THE UNIVERSITY *of* EDINBURGH

Investigation of putative regulatory loci relevant to the pathogenesis of psychiatric illness

Rosie Walker

Doctor of Philosophy
The University of Edinburgh
2013

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.

_____

Rosie Walker

Acknowledgements

I am grateful to many people for making the research presented in this thesis possible. Firstly, to Kathy Evans, my supervisor, for her unfailing support and encouragement over the last few years: I feel very lucky to have been supervised by Kathy. I would also like to thank my second supervisor, Pippa Thomson, for her invaluable insights and inspiring enthusiasm, and Jeremy Hall, my third supervisor. I am grateful to Andrea Christoforou, Sarah Brown, Ellen Grünewald, Gillian Hamilton, Shaun Mackie, Fumiaki Ogawa, Chris Boyd and Alison Hill for their advice on various aspects of my work, and to our German collaborators (Prof. Sven Cichon and colleagues) for their help with the *Neuregulin 1* association study and to Prof. Simon Fisher and members of his group for their assistance with the work on FOXP2. Helen Torrance and Susan Anderson have provided fantastic support in the lab and, with Rosemary, Heather, and Sharon, have dealt with all admin-related tasks with incredible efficiency. Thanks to Stewart and Steve for their help with all computer issues. I would also like to thank Prof. David Porteous for his support and guidance.

I have been very fortunate to have shared my PhD experience with some wonderful fellow students: Becky, Ellen, Elise, Gareth, Lowri, Manu, John and Niamh (and many others I haven't mentioned) thanks for making the PhD so enjoyable. I have been incredibly lucky to live with the best "flatties" I could have ever hoped for during the last few years: thanks to Emma and Gareth.

Thanks to my family who have been incredibly supportive and who have always been at the end of the phone, and to my friends for ensuring I sometimes left the lab. Finally, I would like to thank James for looking after me and putting up with me even at my grumpiest moments.

Table of Contents

List of Figures

List of Tables

Abbreviations

Units

˚C	Degrees Centigrade
bp	Base pair
Da	Kilo Dalton
g	Gram
g	Gravity
l	Litre
m	Metre
M	Molar

Nucleotides

A	Adenine
C	Cytosine
G	Guanine
K	Guanine or thymine
N	Adenine, cytosine, guanine, or thymine
R	Adenine or guanine (purine)
S	Cytosine or guanine
T	Thymine
V	Adenine, cytosine or guanine
W	Adenine or thymine
Y	Cytosine or thymine (pyrimidine)

Amino acids

H	Histidine
L	Leucine
P	Proline
Q	Glutamine
R	Arginine
X	Stop codon

Diagnostic tools

APA	American Psychiatric Association
DSM	Diagnostic and Statistical Manual of Mental Disorders
ICD-10	International Classification of Diseases (10th edition)
SADS-L	Schedule for Affective Disorder and Schizophrenia - Lifetime version
SCID	Structured Clinical Interview for DSM-IV
SCID-I	Structured clinical interview for DSM-III-R

Neuropsychiatric disorders

ADHD	Attention deficit hyperactivity disorder
ASD	Autism spectrum disorder
BPD	Bipolar disorder
DVD	Developmental verbal dyspraxia
SCZ	Schizophrenia

Anatomical regions

CBM	Cerebellum
CNS	Central nervous system
CTX	Cortex
HPC	Hippocampus
PFC	Prefrontal cortex
STM	Striatum

Genes and proteins

AMPA	Alpha-amino-3-hydroxy-5-methyl-4- isoxazole-propionic acid
AP-1	Activator protein 1
cAMP	Cyclic adenosine monophosphate
CREB	cAMP response element-binding
DISC1	Disrupted in schizophrenia 1
EGF	Epidermal growth factor
FOX	Forkhead box
FOXP2	Forkhead box P2
GABA	Gamma-aminobutyric acid
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GSK3	Glycogen synthase kinase 3
HAT	Histone acetyltransferase
HDAC	Histone deacetyltransferase
HPRT1	Hypoxanthine-guanine phosphoribosyltransferase
HRGα	Heregulin alpha
nAChR	Nicotinic acetylcholine receptor
NMDA	N-methyl-D-aspartic acid
NRG1	Neuregulin 1
NRXN	Neurexin
PI3K	Phosphoinositide-3-kinase
SDHA	Succinate dehydrogenase complex, subunit A
SMDF	Sensory and motor neuron-derived factor
SF3A1	Splicing factor 3 subunit 1
TAF	TBP-associated factor
TCF4	T-cell factor 4
TBP	TATA box binding protein
TFIIA-H	Transcription factor II A-H
TSNAX	Transilin-associated protein X
UBC	Ubiquitin C

