILMN_2066858	13	*TNFSF13B*	-1.22	-2.60	0.0286
ILMN_2352090	17	*GPRC5C*	-1.27	-2.60	0.0286
ILMN_1712312	15	*RAB11A*	-1.22	-2.59	0.0287
ILMN_3248443	8	*SNHG6*	-1.22	-2.59	0.0290
ILMN_1810962	6	*PTPRK*	-1.32	-2.58	0.0292
ILMN_3251467	3	*LRRC58*	-1.27	-2.58	0.0292
ILMN_1739576	11	*CYB5R2*	1.49	2.58	0.0296
ILMN_1808238	15	*RBPMS2*	-1.71	-2.57	0.0298
ILMN_1783247	10	*C10orf11*	-1.21	-2.56	0.0302
ILMN_2047885	5	*PCDHB9*	-1.24	-2.56	0.0303
ILMN_1684929	3	*TOPBP1*	1.22	2.56	0.0304
ILMN_1712985	17	*C17orf58*	1.27	2.55	0.0307
ILMN_1782538	10	*VIM**	-1.60	-2.55	0.0309
ILMN_1700515	17	*C17orf58*	1.22	2.55	0.0311
ILMN_1739805	16	*NDE1*	-1.23	-2.55	0.0311
ILMN_1692295	7	*MYO1G*	-1.23	-2.54	0.0312
ILMN_1770922	3	*TMEM45A*	-1.29	-2.53	0.0320
ILMN_1763036	1	*CLCN6*	-1.22	-2.52	0.0322
ILMN_1798181	11	*IRF7*	-1.31	-2.52	0.0324

ProbeID	Chromosome	Gene	Fold-Change	t	Differential Expression *p*-value
ILMN_3265365	9	*CEP78*	1.22	2.51	0.0327
ILMN_1730291	1	*ATP1B1**	-1.63	-2.51	0.0328
ILMN_1679438	4	*MLF1IP*	1.30	2.51	0.0331
ILMN_2112402	22	*PHF5A*	1.34	2.51	0.0331
ILMN_1709634	5	*CMBL*	-1.47	-2.50	0.0334
ILMN_1789999	1	*SLC30A7*	-1.29	-2.50	0.0337
ILMN_1661599	10	*DDIT4*	-1.42	-2.50	0.0338
ILMN_1718629	21	*NRIP1*	1.30	2.49	0.0342
ILMN_1780799	8	*ENPP2**	-1.41	-2.48	0.0344
ILMN_1714364	8	*PTK2*	1.46	2.48	0.0344
ILMN_3243644	12	*LOC100132564*	1.45	2.48	0.0344
ILMN_1771051	3	*RPL29*	1.24	2.48	0.0347
ILMN_3235326		*LOC388796*	1.51	2.48	0.0347
ILMN_1806754	9	*GLDC*	1.50	2.48	0.0347
ILMN_1746408	19	*MIDN*	-1.21	-2.47	0.0350
ILMN_1796497	6	*PIP3-E*	-1.28	-2.47	0.0351
ILMN_1707551	17	*AFMID*	1.20	2.47	0.0353
ILMN_1882512	5		-1.24	-2.46	0.0358
ILMN_2362549	10	*ZWINT*	1.27	2.46	0.0361
ILMN_1745256	5	*CXXC5*	-1.28	-2.45	0.0364
ILMN_2341815	3	*TFG*	1.56	2.45	0.0365
ILMN_1761138	9	*C9orf142*	1.20	2.45	0.0367
ILMN_1796063	11	*TRIM44*	-1.23	-2.44	0.0370
ILMN_1780533	14	*RNASE6*	1.23	2.43	0.0374
ILMN_2044832	20	*NOP56**	1.21	2.42	0.0380
ILMN_2201533	17	*C17orf61*	-1.25	-2.42	0.0386
ILMN_2138801	3	*TP73L*	1.69	2.41	0.0386
ILMN_1665877	2	*RNF149*	-1.29	-2.41	0.0388
ILMN_1771084	16	*ACSM3*	-1.28	-2.41	0.0391
ILMN_2235785	17	*KCNH6*	-1.38	-2.41	0.0392
ILMN_2327860	2	*MAL**	-1.79	-2.40	0.0394
ILMN_2073604	X	*EBP*	1.25	2.40	0.0396
ILMN_3247139	17	*C17orf96*	-1.20	-2.40	0.0396
ILMN_3277365	1	*LOC100133233*	-1.22	-2.39	0.0400
ILMN_2308582	22	*CYB5R3*	-1.24	-2.39	0.0401
ILMN_2367743	17	*TUBG1*	1.20	2.38	0.0405
ILMN_1678766	6	*DYNLT1*	-1.34	-2.38	0.0408
ILMN_1774604	2	*PNKD*	-1.27	-2.38	0.0411

ProbeID	Chromosome	Gene	Fold-Change	t	Differential Expression *p*-value
ILMN_3210741	15	*FABP5L9*	1.26	2.38	0.0412
ILMN_1785198	9	*POLE3*	1.23	2.37	0.0413
ILMN_1801118	16	*C16orf33*	1.33	2.37	0.0413
ILMN_1702177	6	*GLO1*	1.21	2.37	0.0416
ILMN_1812666	13	*DNAJC15*	1.24	2.36	0.0421
ILMN_1652371	7	*KIAA1324L*	1.34	2.36	0.0424
ILMN_1723043	19	*NAPSB*	1.42	2.36	0.0426
ILMN_1712678	15	*RPS27L*	-1.26	-2.35	0.0426
ILMN_1656011	1	*RGS1*	-1.89	-2.35	0.0428
ILMN_2338038	1	*AK3L1*	1.45	2.35	0.0431
ILMN_2373515	14	*HSP90AA1*	1.21	2.35	0.0432
ILMN_3274929	2	*LOC653924*	1.40	2.34	0.0434
ILMN_2143795	2	*MGC4677*	1.30	2.34	0.0435
ILMN_1778991	9	*NFIB*	-1.32	-2.34	0.0438
ILMN_1676629	2	*INSIG2*	-1.21	-2.33	0.0445
ILMN_1742577	10	*GTPBP4*	-1.48	-2.33	0.0447
ILMN_1699695	6	*TNFRSF21*	-1.49	-2.32	0.0453
ILMN_1784602	6	*CDKN1A*	-1.38	-2.32	0.0453
ILMN_1723274	2	*GPR55*	1.30	2.32	0.0455
ILMN_1714197	20	*ACSS2*	1.32	2.31	0.0457
ILMN_1829845	13		-1.35	-2.31	0.0458
ILMN_1794782	21	*ABCG1*	-1.34	-2.31	0.0460
ILMN_1787509	20	*PRIC285*	-1.30	-2.30	0.0466
ILMN_1798886	16	*NUDT21*	1.20	2.30	0.0467
ILMN_1704369	12	*LIMA1*	1.23	2.30	0.0467
ILMN_2396672	10	*ABLIM1*	1.40	2.30	0.0468
ILMN_3219764	16	*LOC441506*	1.21	2.29	0.0474
ILMN_2155719	1	*NBPF10*	1.33	2.29	0.0478
ILMN_1682038	11	*SNORA25*	1.21	2.28	0.0480
ILMN_1695420	9	*CLTA*	1.20	2.28	0.0482
ILMN_1711073	2	*LOC653489*	-1.24	-2.28	0.0484
ILMN_1707503	1	*C1orf144*	1.27	2.28	0.0484
ILMN_1813295	12	*LMO3*	-1.42	-2.28	0.0485
ILMN_1742981	12	*TUBA1A*	1.21	2.27	0.0487
ILMN_1752802	5	*CLPTM1L*	1.21	2.27	0.0490
ILMN_2364700	1	*ENSA*	1.20	2.27	0.0494
ILMN_1769810	3	*ARL6IP5*	-1.22	-2.26	0.0496

Table A1: Differentially expressed genes in t(1;11) carrier LCL RNA.

Summary of genes meeting an absolute expression fold-change ≥ 1.2 in t(1;11) carriers with an associated p-*value ≤ 0.05 for differential expression. In order of column appearance are the HT12 array probe's associated Illumina ID, the probe's target chromosome, gene name, fold-change in t(1;11) carriers, differential expression t-statistic, and unadjusted* p-*value for differential expression between t(1;11) carriers and non-carriers. Genes flagged with a red asterisk (*) are those with multiple non-overlapping probes meeting the criteria for differential expression in t(1;11) carriers. Shown are the top 10 differentially expressed genes ranked by* p-*value.*

GO Term	Description	Class	Enrichment	*p*-value	*q*-value
GO:0043005	neuron projection	Component	1.73	2.29×10^{-9}	3.72×10^{-6}
GO:0042995	cell projection	Component	1.44	3.63×10^{-7}	0.0003
GO:0030425	dendrite	Component	1.43	2.10×10^{-6}	0.0009
GO:0009653	anatomical structure morphogenesis	Process	1.5	6.49×10^{-8}	0.0009
GO:0097458	neuron part	Component	1.96	1.93×10^{-6}	0.001
GO:0009987	cellular process	Process	1.33	5.98×10^{-7}	0.002
GO:0048562	embryonic organ morphogenesis	Process	4.36	4.93×10^{-7}	0.0022
GO:0048856	anatomical structure development	Process	1.07	4.14×10^{-7}	0.0028
GO:0048731	system development	Process	1.72	1.04×10^{-6}	0.0028
GO:0030424	axon	Component	1.12	3.05×10^{-5}	0.0083
GO:0044444	cytoplasmic part	Component	4.26	2.75×10^{-5}	0.0089
GO:0032502	developmental process	Process	1.3	6.12×10^{-6}	0.0093
GO:0010646	regulation of cell communication	Process	1.25	4.91×10^{-6}	0.0096
GO:0044763	single-organism cellular process	Process	1.1	5.69×10^{-6}	0.0097
GO:0050793	regulation of developmental process	Process	1.21	4.54×10^{-6}	0.01
GO:0048812	neuron projection morphogenesis	Process	2.26	8.99×10^{-6}	0.012
GO:0009887	organ morphogenesis	Process	2.15	8.55×10^{-6}	0.017
GO:0051960	regulation of nervous system development	Process	1.25	1.29×10^{-5}	0.014
GO:0023051	regulation of signaling	Process	1.54	1.52×10^{-5}	0.014
GO:0051094	positive regulation of developmental process	Process	1.41	1.43×10^{-5}	0.014
GO:0048583	regulation of response to stimulus	Process	1.25	1.24×10^{-5}	0.014
GO:0050807	regulation of synapse organisation	Process	6.99	2.31×10^{-5}	0.02
GO:0044424	intracellular part	Component	1.07	8.74×10^{-5}	0.02
GO:0070588	calcium ion transmembrane transport	Process	1.21	4.33×10^{-5}	0.021
GO:0051962	positive regulation of nervous system development	Process	1.93	2.97×10^{-5}	0.021
GO:0044708	single-organism behavior	Process	1.68	4.57×10^{-5}	0.022
GO:2000026	regulation of multicellular organismal development	Process	1.38	3.19×10^{-5}	0.022
GO:0045595	regulation of cell differentiation	Process	1.56	3.52×10^{-5}	0.022
GO:2001030	negative regulation of cellular glucuronidation	Process	1.34	4.33×10^{-5}	0.022
GO:0048858	cell projection morphogenesis	Process	1.98	2.90×10^{-5}	0.022
GO:0009888	tissue development	Process	1.63	3.75×10^{-5}	0.022
GO:0050767	regulation of neurogenesis	Process	25.57	3.46×10^{-5}	0.023

GO Term	Description	Class	Enrichment	*p*-value	*q*-value
GO:0031344	regulation of cell projection organisation	Process	25.57	3.98 X 10^{-5}	0.023
GO:1904224	negative regulation of glucuronosyltransferase activity	Process	25.57	4.33 X 10^{-5}	0.023
GO:0044767	single-organism developmental process	Process	2.35	2.87 X 10^{-5}	0.023
GO:1904223	regulation of glucuronosyltransferase activity	Process	1.79	4.33 X 10^{-5}	0.024
GO:0021852	pyramidal neuron migration	Process	17,640	5.67 X 10^{-5}	0.026
GO:0051239	regulation of multicellular organismal process	Process	1.29	6.25 X 10^{-5}	0.028
GO:0032990	cell part morphogenesis	Process	1.84	6.46 X 10^{-5}	0.028
GO:0003197	endocardial cushion development	Process	1.45	7.97 X 10^{-5}	0.028
GO:2001029	regulation of cellular glucuronidation	Process	2.01	7.94 X 10^{-5}	0.029
GO:0045597	positive regulation of cell differentiation	Process	1.75	7.08 X 10^{-5}	0.029
GO:0052697	xenobiotic glucuronidation	Process	46.22	7.94 X 10^{-5}	0.029
GO:2000467	positive regulation of glycogen (starch) synthase activity	Process	22.72	7.90 X 10^{-5}	0.03
GO:0046578	regulation of Ras protein signal transduction	Process	22.72	7.55 X 10^{-5}	0.03
GO:0048598	embryonic morphogenesis	Process	21.83	7.86 X 10^{-5}	0.031
GO:0051240	positive regulation of multicellular organismal process	Process	1.35	9.01 X 10^{-5}	0.031
GO:0048513	organ development	Process	1.43	0.0001	0.034
GO:0034702	ion channel complex	Component	1.81	0.0002	0.035
GO:0072300	positive regulation of metanephric glomerulus development	Process	213.82	0.0001	0.037
GO:0005783	endoplasmic reticulum	Component	5.76	0.0003	0.041
GO:0005768	endosome	Component	1.64	0.0003	0.042
GO:0014069	postsynaptic density	Component	35.18	0.0003	0.043
GO:0048167	regulation of synaptic plasticity	Process	1.79	0.0001	0.044
GO:0044699	single-organism process	Process	1.07	0.0001	0.044
GO:0044216	other organism cell	Component	5.29	0.0003	0.044
GO:0044297	cell body	Component	1.43	0.0002	0.045
GO:0050804	modulation of synaptic transmission	Process	2.5	0.0001	0.045
GO:0043524	negative regulation of neuron apoptotic process	Process	2.63	0.0002	0.046
GO:0033387	putrescine biosynthetic process from ornithine	Process	22.88	0.0002	0.047
GO:0031225	anchored component of membrane	Component	1.38	0.0004	0.048

GO Term	Description	Class	Enrichment	*p*-value	*q*-value
GO:0044463	cell projection part	Component	4.34	0.0004	0.049

Table A2: Summary of GO terms enriched amongst differentially methylated genes in t(1;11) carriers.

For each GO term, table summarises the GO identifier, the GO description, the GO class, enrichment, the enrichment p-values, and enrichment FDR q-values for genes showing the most differential methylation in t(1,11) carriers compared to non-carriers. Enrichment is defined as (b/n) / (B/N), where b = the number of genes at the top of the p-value ranked list that is associated with a given GO term, n = the number of genes at the top of the p-*value-ranked gene list, B = the total number of genes associated with a given GO term, and N = the total number of genes.*

Gene(s)	Feature(s)	Region	Probes	DMR *p*-value
TNXB	Body	Chr6:32063516-32065113	51	2.46×10^{-13}
NA	IGR	Chr3:196704439-196707088	5	7.89×10^{-10}
C11orf75	TSS1500	Chr11:93277097-93277255	3	1.67×10^{-9}
NA;*PRRT1*	IGR, 3'UTR	Chr6:32115866-32116728	14	2.60×10^{-9}
RNF5P1;AGPAT1	TSS1500, TSS200	Chr6:32145233-32145902	20	6.26×10^{-9}
GABRG1	TSS1500, TSS200, 5'UTR	Chr4:46125801-46126455	7	7.27×10^{-9}
KRTAP5-9	TSS1500, TSS200, 5'UTR	Chr11:71259142-71259846	5	3.17×10^{-8}
CYP2E1	Body	Chr10:135343047-135343426	3	5.93×10^{-8}
XRRA1	TSS200	Chr11:74660246-74660274	4	6.07×10^{-8}
RHOD	3'UTR	Chr11:66839183-66839543	3	8.99×10^{-8}
TUBGCP5	TSS1500, TSS200	Chr15:22833108-22833236	3	1.58×10^{-7}
TRAK1	TSS1500, Body, 1stExon	Chr3:42201087-42202130	6	1.64×10^{-7}
PSMB8	3'UTR, Body	Chr6:32807898-32810304	22	3.32×10^{-7}
NA	IGR	Chr1:209526301-209528771	3	3.46×10^{-7}
NA	IGR	Chr6:159359910-159360236	3	1.48×10^{-6}
RTKN	TSS1500	Chr2:74669347-74669415	4	1.94×10^{-6}
NA; *KLHL30*	IGR, TSS1500, TSS200	Chr2:239043942-239047460	9	2.67×10^{-6}
PCSK6	Body	Chr15:101991031-101992367	5	2.8×10^{-6}
MIR663	TSS1500	Chr20:26190290-26190418	3	3.66×10^{-6}
ZNF385D	Body, 1stExon	Chr3:21791767-21792729	4	3.82×10^{-6}
TNF	1stExon	Chr6:31543557-31543767	6	4.05×10^{-6}
BNIP3	Body	Chr10:133793398-133793556	3	9.21×10^{-6}
TMEM131	Body	Chr2:98377310-98378420	3	9.58×10^{-6}
LTA	TSS200, 5'UTR, 1stExon	Chr6:31540040-31540705	11	9.59×10^{-6}
NA	IGR	Chr17:41380728-41383399	5	9.98×10^{-6}
MIXL1	TSS1500	Chr1:226411005-226411085	4	1.06×10^{-5}
C17orf98	Body, 1stExon	Chr17:36997274-36997566	3	1.3×10^{-5}
FAM83A	Body	Chr8:124217614-124217906	3	1.42×10^{-5}
SP140	TSS200, 5'UTR, Body	Chr2:231090159-231091121	5	1.56×10^{-5}
MEST	5'UTR	Chr7:130131819-130131937	8	3×10^{-5}
MOV10L1	TSS200, Body	Chr22:50584735-50585710	7	3.02×10^{-5}
NA	IGR	Chr5:101117938-101120956	5	3.37×10^{-5}

Gene(s)	Feature(s)	Region	Probes	DMR *p*-value
NA	IGR	Chr2:54934730-54938012	4	3.42×10^{-5}
LOC399815	TSS200	Chr10:124638974-124638992	3	3.43×10^{-5}
STX6	TSS200	Chr1:180992101-180992126	3	3.44×10^{-5}
NA	IGR	Chr7:156888064-156890971	5	3.65×10^{-5}
NA	IGR	Chr14:54814555-54817362	4	5.68×10^{-5}
NA	IGR	Chr7:73001195-73003686	3	6.35×10^{-5}
NA	IGR	Chr4:77341749-77342075	3	6.66×10^{-5}
TDH	Body	Chr8:11203736-11204163	4	7.48×10^{-5}
SMYD2	Body	Chr1:214476582-214477761	3	9.03×10^{-5}
IGSF9B	Body	Chr11:133820356-133821733	4	9.05×10^{-5}
TAP1	Body	Chr6:32819956-32820360	7	9.06×10^{-5}
NA	IGR	Chr1:39281563-39282050	4	9.11×10^{-5}
NA	IGR	Chr10:134898718-134899190	3	9.24×10^{-5}
NA	IGR	Chr11:123939376-123941869	3	0.0001
NA	IGR	Chr1:147799913-147804738	9	0.0001
HLA-DPA1	Body, 3'UTR	Chr6:33033283-33037541	16	0.0001
MIR548N	Body	Chr2:179387372-179388545	3	0.0001
LIME1	Body	Chr20:62369366-62369583	3	0.0001
NA	IGR	Chr7:1208305-1210932	3	0.0001
VSTM1	TSS200, Body, 1stExon	Chr19:54566357-54567319	3	0.0001
NA	IGR	Chr5:61027075-61030044	6	0.0001
ZNF677	TSS200	Chr19:53758279-53758315	3	0.0001
NA	IGR	Chr11:8290652-8290978	3	0.0002
ACP5	TSS1500	Chr19:11689791-11689944	5	0.0002
PSKH2	TSS1500, TSS200	Chr8:87082001-87082129	5	0.0002
ATP8A2	Body	Chr13:26586108-26586400	3	0.0002
VAX2	TSS1500	Chr2:71126945-71127001	3	0.0002
LRP1B	TSS200	Chr2:142889366-142889402	3	0.0003
C2orf54	TSS1500	Chr2:241836300-241836518	3	0.0003
KCNIP1; *KCNMB1*	TSS200, Body	Chr5:169815897-169816859	4	0.0003
NA	IGR	Chr8:832849-833211	4	0.0003
C6orf48	5'UTR	Chr6:31802972-31803220	9	0.0004