Reagents and materials

dH_2O	Deionised water
ddH_2O	Double distilled water
DMEM	Dulbecco's Minimal Essential Medium
EDTA	Ethylenediaminetetraacetic acid
ENU	N-ethyl-N-nitrosourea
FBS	Foetal bovine serum
HRP	Horseradish peroxidase
LARII	Luciferase assay reagent II
LB	Lysogeny broth
PBS	Phosphate buffered saline
PBS-T	Phosphate buffered saline-Tween
PVDF	Polyvinyl difluoride
SAP	Shrimp alkaline phosphatase
SDS	Sodium dodecyl sulphate
SOC	Super optimal broth with catabolite repression
TBE	Tris-borate-EDTA

Techniques

5'RACE	5' rapid amplification of cDNA ends
ChIP	Chromatin immunoprecipitation
DLR	Dual luciferase reporter
PCR	Polymerase chain reaction
PET	Positron emission tomography
qRT-PCR	Quantitative reverse transcription-polymerase chain reaction
RT-PCR	Reverse transcription-polymerase chain reaction
SPECT	Single-photon emission computerised tomography

Statistical

ANOVA	Analysis of variance
CI	Confidence Intervals
EM	Expectation-maximisation
HSD	Honestly significant difference
OR	Odds Ratio
SEM	Standard error of the mean
p_{ew}	*p*-value corrected for multiple testing at the experiment-wise level
*p*g	*p*-value for the global test of haplotype significance
*p*i	*p*-value for the individual test of haplotype significance
p_{st}	*p*-value corrected for multiple testing at the single-test level
SPSS	Statistical Package for the Social Sciences

Other

BAC	Bacterial artificial chromosome
BasicAs	Basic Association Sudy
BRE	Transcription factor II B recognition element
CDCV	Common disease common variant
cDNA	Complimentary DNA
CEPH	Centre d'Étude du Polymorphisme Humaine
CEU	CEPH trios of Utah residents with ancestry from northern and western Europe
CNV	Copy number variant
CO_2	Carbon dioxide
CRD	Cysteine rich domain
Ct	Threshold cycle
DNA	Deoxyribonucleic acid
DHS	DNase 1 hypersensitivity site
DPE	Downstream promoter element
E	Embryonic
ENCODE	Encyclopaedia or DNA elements
ERP	Event related potential
ESPERR	Evolutionary and sequence pattern extraction through reduced

	representation
GOI	Gene of interest
GWAS	Genome-wide association study
H3K4me3	Trimethylation of lysine 4 on histone H3
H3K27ac	Acetylation of lysine 27 on histone H3
H3K27me3	Trimethylation of lysine 27 on histone H3
HWE	Hardy-Weinberg Equilibrium
Inr	Initiator
KO	Knock-out
LD	Linkage disequilibrium
LI	Latent inhibition
LOD	Logarithm of odds
LTD	Long-term depression
LTP	Long-term potentiation
MTE	Motif ten element
MRI	Magnetic resonance imaging
miRNA	Micro RNA
mRNA	Messenger RNA
NFR	Nucleosome free region
OD	Optical density
P	Postnatal
PPI	Pre-pulse inhibition
PWM	Positional weight matrix
RNA	Ribonucleic acid
SNP	Single nucleotide polymorphism
TFBS	Transcription factor binding site
TM	Transmembrane domain
TSS	Transcription start site
UCSC	University of California, Santa Cruz
WMH	White matter hyperintensities
WHO	World Health Organisation
WTCRF	Wellcome Trust Clinical Research Facility

List of solutions and buffers

L-Agar

50 g	Tryptone
25g	Yeast extract
50g	Sodium chloride

Make up to five litres with dH_2O, adjust pH to 7.2. Pour into a bottle and add 75g Agar.

LB-Broth

50 g	Tryptone
25g	Yeast extract
25g	Sodium chloride

Orange G loading dye

1g	Ficoll 400 (Sigma)
0.2 ml	50mM EDTA (Fisher)
10 ml	ddH_2O
To colour	Orange G

Phosphate buffered saline (PBS), pH 7.3-7.5

From tablets (Invitrogen), containing:

10mM	Phosphate
150nM	Sodium chloride
500ml	dH_2O

Ponceau S stain

1g	Ponceau (Fisher)
4 ml	Acetic acid (Fisher)

Make up to 200ml with dH_2O

Protein extraction buffer

1% v/v	Triton-X
25% v/v	Glycerol
1 x	PBS, pH 7.3-7.5

A protease inhibitor tablet (Roche) was added at a concentration of 1 tablet per 50ml solution

Protein gel electrophoresis running buffer

50ml	20 x Novex Tris-acetate SDS running buffer (Invitrogen)
950ml	dH_2O

Protein gel electrophoresis transfer buffer

50ml	20 x NuPAGE transfer buffer (Invitrogen)
100ml	Methanol (Fisher)
850ml	dH_2O

Protein sample loading buffer (5x)

0.16M	Tris-chloride, pH 6.8
5% v/v	SDS
25% v/v	Glycerol
6% v/v	β-mercaptoethanol
0.02% w/v	Bromophenol blue

TBE buffer (5x)

54g	Tris (hydroxymethyl) aminomethane (Sigma)
27.5g	Boric acid (Sigma)
20ml	0.5M EDTA (Sigma)

Make up to one litre with dH_2O

Abstract

The genetic contribution to the aetiology of psychiatric illness is well-established; however, few variants that alter the encoded protein have been irrefutably identified as causative, leading to the hypothesis that variants affecting gene regulation may play a pathogenic role. This thesis focuses on two genes, *Neuregulin 1* (*NRG1*) and *Disrupted in Schizophrenia 1* (*DISC1*), for which there is strong genetic evidence for involvement in psychiatric illness, as well as evidence for altered expression in patients.

Association analysis was carried out to assess the involvement of six intronic *NRG1* single nucleotide polymorphisms (SNPs) in schizophrenia and bipolar disorder in two independent samples from the Scottish (Scottish 2; n = 307 control subjects, 303 schizophrenic patients, and 239 bipolar disorder patients and German populations (n = 397 control subjects, 396 schizophrenic patients, and 400 bipolar disorder patients). These SNPs form two haplotypes, one encompassing the 5' and promoter region of the gene and the other located at the 3' end of the gene, that were previously associated with schizophrenia and bipolar disorder in a Scottish sample (Scottish 1). The location of these haplotypes, together with the prior evidence for altered *NRG1* expression in schizophrenia, suggested the potential involvement of regulatory variants. On combining the Scottish 1 and Scottish 2 samples (combined n = 765 control subjects, 682 schizophrenic patients and 601 bipolar disorder patients), a two-SNP haplotype spanning both coding and non-coding regions in the 3' region was associated with schizophrenia (p = 0.0037, OR=1.3, 95% CI: 1.1-1.6) and the combined schizophrenia and bipolar disorder case group (p = 0.0080, OR=1.2, 95% CI: 1.1-1.5), with both these associations remaining significant after permutation analysis (p = 0.022 and p = 0.044, respectively).

To further understanding of how *DISC1*, a leading candidate gene for schizophrenia that has also been implicated in other psychiatric disorders, is regulated the previously uncharacterised promoter region was assessed both bioinformatically and *in vitro* using the dual luciferase reporter assay. The region was found to lack canonical promoter motifs but to contain a CpG island, consistent with *DISC1*'s ubiquitous pattern of expression. A region located 300bp to -177bp relative to the transcription start site (TSS) was identified as contributing positively to *DISC1* promoter activity, whilst a region -982bp to -301bp relative to the TSS was found to confer a repressive effect. FOXP2, a transcription factor which is mutated in a rare speech and language disorder and implicated in autism pathogenesis, was found to repress transcription from the *DISC1* promoter. Two pathogenic FOXP2 point

mutations reduced this transcriptional repression. Preliminary evidence for a bi-directional regulatory relationship between DISC1 and FOXP2 was observed: a mouse model of schizophrenia that carries a *Disc1* L100P amino acid substitution and shows altered developmental *Disc1* expression was also found to show altered developmental expression of *Foxp2*.