Gene(s)	Feature(s)	Region	Probes	DMR p-value
NHEDC1	TSS200, 1stExon	Chr4:103940811-103940897	3	0.0004
HEATR2	Body, 3'UTR	Chr7:824351-826059	6	0.0004
NA	IGR	Chr10:115860117-115860473	3	0.0004
OR2L13	1stExon	Chr1:248100542-248100657	4	0.0005
RPTOR	Body	Chr17:78809217-78810354	3	0.0005
ADARB2	Body	Chr10:1452489-1453586	3	0.0005
KLHL29	Body	Chr2:23885619-23886838	3	0.0005
NA	IGR	Chr6:33870717-33874881	16	0.0006
C2CD4D	Body	Chr1:151810808-151810978	4	0.0006
KATNAL2	5'UTR	Chr18:44561939-44562173	5	0.0006
NA	IGR	Chr1:46632418-46632744	4	0.0007
NA	IGR	Chr1:117316713-117319422	6	0.0007
RHOU	Body	Chr1:228872106-228872264	3	0.0007
ADAMTS2	Body	Chr5:178594504-178594662	3	0.0007
NFE2L1	TSS1500	Chr17:46124661-46124789	3	0.0007
TBCD	Body	Chr17:80759197-80760159	3	0.0007
RP1	Body	Chr8:55533808-55534034	4	0.0007
DPYSL3	Body	Chr5:146832152-146832444	5	0.0008
DZIP1L	TSS1500	Chr3:137834643-137834771	3	0.0009
NA	IGR	Chr12:29302148-29302314	3	0.0009
NA	IGR	Chr12:114916775-114919883	6	0.0010
MSX1	Body	Chr4:4864284-4864576	6	0.0010
NA	IGR	Chr6:32848233-32852184	3	0.0011
MAPT	1stExon	Chr17:43971863-43971971	4	0.0011
ERMAP	5'UTR, Body	Chr1:43295986-43296972	4	0.0012
TRIM10	3'UTR, Body	Chr6:30118914-30123074	24	0.0012
NA; *RPRM*	IGR, 3'UTR	Chr2:154333734-154334094	5	0.0013
NA	IGR	Chr1:160950485-160953893	5	0.0013
NA	IGR	Chr10:133528532-133530912	4	0.0016
HOXC11	3'UTR	Chr12:54369487-54373005	6	0.0017
TSPAN19;*LRRIQ1*	TSS200, 5'UTR	Chr12:85430071-85430601	8	0.0018
TRIM15	3'UTR	Chr6:30140125-30140546	8	0.0019

Gene(s)	Feature(s)	Region	Probes	DMR p-value
PRICKLE1	Body, 3'UTR	Chr12:42850912-42854914	7	0.0019
STAT3	Body	Chr17:40489088-40490266	5	0.0021
NA; *COX19*	IGR, 3'UTR	Chr7:1001886-1005509	9	0.0022
NA	IGR	Chr8:53325115-53327858	8	0.0023
S100A13; *S100A14*	TSS1500, 3'UTR	Chr1:153589312-153593314	5	0.0036
MICB	Body	Chr6:31466095-31466266	4	0.0039
NA	IGR	Chr6:30094973-30095632	25	0.0040
SLC6A12	TSS200, 5'UTR	Chr12:322249-322914	9	0.0044
TNXB	Body	Chr6:32054900-32055225	8	0.0063
NA	IGR	Chr5:3102528-3106063	4	0.0064
TMCO3	Body	Chr13:114192835-114194543	11	0.0064
ASCL2	TSS1500	Chr11:2293053-2293275	9	0.0064
VARS2	Body	Chr6:30883736-30885444	14	0.0074
B3GALT4	1stExon	Chr6:33245445-33245846	20	0.0090
NOM1	Body	Chr7:156755232-156756616	6	0.0092
ALPL	5'UTR	Chr1:21877057-21877844	6	0.0105
HLA-DPA1	Body	Chr6:33040054-33041250	8	0.0121
EHMT2	Body	Chr6:31858853-31861146	20	0.0124
NA	IGR	Chr12:131700432-131703053	5	0.0166
KLC2	3'UTR	Chr11:66034892-66035327	14	0.0169
CNTN4	Body	Chr3:3079716-3080819	4	0.0224
GNL1	3'UTR	Chr6:30507466-30511643	15	0.0289
PRDM16	Body	Chr1:3209038-3210391	5	0.0380
HCP5	TSS1500, TSS200, 1stExon, Body, 3'UTR	Chr6:31429311-31433970	18	0.0392
NA	IGR	Chr5:170762452-170765146	9	0.0408
ARHGAP25	TSS1500, 5'UTR, Body	Chr2:69001334-69002335	10	0.0428
HSPA1L	5'UTR, Body	Chr6:31778918-31782134	19	0.0498

Table A3: Summary of t(1;11)-associated DMRs identified by the probe lasso algorithm.

Summary of the genes (if applicable), genomic features, hg19 genomic coordinates, number of probes and p-*value associated with each DMR. "NA" in the "Gene" column represents intergenic regions (i.e. regions not annotated to a RefSeq gene). Genomic features are coded "IGR" for intergenic regions, "TSS200" and "TSS1500" for probes occurring within 200 and 1500 of a gene's transcription start site, respectively; "5'UTR" and "3'UTR" for probes occurring within a gene's 5' and 3' untranslated region, respectively, "1stExon" for probes occurring within the first exon of a gene, and "Body" for probes occurring within the gene body.*

GO Term	Description	Class	Enrichment	*p*-value	*q*-value
GO:0071556	integral component of lumenal side of endoplasmic reticulum membrane	Component	82.96	3.55×10^{-8}	5.89×10^{-5}
GO:0042611	MHC protein complex	Component	140.38	1.45×10^{-7}	0.0001
GO:0030658	transport vesicle membrane	Component	30.85	5.56×10^{-7}	0.0003
GO:0048518	positive regulation of biological process	Process	1.19	8.43×10^{-8}	0.0012
GO:0012507	ER to Golgi transport vesicle membrane	Component	68.76	3.41×10^{-6}	0.0014
GO:0048522	positive regulation of cellular process	Process	1.2	5.73×10^{-7}	0.0020
GO:0042605	peptide antigen binding	Function	4.97	2.29×10^{-6}	0.0024
GO:0042995	cell projection	Component	1.38	7.47×10^{-6}	0.0025
GO:0051960	regulation of nervous system development	Process	1.6	5.64×10^{-7}	0.0026
GO:0048856	anatomical structure development	Process	1.28	4.57×10^{-7}	0.0032
GO:0005150	interleukin-1, Type I receptor binding	Function	655.11	2.24×10^{-6}	0.0032
GO:0002486	antigen processing and presentation of endogenous peptide antigen via MHC class I via ER pathway, TAP-independent	Process	1,310.22	1.22×10^{-6}	0.0034
GO:0008092	cytoskeletal protein binding	Function	1.51	4.26×10^{-6}	0.0036
GO:0003779	actin binding	Function	1.81	9.08×10^{-7}	0.0039
GO:0004653	polypeptide N-acetylgalactosaminyltransferase activity	Function	9.54	5.54×10^{-6}	0.0039
GO:0046977	TAP binding	Function	184.89	1.86×10^{-6}	0.0040
GO:0030054	cell junction	Component	1.43	1.54×10^{-5}	0.0043
GO:0048869	cellular developmental process	Process	1.28	2.15×10^{-6}	0.0050
GO:0014069	postsynaptic density	Component	2.09	2.68×10^{-5}	0.0064
GO:0008376	acetylgalactosaminyltransferase activity	Function	6.81	1.05×10^{-5}	0.0064
GO:0030534	adult behavior	Process	2.32	3.34×10^{-6}	0.0066
GO:0035315	hair cell differentiation	Process	12.41	4.34×10^{-6}	0.0067
GO:0060284	regulation of cell development	Process	1.52	4.83×10^{-6}	0.0067
GO:0050767	regulation of neurogenesis	Process	1.59	5.39×10^{-6}	0.0068
GO:0042613	MHC class II protein complex	Component	6.68	3.28×10^{-5}	0.0068
GO:0009653	anatomical structure morphogenesis	Process	1.39	4.03×10^{-6}	0.0070
GO:0005856	cytoskeleton	Component	1.51	4.09×10^{-5}	0.0075

GO Term	Description	Class	Enrichment	*p*-value	*q*-value
GO:0032502	developmental process	Process	1.18	6.69×10^{-6}	0.0077
GO:0019885	antigen processing and presentation of endogenous peptide antigen via MHC class I	Process	129.42	7.84×10^{-6}	0.0077
GO:0002484	antigen processing and presentation of endogenous peptide antigen via MHC class I via ER pathway	Process	786.13	8.87×10^{-6}	0.0082
GO:0060333	interferon-gamma-mediated signaling pathway	Process	4.05	7.84×10^{-6}	0.0083
GO:0030662	coated vesicle membrane	Component	32.71	7.63×10^{-5}	0.0091
GO:0030176	integral component of endoplasmic reticulum membrane	Component	20.35	6.04×10^{-5}	0.0091
GO:0042612	MHC class I protein complex	Component	393.07	5.53×10^{-5}	0.0092
GO:0032395	MHC class II receptor activity	Function	8.19	1.74×10^{-5}	0.0093
GO:0002483	antigen processing and presentation of endogenous peptide antigen	Process	117.66	1.09×10^{-5}	0.0094
GO:0030666	endocytic vesicle membrane	Component	24.96	6.86×10^{-5}	0.0095
GO:0031227	intrinsic component of endoplasmic reticulum membrane	Component	19.43	7.50×10^{-5}	0.0096
GO:0042270	protection from natural killer cell mediated cytotoxicity	Process	655.11	1.51×10^{-5}	0.0123
GO:0048708	astrocyte differentiation	Process	5.88	1.72×10^{-5}	0.0132
GO:0019883	antigen processing and presentation of endogenous antigen	Process	99.56	2.05×10^{-5}	0.0149
GO:0044767	single-organism developmental process	Process	1.18	2.30×10^{-5}	0.0152
GO:0048731	system development	Process	1.55	2.29×10^{-5}	0.0158
GO:0002711	positive regulation of T cell mediated immunity	Process	103.64	2.52×10^{-5}	0.0158
GO:0097458	neuron part	Component	1.34	0.0002	0.0182
GO:0043005	neuron projection	Component	1.45	0.0002	0.0183
GO:0009893	positive regulation of metabolic process	Process	1.2	3.85×10^{-5}	0.0231
GO:0031344	regulation of cell projection organisation	Process	1.61	4.16×10^{-5}	0.0239
GO:0002480	antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-independent	Process	436.74	4.40×10^{-5}	0.0243
GO:0045936	negative regulation of phosphate metabolic process	Process	1.57	5.95×10^{-5}	0.0249

GO Term	Description	Class	Enrichment	*p*-value	*q*-value
GO:0010563	negative regulation of phosphorus metabolic process	Process	1.57	5.95×10^{-5}	0.0257
GO:1903403	negative regulation of renal phosphate excretion	Process	17,688.00	5.65×10^{-5}	0.0260
GO:0051130	positive regulation of cellular component organisation	Process	1.4	5.84×10^{-5}	0.0260
GO:1903402	regulation of renal phosphate excretion	Process	17,688.00	5.65×10^{-5}	0.0269
GO:0003421	growth plate cartilage axis specification	Process	17,688.00	5.65×10^{-5}	0.0279
GO:0045953	negative regulation of natural killer cell mediated cytotoxicity	Process	393.07	5.53×10^{-5}	0.0283
GO:0002716	negative regulation of natural killer cell mediated immunity	Process	393.07	5.53×10^{-5}	0.0294
GO:0042491	auditory receptor cell differentiation	Process	13.7	8.05×10^{-5}	0.0318
GO:0032989	cellular component morphogenesis	Process	1.59	7.86×10^{-5}	0.0319
GO:0005883	neurofilament	Component	6.48	0.0003	0.0323
GO:0051961	negative regulation of nervous system development	Process	1.84	8.91×10^{-5}	0.0342
GO:0002709	regulation of T cell mediated immunity	Process	69.09	0.0001	0.0350
GO:0016045	detection of bacterium	Process	302.36	0.0001	0.0353
GO:0001911	negative regulation of leukocyte mediated cytotoxicity	Process	302.36	0.0001	0.0362
GO:0007411	axon guidance	Process	1.53	9.99×10^{-5}	0.0363
GO:0097485	neuron projection guidance	Process	1.53	9.99×10^{-5}	0.0373
GO:0044763	single-organism cellular process	Process	1.09	0.0001	0.0389
GO:0045665	negative regulation of neuron differentiation	Process	3.03	0.0001	0.0392
GO:0007166	cell surface receptor signaling pathway	Process	1.25	0.0001	0.0395
GO:0003382	epithelial cell morphogenesis	Process	3.17	0.0001	0.0396
GO:0000902	cell morphogenesis	Process	1.89	0.0001	0.0399
GO:0031342	negative regulation of cell killing	Process	262.04	0.0001	0.0399
GO:0031225	anchored component of membrane	Component	2.48	0.0004	0.0400
GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	Process	44.82	0.0001	0.0404

GO Term	Description	Class	Enrichment	*p*-value	*q*-value
GO:0098543	detection of other organism	Process	262.04	0.0001	0.0407
GO:0050768	negative regulation of neurogenesis	Process	2.75	0.0001	0.0413
GO:0045664	regulation of neuron differentiation	Process	1.55	0.0001	0.0419
GO:0007010	cytoskeleton organisation	Process	1.47	0.0002	0.0427
GO:0002478	antigen processing and presentation of exogenous peptide antigen	Process	12.99	0.0002	0.0433
GO:0031061	negative regulation of histone methylation	Process	20.69	0.0002	0.0433
GO:0010721	negative regulation of cell development	Process	1.75	0.0002	0.0436
GO:0005737	cytoplasm	Component	1.14	0.0006	0.0465
GO:0019884	antigen processing and presentation of exogenous antigen	Process	19.47	0.0002	0.0465
GO:0015629	actin cytoskeleton	Component	1.79	0.0005	0.0474
GO:0001916	positive regulation of T cell mediated cytotoxicity	Process	231.22	0.0002	0.0485
GO:0048002	antigen processing and presentation of peptide antigen	Process	18.93	0.0002	0.0494
GO:0043197	dendritic spine	Component	2.82	0.0006	0.0496

Table A4: Summary of GO terms enriched for differentially methylated genes in t(1;11) carriers with a psychotic disorder.

For each GO term, table summarises the GO identifier, the GO description, the GO class, enrichment, the enrichment p-values, and enrichment FDR q-values for genes showing the most differential methylation in t(1,11) carriers with psychotic disorders compared to carriers with a non-psychotic disorder. Enrichment is defined as (b/n) / (B/N), where b = the number of genes at the top of the p-value ranked list that is associated with a given GO term, n = the number of genes at the top of the p*-value-ranked gene list, B = the total number of genes associated with a given GO term, and N = the total number of genes.*

Gene(s)	Feature	Region	Probe Count	DMR *p*-value
RPTOR	Body	Chr17:78865119-78867431	8	1.55×10^{-21}
NA, *MIR886*	IGR, TSS200, Body	Chr5:135415531-135416414	13	1.18×10^{-19}
PTPRN2	Body	Chr7:158045532-158046806	6	2.13×10^{-15}
TNXB	Body	Chr6:32064430-32064738	13	1.49×10^{-13}
NA	IGR	Chr6:29721548-29725160	4	1.6×10^{-13}
NA	IGR	Chr13:113295067-113297572	4	4.17×10^{-13}
SLC38A4	TSS200, 5'UTR, 1stExon	Chr12:47219385-47219867	6	4.95×10^{-12}
HCG27	TSS1500	Chr6:31164851-31165031	8	3.24×10^{-11}
NA	IGR	Chr6:29520536-29521310	15	4.37×10^{-11}
NA	IGR	Chr6:25882428-25882752	3	7.78×10^{-11}
NA	IGR	Chr3:196704504-196707023	5	1.4×10^{-9}
C1orf173	TSS200, 1stExon, Body	Chr1:75138516-75139587	9	2.29×10^{-9}
TAGLN	TSS1500, TSS200, 5'UTR	Chr11:117069672-117070287	6	2.33×10^{-9}
TNNT1	TSS200, 5'UTR, 1stExon	Chr19:55660273-55660755	5	2.82×10^{-9}
C11orf21, *TSPAN32*	Body, 1stExon	Chr11:2321881-2324344	29	3.68×10^{-9}
PKNOX2	5'UTR	Chr11:125105815-125106376	3	6.05×10^{-9}
TRIM31	Body	Chr6:30071309-30071756	13	1.18×10^{-8}
NA	IGR	Chr12:10094777-10097919	6	1.91×10^{-8}
NA	IGR	Chr6:31275481-31276043	11	3.37×10^{-8}
NA	IGR	Chr2:730996-731756	6	4.95×10^{-8}
SLC23A1	Body	Chr5:138714244-138714417	3	5.16×10^{-8}
NA	IGR	Chr2:54934795-54937947	4	5.33×10^{-8}
JARID2	Body	Chr6:15504475-15505793	6	1.37×10^{-7}
KIAA1875	Body	Chr8:145162890-145163180	4	1.43×10^{-7}
NA	IGR	Chr6:28956177-28956469	13	1.8×10^{-7}
NA	IGR	Chr4:187421843-187422201	5	1.85×10^{-7}
NA	IGR	Chr1:117316713-117319357	6	2.96×10^{-7}
NA	IGR	Chr1:28573735-28573959	4	3.87×10^{-7}
ANGPT2	TSS1500, TSS200, 5'UTR, 1stExon, Body	Chr8:6418918-6422358	11	3.87×10^{-7}
CRISP2	TSS200, 5'UTR	Chr6:49680937-49681474	7	3.93×10^{-7}
NA	IGR	Chr6:170337585-170338059	3	3.95×10^{-7}
NA	IGR	Chr12:68880247-68882552	3	3.97×10^{-7}
NA	IGR	Chr13:110385027-110387392	3	5.14×10^{-7}
ATP6V1C1	TSS1500	Chr8:104032756-104032939	3	5.24×10^{-7}
B3GNT3	TSS1500	Chr19:17905578-17905634	3	5.49×10^{-7}
VWA5B2	Body	Chr3:183958890-183959196	7	6.92×10^{-7}
NA	IGR	Chr1:149168848-149171477	5	7.77 x 10-7
ZNF528	TSS1500,TSS200	Chr19:52900819-52900977	6	9.28 x 10-7
DDR1	TSS1500	Chr6:30850988-30851701	9	9.63 x 10-7
CRB2	Body	Chr9:126135156-126135444	3	1.26 x 10-6