These results further understanding of two genes whose altered expression might contribute to the pathogenesis of psychiatric illness.

Chapter 1

Introduction

Chapter 1: Introduction

1.1 Overview of psychiatric illness

Millions of people worldwide are affected by psychiatric illness: in their 2001 report, the World Health Organisation (WHO, 2001) reported neuropsychiatric conditions to have a point prevalence of 10% in adults[1]. When considering 12 month prevalence, the percentage affected rises to approximately 38% in Europe[2] and 30% in the USA[3] (Kessler et al., 1994; Wittchen et al., 2011). A recent report on the financial consequences of brain disorders concluded that these conditions cost Europe almost €800 billion each year (Gustavsson et al., 2011). Mood and psychotic disorders alone were estimated to account for €209.4 billion of this expenditure. For comparison, Europe's expenditure on cancer is thought to lie in the range €150 - €250 billion (Gustavsson et al., 2011). This thesis will focus on two of the most severe psychiatric conditions: schizophrenia and bipolar disorder.

Schizophrenia is characterised by a constellation of positive, negative, and cognitive symptoms (van Os and Kapur, 2009). Positive symptoms are symptoms that are not usually experienced by individuals without the disorder. These include hallucinations, delusions, thought disorder, and disorders of movement. Negative symptoms, on the other hand, represent a loss of normal function and include social withdrawal, blunted affect, poverty of speech, and a lack of motivation. Abnormalities in cognitive function, which include deficits in working memory, long-term memory, and executive function, are considered a core component of the disorder as they are present before the onset of the positive and negative symptoms, before the onset of antipsychotic treatment, and are also seen in the unaffected relatives of patients (Barch and Ceaser, 2012; Callicott et al., 2003; Saykin et al., 1994; Snitz et al., 2006). Furthermore, deficits in cognitive function have been shown to be a good predictor of functional outcome (Green et al., 2000; Silver et al., 2003). Given the centrality of cognitive dysfunction to the schizophrenia phenotype, it is surprising that the Diagnostic

[1] The conditions considered were unipolar depression, bipolar disorder, and schizophrenia, epilepsy, alcohol and selected drug use disorders, Alzheimer's and other dementias, post-traumatic stress disorder, obsessive and compulsive disorder, panic disorder, and primary insomnia.
[2] The conditions considered were alcohol and drug dependence, psychotic disorder, major depressive disorder, bipolar disorder, anxiety disorders, somatoform disorders, eating disorders, personality disorders, childhood and adolescent disorders, mental retardation, sleep disorders, and dementia.
[3] The conditions considered were major depressive disorder, mania, dsythymia, panic disorder, social phobia, agoraphobia, generalised anxiety disorder, alcohol and drug dependence, antisocial personality disorder, and nonaffective psychosis, which comprises schizophrenia, schizophreniform disorder, schizoaffective disorder, delusional disorder and atypical psychosis.

and Statistical Manual of Mental Disorder, Fourth Edition (DSM-IV) (APA, 1994) does not include cognitive deficits amongst the criteria for schizophrenia. It has been suggested that the inclusion of cognitive impairment in the diagnostic criteria for schizophrenia in the forthcoming DSM-V would improve prognosis and treatment outcomes (Keefe and Fenton, 2007).

Several subtypes of schizophrenia are recognised by the DSM-IV, including paranoid type, disorganised type, and catatonic type. It has been argued that these subtypes fail to capture the true heterogeneity of schizophrenia and that they have low diagnostic stability; it has, therefore been proposed that schizophrenia subtypes should not be included in the DSM-V (American Psychiatric Association (APA) DSM-V Development http://www.dsm5.org/ProposedRevisions/Pages/proposedrevision.aspx?rid=411#). Instead, it is proposed that the DSM-V should use psychopathological dimensions (American Psychiatric Association DSM-V Development); this suggestion is consistent with the findings of (Cuesta et al., 2007) who found no evidence for taxa within psychosis based on the severity of positive, negative, and disorganisation symptoms.