Gene(s)	Feature	Region	Probe Count	DMR *p*-value
NA	IGR	Chr10:125032645-125035943	10	1.66 x 10-6
VPS16	Body	Chr20:2844211-2845161	6	1.75 x 10-6
PAX8	Body	Chr2:113992699-113993196	6	1.76 x 10-6
NA	IGR	Chr17:41438248-41438484	3	1.87 x 10-6
NA	IGR	Chr1:35586487-35586811	3	1.87 x 10-6
TOP1MT	Body	Chr8:144403361-144403517	3	1.88 x 10-6
NA	IGR	Chr6:19804549-19804946	3	1.91 x 10-6
BLCAP	5'UTR	Chr20:36148593-36148721	7	2.29 x 10-6
NA	IGR	Chr6:164028178-164030681	3	2.29 x 10-6
C6orf138	TSS200	Chr6:48036584-48036634	6	2.34 x 10-6
TBC1D16	Body	Chr17:77924521-77924809	3	2.39 x 10-6
PRTN3	TSS1500, TSS200, 1stExon, Body	Chr19:840230-841934	6	3.09 x 10-6
TNNT3	Body	Chr11:1948954-1949191	4	3.2 x 10-6
NA	IGR	Chr4:6689577-6692218	4	3.21 x 10-6
NA	IGR	Chr7:27137097-27140727	7	3.59 x 10-6
ITGAE	1stExon	Chr17:3704392-3704573	4	3.95 x 10-6
NA	IGR	Chr7:73155464-73159136	4	3.99 x 10-6
WFIKKN2	TSS200, 1stExon, Body	Chr17:48912518-48912700	5	4.04 x 10-6
HLCS	TSS1500, TSS200	Chr21:38362714-38362782	4	4.42 x 10-6
NA	IGR	Chr7:155149556-155152552	5	5.02 x 10-6
NA	IGR	Chr13:22613924-22616459	5	5.3 x 10-6
SFTA2	3'UTR, Body	Chr6:30898957-30900097	7	5.32 x 10-6
NA	IGR	Chr8:67454513-67454749	4	5.32 x 10-6
LRRC17	TSS200, 5'UTR	Chr7:102553128-102553698	5	6.92 x 10-6
NA	IGR	Chr16:53407596-53407970	4	7.22 x 10-6
STK19	TSS1500	Chr6:31939217-31939385	4	7.64 x 10-6
C10orf11	Body	Chr10:77871474-77871819	3	7.67 x 10-6
NA	IGR	Chr16:3210063-3210387	3	8.02 x 10-6
HLA-DQB1	Body	Chr6:32632883-32633301	10	8.11 x 10-6
NA	IGR	Chr11:41480233-41482703	5	9.04 x 10-6
NA	IGR	Chr3:42977807-42978035	5	9.76 x 10-6
DLGAP2	Body	Chr8:1616237-1616594	3	9.82 x 10-6
TPM4	Body	Chr19:16186762-16186951	3	1.01 x 10-5
NA	IGR	Chr2:121775189-121777729	3	1.02 x 10-5
RNU5E	Body	Chr5:80690127-80690303	3	1.21 x 10-5
NA	IGR	Chr3:195576915-195579384	6	1.31 x 10-5
NA	IGR	Chr7:20818495-20819006	6	1.31 x 10-5
HLA-DQB2	Body	Chr6:32728986-32729418	8	1.65 x 10-5
NA	IGR	Chr1:19110816-19111140	3	1.8 x 10-5
SLC36A3	Body	Chr5:150677714-150678610	3	1.91 x 10-5
NMUR1	Body	Chr2:232393052-232393340	3	1.92 x 10-5

Gene(s)	Feature	Region	Probe Count	DMR *p*-value
SGK1	TSS1500, Body	Chr6:134497031-134497391	6	2.08 x 10-5
ZBTB47	Body	Chr3:42700441-42700822	3	2.2 x 10-5
NA	IGR	Chr5:149866606-149870608	5	2.24 x 10-5
ENOSF1	TSS200	Chr18:712722-712746	3	2.54 x 10-5
C22orf9	Body	Chr22:45607980-45609161	10	2.61 x 10-5
C6orf47	1stExon	Chr6:31627549-31627782	6	2.64 x 10-5
NA	IGR	Chr13:106062013-106064297	3	2.8 x 10-5
NA	IGR	Chr6:28602381-28602829	12	2.9 x 10-5
NA	IGR	Chr6:29768148-29768332	6	2.94 x 10-5
NA	IGR	Chr13:29327854-29330351	3	3.13 x 10-5
NA	IGR	Chr15:98446452-98448821	3	3.29 x 10-5
ARHGEF10	5'UTR	Chr8:1788928-1789704	6	3.41 x 10-5
VSTM1	TSS200, 1stExon, Body	Chr19:54566390-54567286	3	3.54 x 10-5
WBSCR17	TSS1500	Chr7:70597037-70597093	3	3.62 x 10-5
SPRED3	Body, 3'UTR	Chr19:38886514-38886811	4	3.65 x 10-5
ASB16	TSS200, 5'UTR	Chr17:42247858-42248340	4	3.66 x 10-5
AOAH	Body	Chr7:36700468-36701464	4	3.98 x 10-5
SLC9A3R2	Body	Chr16:2081837-2084020	5	4.41 x 10-5
NA	IGR	Chr4:187985370-187985534	3	4.43 x 10-5
ECM1	1stExon, Body	Chr1:150480408-150481304	3	4.61 x 10-5
TMTC4	Body	Chr13:101314806-101315848	3	4.68 x 10-5
NA	IGR	Chr4:174429164-174429536	5	5.19 x 10-5
RARA	TSS200, 5'UTR	Chr17:38465269-38465751	6	5.58 x 10-5
NA	IGR	Chr16:53542860-53545446	4	5.63 x 10-5
RASA3	Body	Chr13:114875026-114875314	3	5.64 x 10-5
NA	IGR	Chr2:66656160-66659666	5	6.68 x 10-5
ZMAT2	TSS1500, TSS200, 1stExon, Body	Chr5:140079646-140080693	11	6.85 x 10-5
HTR2A	TSS1500	Chr13:47472142-47472468	8	6.9 x 10-5
NA	IGR	Chr3:87138478-87138802	3	7.34 x 10-5
NA	IGR	Chr12:7781126-7781450	3	7.36 x 10-5
NA	IGR	Chr2:3485268-3488537	10	7.71 x 10-5
NA	IGR	Chr8:48675485-48675933	4	7.77 x 10-5
NA	IGR	Chr10:134331152-134333567	3	7.93 x 10-5
NA	IGR	Chr1:9222878-9225501	3	8.11 x 10-5
PTPRN2	Body	Chr7:157955906-157956305	4	8.12 x 10-5
NA	IGR	Chr7:148662823-148665257	3	8.14 x 10-5
RADIL	Body	Chr7:4885073-4886091	3	8.2 x 10-5
TET1	5'UTR	Chr10:70321759-70322074	4	8.69 x 10-5
NA	IGR	Chr19:14444105-14444479	4	8.7 x 10-5
KIAA1614	Body	Chr1:180882511-180882825	3	8.79 x 10-5
ZMYND15	Body, 3'UTR	Chr17:4648132-4649524	7	9.21 x 10-5

Gene(s)	Feature	Region	Probe Count	DMR *p*-value
NA	IGR	Chr14:97058702-97059026	4	9.55 x 10-5
NA	IGR	Chr2:89158063-89160390	3	0.0001
LOC84931	TSS1500, TSS200, Body	Chr2:121223086-121224357	8	0.0001
GRM2	TSS1500, TSS200	Chr3:51740847-51740903	3	0.0001
TNF	1stExon	Chr6:31543559-31543734	5	0.0001
NA	IGR	Chr12:19935352-19937788	3	0.0001
NA	IGR	Chr10:7516090-7518958	4	0.0002
AHNAK	Body	Chr11:62272524-62273866	3	0.0002
CRY2	TSS1500	Chr11:45868386-45868512	4	0.0002
EPB49	1stExon	Chr8:21916725-21916930	3	0.0002
NA	IGR	Chr13:114907148-114911086	3	0.0002
NA	IGR	Chr7:155831867-155834117	4	0.0002
NA	IGR	Chr1:247802541-247802865	4	0.0002
LTB4R2, LTB4R	5'UTR, Body	Chr14:24780604-24780769	3	0.0002
TRIM69	TSS1500	Chr15:45027975-45028269	3	0.0002
NA	IGR	Chr5:4229421-4231997	6	0.0002
PRSSL1	TSS200, 5'UTR, Body	Chr19:693985-695689	6	0.0002
C19orf77	TSS200, 5'UTR	Chr19:3480393-3480623	3	0.0002
LDHC	TSS1500	Chr11:18433392-18433672	3	0.0002
ITFG3	5'UTR	Chr16:302924-303433	3	0.0002
KIF13A	Body	Chr6:17984556-17986260	3	0.0003
NA	IGR	Chr6:106441279-106441668	5	0.0003
TRAPPC9	Body	Chr8:140945334-140946291	3	0.0003
NA	IGR	Chr6:31037468-31039838	6	0.0003
NA	IGR	Chr11:69259291-69262797	7	0.0003
NA	IGR	Chr2:8421111-8424247	5	0.0003
NA	IGR	Chr19:13949415-13952921	3	0.0004
FAM83A	Body	Chr8:124217616-124217904	3	0.0004
NA	IGR	Chr13:114060687-114066289	13	0.0004
GAS7	TSS1500, Body	Chr17:9939556-9940675	4	0.0004
MAD1L1	Body	Chr7:2060038-2060194	4	0.0005
SFRP2	TSS1500	Chr4:154711557-154711738	5	0.0005
C19orf57	5'UTR	Chr19:14016712-14016810	4	0.0005
NA	IGR	Chr10:135040088-135040412	3	0.0006
RASA3	Body	Chr13:114809779-114811646	4	0.0006
NA	IGR	Chr15:89156690-89159521	5	0.0006
TXNDC11	Body	Chr16:11835324-11835612	3	0.0006
LTBP1	TSS200, Body	Chr2:33358750-33359646	5	0.0006
DIO2	1stExon	Chr14:80677579-80677737	3	0.0006
NA	IGR	Chr12:89748564-89749094	8	0.0007
NA	IGR	Chr8:53325249-53328204	8	0.0007

Gene(s)	Feature	Region	Probe Count	DMR *p*-value
NA	IGR	Chr5:71851339-71853987	4	0.0008
BEND7	Body	Chr10:13481398-13482392	3	0.0008
CD8A	TSS200	Chr2:87018941-87018975	3	0.0008
NA	IGR	Chr10:96989418-96992630	7	0.0009
NA	IGR	Chr10:112290169-112290493	3	0.0010
NA	IGR	Chr6:170555977-170559101	7	0.0010
NOP56	1stExon, Body	Chr20:2633246-2633402	4	0.0010
NA	IGR	Chr13:113273357-113275660	3	0.0010
NA	IGR	Chr6:27724063-27726533	3	0.0011
EYA4	TSS1500	Chr6:133562239-133562295	5	0.0011
NA	IGR	Chr6:170552720-170554970	7	0.0011
TNXB	Body	Chr6:32014062-32016684	45	0.0011
NA	IGR	Chr4:572257-575763	3	0.0012
POU2AF1	TSS1500, TSS200	Chr11:111250323-111250539	4	0.0013
NA	IGR	Chr10:81967374-81967828	3	0.0013
NA	IGR	Chr15:97320021-97323629	11	0.0015
NA	IGR	Chr15:31514986-31517606	9	0.0017
ZG16B	TSS1500, TSS200, 1stExon, Body	Chr16:2879911-2880879	6	0.0020
SLC29A1	TSS1500, TSS200	Chr6:44186968-44187094	5	0.0022
PTPRN2	Body	Chr7:157811395-157812366	4	0.0022
C17orf90, *CCDC137*	TSS1500, TSS200, 1stExon	Chr17:79633717-79633844	6	0.0022
SCHIP1	Body	Chr3:159557104-159558245	4	0.0022
NA	IGR	Chr16:85599110-85601764	4	0.0026
GPX5	Body	Chr6:28498931-28500273	6	0.0026
VARS2	Body	Chr6:30882497-30882852	4	0.0026
WDFY4	Body	Chr10:50142889-50143931	5	0.0028
PSORS1C1, *CDSN*	TSS1500, TSS200, 5'UTR, 1stExon, Body, 3'UTR	Chr6:31080938-31085323	16	0.0038
RPH3AL	Body	Chr17:153647-155523	5	0.0041
CNKSR1	TSS1500, TSS200, Body	Chr1:26503543-26504467	10	0.0044
PPAP2C	TSS1500	Chr19:292208-292334	4	0.0046
UBE2I, NA	IGR, Body, 3'UTR	Chr16:1372915-1376938	10	0.0048
NA	IGR	Chr14:76733243-76735730	6	0.0048
C6orf136	Body	Chr6:30615599-30615887	7	0.0051
BTNL2	TSS200, Body	Chr6:32374306-32375202	4	0.0063
NA	IGR	Chr6:30419411-30423402	21	0.0063
RPL13AP5, *SNORD34*, *SNORD35A*	TSS200, Body	Chr19:49993013-49994725	12	0.0064
EXPH5	TSS1500, Body	Chr11:108408514-108409429	6	0.0065
TNFRSF9	TSS1500, TSS200, 5'UTR, Body	Chr1:7998726-8001902	6	0.0073
PPT2	Body	Chr6:32122609-32123110	10	0.0074

Gene(s)	Feature	Region	Probe Count	DMR *p*-value
SLC22A18AS, *SLC22A18*	TSS200, 5'UTR, Body	Chr11:2919203-2921289	21	0.0074
NUP188	Body	Chr9:131710373-131710529	4	0.0084
HLA-DPB2	Body	Chr6:33095685-33096760	6	0.0087
PURG, *WRN*	TSS1500, TSS200	Chr8:30890445-30890661	8	0.0098
HLA-DPB2	Body	Chr6:33091164-33092578	16	0.0101
KLHL29	Body	Chr2:23839834-23841003	7	0.0110
WIPI2	3'UTR	Chr7:5270698-5275408	14	0.0112
NA	IGR	Chr11:1364021-1366533	8	0.0112
P2RX7	TSS200, 5'UTR, 1stExon	Chr12:121570436-121570993	4	0.0112
NA, *MC1R*	IGR, TSS1500	Chr16:89980515-89984172	8	0.0143
TNXB	Body	Chr6:32048325-32050029	23	0.0147
DIP2C	Body	Chr10:670554-672258	7	0.0155
SLAMF8	TSS200, 5'UTR	Chr1:159796376-159796899	4	0.0160
PHYHIP	Body	Chr8:22084825-22088001	7	0.0160
MCF2L	Body	Chr13:113689103-113690807	6	0.0160
FXYD1	TSS1500, TSS200, 5'UTR	Chr19:35628518-35631893	14	0.0162
NA	IGR	Chr6:170751560-170755066	6	0.0197
KRT222	TSS200, 5'UTR, 1stExon, Body	Chr17:38821149-38821631	7	0.0197
NA	IGR	Chr6:31408342-31410882	13	0.0219
NA	IGR	Chr10:33293233-33295902	5	0.0228
RPTOR	Body	Chr17:78853114-78855125	4	0.0255
NA	IGR	Chr6:29895013-29895349	7	0.0266
TBCD	Body	Chr17:80832541-80834245	8	0.0277
NA	IGR	Chr12:115131136-115135488	43	0.0298
OSBPL5, NA	IGR, TSS1500	Chr11:3187454-3192429	24	0.0304
NA	IGR	Chr11:133445060-133447540	4	0.0330
SNCA	5'UTR	Chr4:90757337-90757567	5	0.0350
COL11A2	3'UTR, Body	Chr6:33131041-33133848	41	0.0354
PCGF3	5'UTR	Chr4:717423-722458	15	0.0365
TNXB	Body	Chr6:32037729-32038989	21	0.0391
PROCR	TSS1500, SS200, 5'UTR, 1stExon	Chr20:33758333-33761509	10	0.0446
NA	IGR	Chr6:31461666-31461903	6	0.0465
NXN	Body	Chr17:800064-801197	8	0.0478
TRIM31	Body, 5'UTR	Chr6:30079782-30080943	10	0.0486
TRIM15	Body	Chr6:30139901-30140074	5	0.0496

Table A5: Psychosis-associated DMRs identified by the probe lasso algorithm.

Table summarises the genes (if applicable), genomic features, hg19 genomic coordinates, number of probes and p-value associated with each DMR. "NA" in the "Gene" column represents intergenic regions (i.e. regions not annotated to a RefSeq gene). Genomic features are coded "IGR" for intergenic regions, "TSS200" and "TSS1500" for probes occurring within 200 and 1500 of a gene's transcription start site, respectively; "5'UTR" and "3'UTR" for probes occurring within a gene's 5' and 3' untranslated region, respectively, "1stExon" for probes occurring within the first exon of a gene, and "Body" for probes occurring within the gene body.

GO Term	Description	Class	Enrichment	P-value	FDR q-value
GO:0044459	Plasma membrane part	Component	433/2385	2.09×10^{-15}	3.57×10^{-12}
GO:0009887	Organ morphogenesis	Process	114/424	3.32×10^{-15}	4.72×10^{-11}
GO:0048856	Anatomical structure development	Process	516/2969	3.81×10^{-14}	2.70×10^{-10}
GO:0032502	Developmental process	Process	654/3896	3.86×10^{-13}	1.82×10^{-9}
GO:0005886	Plasma membrane	Component	596/3766	2.49×10^{-11}	2.12×10^{-8}
GO:0048731	System development	Process	141/629	7.86×10^{-12}	2.79×10^{-8}
GO:0009653	Anatomical structure morphogenesis	Process	245/1229	1.61×10^{-11}	4.58×10^{-8}
GO:0032501	Multicellular organismal process	Process	518/3067	1.37×10^{-10}	3.24×10^{-7}
GO:0044767	Single-organism developmental process	Process	567/3416	2.31×10^{-10}	4.69×10^{-7}
GO:0048562	Embryonic organ morphogenesis	Process	41/115	3.37×10^{-10}	5.99×10^{-7}
GO:0050804	Modulation of synaptic transmission	Process	68/269	5.94×10^{-10}	9.36×10^{-7}
GO:0043235	Receptor complex	Component	78/306	1.65×10^{-9}	9.38×10^{-7}
GO:0044707	Single-multicellular organism process	Process	354/2319	1.72×10^{-9}	2.44×10^{-6}
GO:0031226	Intrinsic component of plasma membrane	Component	240/1334	7.32×10^{-9}	3.13×10^{-6}
GO:0098797	Plasma membrane protein complex	Component	99/485	9.34×10^{-9}	3.19×10^{-6}
GO:0048869	Cellular developmental process	Process	398/2379	3.48×10^{-9}	4.49×10^{-6}
GO:0097458	Neuron part	Component	226/1215	2.50×10^{-8}	7.11×10^{-6}
GO:0044425	Membrane part	Component	920/6105	3.33×10^{-8}	8.12×10^{-6}
GO:0044456	Synapse part	Component	120/554	5.58×10^{-8}	1.19×10^{-5}
GO:0005887	Integral component of plasma membrane	Component	223/1278	6.75×10^{-8}	1.28×10^{-5}
GO:0051239	Regulation of multicellular organismal process	Process	414/2453	1.46×10^{-8}	1.73×10^{-5}
GO:0098590	Plasma membrane region	Component	161/877	1.11×10^{-7}	1.89×10^{-5}
GO:0009952	Anterior/posterior pattern specification	Process	43/133	2.62×10^{-8}	2.86×10^{-5}
GO:0044763	Single-organism cellular process	Process	1232/8467	5.81×10^{-8}	5.90×10^{-5}
GO:0034702	Ion channel complex	Component	59/260	4.09×10^{-7}	6.35×10^{-5}
GO:0043565	Sequence-specific DNA binding	Function	187/978	2.09×10^{-8}	9.02×10^{-5}
GO:0051094	Positive regulation of developmental process	Process	185/994	1.04×10^{-7}	9.86×10^{-5}
GO:0008066	Glutamate receptor activity	Function	15/26	4.66×10^{-8}	1.01E-04
GO:0030054	Cell junction	Component	197/1085	1.05 x 10-6	1.49E-04
GO:0043005	Neuron projection	Component	120/703	1.25 x 10-6	1.64E-04
GO:0045202	Synapse	Component	58/231	1.50 x 10-6	1.83E-04
GO:0098802	Plasma membrane receptor complex	Component	25/161	1.71 x 10-6	1.94E-04

GO Term	Description	Class	Enrichment	P-value	FDR q-value
GO:0007389	Pattern specification process	Process	82/357	2.49 x 10^{-7}	2.21E-04
GO:0048704	Embryonic skeletal system morphogenesis	Process	28/78	3.23 x 10^{-7}	2.55E-04
GO:0048705	Skeletal system morphogenesis	Process	34/106	3.14 x 10^{-7}	2.62E-04
GO:0098878	Neurotransmitter receptor complex	Component	17/41	2.65 x 10-6	2.66E-04
GO:0008328	Ionotropic glutamate receptor complex	Component	17/41	2.65 x 10-6	2.83E-04
GO:0030154	Cell differentiation	Process	274/1605	4.70 x 10^{-7}	3.51E-04
GO:0051240	Positive regulation of multicellular organismal process	Process	231/1325	5.35 x 10^{-7}	3.80E-04
GO:0000977	RNA polymerase II regulatory region sequence-specific DNA binding	Function	115/556	4.00 x 10^{-7}	4.33E-04
GO:0048167	Regulation of synaptic plasticity	Process	36/135	6.44 x 10^{-7}	4.35E-04
GO:0001012	RNA polymerase II regulatory region DNA binding	Function	115/559	5.04 x 10^{-7}	4.36E-04
GO:0015267	Channel activity	Function	84/435	7.12 x 10^{-7}	4.40E-04
GO:0045597	Positive regulation of cell differentiation	Process	154/784	7.25 x 10^{-7}	4.68E-04
GO:0044699	Single-organism process	Process	1341/9882	7.79 x 10^{-7}	4.81E-04
GO:0005216	Ion channel activity	Function	80/395	7.00 x 10^{-7}	5.05E-04
GO:0022803	Passive transmembrane transporter activity	Function	85/436	3.68 x 10^{-7}	5.31E-04
GO:0050803	Regulation of synapse structure or activity	Process	36/137	9.02 x 10^{-7}	5.33E-04
GO:0003002	Regionalisation	Process	57/221	1.03 x 10-6	5.83E-04
GO:0023052	Signaling	Process	124/608	1.16 x 10-6	6.31E-04
GO:0044700	Single organism signaling	Process	123/605	1.55 x 10-6	6.65E-04
GO:2000026	Regulation of multicellular organismal development	Process	266/1584	1.60 x 10-6	6.70E-04
GO:0007268	Chemical synaptic transmission	Process	68/283	1.53 x 10-6	6.81E-04
GO:0048598	Embryonic morphogenesis	Process	85/384	1.30 x 10-6	6.82E-04
GO:0098916	Anterograde trans-synaptic signaling	Process	68/283	1.53 x 10-6	7.03E-04
GO:0022838	Substrate-specific channel activity	Function	81/407	1.34 x 10-6	7.22E-04
GO:0099537	Trans-synaptic signaling	Process	68/283	1.53 x 10-6	7.26E-04
GO:0097060	Synaptic membrane	Component	54/245	8.68 x 10-6	7.41E-04
GO:0099536	Synaptic signaling	Process	68/283	1.53 x 10-6	7.51E-04
GO:0031224	Intrinsic component of membrane	Component	722/4796	7.97 x 10-6	7.56E-04
GO:0045211	Postsynaptic membrane	Component	32/190	8.51 x 10-6	7.64E-04

GO Term	Description	Class	Enrichment	P-value	FDR q-value
GO:0003008	System process	Process	228/1269	1.52 x 10-6	7.69E-04
GO:0015291	Secondary active transmembrane transporter activity	Function	55/220	2.89 x 10-6	0.0011
GO:0022891	Substrate-specific transmembrane transporter activity	Function	165/872	2.43 x 10-6	0.0012
GO:0010646	Regulation of cell communication	Process	432/2788	2.91 x 10-6	0.0012
GO:0002009	Morphogenesis of an epithelium	Process	61/270	3.01 x 10-6	0.0012
GO:0022836	Gated channel activity	Function	64/304	2.77 x 10-6	0.0012
GO:0048732	Gland development	Process	51/235	3.22 x 10-6	0.0012
GO:0000976	Transcription regulatory region sequence-specific DNA binding	Function	123/631	3.63 x 10-6	0.0013
GO:0022610	Biological adhesion	Process	177/953	3.75 x 10-6	0.0014
GO:0048468	Cell development	Process	106/554	4.01 x 10-6	0.0015
GO:0098796	Membrane protein complex	Component	154/968	1.81 x 10-5	0.0015
GO:1902495	Transmembrane transporter complex	Component	59/292	2.01 x 10-5	0.0015
GO:1990351	Transporter complex	Component	59/298	1.94 x 10-5	0.0015
GO:0007155	Cell adhesion	Process	176/949	4.50 x 10-6	0.0016
GO:1990837	Sequence-specific double-stranded DNA binding	Function	127/660	4.97 x 10-6	0.0017
GO:0001228	Transcriptional activator activity, RNA polymerase II transcription regulatory region sequence-specific binding	Function	69/306	5.61 x 10-6	0.0017
GO:0016324	Apical plasma membrane	Component	51/272	3.03 x 10-5	0.0022
GO:0007215	Glutamate receptor signaling pathway	Process	16/39	6.65 x 10-6	0.0023
GO:0023051	Regulation of signaling	Process	435/2830	6.64 x 10-6	0.0023
GO:0001501	Skeletal system development	Process	41/155	7.14 x 10-6	0.0024
GO:0050877	Neurological system process	Process	154/815	7.92 x 10-6	0.0026
GO:0034703	Cation channel complex	Component	42/164	4.37 x 10-5	0.0030
GO:0015293	Symporter activity	Function	35/137	1.05 x 10-5	0.0030
GO:0042995	Cell projection	Component	209/1281	4.78 x 10-5	0.0031
GO:0035113	Embryonic appendage morphogenesis	Process	26/81	1.03 x 10-5	0.0032
GO:0030326	Embryonic limb morphogenesis	Process	26/81	1.03 x 10-5	0.0033
GO:0004970	Ionotropic glutamate receptor activity	Function	10/18	1.32 x 10-5	0.0036
GO:0015075	Ion transmembrane transporter activity	Function	140/754	1.67 x 10-5	0.0038
GO:0022857	Transmembrane transporter activity	Function	172/943	1.50 x 10-5	0.0038
GO:0005516	Calmodulin binding	Function	37/171	1.60 x 10-5	0.0038

GO Term	Description	Class	Enrichment	P-value	FDR q-value
GO:0097485	Neuron projection guidance	Process	45/188	1.27 x 10-5	0.0038
GO:0098609	Cell-cell adhesion	Process	112/560	1.36 x 10-5	0.0040
GO:0014069	Postsynaptic density	Component	20/132	6.71 x 10-5	0.0041
GO:0021615	Glossopharyngeal nerve morphogenesis	Process	4/4	1.46 x 10-5	0.0042
GO:0005230	Extracellular ligand-gated ion channel activity	Function	21/69	1.96 x 10-5	0.0042
GO:0099572	Postsynaptic specialisation	Component	20/132	6.71 x 10-5	0.0043
GO:0007154	Cell communication	Process	126/746	1.52 x 10-5	0.0043
GO:0048513	Animal organ development	Process	202/1137	1.77 x 10-5	0.0049
GO:0005234	Extracellular-glutamate-gated ion channel activity	Function	10/19	2.42 x 10-5	0.0050
GO:0000975	Regulatory region DNA binding	Function	142/780	2.57 x 10-5	0.0050
GO:0031012	Extracellular matrix	Component	72/353	8.64 x 10-5	0.0051
GO:0044449	Contractile fiber part	Component	42/190	9.03 x 10-5	0.0051
GO:0008092	Cytoskeletal protein binding	Function	138/779	2.96 x 10-5	0.0053
GO:0001067	Regulatory region nucleic acid binding	Function	142/781	2.86 x 10-5	0.0054
GO:0032879	Regulation of localisation	Process	369/2284	2.08 x 10-5	0.0057
GO:0048729	Tissue morphogenesis	Process	69/336	2.16 x 10-5	0.0057
GO:0007411	Axon guidance	Process	44/186	2.22 x 10-5	0.0057
GO:0007187	G-protein coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger	Process	33/162	2.14 x 10-5	0.0057
GO:0003690	Double-stranded DNA binding	Function	128/730	3.45 x 10-5	0.0060
GO:0044212	Transcription regulatory region DNA binding	Function	141/777	3.62 x 10-5	0.0060
GO:0007267	Cell-cell signaling	Process	87/545	2.67 x 10-5	0.0068
GO:0015276	Ligand-gated ion channel activity	Function	19/133	5.34 x 10-5	0.0070
GO:0007188	Adenylate cyclase-modulating G-protein coupled receptor signaling pathway	Process	29/136	2.84 x 10-5	0.0071
GO:0035108	Limb morphogenesis	Process	29/95	2.94 x 10-5	0.0071
GO:0035107	Appendage morphogenesis	Process	29/95	2.94 x 10-5	0.0072
GO:0022834	Ligand-gated channel activity	Function	19/133	5.34 x 10-5	0.0072
GO:0022835	Transmitter-gated channel activity	Function	11/24	4.91 x 10-5	0.0073
GO:0046873	Metal ion transmembrane transporter activity	Function	75/397	5.11 x 10-5	0.0074
GO:0004972	NMDA glutamate receptor activity	Function	5/8	5.28 x 10-5	0.0074

GO Term	Description	Class	Enrichment	P-value	FDR q-value
GO:0022824	Transmitter-gated ion channel activity	Function	11/24	4.91 x 10-5	0.0076
GO:0000981	RNA polymerase II transcription factor activity, sequence-specific DNA binding	Function	113/599	4.76 x 10-5	0.0076
GO:0051046	Regulation of secretion	Process	118/634	3.32 x 10-5	0.0079
GO:0061138	Morphogenesis of a branching epithelium	Process	37/137	3.65 x 10-5	0.0085
GO:0031852	Mu-type opioid receptor binding	Function	3/3	7.45 x 10-5	0.0095
GO:0044420	Extracellular matrix component	Component	29/120	1.83 x 10-4	0.0101
GO:0051049	Regulation of transport	Process	280/1685	4.52 x 10-5	0.0104
GO:0016020	Membrane	Component	999/6965	1.97 x 10-4	0.0105
GO:0030425	Dendrite	Component	60/301	2.26 x 10-4	0.0117
GO:0015081	Sodium ion transmembrane transporter activity	Function	31/128	9.59 x 10-5	0.0118
GO:0045666	Positive regulation of neuron differentiation	Process	63/295	5.24 x 10-5	0.0118
GO:0051410	Detoxification of nitrogen compound	Process	2/5	5.70 x 10-5	0.0126
GO:0098742	Cell-cell adhesion via plasma-membrane adhesion molecules	Process	24/197	5.80 x 10-5	0.0127
GO:0030667	Secretory granule membrane	Component	21/69	0.0003	0.0138
GO:0022892	Substrate-specific transporter activity	Function	182/1043	0.0001	0.0138
GO:0050793	Regulation of developmental process	Process	315/1930	6.51 x 10-5	0.0140
GO:0008514	Organic anion transmembrane transporter activity	Function	23/84	0.0001	0.0141
GO:0007156	Homophilic cell adhesion via plasma membrane adhesion molecules	Process	20/149	0.0001	0.0141
GO:0051962	Positive regulation of nervous system development	Process	83/418	0.0001	0.0143
GO:2000667	Positive regulation of interleukin^{-13} secretion	Process	5/6	0.0001	0.0149
GO:0048583	Regulation of response to stimulus	Process	480/3366	0.0001	0.0154
GO:0005578	Proteinaceous extracellular matrix	Component	38/251	0.0003	0.0160
GO:0016021	Integral component of membrane	Component	690/4674	0.0004	0.0175
GO:0071837	HMG box domain binding	Function	10/18	0.0002	0.0177
GO:0009581	Detection of external stimulus	Process	27/107	9.05 x 10-5	0.0181
GO:0017146	NMDA selective glutamate receptor complex	Component	5/11	0.0004	0.0186
GO:0008509	Anion transmembrane transporter activity	Function	39/192	0.0002	0.0195
GO:0035239	Tube morphogenesis	Process	34/207	0.0001	0.0197

GO Term	Description	Class	Enrichment	P-value	FDR q-value
GO:0042221	Response to chemical	Process	322/2164	0.0001	0.0197
GO:0090596	Sensory organ morphogenesis	Process	18/48	0.0001	0.0198
GO:0005509	Calcium ion binding	Function	122/667	0.0002	0.0201
GO:0051960	Regulation of nervous system development	Process	129/717	0.0001	0.0204
GO:0048592	Eye morphogenesis	Process	16/40	0.0001	0.0205
GO:2000664	Positive regulation of interleukin-5 secretion	Process	4/4	0.0001	0.0221
GO:0051047	Positive regulation of secretion	Process	60/346	0.0001	0.0226
GO:0021707	Cerebellar granule cell differentiation	Process	3/7	0.0001	0.0227
GO:0007417	Central nervous system development	Process	19/111	0.0001	0.0250
GO:0001763	Morphogenesis of a branching structure	Process	36/146	0.0002	0.0250
GO:0048168	Regulation of neuronal synaptic plasticity	Process	16/47	0.0001	0.0251
GO:0008631	Intrinsic apoptotic signaling pathway in response to oxidative stress	Process	9/17	0.0002	0.0251
GO:0019369	Arachidonic acid metabolic process	Process	11/48	0.0001	0.0252
GO:0055085	Transmembrane transport	Process	160/1036	0.0001	0.0252
GO:0009582	Detection of abiotic stimulus	Process	27/110	0.0002	0.0253
GO:0001505	Regulation of neurotransmitter levels	Process	34/130	0.0001	0.0254
GO:0048646	Anatomical structure formation involved in morphogenesis	Process	141/776	0.0002	0.0259
GO:0043194	Axon initial segment	Component	7/12	0.0006	0.0264
GO:0045664	Regulation of neuron differentiation	Process	98/531	0.0002	0.0288
GO:0010720	Positive regulation of cell development	Process	86/452	0.0002	0.0288
GO:0031430	M band	Component	8/19	0.0007	0.0307
GO:0031235	Intrinsic component of the cytoplasmic side of the plasma membrane	Component	7/13	0.0007	0.0312
GO:0048522	Positive regulation of cellular process	Process	582/4317	0.0002	0.0318
GO:0035235	Ionotropic glutamate receptor signaling pathway	Process	10/23	0.0002	0.0320
GO:0008285	Negative regulation of cell proliferation	Process	110/613	0.0002	0.0335
GO:0001642	Group III metabotropic glutamate receptor activity	Function	4/4	0.0003	0.0336
GO:0004419	Hydroxymethylglutaryl-coa lyase activity	Function	2/2	0.0003	0.0338

GO Term	Description	Class	Enrichment	P-value	FDR q-value
GO:0044463	Cell projection part	Component	148/867	0.0008	0.0344
GO:1902710	GABA receptor complex	Component	4/15	0.0009	0.0356
GO:0051430	Corticotropin-releasing hormone receptor 1 binding	Function	2/4	0.0004	0.0359
GO:1902711	GABA-A receptor complex	Component	4/15	0.0009	0.0364
GO:0071625	Vocalisation behavior	Process	7/10	0.0002	0.0365
GO:0000987	Core promoter proximal region sequence-specific DNA binding	Function	72/354	0.0004	0.0366
GO:0060078	Regulation of postsynaptic membrane potential	Process	17/53	0.0002	0.0367
GO:0048518	Positive regulation of biological process	Process	635/4755	0.0003	0.0370
GO:0010575	Positive regulation of vascular endothelial growth factor production	Process	11/25	0.0003	0.0372
GO:0005581	Collagen trimer	Component	24/85	0.0010	0.0382
GO:0001503	Ossification	Process	31/111	0.0003	0.0409
GO:1903793	Positive regulation of anion transport	Process	12/33	0.0003	0.0410
GO:0001159	Core promoter proximal region DNA binding	Function	72/357	0.0004	0.0430
GO:0006836	Neurotransmitter transport	Process	27/93	0.0003	0.0430
GO:0021602	Cranial nerve morphogenesis	Process	7/11	0.0003	0.0431
GO:0044708	Single-organism behavior	Process	64/360	0.0003	0.0432
GO:0005272	Sodium channel activity	Function	13/36	0.0005	0.0433
GO:0050806	Positive regulation of synaptic transmission	Process	28/102	0.0003	0.0433
GO:0007605	Sensory perception of sound	Process	32/127	0.0003	0.0441
GO:0006811	Ion transport	Process	179/1091	0.0003	0.0443
GO:0060575	Intestinal epithelial cell differentiation	Process	5/6	0.0003	0.0443
GO:0045785	Positive regulation of cell adhesion	Process	55/353	0.0003	0.0445
GO:0045595	Regulation of cell differentiation	Process	217/1407	0.0004	0.0478
GO:0048251	Elastic fiber assembly	Process	4/6	0.0004	0.0479
GO:0045582	Positive regulation of T cell differentiation	Process	17/65	0.0004	0.0492
GO:1902578	Single-organism localisation	Process	383/2459	0.0004	0.0494

Table A6: Summary of GO terms found to be enriched amongst the most differentially methylated genes in iPSC-derived neurons of t(1;11) carriers.

Legend is presented on the next page.