Patients with bipolar affective disorder experience repeated episodes of mania and depression. Symptoms of mania include elated mood, without an obvious cause, rapid speech, increased energy, and disordered thoughts. Depressive episodes are characterised by feelings of hopelessness, low self-esteem, decreased energy, and insomnia. In common with schizophrenia, the DSM-IV recognises different subtypes of bipolar disorder. These include bipolar I disorder, characterised by at least one manic or mixed (comprising both manic and major depressive symptoms) episode, bipolar II disorder, which includes at least one major depressive episode and at least one hypomanic episode, and cyclothymia, a milder form of bipolar disorder, which is characterised by a fluctuating mood that comprises episodes of hypomania and episodes of depression that do not meet the diagnostic criteria for a major depressive episode.

The precise estimation of the lifetime prevalence of these disorders is complicated by variation between samples, which is, at least in part, attributable to the heterogeneity of the disorders. From a review of studies carried out between 1965 and 2002, Saha et al. (2005) reported the lifetime prevalence of schizophrenia to be between 0.3 and 2%, with an average of approximately 0.7%. The majority of patients are diagnosed with schizophrenia in their late adolescence or early twenties, and there are several reports of earlier onset in males than

females (Angermeyer and Kuhn, 1988), although a recent meta-analysis has revealed this gender difference to be only about 1.5 years, smaller than previously believed (Eranti et al., 2012). There is a lack of consensus regarding the proportion of males to females affected by the disorder: initial studies concluded that the risk of developing schizophrenia was equal for males and females (Wyatt et al., 1988), while subsequent studies have indicated a higher prevalence of the disorder amongst males (Aleman et al., 2003; McGrath, 2005). Bipolar disorder occurs with similar lifetime prevalence to schizophrenia, with recent findings from the World Mental Health Survey Initiative indicating that bipolar I disorder has a lifetime prevalence of 0.6% and bipolar II disorder has a lifetime prevalence of 0.4% (Merikangas et al., 2011). Gender differences have been reported in the presentation of bipolar disorder: while approximately equal numbers of men and women are affected by the disorder (Gold, 1998), women have been reported as experiencing depression, mixed episodes, and rapid cycling more often than men, as well as having a later age of onset (Robb et al., 1998).

The accuracy of estimates of the prevalence of both schizophrenia and bipolar disorder is limited by the lack of definitive biomarkers for psychiatric illness. Moreover, as the symptoms of schizophrenia and bipolar disorder are heterogeneous and overlapping, debate exists over the extent to which these disorders should be considered separate entities. The distinction of schizophrenia and bipolar disorder as separate clinical conditions can be traced back to Emil Kraepelin (1919) who described two conditions, crystallised dementia praecox, now referred to as schizophrenia, and manic depressive illness, now known as bipolar disorder. Craddock and Owen (2005; 2007) question the validity of this traditional diagnostic distinction by highlighting several genes implicated in both disorders. They predict that advances in genetic epidemiology mark "The beginning of the end for the Kraepelinian dichotomy". Crow (2008), in contrast, highlights the fact that the beginning of the downfall of the Kraepelinian dichotomy started with Kraepelin himself, who stated that "It is becoming increasingly clear that we cannot distinguish satisfactorily between these two illnesses and this brings home the suspicion that our formulation of the problem may be incorrect.", (Kraepelin, 1920).

1.2 Evidence for a genetic basis to schizophrenia and bipolar disorder

A genetic contribution to the aetiology of psychiatric illness has been demonstrated by family, twin and adoption studies (Cardno and Gottesman, 2000; Ingraham and Kety, 2000; Lichtenstein et al., 2009; McGuffin, 2004; Sullivan et al., 2003; Sullivan et al., 2000). For

schizophrenia, the concordance rate between monozygotic twins is higher than between dizygotic twins: in a review of the literature, Cardno and Gottesman (2000) report concordance rates of 41-65% in monozygotic twins and concordance rates of 0-28% in dizygotic twins. The heritability of schizophrenia was estimated as being approximately 80-85% (Cardno and Gottesman, 2000). First degree relatives of patients with schizophrenia have an increased risk of developing schizophrenia of approximately 10-fold compared to other members of the population (Lichtenstein et al., 2009), and adoption studies have found that adopted children's risk of developing schizophrenia is related to the affected status of their biological but not adoptive parents (Ingraham and Kety, 2000).