For each GO term, table summarises the GO identifier, the GO description, the GO class, enrichment, the enrichment p-values, and enrichment FDR q-values for genes showing the most differential methylation in t(1,11) carriers compared to non-carriers. Enrichment is defined as (b/n) / (B/N), where b = the number of genes at the top of the p-value ranked list that is associated with a given GO term, n = the number of genes at the top of the p-*value-ranked gene list, B = the total number of genes associated with a given GO term, and N = the total number of genes.*

Gene	Features	Region	No. Probes	DMR *p*
FLOT1, *IER3*	TSS1500, 3'UTR, Body	Chr6:30710912-30711968	26	7.72×10^{-29}
GNASAS, *GNAS*	3'UTR, TSS1500	Chr20:57425870-57427652	51	1.96×10^{-20}
CCDC130	TSS1500, TSS200	Chr19:13858480-13858585	5	1.46×10^{-15}
SNORD116-15 - SNORD116-19	TSS1500, TSS200, Body	Chr15:25325510-25330514	16	1.46×10^{-15}
C22orf32, *SMDT1*	TSS200, 1stExon	Chr22:42475680-42475844	6	4.10×10^{-15}
MIR4458HG, *LOC729506*	Body	Chr5:8457127-8457933	7	1.89×10^{-14}
DHRS4, *C14orf167*	Body, TSS200	Chr14:24422520-24423061	7	1.62×10^{-13}
C13orf38, *CCDC169*	5'UTR, 1stExon, TSS200	Chr13:36871465-36872189	10	6.16×10^{-13}
GDNF	Body, 5'UTR	Chr5:37834742-37835348	9	1.18×10^{-12}
RFPL2	TSS1500	Chr22:32601040-32601418	4	2.08×10^{-12}
SNORD116-3, *SNORD116-8 - SNORD116-12*	TSS1500, TSS200, Body	Chr15:25314660-25322096	25	5.52×10^{-12}
LOC286083	TSS1500, TSS200	Chr8:1247657-1254583	9	5.75×10^{-12}
LRRIQ3, *TNNI3K*, *FPGT*, *FPGT-TNNI3K*	TSS200, 1stExon, Body	Chr1:74663818-74664684	13	5.75×10^{-12}
DYX1C1	Body, 5'UTR	Chr15:55790403-55790649	4	7.20×10^{-12}
SUCLG2	TSS200, Body, TSS1500	Chr3:67704958-67705415	5	1.26×10^{-11}
LYNX1	5'UTR	Chr8:143858411-143858899	10	2.99×10^{-11}
GABRB3	5'UTR, Body	Chr15:26874257-26875242	10	5.50×10^{-11}
OPCML	1stExon, TSS200	Chr11:133402081-133402540	7	9.69×10^{-11}
SLC25A4	1stExon, Body	Chr4:186064419-186064876	4	2.01×10^{-10}
FAM115A, *TCAF1*	5'UTR	Chr7:143582420-143582830	7	2.59×10^{-10}
ATPBD4, *DPH6-AS1*, *DPH6*	Body, TSS200, TSS1500	Chr15:35837611-35838829	11	3.10×10^{-10}
TRIM58	1stExon	Chr1:248020638-248020745	3	4.74×10^{-10}
NA	IGR, TSS200	Chr1:24740123-24740365	5	6.18×10^{-10}
NA	IGR	Chr16:86230282-86235475	8	1.34×10^{-9}
SYCP1	TSS200, 1stExon	Chr1:115397370-115397669	8	1.85×10^{-9}
AK2	TSS200, TSS1500	Chr1:33502644-33502906	7	2.45×10^{-9}
SNCA	5'UTR	Chr4:90757170-90757809	6	2.76×10^{-9}
C4orf39, *TRIM61*	Body, 1stExon	Chr4:165877646-165878196	6	2.76×10^{-9}
RAG2, *C11orf74*	TSS1500, Body, 5'UTR	Chr11:36614033-36617183	15	2.76×10^{-9}
NBR1, *TMEM106A*	3'UTR, TSS1500	Chr17:41363234-41363894	9	2.89×10^{-9}
IFT74	5'UTR, TSS200	Chr9:26955771-26956380	8	3.58×10^{-9}
SNORD116-2, *SNORD116-6*, *SNORD116-5*	TSS1500, TSS200	Chr15:25308804-25311885	6	3.89×10^{-9}
NMBR	TSS200, TSS1500	Chr6:142410082-142410331	5	4.78×10^{-9}
GPR19	5'UTR, 1stExon, TSS200	Chr12:12848796-12849157	5	8.48×10^{-9}
NPY	TSS200, 1stExon	Chr7:24323786-24323893	3	9.22×10^{-9}
DYNLRB2, *LOC102724084*	Body, 1stExon	Chr16:80574806-80575123	5	1.24×10^{-8}

Gene	Features	Region	No. Probes	DMR *p*
NA	IGR	Chr3:138658825-138659056	3	1.40 x 10^{-8}
TXNRD2, *COMT*	Body, 1stExon, 5'UTR, TSS200	Chr22:19929008-19929737	10	2.34 x 10^{-8}
NA	IGR	Chr3:64253489-64253933	6	2.58 x 10^{-8}
ZNF267	TSS200, 1stExon	Chr16:31885037-31885144	3	3.96 x 10^{-8}
MIR3663HG, *MIR3663*	Body, TSS200	Chr10:118928281-118928738	3	3.96 x 10^{-8}
NR2F1, *MIR548AO*	Body, 3'UTR	Chr5:92929213-92929839	5	4.25 x 10^{-8}
NA	IGR	Chr15:56297839-56301176	4	4.49 x 10^{-8}
LOC151174, *LOC643387*	Body	Chr2:239139682-239140295	6	4.68 x 10^{-8}
ACAA1, *MYD88*	1stExon, TSS1500	Chr3:38180042-38180158	3	5.02 x 10^{-8}
SVOPL	Body	Chr7:138349032-138349263	3	7.46 x 10^{-8}
NKD1	Body	Chr16:50586910-50589287	6	9.75 x 10^{-8}
NBEAL1	TSS1500	Chr2:203879114-203879307	3	9.75 x 10^{-8}
NSMCE1, *FLJ21408*	5'UTR	Chr16:27279315-27279964	5	1.24 x 10^{-7}
NA	IGR	Chr8:33862743-33869095	8	1.56 x 10^{-7}
SNORD116-1	TSS1500, TSS200, Body	Chr15:25295628-25297627	7	1.58 x 10^{-7}
GABPB2	TSS200, TSS1500	Chr1:151042707-151042918	5	1.82 x 10^{-7}
LOC153684	Body	Chr5:43044255-43046853	6	2.12 x 10^{-7}
*OR2L13**	1stExon	Chr1:248100531-248100667	4	3.26 x 10^{-7}
*PRDM16**	Body	Chr1:3301199-3304640	11	3.60 x 10^{-7}
TSNARE1	5'UTR	Chr8:143473719-143476331	8	4.17 x 10^{-7}
NA	IGR	Chr10:131216295-131219549	3	9.77 x 10^{-7}
TNFSF11	5'UTR	Chr13:43147443-43148394	8	1.13 x 10^{-6}
PAR5, *SNORD64*	TSS1500, Body	Chr15:25229200-25231867	7	1.24 x 10^{-6}
MPZL3, *MPZL2*	Body, 1stExon, TSS200, TSS1500, 3'UTR	Chr11:118121715-118125529	16	1.45 x 10^{-6}
KCNQ1DN	TSS1500	Chr11:2890586-2890752	12	1.47 x 10^{-6}
KIAA1462	TSS200 , TSS1500	Chr10:30348608-30348805	4	1.47 x 10^{-6}
NA	IGR	Chr8:74279957-74284471	5	1.58 x 10^{-6}
*LOC399815**, *FAM24B*	1stExon, Body	Chr10:124639051-124639365	8	1.81 x 10^{-6}
SNORD116-28	TSS1500, Body	Chr15:25348817-25350835	4	2.54 x 10^{-6}
NRROS, *LRRC33*	1stExon, TSS200	Chr3:196366527-196366658	3	2.65 x 10^{-6}
NA	IGR	Chr4:12797269-12801122	6	2.69 x 10^{-6}
TXNRD1, *EID3*	5'UTR	Chr12:104697435-104697623	5	2.69 x 10^{-6}
PDLIM3	3'UTR, Body	Chr4:186422174-186426753	6	3.60 x 10^{-6}
SNORD116-23, *SNORD116-24*	Body, TSS1500, TSS200, Body	Chr15:25336103-25340250	10	3.94 x 10^{-6}
LOC646405	Body	Chr13:25506129-25506611	4	4.13 x 10^{-6}
C3orf67	TSS1500	Chr3:59035879-59036089	5	5.65 x 10^{-6}
KIF17	TSS200, TSS1500	Chr1:21044642-21044733	4	6.08 x 10^{-6}
NA	IGR	Chr13:45490318-45494868	8	6.11 x 10^{-6}
PAX6	Body	Chr11:31821192-31821680	5	7.34 x 10^{-6}

Gene	Features	Region	No. Probes	DMR *p*
HOXA5	TSS1500	Chr7:27183907-27184072	6	7.41×10^{-6}
KHDC1	1stExon	Chr6:73972688-73972873	4	8.81×10^{-6}
FOXG1-AS1, *FOXG1*	TSS1500, TSS200	Chr14:29234379-29234962	5	9.08×10^{-6}
WDR72	1stExon, Body, TSS200	Chr15:54051767-54051999	5	9.88×10^{-6}
FMO5	1stExon	Chr1:146696812-146697092	5	1.17×10^{-5}
KDM6B, *TMEM88*	3'UTR, TSS1500, TSS200	Chr17:7757665-7758368	10	1.17×10^{-5}
C1orf87	TSS200, TSS1500	Chr1:60539574-60539771	3	1.22×10^{-5}
PWWP2B	Body	Chr10:134211628-134212136	3	1.35×10^{-5}
LOC648987	Body	Chr5:43037295-43037957	5	1.44×10^{-5}
MGMT	Body	Chr10:131498562-131507260	18	1.54×10^{-5}
C6orf174	Body	Chr6:127835947-127836264	4	1.54×10^{-5}
SHF	3'UTR	Chr15:45459463-45459997	3	1.54×10^{-5}
SLC35E3	TSS1500, TSS200	Chr12:69139606-69139841	8	1.54×10^{-5}
CLNK	Body	Chr4:10599367-10602421	5	1.59×10^{-5}
NA	IGR	Chr17:8102030-8105175	3	1.84×10^{-5}
PABPC4L	5'UTR, TSS200, TSS1500	Chr4:135121591-135123536	6	2.09×10^{-5}
MARVELD3	Body	Chr16:71660679-71661136	4	2.42×10^{-5}
GAL3ST3	TSS1500	Chr11:65816838-65817081	3	2.45×10^{-5}
NA	IGR	Chr5:175558908-175562712	3	2.45×10^{-5}
SBK2	TSS1500, TSS200	Chr19:56048356-56048614	5	2.45×10^{-5}
RUNDC3A	Body	Chr17:42392562-42392802	3	2.61×10^{-5}
NA	IGR	Chr1:165086860-165087091	3	2.84×10^{-5}
SORCS1	Body	Chr10:108728446-108730917	4	2.85×10^{-5}
KIAA0556	Body	Chr16:27728635-27731609	8	2.96×10^{-5}
UNC5D	Body	Chr8:35400059-35402907	5	3.01×10^{-5}
NA	IGR	Chr4:138464719-138468410	6	3.29×10^{-5}
PRR15	Body, 3'UTR	Chr7:29606081-29606615	3	3.29×10^{-5}
LINC00887, *LOC100131551*	Body, TSS200, TSS1500	Chr3:194029837-194031210	6	3.31×10^{-5}
SYNPO2L	Body	Chr10:75407143-75407600	3	3.42×10^{-5}
NA	IGR	Chr6:27511308-27515019	6	3.57×10^{-5}
DIABLO, *LOC101593348*	TSS200, 5'UTR	Chr12:122711109-122711266	5	3.68×10^{-5}
SPDL1, *CCDC99*	TSS1500, TSS200	Chr5:169010192-169010471	6	3.68×10^{-5}
NA	IGR	Chr8:103818745-103819034	3	3.70×10^{-5}
37681	Body	Chr5:126210060-126212351	3	3.91×10^{-5}
NA	IGR	Chr8:72470408-72471273	5	4.00×10^{-5}
PACS2, *TEX22*	3'UTR	Chr14:105864044-105864696	4	4.05×10^{-5}
URI1	Body	Chr19:30493289-30496076	6	4.06×10^{-5}
MAP3K13	TSS200	Chr3:185000532-185000875	6	4.26×10^{-5}
DLGAP2-AS1, *DLGAP2*	Body	Chr8:1568905-1570904	5	4.26×10^{-5}
CLNK	Body	Chr4:10657763-10660447	5	4.52×10^{-5}

Gene	Features	Region	No. Probes	DMR *p*
IFIH1	TSS200, TSS1500	Chr2:163175190-163175547	7	4.59 x 10^{-5}
GATA4	TSS1500	Chr8:11531474-11535523	4	4.91 x 10^{-5}
RAD50	1stExon, 5'UTR	Chr5:131893120-131893481	3	4.91 x 10^{-5}
MGMT	Body	Chr10:131423627-131426281	5	5.01 x 10^{-5}
BBS4, *HIGD2B*	TSS1500, Body	Chr15:72977622-72978268	6	5.53 x 10^{-5}
ZNF486	TSS200, 1stExon, Body	Chr19:20277423-20279422	6	5.56 x 10^{-5}
KCNC3	TSS200	Chr19:50836794-50837025	3	5.68 x 10^{-5}
C1orf101	TSS1500, TSS200	Chr1:244624348-244624541	4	5.69 x 10^{-5}
ZNF528	TSS1500, TSS200	Chr19:52900804-52901010	7	5.77 x 10^{-5}
NA	IGR	Chr18:76266256-76266487	3	5.81 x 10^{-5}
LBXCOR1	Body	Chr15:68126062-68126293	3	6.07 x 10^{-5}
MGMT	Body	Chr10:131459030-131461633	3	6.11 x 10^{-5}
KIAA1324, *C1orf194*	1stExon	Chr1:109656628-109656856	4	6.20 x 10^{-5}
BRCA1, *NBR2*	TSS1500	Chr17:41278098-41278221	4	6.20 x 10^{-5}
NA	IGR	Chr2:47499600-47499831	3	6.30 x 10^{-5}
HLCS	TSS200, TSS1500	Chr21:38362699-38362796	4	6.43 x 10^{-5}
NA	IGR	Chr8:1198261-1202123	12	6.50 x 10^{-5}
NA	IGR	Chr5:2684024-2688626	4	6.75 x 10^{-5}
NA	IGR	Chr10:43913915-43918051	9	6.96 x 10^{-5}
NA	IGR	Chr6:166711265-166711603	3	7.10 x 10^{-5}
QPCT	Body	Chr2:37572194-37572651	3	7.14 x 10^{-5}
HIST1H3G, *HIST1H2BI*	TSS1500	Chr6:26271651-26271844	5	7.54 x 10^{-5}
TMEM246	TSS200, TSS1500	Chr9:104249644-104249855	4	7.59 x 10^{-5}
NFE2	TSS200, TSS1500	Chr12:54694871-54695160	6	7.70 x 10^{-5}
FN3K, *TBCD**	Body, 3'UTR	Chr17:80708364-80708898	3	8.15 x 10^{-5}
COL5A2	Body, 1stExon, TSS200, TSS1500	Chr2:190043091-190045109	9	8.16 x 10^{-5}
MFNG	TSS200	Chr22:37882512-37882593	3	9.01 x 10^{-5}
LINC01007	Body	Chr7:101206410-101209723	4	9.19 x 10^{-5}
NA	IGR	Chr6:10883574-10884431	5	9.36 x 10^{-5}
PLD6	1stExon, TSS200	Chr17:17109586-17109693	4	0.0001
CHDH, *IL17RB*	Body	Chr3:53880624-53881100	4	0.0001
P3H4, *FKBP10*, *SC65*	1stExon	Chr17:39969159-39969376	5	0.0001
RCCD1	5'UTR	Chr15:91498772-91499203	3	0.0001
VTRNA1-3	TSS1500, TSS200	Chr5:140104488-140108652	10	0.0001
TSSC1	Body	Chr2:3286306-3286517	3	0.0001
ASPSCR1	Body	Chr17:79952299-79952756	3	0.0001
ZIC4	3'UTR, Body	Chr3:147105152-147107234	8	0.0001
NA	IGR	Chr1:2765178-2768849	5	0.0001
AURKC	TSS200, 1stExon	Chr19:57742340-57742497	5	0.0001
SNRPN	TSS1500, TSS200	Chr15:25068557-25070194	8	0.0001

Gene	Features	Region	No. Probes	DMR *p*
MDK	TSS1500	Chr11:46402504-46402697	4	0.0001
WDR90	Body	Chr16:710804-711261	4	0.0001
HIST1H2BI	TSS200	Chr6:26272999-26273192	5	0.0001
C17orf49, *RNASEK-C17orf49*, *MIR497HG*	Body	Chr17:6918723-6919180	3	0.0001
L3MBTL1	Body	Chr20:42165996-42167995	3	0.0002
NELL1	Body	Chr11:20692638-20693112	6	0.0002
NA	IGR	Chr12:132168689-132169421	4	0.0002
EVX1, *EVX1-AS*	3'UTR	Chr7:27286392-27287285	5	0.0002
FAAP20	1stExon; TSS200	Chr1:2143710-2144293	6	0.0002
RARA	Body	Chr17:38501385-38501596	4	0.0002
RFWD3	TSS1500	Chr16:74701311-74701583	3	0.0002
SERHL2	TSS1500	Chr22:42949429-42949622	4	0.0002
NA	IGR	Chr2:11529070-11532247	5	0.0002
WDR27	Body	Chr6:170068156-170068686	3	0.0002
NA	IGR	Chr1:244393574-244396892	3	0.0002
RGS22	Body	Chr8:101117673-101118130	6	0.0002
NA	IGR	Chr22:35387324-35391041	3	0.0002
NA	IGR	Chr18:76476379-76480717	4	0.0002
NA	IGR	Chr12:131215610-131218928	3	0.0002
NA	IGR	Chr2:127653621-127657985	4	0.0002
ESYT3	TSS1500	Chr3:138153030-138153266	3	0.0002
HLA-G	Body	Chr6:29795909-29796210	6	0.0002
NA	IGR	Chr11:69286177-69286467	5	0.0002
CWH43	Body	Chr4:48988400-48988707	5	0.0002
GCM2	TSS200, TSS1500	Chr6:10882229-10882422	3	0.0002
ANGPTL2, *RALGPS1*	5'UTR, 1stExon, Body	Chr9:129883494-129886937	12	0.0002
MCM5	TSS1500	Chr22:35795207-35795400	5	0.0002
FBXL16	Body	Chr16:745819-746360	4	0.0002
CACNB3	TSS1500, 1stExon	Chr12:49207818-49208401	5	0.0002
NA	IGR	Chr8:143119205-143125208	7	0.0002
KLHL8	TSS1500	Chr4:88142178-88142383	3	0.0002
FAM19A3	TSS1500	Chr1:113261680-113261765	3	0.0002
RAI1	5'UTR	Chr17:17694446-17696389	6	0.0002
NA	IGR	Chr2:21616546-21620941	6	0.0002
HERC5	TSS1500 , TSS200	Chr4:89377875-89378148	5	0.0002
CACNA2D2	Body	Chr3:50491746-50493882	3	0.0002
RUNX1	Body	Chr21:36399120-36399331	3	0.0003
L3MBTL1	TSS200	Chr20:42136013-42136244	4	0.0003
SFRP1	Body	Chr8:41164377-41165032	4	0.0003
CR1L	Body	Chr1:207842497-207843032	4	0.0003

Gene	Features	Region	No. Probes	DMR *p*
TNFRSF11A	3'UTR	Chr18:60052196-60052730	3	0.0003
USP3	Body	Chr15:63847612-63849744	4	0.0003
TMED7-TICAM2, *TICAM2*, *LOC101927100*	Body, TSS1500	Chr5:114937768-114938107	5	0.0003
WFIKKN2	TSS1500	Chr17:48911036-48911325	4	0.0003
NA	IGR	Chr7:127911851-127912082	3	0.0003
MGMT	Body	Chr10:131333690-131335865	3	0.0003
RASA3	TSS1500	Chr13:114898366-114898451	3	0.0003
FAM107B	5'UTR, Body	Chr10:14643903-14644416	3	0.0003
SYNGR3	Body	Chr16:2041573-2042030	3	0.0003
KIFC3	TSS200, TSS1500	Chr16:57836613-57836734	7	0.0003
C21orf119, *URB1*	1stExon, Body	Chr21:33765223-33765680	8	0.0003
NA	IGR	Chr4:11635992-11639073	3	0.0003
MAS1L	1stExon	Chr6:29454770-29455056	3	0.0003
CPN1	Body	Chr10:101825012-101825223	4	0.0004
UNC5D	Body	Chr8:35233678-35236593	3	0.0004
OR2T11	TSS200	Chr1:248790418-248790499	3	0.0004
NA	IGR	Chr2:22752740-22756406	3	0.0004
ACCN1	TSS200, TSS1500	Chr17:32483984-32484077	4	0.0004
NA	IGR	Chr14:106345374-106348606	3	0.0004
GPX1	TSS1500	Chr3:49396096-49396328	3	0.0004
NA	IGR	Chr7:45298075-45301306	3	0.0005
PCSK2	TSS1500	Chr20:17206677-17206786	3	0.0005
CA3	TSS1500	Chr8:86350471-86350688	3	0.0005
LOC100188947, *HECTD2-AS1*	Body	Chr10:93333974-93336676	5	0.0005
TSNARE1	TSS200, TSS1500	Chr8:143484773-143484880	4	0.0005
MGMT	Body	Chr10:131405583-131408553	3	0.0005
NA	IGR	Chr8:711621-714723	4	0.0005
NA	IGR	Chr6:100914986-100915217	3	0.0005
DLX3	3'UTR	Chr17:48064472-48067553	3	0.0005
CST6	1stExon, Body	Chr11:65779618-65780075	3	0.0006
FABP3	TSS1500	Chr1:31846346-31846686	3	0.0006
LOC100129550	TSS200	Chr3:122605307-122605388	3	0.0006
CAMTA1	Body	Chr1:7740137-7740391	4	0.0006
HOXC12	Body	Chr12:54349924-54350522	3	0.0006
TTC7B	Body	Chr14:91021269-91023407	3	0.0006
RTN2	Body	Chr19:45996231-45996812	5	0.0006
NA	IGR	Chr2:177003443-177004306	4	0.0006
HLA-L	Body	Chr6:30228137-30228659	4	0.0006
NA	IGR	Chr17:53509189-53512861	3	0.0006
PRKCD	TSS1500	Chr3:53194601-53194794	3	0.0006

Gene	Features	Region	No. Probes	DMR *p*
MBD6, DDIT3	TSS1500	Chr12:57915458-57915732	4	0.0006
NA	IGR	Chr2:241289146-241292368	6	0.0006
DPP10	Body	Chr2:115920596-115921053	3	0.0007
NA	IGR	Chr17:41446167-41446398	4	0.0007
KCNQ4	1stExon , Body	Chr1:41250011-41250222	3	0.0007
LOC101928414	TSS1500, TSS200	Chr15:45569676-45573176	8	0.0007
AMICA	1stExon, Body, 5'UTR	Chr11:118083920-118086554	10	0.0007
NA	IGR	Chr20:48535779-48539395	4	0.0007
WNT10A	Body	Chr2:219746625-219747307	3	0.0007
NA	IGR	Chr11:2884002-2884260	3	0.0007
RCAN3	Body , 3'UTR	Chr1:24861708-24861926	4	0.0007
EPB42	Body	Chr15:43509478-43512449	4	0.0007
ACTN4	Body	Chr19:39218124-39218639	3	0.0007
TSNARE1	Body	Chr8:143327572-143330725	7	0.0007
SORCS1	Body	Chr10:108459467-108461516	3	0.0008
NA	IGR	Chr17:75524797-75525119	3	0.0008
ERICH1-AS1	Body	Chr8:819402-822386	3	0.0008
NA	IGR	Chr12:120835662-120835893	3	0.0008
NA	IGR	Chr7:142420271-142423352	5	0.0009
HOXA11AS, HOXA11	Body	Chr7:27225116-27225501	9	0.0009
MGMT	Body	Chr10:131353782-131356301	7	0.0009
SLC6A3	Body	Chr5:1408937-1411155	4	0.0010
ZC3H12D	Body	Chr6:149772663-149773120	4	0.0010
TPO	Body	Chr2:1486895-1489637	5	0.0010
NKX6-2	TSS1500	Chr10:134600608-134600743	5	0.0010
NA	IGR	Chr10:130757981-130761062	5	0.0010
FOLH1	1stExon	Chr11:49230020-49230179	5	0.0010
FOXL2	1stExon	Chr3:138662648-138663231	4	0.0011
TBC1D16	Body	Chr17:77982111-77982705	3	0.0011
NA	IGR	Chr10:133206945-133211351	7	0.0011
WDR52	5'UTR, 1stExon	Chr3:113160002-113160363	3	0.0011
BPIL2, BPIFC	Body, TSS1500	Chr22:32852147-32854146	4	0.0012
NA	IGR	Chr6:29413214-29417635	5	0.0012
ZNF385A, LOC102724050	Body	Chr12:54779066-54781194	8	0.0012
PAOX	Body	Chr10:135192881-135193105	3	0.0012
ALKBH3	TSS1500	Chr11:43898530-43901398	4	0.0012
ZNF729	TSS200, 1stExon	Chr19:22468931-22469514	5	0.0012
MGMT	3'UTR, Body	Chr10:131563371-131572287	33	0.0013
HTR3D	5'UTR, 1stExon	Chr3:183749242-183751251	6	0.0013
MGC2752, CENPBD1P1	Body	Chr19:59092383-59092840	3	0.0013
DUSP22	TSS200, TSS1500	Chr6:291785-291978	3	0.0014

Gene	Features	Region	No. Probes	DMR *p*
CABIN1	Body	Chr22:24551449-24551957	4	0.0014
ARL6IP1	TSS1500	Chr16:18813216-18813320	3	0.0014
PAQR8	TSS1500, TSS200	Chr6:52226601-52226803	4	0.0014
NA	IGR	Chr8:47117778-47120859	4	0.0016
BEND7	3'UTR, Body	Chr10:13479534-13484810	11	0.0016
SH3GL3	5'UTR, Body	Chr15:84158516-84160969	4	0.0017
*NFE2L1**	Body	Chr17:46130332-46133017	5	0.0017
PLA2G16	TSS200, TSS1500	Chr11:63382046-63382239	6	0.0017
GRM7	1stExon	Chr3:6902965-6903072	3	0.0017
NA	IGR	Chr10:130828104-130833933	20	0.0017
IL1F10	5'UTR	Chr2:113828621-113830684	5	0.0017
SOX9, SOX9-AS1	Body	Chr17:70115602-70116059	4	0.0018
KCNQ5	TSS1500	Chr6:73330200-73330393	3	0.0018
LINC01449	Body	Chr7:41146516-41149597	3	0.0018
RUFY4	Body	Chr2:218936172-218939005	10	0.0019
SLC16A7	5'UTR	Chr12:59990383-59991085	4	0.0019
NA	IGR	Chr12:119311119-119314922	3	0.0021
DLGAP2	TSS1500, TSS200, 5'UTR	Chr8:1448815-1450849	9	0.0021
NA	IGR	Chr3:13323451-13323757	3	0.0022
RAPGEFL1	Body	Chr17:38347587-38348044	5	0.0022
ZNF354A	TSS1500	Chr5:178157859-178158141	5	0.0022
*TRIM10**	3'UTR, Body	Chr6:30117358-30122147	15	0.0023
NA	IGR	Chr8:1136005-1136829	8	0.0026
SYNGAP1	Body	Chr6:33393493-33394012	11	0.0028
NMRAL1, HMOX2	TSS200	Chr16:4526196-4526263	3	0.0028
NA	IGR	Chr1:179696392-179699641	7	0.0029
DDX54	Body	Chr12:113598306-113601158	6	0.0030
NA	IGR	Chr2:97134967-97138795	4	0.0030
ZNF790, ZNF345	5'UTR	Chr19:37342251-37342765	7	0.0031
EYS	Body	Chr6:65085402-65088319	4	0.0031
PRKG2	1stExon, TSS200, 5'UTR, TSS1500	Chr4:82125889-82127526	5	0.0032
CASZ1	Body	Chr1:10731063-10734510	13	0.0032
SNRPN	5'UTR	Chr15:25100155-25101792	7	0.0033
SV2B	5'UTR	Chr15:91767439-91769754	5	0.0036
LINC01574	TSS1500, TSS200	Chr5:176169846-176170429	6	0.0036
LOC646999	Body	Chr7:39648995-39649464	5	0.0038
NA	IGR	Chr6:3051955-3056424	12	0.0040
MGMT	Body	Chr10:131558596-131561420	9	0.0040
C15orf29	TSS1500, TSS200	Chr15:34502428-34502676	6	0.0040
LOC101927274, LOC4404466	Body, TSS1500	Chr17:49411301-49415610	6	0.0041

Gene	Features	Region	No. Probes	DMR *p*
NA	IGR	Chr6:8341845-8346144	5	0.0041
NA	IGR	Chr16:86669407-86673086	4	0.0043
C22orf26, *LOC150381*	Body, 1stExon	Chr22:46449201-46449778	5	0.0043
NA	IGR	Chr5:140704206-140707792	6	0.0044
C3orf30, *IGSF11*	5'UTR, 1stExon, TSS200	Chr3:118863646-118865447	17	0.0044
IGFBP7-AS1, *IGFBP7*	Body	Chr4:57975509-57976139	6	0.0045
UNC5D	Body	Chr8:35579713-35582594	6	0.0049
NUDT12	5'UTR, 1stExon, TSS200, TSS1500	Chr5:102897404-102899041	11	0.0050
NA	IGV	Chr10:133483051-133486730	7	0.0052
SPRED2	1stExon, Body	Chr2:65592933-65595889	10	0.0053
HRH2	TSS1500	Chr5:175084555-175084839	4	0.0053
CAPN14	TSS1500	Chr2:31441134-31444215	6	0.0057
NA	IGV	Chr16:34966298-34971523	14	0.0057
FAM55A, *NXPE1*	5'UTR, TSS200, TSS1500	Chr11:114429466-114431103	4	0.0062
LOC100506406, *CAPSL*	5'UTR, TSS1500	Chr5:35937856-35939510	4	0.0064
CCL24	Body, 1stExon, TSS200, TSS1500	Chr7:75441678-75443677	10	0.0064
VASN, *CORO7*, *CORO7-PAM16*	Body	Chr16:4421216-4421831	6	0.0065
NA	IGV	Chr2:89156388-89161703	13	0.0069
NPBWR1	TSS1500	Chr8:53850517-53851224	4	0.0069
WDR72	5'UTR	Chr15:54024598-54026992	4	0.0069
ZNF257	TSS1500, TSS200, Body	Chr19:22234850-22236849	8	0.0070
ANKRD53	3'UTR	Chr2:71210706-71213389	4	0.0074
OXSM, *NGLY1*	TSS1500, TSS200, 5'UTR, Body	Chr3:25830824-25832767	15	0.0078
NA	IGV	Chr6:170410016-170413493	6	0.0086
CYP51A1-AS1	Body	Chr7:91808513-91809345	9	0.0091
LOC100506272	Body	Chr4:188512911-188515758	5	0.0094
NA	IGV	Chr8:1271087-1276226	12	0.0094
NA	IGV	Chr12:131269782-131274762	8	0.0096
COASY	Body	Chr17:40715116-40715386	5	0.0098
C2, *CFB*	Body, 3'UTR, 1stExon	Chr6:31912016-31916531	22	0.0098
OR2H1	5'UTR, Body	Chr6:29429272-29432460	10	0.0100
SNRPN, *SNURF*	5'UTR	Chr15:25198847-25199344	4	0.0103
MYOM2, *MIR7160*	Body	Chr8:2022251-2025299	8	0.0103
MGMT	Body	Chr10:131449224-131451999	5	0.0103
HLA-DQB2	TSS1500	Chr6:32731678-32734276	6	0.0104
SLC34A2	5'UTR	Chr4:25657712-25658215	4	0.0106
PRLR	5'UTR, 1stExon, TSS200	Chr5:35229347-35231161	11	0.0110

Gene	Features	Region	No. Probes	DMR *p*
NA	IGV	Chr5:43006091-43011083	9	0.0110
MGMT	Body	Chr10:131360165-131362539	6	0.0111
ACOX3	Body	Chr4:8394941-8397813	7	0.0113
NA	IGV	Chr6:68597160-68601633	11	0.0113
DLGAP2	5'UTR	Chr8:1462084-1463943	4	0.0113
LEMD3	TSS1500	Chr12:65562806-65563059	5	0.0115
TRIM40	Body	Chr6:30112074-30115801	20	0.0116
GPR116, *ADGRF5*	5'UTR	Chr6:46889401-46891760	8	0.0117
DLGAP2	Body	Chr8:1623580-1627554	11	0.0118
SULT2B1	Body, 1stExon	Chr19:49076983-49079118	4	0.0118
ZDHHC8P	Body, TSS1500	Chr22:23743094-23745724	5	0.0123
PSMB9	Body	Chr6:32824677-32826383	13	0.0128
IRF8	5'UTR, Body	Chr16:85934624-85937410	10	0.0128
TBX5	5'UTR	Chr12:114843806-114844343	9	0.0131
MAP2K5	Body	Chr15:67839376-67842406	6	0.0137
NT5C	TSS1500, TSS200	Chr17:73127982-73128175	6	0.0138
NME2, *NME1-NME2*	Body	Chr17:49244488-49245025	4	0.0138
*TAP1**, *PSMB9*	Body, TSS1500	Chr6:32820192-32820649	11	0.0140
DPP10-AS1	Body	Chr2:115918369-115919118	7	0.0141
NA	IGV	Chr16:86754036-86757172	8	0.0152
KLK11	5'UTR, TSS1500, TSS200	Chr19:51529932-51531941	8	0.0158
NA	IGV	Chr1:2819124-2823189	11	0.0171
VAX1	3'UTR	Chr10:118890822-118891356	4	0.0172
NA	IGV	Chr8:808950-812979	10	0.0174
IRX4, *CTD-2194D22.4*	TSS1500	Chr5:1887741-1888324	4	0.0174
SEMA3G	Body, 1stExon, TSS200, TSS1500	Chr3:52477874-52479873	9	0.0174
NA	IGV	Chr10:131073007-131076843	9	0.0178
NA	IGV	Chr6:27634837-27638057	11	0.0180
NA	IGV	Chr7:156294950-156298855	7	0.0192
ZBTB20	5'UTR	Chr3:114789424-114792059	6	0.0194
URAHP, *LOC100130015*	Body, TSS200, TSS1500	Chr16:90112696-90115172	11	0.0198
WHSC1L1	3'UTR, Body	Chr8:38130862-38134370	6	0.0216
DLX6AS, *DLX6-AS1*	Body	Chr7:96625788-96628252	11	0.0231
NAF1	TSS1500	Chr4:164088381-164088661	4	0.0238
HOXB5, *LOC404266*	Body	Chr17:46669349-46669671	6	0.0239
GABBR1	Body	Chr6:29598931-29599618	9	0.0242
MEIS1	Body	Chr2:66672008-66672565	5	0.0245
GOLT1A	Body, 1stExon, TSS200, TSS1500	Chr1:204181957-204183956	13	0.0259
CDH22	Body	Chr20:44848207-44851180	7	0.0268

Gene	Features	Region	No. Probes	DMR *p*
MKRN3	TSS1500, TSS200	Chr15:23810093-23810382	5	0.0269
ST8SIA6, *ST8SIA6-AS1*	Body	Chr10:17439426-17442047	4	0.0271
NA	IGV	Chr19:7411023-7415602	8	0.0276
CPSF3L	Body	Chr1:1249982-1252004	6	0.0278
NA	IGV	Chr10:130694531-130698677	4	0.0291
ADCY1	TSS1500	Chr7:45613667-45613752	4	0.0295
EGFR	Body	Chr7:55144727-55148426	9	0.0297
ERICH1-AS1	Body	Chr8:858574-862612	9	0.0297
NA	IGV	Chr1:2730034-2734601	8	0.0319
NA	IGV	Chr8:47013884-47018323	7	0.0327
LRRTM2, *CTNNA1*	Body, 1stExon	Chr5:138209749-138212908	17	0.0351
GSTK1	TSS1500, TSS200, 1stExon, Body	Chr7:142959778-142961868	9	0.0354
NA	IGV	Chr4:11372499-11375675	4	0.0361
BTNL2	Body	Chr6:32369520-32372505	15	0.0366
SMIM5, *RECQL5*, *LOC643008*	TSS1500, Body	Chr17:73629012-73629691	7	0.0387
CTD-3080P12.3	Body, TSS200	Chr5:1177618-1180699	6	0.0388
NA	IGV	Chr15:77374694-77379194	8	0.0401
TREX1, *ATRIP*	Body, TSS200, 1stExon	Chr3:48505717-48507912	15	0.0414
LOC149837, *LINC00654*	Body, TSS200, TSS1500	Chr20:5484144-5486143	9	0.0425
NA	IGV	Chr2:18010334-18013922	5	0.0428
VGLL4	Body	Chr3:11674300-11676659	6	0.0433
DOPEY2	TSS1500, TSS200, 5'UTR, Body	Chr21:37536104-37538122	5	0.0433
NA	IGV	Chr11:64653573-64657364	4	0.0446
ABHD12B	5'UTR	Chr14:51338993-51339396	4	0.0446
NA	IGV	Chr20:42098438-42103116	5	0.0449
LOC100130357	TSS1500, TSS200	Chr6:13294608-13297689	10	0.0453
GNA12	Body	Chr7:2846477-2848922	6	0.0455
VANGL1	5'UTR, Body	Chr1:116192821-116195193	5	0.0462
LINC01346	TSS1500, TSS200, Body	Chr1:3998584-4002140	8	0.0462
RORA	Body	Chr15:61025728-61028719	4	0.0467
MYLK2	TSS1500, TSS200, 5'UTR, Body	Chr20:30406329-30408300	12	0.0468
NA	IGV	Chr8:1281291-1285648	7	0.0497

Table A7: DMRs identified between t(1;11) carriers and non-carriers in iPSC-derived neurons.