Compared with schizophrenia, fewer studies have been carried out to ascertain the genetic contribution to bipolar disorder; however, those that have been performed suggest a strong genetic basis. In a twin study in which the diagnostic groups of bipolar I disorder and bipolar II disorder were considered together, McGuffin et al. (2003) reported concordance rates of 40% for monozygotic twins and 5.4% for dizygotic twins, which resulted in a heritability estimate of 85%. Similar concordance rates were observed in a study in which only bipolar I disorder was considered: in this study, the concordance rate for monozygotic twins was 43% for monozygotic twins and 6% for dizygotic twins, resulting in a heritability estimate of 93% (Kieseppa et al., 2004). Consistent with the genetic contribution suggested by twin studies, the first degree relatives of patients with bipolar disorder have an elevated risk of bipolar disorder of approximately seven-fold compared to the general population (Lichtenstein et al., 2009). Only two adoption studies have been carried out in which the modern definition of bipolar disorder was used: both of these studies found evidence suggestive of a genetic basis to the disorder; however, for both studies, the sample size was relatively small (Mendlewicz and Rainer, 1977; Wender et al., 1986). Moreover, both studies used definitions of illness that extended beyond bipolar disorder: Mendlewicz and Rainer (1977) identified parents as "ill" when they were diagnosed with bipolar disorder, unipolar disorder, schizoaffective disorder, or cyclothymia, and Wender et al. (1986) considered both probands and parents as affected when they were diagnosed with either bipolar disorder or unipolar disorder.

The use of broad diagnostic categories by both Mendlewicz and Rainer (1977) and Wender et al. (1986) hints at an observation that has important consequences when considering the genetic basis of psychiatric illness. This observation is that psychiatric illnesses do not "breed true": close relatives of individuals diagnosed with one disorder are themselves not only at greater risk of developing the same disorder but other psychiatric disorders too.

Relatives of patients with bipolar disorder are at an elevated risk of unipolar disorder compared to relatives of control subjects (Merikangas et al., 2002), and in a twin study in which the probands were diagnosed with bipolar disorder, McGuffin et al. (2003) found that the inclusion of co-twins with either bipolar disorder or unipolar disorder resulted in an increase in the heritability estimate compared to when co-twins with only bipolar disorder were considered.

While the co-aggregation of affective disorders in families has been recognised for several years, the evidence for the co-aggregation of schizophrenia and bipolar disorder has been more contentious. This debate has clear implications for the diagnostic distinction between schizophrenia and bipolar disorder. Early studies suggesting that schizophrenia and bipolar disorder do breed true supported their diagnostic distinction (Berrettini, 2003); however a recent study, the largest family study of schizophrenia and bipolar disorder to date, has challenged this assertion. Lichtenstein et al. (2009) found the first degree relatives of probands diagnosed with either schizophrenia or bipolar disorder to have an increased risk of developing either disorder. This observation is in keeping with Cardno et al.'s (2002) finding from the study of twins diagnosed with schizophrenia, schizoaffective disorder, or manic disorder that the genetic contribution to the three disorders is correlated and that there are both disorder-specific and common genetic contributions. The existence of a shared genetic component is further corroborated by the findings of the International Schizophrenia Consortium who have demonstrated that polygenic risk alleles identified in a schizophrenic sample are associated with risk for bipolar disorder in two independent samples (Purcell et al., 2009).

1.3 The role of non-genetic factors in the aetiology of schizophrenia and bipolar disorder

The observations that (i) monozygotic twins show less than perfect concordance for schizophrenia and bipolar disorder and (ii) concordance rates for dizygotic twins are less than half the monozygotic rate hint at the involvement of additional non-genetic influences. Furthermore, the risks to offspring of both the affected and unaffected twins from a pair of monozygotic twins who are discordant for schizophrenia are almost equally elevated (Kringlen and Cramer, 1989). Epidemiological studies have identified several environmental factors conferring risk for schizophrenia and bipolar disorder. For schizophrenia, prenatal stress, such as bereavement or famine, obstetric complications, infection, winter births, and