From left to right, columns summarise the DMR-containing genes, the DMR's underlying genomic features, the Hg19 genomic coordinates of each DMR, the number of probes within each DMR, and the p*-value for differential methylation in t(1;11) carriers. Genes highlighted with a red asterisk (*) indicate those that also contained t(1;11)-associated DMRs identified in blood.*

SNP1	SNP2	SNP3	Gene1	Gene2	Gene3	LRT *p*-value	R^2	*q*-value
rs12730369	rs1336979	rs6541281	*DISC1*	*SORCS1*	*DISC1*	0.0001	0.08	0.36
rs1565415	rs2107182	rs823162	*SORCS3*	*SORCS2*	*DISC1*	0.0002	0.08	0.79
rs17466832	rs6541281	rs823162	*SORCS2*	*DISC1*	*DISC1*	0.0003	0.08	0.98
rs6835799	rs7440772	rs823162	*SORCS2*	*SORCS2*	*DISC1*	0.0003	0.08	1
rs1565415	rs756255	rs823162	*SORCS3*	*SORCS2*	*DISC1*	0.0004	0.08	1
rs1336979	rs2269852	rs4918288	*SORCS1*	*SORCS2*	*SORCS1*	0.0004	0.08	1
rs1565415	rs2269852	rs823162	*SORCS3*	*SORCS2*	*DISC1*	0.0005	0.08	1
rs2107182	rs6541281	rs823162	*SORCS2*	*DISC1*	*DISC1*	0.0006	0.08	1
rs10937826	rs1557816	rs6541281	*SORCS2*	*SORCS2*	*DISC1*	0.0007	0.08	1
rs6541281	rs756255	rs823162	*DISC1*	*SORCS2*	*DISC1*	0.0009	0.08	1
rs10884100	rs1251753	rs1336979	*SORCS3*	*SORCS1*	*SORCS1*	0.0010	0.08	1
rs2269852	rs6541281	rs823162	*SORCS2*	*DISC1*	*DISC1*	0.0016	0.08	1
rs1336979	rs610785	rs756255	*SORCS1*	*SORCS1*	*SORCS2*	0.0017	0.08	1
rs11932646	rs2107182	rs4350297	*SORCS2*	*SORCS2*	*SORCS3*	0.0017	0.08	1
rs610785	rs665679	rs756255	*SORCS1*	*SORCS1*	*SORCS2*	0.0017	0.08	1
rs10937826	rs7667970	rs9432040	*SORCS2*	*SORCS2*	*DISC1*	0.0020	0.08	1
rs1557816	rs2269852	rs2295959	*SORCS2*	*SORCS2*	*DISC1*	0.0021	0.08	1
rs11932646	rs4350297	rs756255	*SORCS2*	*SORCS3*	*SORCS2*	0.0021	0.08	1
rs1336979	rs1557816	rs4918288	*SORCS1*	*SORCS2*	*SORCS1*	0.0025	0.08	1
rs2107182	rs610785	rs6541281	*SORCS2*	*SORCS1*	*DISC1*	0.0029	0.08	1
rs2269850	rs4918288	rs665679	*SORCS2*	*SORCS1*	*SORCS1*	0.0029	0.08	1
rs4350297	rs6835799	rs823162	*SORCS3*	*SORCS2*	*DISC1*	0.0036	0.08	1
rs12730369	rs1557816	rs9432040	*DISC1*	*SORCS2*	*DISC1*	0.0036	0.08	1

SNP1	SNP2	SNP3	Gene1	Gene2	Gene3	LRT *p*-value	R^2	*q*-value
rs4689869	rs4918288	rs756255	*SORCS2*	*SORCS1*	*SORCS2*	0.0037	0.08	1
rs1565415	rs2269850	rs823162	*SORCS3*	*SORCS2*	*DISC1*	0.0039	0.08	1
rs1251753	rs1557816	rs6835799	*SORCS1*	*SORCS2*	*SORCS2*	0.0043	0.08	1
rs10884100	rs12730369	rs17466832	*SORCS3*	*DISC1*	*SORCS2*	0.0045	0.08	1
rs10032900	rs2269850	rs7897974	*SORCS2*	*SORCS2*	*SORCS1*	0.0045	0.08	1
rs1565415	rs2269852	rs610785	*SORCS3*	*SORCS2*	*SORCS1*	0.0052	0.08	1
rs1557816	rs4918288	rs665679	*SORCS2*	*SORCS1*	*SORCS1*	0.0054	0.08	1
rs11932646	rs6541281	rs7897974	*SORCS2*	*DISC1*	*SORCS1*	0.0054	0.08	1
rs10884100	rs1565415	rs2269852	*SORCS3*	*SORCS3*	*SORCS2*	0.0056	0.08	1
rs2269852	rs4689869	rs665679	*SORCS2*	*SORCS2*	*SORCS1*	0.0059	0.08	1
rs1557816	rs2269850	rs4918288	*SORCS2*	*SORCS2*	*SORCS1*	0.0060	0.08	1
rs10937826	rs4613570	rs6541281	*SORCS2*	*SORCS2*	*DISC1*	0.0061	0.08	1
rs2269852	rs4918288	rs665679	*SORCS2*	*SORCS1*	*SORCS1*	0.0063	0.08	1
rs10884100	rs2269852	rs7667970	*SORCS3*	*SORCS2*	*SORCS2*	0.0068	0.08	1
rs1557816	rs7440772	rs823162	*SORCS2*	*SORCS2*	*DISC1*	0.0070	0.08	1
rs4350297	rs6835799	rs7667970	*SORCS3*	*SORCS2*	*SORCS2*	0.0073	0.08	1
rs4918288	rs7440772	rs9432040	*SORCS1*	*SORCS2*	*DISC1*	0.0074	0.08	1
rs1557816	rs4613570	rs6541281	*SORCS2*	*SORCS2*	*DISC1*	0.0080	0.08	1
rs1565415	rs2269850	rs610785	*SORCS3*	*SORCS2*	*SORCS1*	0.0086	0.08	1
rs10884100	rs17466832	rs2269850	*SORCS3*	*SORCS2*	*SORCS2*	0.0088	0.08	1
rs4350297	rs4689869	rs7897974	*SORCS3*	*SORCS2*	*SORCS1*	0.0093	0.08	1
rs1336979	rs6541281	rs7440772	*SORCS1*	*DISC1*	*SORCS2*	0.0105	0.08	1
rs10937826	rs12730369	rs9432040	*SORCS2*	*DISC1*	*DISC1*	0.0106	0.08	1

SNP1	SNP2	SNP3	Gene1	Gene2	Gene3	LRT *p*-value	R^2	*q*-value
rs1251753	rs4918288	rs7440772	*SORCS1*	*SORCS1*	*SORCS2*	0.0107	0.08	1
rs1557816	rs17466832	rs610785	*SORCS2*	*SORCS2*	*SORCS1*	0.0109	0.08	1
rs10884100	rs12730369	rs756255	*SORCS3*	*DISC1*	*SORCS2*	0.0110	0.08	1
rs10884100	rs7440772	rs756255	*SORCS3*	*SORCS2*	*SORCS2*	0.0117	0.08	1
rs10937826	rs6541281	rs9432040	*SORCS2*	*DISC1*	*DISC1*	0.0117	0.08	1
rs1557816	rs2269850	rs2295959	*SORCS2*	*SORCS2*	*DISC1*	0.0117	0.08	1
rs610785	rs6541281	rs756255	*SORCS1*	*DISC1*	*SORCS2*	0.0120	0.08	1
rs4613570	rs7440772	rs823162	*SORCS2*	*SORCS2*	*DISC1*	0.0122	0.09	1
rs10884100	rs1565415	rs823162	*SORCS3*	*SORCS3*	*DISC1*	0.0125	0.08	1
rs11932646	rs2269850	rs4350297	*SORCS2*	*SORCS2*	*SORCS3*	0.0127	0.08	1
rs1336979	rs17466832	rs7440772	*SORCS1*	*SORCS2*	*SORCS2*	0.0127	0.08	1
rs6541281	rs665679	rs6835799	*DISC1*	*SORCS1*	*SORCS2*	0.0127	0.08	1
rs10937826	rs2269850	rs7897974	*SORCS2*	*SORCS2*	*SORCS1*	0.0128	0.08	1
rs1251753	rs1557816	rs2295959	*SORCS1*	*SORCS2*	*DISC1*	0.0130	0.08	1
rs10884100	rs1557816	rs2107182	*SORCS3*	*SORCS2*	*SORCS2*	0.0132	0.08	1
rs12730369	rs17466832	rs2269850	*DISC1*	*SORCS2*	*SORCS2*	0.0132	0.08	1
rs10032900	rs2269852	rs7897974	*SORCS2*	*SORCS2*	*SORCS1*	0.0135	0.08	1
rs11932646	rs1565415	rs665679	*SORCS2*	*SORCS3*	*SORCS1*	0.0136	0.08	1
rs2269850	rs6541281	rs823162	*SORCS2*	*DISC1*	*DISC1*	0.0137	0.08	1
rs2107182	rs610785	rs665679	*SORCS2*	*SORCS1*	*SORCS1*	0.0138	0.08	1
rs1336979	rs4689869	rs7440772	*SORCS1*	*SORCS2*	*SORCS2*	0.0138	0.08	1
rs1565415	rs4689869	rs823162	*SORCS3*	*SORCS2*	*DISC1*	0.0140	0.08	1
rs1251753	rs2269852	rs9432040	*SORCS1*	*SORCS2*	*DISC1*	0.0143	0.08	1

SNP1	SNP2	SNP3	Gene1	Gene2	Gene3	LRT *p*-value	R^2	*q*-value
rs1251753	rs7440772	rs9432040	*SORCS1*	*SORCS2*	*DISC1*	0.0143	0.08	1
rs610785	rs665679	rs823162	*SORCS1*	*SORCS1*	*DISC1*	0.0146	0.08	1
rs1565415	rs17466832	rs4689869	*SORCS3*	*SORCS2*	*SORCS2*	0.0147	0.08	1
rs1565415	rs6541281	rs9432040	*SORCS3*	*DISC1*	*DISC1*	0.0147	0.08	1
rs10032900	rs7667970	rs9432040	*SORCS2*	*SORCS2*	*DISC1*	0.0156	0.08	1
rs2269852	rs4350297	rs665679	*SORCS2*	*SORCS3*	*SORCS1*	0.0158	0.08	1
rs10032900	rs12730369	rs1565415	*SORCS2*	*DISC1*	*SORCS3*	0.0164	0.08	1
rs1557816	rs4689869	rs665679	*SORCS2*	*SORCS2*	*SORCS1*	0.0166	0.08	1
rs4637403	rs610785	rs9432040	*SORCS2*	*SORCS1*	*DISC1*	0.0166	0.08	1
rs11932646	rs756255	rs7897974	*SORCS2*	*SORCS2*	*SORCS1*	0.0168	0.08	1
rs4918288	rs823162	rs9432040	*SORCS1*	*DISC1*	*DISC1*	0.0171	0.08	1
rs10937826	rs1251753	rs12730369	*SORCS2*	*SORCS1*	*DISC1*	0.0178	0.08	1
rs4350297	rs4637403	rs9432040	*SORCS3*	*SORCS2*	*DISC1*	0.0180	0.08	1
rs1557816	rs4613570	rs665679	*SORCS2*	*SORCS2*	*SORCS1*	0.0184	0.08	1
rs17466832	rs7667970	rs823162	*SORCS2*	*SORCS2*	*DISC1*	0.0192	0.08	1
rs12730369	rs1565415	rs6541281	*DISC1*	*SORCS3*	*DISC1*	0.0195	0.08	1
rs610785	rs6541281	rs9432040	*SORCS1*	*DISC1*	*DISC1*	0.0195	0.08	1
rs6541281	rs665679	rs7440772	*DISC1*	*SORCS1*	*SORCS2*	0.0196	0.08	1
rs1251753	rs1557816	rs4689869	*SORCS1*	*SORCS2*	*SORCS2*	0.0198	0.08	1
rs665679	rs7440772	rs823162	*SORCS1*	*SORCS2*	*DISC1*	0.0199	0.08	1
rs4350297	rs7440772	rs823162	*SORCS3*	*SORCS2*	*DISC1*	0.0201	0.08	1
rs4613570	rs610785	rs823162	*SORCS2*	*SORCS1*	*DISC1*	0.0201	0.09	1
rs11932646	rs4350297	rs4689869	*SORCS2*	*SORCS3*	*SORCS2*	0.0206	0.08	1

SNP1	SNP2	SNP3	Gene1	Gene2	Gene3	LRT *p*-value	R^2	*q*-value
rs4350297	rs756255	rs823162	*SORCS3*	*SORCS2*	*DISC1*	0.0213	0.08	1
rs10884100	rs17466832	rs2295959	*SORCS3*	*SORCS2*	*DISC1*	0.0222	0.08	1
rs1251753	rs1336979	rs9432040	*SORCS1*	*SORCS1*	*DISC1*	0.0226	0.08	1
rs12730369	rs1565415	rs4918288	*DISC1*	*SORCS3*	*SORCS1*	0.0228	0.08	1
rs1336979	rs2107182	rs610785	*SORCS1*	*SORCS2*	*SORCS1*	0.0230	0.08	1
rs1336979	rs2269850	rs4918288	*SORCS1*	*SORCS2*	*SORCS1*	0.0234	0.08	1
rs2269850	rs6835799	rs823162	*SORCS2*	*SORCS2*	*DISC1*	0.0243	0.08	1
rs2295959	rs610785	rs7440772	*DISC1*	*SORCS1*	*SORCS2*	0.0246	0.08	1
rs1251753	rs610785	rs6541281	*SORCS1*	*SORCS1*	*DISC1*	0.0251	0.08	1
rs2269852	rs6541281	rs665679	*SORCS2*	*DISC1*	*SORCS1*	0.0259	0.08	1
rs4918288	rs6835799	rs823162	*SORCS1*	*SORCS2*	*DISC1*	0.0262	0.08	1
rs10032900	rs1251753	rs12730369	*SORCS2*	*SORCS1*	*DISC1*	0.0263	0.08	1
rs1557816	rs4613570	rs7897974	*SORCS2*	*SORCS2*	*SORCS1*	0.0264	0.08	1
rs10937826	rs12730369	rs1565415	*SORCS2*	*DISC1*	*SORCS3*	0.0264	0.08	1
rs1251753	rs1557816	rs7440772	*SORCS1*	*SORCS2*	*SORCS2*	0.0264	0.08	1
rs1251753	rs6835799	rs9432040	*SORCS1*	*SORCS2*	*DISC1*	0.0272	0.08	1
rs12730369	rs4918288	rs6541281	*DISC1*	*SORCS1*	*DISC1*	0.0274	0.08	1
rs17466832	rs4637403	rs7897974	*SORCS2*	*SORCS2*	*SORCS1*	0.0278	0.08	1
rs12730369	rs1336979	rs1565415	*DISC1*	*SORCS1*	*SORCS3*	0.0286	0.08	1
rs1557816	rs2269850	rs823162	*SORCS2*	*SORCS2*	*DISC1*	0.0288	0.08	1
rs4637403	rs7667970	rs7897974	*SORCS2*	*SORCS2*	*SORCS1*	0.0294	0.08	1
rs2269850	rs2269852	rs2295959	*SORCS2*	*SORCS2*	*DISC1*	0.0296	0.08	1
rs1557816	rs4918288	rs7897974	*SORCS2*	*SORCS1*	*SORCS1*	0.0298	0.08	1

SNP1	SNP2	SNP3	Gene1	Gene2	Gene3	LRT *p*-value	R2	*q*-value
rs10884100	rs11932646	rs2107182	*SORCS3*	*SORCS2*	*SORCS2*	0.0300	0.08	1
rs10884100	rs2269852	rs4613570	*SORCS3*	*SORCS2*	*SORCS2*	0.0301	0.08	1
rs2295959	rs4918288	rs610785	*DISC1*	*SORCS1*	*SORCS1*	0.0306	0.08	1
rs10937826	rs12730369	rs6541281	*SORCS2*	*DISC1*	*DISC1*	0.0309	0.08	1
rs1557816	rs2269852	rs4918288	*SORCS2*	*SORCS2*	*SORCS1*	0.0309	0.08	1
rs10884100	rs2107182	rs7440772	*SORCS3*	*SORCS2*	*SORCS2*	0.0313	0.08	1
rs10032900	rs12730369	rs4613570	*SORCS2*	*DISC1*	*SORCS2*	0.0315	0.08	1
rs10937826	rs12730369	rs4637403	*SORCS2*	*DISC1*	*SORCS2*	0.0318	0.08	1
rs6835799	rs7897974	rs9432040	*SORCS2*	*SORCS1*	*DISC1*	0.0319	0.08	1
rs4637403	rs665679	rs9432040	*SORCS2*	*SORCS1*	*DISC1*	0.0326	0.08	1
rs10884100	rs17466832	rs2107182	*SORCS3*	*SORCS2*	*SORCS2*	0.0330	0.08	1
rs10884100	rs756255	rs7667970	*SORCS3*	*SORCS2*	*SORCS2*	0.0332	0.08	1
rs2107182	rs6541281	rs665679	*SORCS2*	*DISC1*	*SORCS1*	0.0333	0.08	1
rs1557816	rs17466832	rs665679	*SORCS2*	*SORCS2*	*SORCS1*	0.0334	0.08	1
rs17466832	rs4918288	rs823162	*SORCS2*	*SORCS1*	*DISC1*	0.0337	0.08	1
rs10032900	rs4918288	rs7440772	*SORCS2*	*SORCS1*	*SORCS2*	0.0341	0.08	1
rs12730369	rs610785	rs6541281	*DISC1*	*SORCS1*	*DISC1*	0.0342	0.08	1
rs1336979	rs4918288	rs9432040	*SORCS1*	*SORCS1*	*DISC1*	0.0343	0.08	1
rs1557816	rs6541281	rs7440772	*SORCS2*	*DISC1*	*SORCS2*	0.0343	0.08	1
rs10032900	rs1557816	rs6541281	*SORCS2*	*SORCS2*	*DISC1*	0.0344	0.08	1
rs2107182	rs7440772	rs823162	*SORCS2*	*SORCS2*	*DISC1*	0.0347	0.08	1
rs1251753	rs1565415	rs6835799	*SORCS1*	*SORCS3*	*SORCS2*	0.0352	0.08	1
rs11932646	rs610785	rs665679	*SORCS2*	*SORCS1*	*SORCS1*	0.0354	0.08	1

SNP1	SNP2	SNP3	Gene1	Gene2	Gene3	LRT *p*-value	R2	*q*-value
rs1565415	rs4350297	rs9432040	*SORCS3*	*SORCS3*	*DISC1*	0.0354	0.08	1
rs10937826	rs1565415	rs7667970	*SORCS2*	*SORCS3*	*SORCS2*	0.0359	0.08	1
rs4613570	rs823162	rs9432040	*SORCS2*	*DISC1*	*DISC1*	0.0363	0.08	1
rs1251753	rs1565415	rs7897974	*SORCS1*	*SORCS3*	*SORCS1*	0.0365	0.08	1
rs10884100	rs10937826	rs6835799	*SORCS3*	*SORCS2*	*SORCS2*	0.0366	0.08	1
rs2269852	rs6541281	rs7667970	*SORCS2*	*DISC1*	*SORCS2*	0.0368	0.08	1
rs10937826	rs12730369	rs4613570	*SORCS2*	*DISC1*	*SORCS2*	0.0370	0.08	1
rs2107182	rs665679	rs6835799	*SORCS2*	*SORCS1*	*SORCS2*	0.0371	0.08	1
rs17466832	rs4637403	rs823162	*SORCS2*	*SORCS2*	*DISC1*	0.0374	0.08	1
rs2269852	rs4350297	rs9432040	*SORCS2*	*SORCS3*	*DISC1*	0.0374	0.08	1
rs1251753	rs2107182	rs756255	*SORCS1*	*SORCS2*	*SORCS2*	0.0375	0.08	1
rs10884100	rs11932646	rs756255	*SORCS3*	*SORCS2*	*SORCS2*	0.0378	0.08	1
rs10937826	rs2269852	rs7897974	*SORCS2*	*SORCS2*	*SORCS1*	0.0379	0.08	1
rs10032900	rs12730369	rs9432040	*SORCS2*	*DISC1*	*DISC1*	0.0382	0.08	1
rs1565415	rs4637403	rs4689869	*SORCS3*	*SORCS2*	*SORCS2*	0.0382	0.08	1
rs2107182	rs665679	rs823162	*SORCS2*	*SORCS1*	*DISC1*	0.0382	0.08	1
rs4613570	rs6541281	rs7897974	*SORCS2*	*DISC1*	*SORCS1*	0.0383	0.08	1
rs10884100	rs17466832	rs756255	*SORCS3*	*SORCS2*	*SORCS2*	0.0392	0.08	1
rs1565415	rs2295959	rs756255	*SORCS3*	*DISC1*	*SORCS2*	0.0394	0.08	1
rs10937826	rs11932646	rs6541281	*SORCS2*	*SORCS2*	*DISC1*	0.0396	0.08	1
rs2269852	rs2295959	rs7667970	*SORCS2*	*DISC1*	*SORCS2*	0.0400	0.08	1
rs10032900	rs4637403	rs9432040	*SORCS2*	*SORCS2*	*DISC1*	0.0401	0.08	1
rs12730369	rs4637403	rs7897974	*DISC1*	*SORCS2*	*SORCS1*	0.0405	0.08	1

SNP1	SNP2	SNP3	Gene1	Gene2	Gene3	LRT *p*-value	R2	*q*-value
rs10884100	rs11932646	rs2269850	*SORCS3*	*SORCS2*	*SORCS2*	0.0409	0.08	1
rs12730369	rs6835799	rs823162	*DISC1*	*SORCS2*	*DISC1*	0.0410	0.08	1
rs610785	rs7667970	rs823162	*SORCS1*	*SORCS2*	*DISC1*	0.0416	0.08	1
rs2269850	rs610785	rs6541281	*SORCS2*	*SORCS1*	*DISC1*	0.0418	0.08	1
rs10032900	rs2107182	rs7897974	*SORCS2*	*SORCS2*	*SORCS1*	0.0423	0.08	1
rs665679	rs6835799	rs756255	*SORCS1*	*SORCS2*	*SORCS2*	0.0426	0.08	1
rs1336979	rs2269852	rs6541281	*SORCS1*	*SORCS2*	*DISC1*	0.0426	0.08	1
rs7440772	rs756255	rs823162	*SORCS2*	*SORCS2*	*DISC1*	0.0428	0.09	1
rs10884100	rs1557816	rs1565415	*SORCS3*	*SORCS2*	*SORCS3*	0.0432	0.08	1
rs1336979	rs2269852	rs4350297	*SORCS1*	*SORCS2*	*SORCS3*	0.0435	0.08	1
rs10937826	rs1336979	rs7897974	*SORCS2*	*SORCS1*	*SORCS1*	0.0436	0.08	1
rs1251753	rs1557816	rs4918288	*SORCS1*	*SORCS2*	*SORCS1*	0.0438	0.08	1
rs1251753	rs2295959	rs4613570	*SORCS1*	*DISC1*	*SORCS2*	0.0439	0.08	1
rs1251753	rs17466832	rs2295959	*SORCS1*	*SORCS2*	*DISC1*	0.0441	0.08	1
rs12730369	rs4918288	rs7440772	*DISC1*	*SORCS1*	*SORCS2*	0.0444	0.08	1
rs10884100	rs1336979	rs4613570	*SORCS3*	*SORCS1*	*SORCS2*	0.0446	0.08	1
rs1336979	rs2107182	rs6541281	*SORCS1*	*SORCS2*	*DISC1*	0.0448	0.08	1
rs2295959	rs756255	rs7667970	*DISC1*	*SORCS2*	*SORCS2*	0.0453	0.08	1
rs2107182	rs610785	rs823162	*SORCS2*	*SORCS1*	*DISC1*	0.0454	0.08	1
rs10937826	rs4613570	rs610785	*SORCS2*	*SORCS2*	*SORCS1*	0.0457	0.08	1
rs10884100	rs1565415	rs6835799	*SORCS3*	*SORCS3*	*SORCS2*	0.0464	0.08	1
rs2269850	rs610785	rs665679	*SORCS2*	*SORCS1*	*SORCS1*	0.0464	0.08	1
rs11932646	rs1565415	rs4637403	*SORCS2*	*SORCS3*	*SORCS2*	0.0466	0.08	1

rs12730369	rs1336979	rs4689869	*DISC1*	*SORCS1*	*SORCS2*	0.0472	0.09	1
rs12730369	rs17466832	rs2269852	*DISC1*	*SORCS2*	*SORCS2*	0.0473	0.08	1
rs12730369	rs1565415	rs4613570	*DISC1*	*SORCS3*	*SORCS2*	0.0475	0.08	1

Table A8: Three-SNP interactions nominally significantly associated with crystallised intelligence.

Shown is a summary of nominally significant genetic interactions associated with crystallised intelligence ($p \leq 0.05$*). From left to right, columns show the SNP identifiers and their corresponding genes the likelihood ratio test* p-*value for the difference in goodness-of-fit of the interaction model versus the null model, omitting the interaction term; the proportion of variance explained by the model (*R^2*), and the Benjamini-Hochberg FDR* q-*value for the SNP interaction following correction for 3168 tests.*

Appendix II

Publications

Genomics Data 9 (2016) 22–24

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Data in Brief

Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip

Daniel L. McCartney [a], Rosie M. Walker [a], Stewart W. Morris [a], Andrew M. McIntosh [a,b,c], David J. Porteous [a,c], Kathryn L. Evans [a,c,*]

[a] Medical Genetics Section, Centre for Genomic and Experimental Medicine, MRC Institute of Genetics and Molecular Medicine, Western General Hospital, The University of Edinburgh, Crewe Road, Edinburgh EH4 2XU, UK
[b] Division of Psychiatry, Royal Edinburgh Hospital, The University of Edinburgh, Edinburgh EH10 5HF, UK
[c] Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, The University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK

ARTICLE INFO

Article history:
Received 21 April 2016
Received in revised form 23 May 2016
Accepted 24 May 2016
Available online 26 May 2016

Keywords:
DNA methylation
Infinium MethylationEPIC BeadChip
Cross-hybridising probes
Polymorphic CpG
Quality control

ABSTRACT

Genome-wide analysis of DNA methylation has now become a relatively inexpensive technique thanks to array-based methylation profiling technologies. The recently developed Illumina Infinium MethylationEPIC BeadChip interrogates methylation at over 850,000 sites across the human genome, covering 99% of RefSeq genes. This array supersedes the widely used Infinium HumanMethylation450 BeadChip, which has permitted insights into the relationship between DNA methylation and a wide range of conditions and traits. Previous research has identified issues with certain probes on both the HumanMethylation450 BeadChip and its predecessor, the Infinium HumanMethylation27 BeadChip, which were predicted to affect array performance. These issues concerned probe-binding specificity and the presence of polymorphisms at target sites. Using *in silico* methods, we have identified probes on the Infinium MethylationEPIC BeadChip that are predicted to (i) measure methylation at polymorphic sites and (ii) hybridise to multiple genomic regions. We intend these resources to be used for quality control procedures when analysing data derived from this platform.

Specifications [*standardized info for the reader*]	
Organism/cell line/tissue	Homo sapiens genome sequence (Hg19)
Sex	n/a
Sequencer or array type	Illumina Infinium MethylationEPIC Array
Data format	Analysed: Table of polymorphic targets and lists of crosshybridising probes
Experimental factors	Infinium MethylationEPIC probe data, 1000 genomes phase 3 data, UCSC genome browser human reference genome sequence
Experimental features	In silico alignment of Infinium MethylationEPIC probe sequences to bisulfite converted genome sequences (Hg19) and cross-referencing of probe target coordinates to 1000 genomes project phase 3 data.
Consent	Raw data available from Illumina, UCSC genome browser and 1000 genomes project
Sample source location	n/a

1. Direct link to deposited data

http://dx.doi.org/10.1016/j.gdata.2016.05.012

* Corresponding author at: Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, The University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK.
E-mail address: Kathy.Evans@igmm.ed.ac.uk (K.L. Evans).

2. Introduction

DNA methylation is an epigenetic mark typically occurring at cytosine-guanine dinucleotides (CpGs). Changes in DNA methylation are observed in normal development, in response to environmental stimuli, and in certain disease states [1]. DNA methylation is linked to transcriptional activity, rendering it a key regulatory motif [2]. Recent years have seen the development of high-throughput DNA methylation profiling techniques including whole-genome bisulphite sequencing (WGBS), methylated DNA immunoprecipitation (meDIP) and microarray-based technologies [3]. The Infinium HumanMethylation450 BeadChip, developed by Illumina, has offered an attractive array-based option to researchers, as it interrogates methylation at over 485,000 sites across the genome at single-base resolution at a relatively low cost (Bibikova et al., 2011 [4]). However, issues with probe-binding specificity and polymorphic targets have been identified which may compromise data integrity if not adequately addressed (Chen et al., 2013 [5]).

The Infinium HumanMethylation450 BeadChip has recently been superseded by the Infinium MethylationEPIC BeadChip. This array interrogates DNA methylation at over 850,000 sites, including >90% of the HumanMethylation450 array's targets. This substantial increase in coverage, coupled with a continuing trend for interest in the role of DNA methylation, is likely to result in wide-spread use of this array.

http://dx.doi.org/10.1016/j.gdata.2016.05.012

As such, it is essential that its potential shortcomings are thoroughly understood. In order to generate a resource that will be of use to researchers using the MethylationEPIC BeadChip we have identified probes that may perform sub-optimally. This work, therefore, represents an update of Chen et al.'s [5] previous characterisation of the Infinium HumanMethylation450 BeadChip.

Like its predecessor, the MethylationEPIC BeadChip uses two types of probe chemistry (Type I and Type II) to interrogate methylation. The differences between the two chemistries and the situations in which they are used have been described fully in previous publications [6]. Briefly, Type I assays use separate probes for unmethylated and methylated target sites while Type II assays use a single probe. Both assay types differentiate methylation state *via* single base extension of a fluorescent-labelled nucleotide.

Taking the differences between Type I and Type II assays into consideration, we have performed *in silico* analyses to identify probes on the Infinium MethylationEPIC BeadChip that are predicted to hybridise to multiple genomic regions, as well as probes where signal may be affected by polymorphisms at the target site, which could alter probe binding. Both of these factors should be taken into account when performing quality control of data produced using this technology.

3. Methods

3.1. Identification of probes with a polymorphic target

Probes potentially affected by polymorphisms at the target site were identified following methods described previously [5].

The signal-generating process of single-base extension requires end-nucleotide matching for both Type I and Type II probes. Therefore, we limited our query to target CpGs and sites of single-base extension, as polymorphisms at these sites are most likely to generate spurious signals.

Using information from the Infinium MethylationEPIC BeadChip manifest file (MethylationEPIC_v-1-0_B1.csv; date of download: 8 February 2016), we generated a list of genomic coordinates (hg19, GRCh37) of the target cytosine base (C) and guanine base (G) for all probes on the array. For Infinium Type I probes we also included the base before the target CpG, as this is the site of single base extension for these probes. We cross-referenced these coordinates to those of variants listed by the 1000 Genomes Project (phase 3) [7] to generate a list of probes affected by polymorphisms at the target CpG and/or site of single-base extension.

3.2. Identification of probes with non-specific hybridisation potential

Probes with the potential to cross-hybridise were identified following methods described previously [5].

3.2.1. Generation of probe sequences for in silico analyses

Many Infinium Type II probe sequences contain an "R" nucleotide representing either an adenine (A) or guanine (G) base, depending on whether the underlying target cytosine is methylated or unmethylated. All possible combinations of Type II probe sequences were generated, and combined with a list of the Type I probe sequences.

3.2.2. Generation of genomic comparison sequences for in silico analyses

The GRCh37 release of the human genome sequence was downloaded from the University of California, Santa Cruz (UCSC) Genome Browser website (https://genome.ucsc.edu/) as a reference, excluding alternative assemblies (*e.g.* chr17_ctg5_hap1) to avoid duplicated results (date of download: 11 January 2016). From this, we generated four modified reference genome sequences. A bisulphite-converted methylated forward genome sequence was generated *in silico* by converting all non-CpG cytosine bases to thymine (T) bases in the reference sequence. The same process was performed for the reverse complement of the reference sequence to generate a bisulphite-converted methylated reverse sequence of the human genome. Bisulphite-converted unmethylated forward and reverse sequences were generated by converting all C bases to T in the forward reference sequence and its reverse complement.

Using the BLAST-like alignment tool (BLAT) [8], we aligned the probe sequences described above to the four modified reference genome sequences, as well as their reverse complements. The BLAT parameters used were: *stepSize* = 5, *minScore* = 0, *minIdentity* = 0 and *repMatch* = 1,000,000,000. Probes were defined as being at high-risk of non-specific binding if there was a gap-free match of 47 or more nucleotides, which had to include the end base of the query sequence, at an off-target locus.

4. Results

4.1. Infinium MethylationEPIC BeadChip probes with polymorphic targets

Coordinates for 866,836 probes were obtained from the Infinium MethylationEPIC BeadChip manifest downloaded on 8th February 2016. Excluding control probes, the manifest file contained 142,262 Type I probes (426,786 potential signal-affecting positions), and 724,574 Type II probes (1,449,148 potential signal-affecting positions), giving a total of 1,875,934 sites which were interrogated for genetic variation.

We identified 340,327 sites with either single nucleotide polymorphisms (SNPs), insertions or deletions (indels), or structural variation. These sites were targeted by 297,744 unique probes: 34% of the total probe content of the Infinium MethylationEPIC BeadChip. Of these, 23,399 probes (2.7%) targeted polymorphic sites with a minor allele frequency (MAF) of ≥5% in at least one population studied. A table of probes affected by polymorphisms, with minor allele frequencies corresponding to African, admixed American, European, South Asian, and East Asian populations (AFR, AMR, EUR, SAS, EAS; respectively) is available in the supplementary information of this paper (Supplementary Table 1).

4.2. Infinium MethylationEPIC BeadChip probes with cross-hybridisation potential

A total of 1,752,932 potential probe sequences, each 50 bases in length, were aligned to *in silico* bisulphite-converted forward and reverse methylated and unmethylated reference genomes, and their corresponding complementary strands in BLAT (*i.e.* eight single-stranded genomes in total). We identified 44,210 probes (11,772 Type I probes and 32,438 Type II probes) with ≥47 nucleotide off-target matches including the end base, which were defined as potentially cross-hybridising. A list of these probes is available in the supplementary information of this paper (Supplementary Tables 2–3).

Consistent with findings on the Infinium HumanMethylation450 BeadChip [5], a larger proportion of non-CpG-targeting probes (Probe ID prefix = "ch") were identified as potentially cross-hybridising compared to CpG-targeting probes (Probe ID prefix = "cg"). Of 863,904 CpG-targeting probes present on the array, 42,558 (4.9% of total CpG-targeting probes) were identified as potentially cross-hybridising (Supplementary Table 2). In contrast, of 2932 non-CpG targeting probes, we found only 1280 to bind specifically to their targets while the remaining 1652 were potentially cross-hybridising (56% of total non-CpG targeting probes; Supplementary Table 3), based on the information provided in the Illumina manifest.

5. Discussion

In order to identify probes that might compromise the performance of the Illumina Infinium MethylationEPIC BeadChip, we have generated

lists of probes that may be affected by non-specific binding and/or polymorphisms at the target site.

Our *in silico* analyses identified 44,210 probes (5.1% total probe content) with potential off-target binding sites and 23,399 probes (2.7% total probe content) whose target site contains a polymorphism with a MAF ≥ 0.05 in at least one population studied, which may lead to artefactual signal due to impaired probe-binding. We recommend that users take these probes into consideration when analysing data on this platform, applying the appropriate filtering criteria in a population-specific manner, where possible. We recognise that there may be some situations where retaining probes mapping to polymorphic target sites will be desirable. For example, a difference in methylation due to a SNP that creates or destroys a CpG at a target site may be informative if it confers a change in disease risk.

Chen et al. (2013) [5] demonstrated that autosomal probes defined as potentially cross-hybridising according to their criterion of an off-target match of 47/50 bases, including the end nucleotide, showed an enrichment for off-target binding sites on the sex chromosomes. Failure to exclude these probes could, therefore, result in the spurious conclusion that these loci are differentially methylated between males and females. Following their methods, we have identified probes on the Infinium MethylationEPIC BeadChip with the potential to hybridise to multiple genomic regions, thus generating off-target signal. We suggest the exclusion of these probes prior to data analysis. Although the exclusion of potentially cross-hybridising probes defined using this method is likely to result in an improvement in the validity of the results obtained from the array, it is likely that the actual extent of off-target binding will vary by locus. Factors such as local sequence composition, including the presence of polymorphisms underlying the probe sequence, are likely to play a role in determining the likelihood of cross-hybridisation. It is, therefore, recommended that any results of interest that may have been generated due to cross-hybridisation are checked using an alternative technique, such as pyrosequencing of bisulphite-converted DNA.

In summary, we have produced lists of probes on the new Illumina Infinium MethylationEPIC BeadChip that measure methylation at sites affected by polymorphisms and/or have the potential to cross-hybridise. Based on the wide-spread use of the HumanMethylation450 BeadChip, we predict that the Illumina Infinium MethylationEPIC BeadChip will play a central role in epigenome-wide association studies (EWAS) over the next few years. As such, it is essential that factors affecting the performance of the array, such as probe specificity and sequence polymorphisms, which we have demonstrated to potentially affect a substantial proportion of probes, are taken into consideration. We recommend that the resources supplied with this paper be used in conjunction with additional standard quality control measures, such as excluding probes with low signal-to-background ratios, omission of samples with a high proportion of such probes, and appropriate data normalisation strategies (for review see Wilhelm-Benartzi et al., 2013 [9]), in order to maximise the likelihood of producing meaningful results.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2016.05.012.

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgements

DLMcC is funded by a Ph.D. studentship awarded by Mental Health Research UK (http://www.mhruk.org/). KLE would like to acknowledge the support of the Brain & Behavior Research Foundation through a NARSAD Independent Investigator Award. AMM would like to acknowledge the support of The Health Foundation through a Clinician Scientist Fellowship and a NARSAD Independent Investigator Award from the Brain & Behavior Research Foundation and Wellcome Trust Strategic Support (104036/Z/14/Z). KLE, AMM and DJP are members of The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology (CCACE), part of the cross council Lifelong Health and Wellbeing Initiative (G0700704/84698). Funding of CCACE from the BBSRC, EPSRC, ESRC and MRC is gratefully acknowledged.

References

[1] S. Gopalakrishnan, B.O. Van Emburgh, K.D. Robertson, DNA methylation in development and human disease. Mutat. Res. 647 (2008) 30–38.
[2] Z.D. Smith, A. Meissner, DNA methylation: roles in mammalian development. Nat. Rev. Genet. 14 (2013) 204–220.
[3] T. Zuo, B. Tycko, T.-M. Liu, H.-J.L. Lin, T.H.-M. Huang, Methods in DNA methylation profiling. Epigenomics 1 (2009) 331–345.
[4] M. Bibikova, et al., High density DNA methylation array with single CpG site resolution. Genomics 98 (2011) 288–295.
[5] Y.A. Chen, et al., Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 8 (2013) 203–209.
[6] S. Dedeurwaerder, et al., Evaluation of the Infinium methylation 450K technology. Epigenomics 3 (2011) 771–784.
[7] 1000 Genomes Project Consortium, A global reference for human genetic variation. Nature 526 (2015) 68–74.
[8] W.J. Kent, BLAT - the BLAST-like alignment tool. Genome Res. 12 (2002) 656–664.
[9] C.S. Wilhelm-Benartzi, et al., Review of processing and analysis methods for DNA methylation array data. Br. J. Cancer 109 (2013) 1394–1402.