being born into an urban environment have all been shown to increase the risk of developing the disorder (Sullivan, 2005). There is strong evidence that gestational exposure to infection increases risk for schizophrenia. Longitudinal study of a birth cohort has found an increased incidence of schizophrenia amongst the offspring of mothers who had experienced the influenza virus, herpes simplex virus type 2, or had elevated levels of toxoplasma gondi during pregnancy (Babulas et al., 2006; Brown et al., 2004). Consistent with these findings, mothers whose offspring developed schizophrenia were found to have elevated levels of the pro-inflammatory cytokine interleukin-8 during the second and early third trimester of pregnancy (Brown et al., 2005). The role of maternal infection is supported by the study of rodent models: prenatal immune challenge is associated with cognitive, behavioural and pharmacological abnormalities relevant to schizophrenia in the adult offspring (Meyer and Feldon, 2009). Moreover, mice expressing a dominant negative form of human *Disrupted in Schizophrenia 1*(*DISC1*), a leading candidate gene for schizophrenia, who are injected neonatally with poly I:C, which promotes the production of pro-inflammatory cytokines, show synergistic deficits in memory function (Ibi et al., 2010), which are characteristic of the cognitive defects observed in schizophrenia. Additionally, mice expressing a mutant truncated form of human *DISC1* who were subjected to prenatal exposure to poly I:C, displayed a postnatal phenotype that recapitulated many features of psychiatric disorders including depressive-like behaviour, anxiety, abnormal social interaction, attenuated serotonin neurotransmission, decreased amygdala volume, reduced granule cell dendritic spine density in the hippocampus and changes in gene expression (Abazyan et al., 2011).

Environmental variables are also known to influence the onset of bipolar disorder. Several studies have reported an increased incidence of stressful life events prior to the onset of individual episodes. (Alloy et al., 2005) and Etain et al. (2008) concluded from a review of the literature that childhood trauma is a risk factor for the onset and clinical expression of bipolar disorder. Investigation of the role of maternal infection in the pathogenesis of bipolar disorder has been less extensive than for schizophrenia; however, based on a limited number of studies, it appears that maternal infection does not confer risk for bipolar disorder (Mortensen et al., 2011; Mortensen et al., 2003; Stober et al., 1997).

One mechanism by which environmental influences might exert a pathogenic effect is via an effect at the epigenetic level. Epigenetic mechanisms that can effect gene expression have been found to be influenced by environmental factors such as diet, drugs and stress (Oh and Petronis, 2008). In a genome-wide analysis of blood DNA methylation in monozygotic twins

discordant for either schizophrenia or bipolar disorder, several loci were found to be differentially methylated between affected and unaffected twin pairs (Dempster et al., 2011). Moreover, pathway analysis of the dysregulated genes revealed enrichment for pathways previously implicated in psychiatric illness and neurodevelopment, thus highlighting the potential of this approach to extend our understanding of pathogenic mechanisms.

It is important to note that epigenetic differences between monozygotic twins do not necessarily reflect the actions of environmental influences: it is possible that they reflect the influence of stochastic variables. It has been suggested that differences in gene expression levels, which are known to randomly fluctuate (Kaern et al., 2005), might be crystallised during embryogenesis by epigenetic chromatin modifications and clonally inherited (Mitchell, 2007). Furthermore, in addition to the epigenetically mediated influence of random variation, it is possible that stochastic events might also impact on other processes involved in brain development and thus risk for schizophrenia (Mitchell, 2007; Woolf, 1997).

1.4 Pathological hallmarks of schizophrenia and bipolar disorder

Whilst no physiological or pharmacological hallmark of schizophrenia or bipolar disorder occurs with the required specificity to qualify as a diagnostic marker, several abnormalities have been replicated in enough studies to provide useful insights into the aetiology of these disorders. Several abnormalities are common to both conditions, while others are more specific in their appearance. Taken together, the identified abnormalities suggest a strong neurodevelopmental basis to schizophrenia (Fatemi and Folsom, 2009) and the involvement of neurodevelopmental factors in bipolar disorder (Sanches et al., 2008).

1.4.1 Structural abnormalities

Brain imaging has permitted huge advances in our understanding of the state of the brain in schizophrenia and bipolar disorder in the living patient, avoiding many confounds associated with the study of post-mortem brains. Magnetic resonance imaging (MRI) studies of both conditions have been numerous and not always unanimous in their findings. Meta-analyses have, however, supported the existence of several structural abnormalities. In schizophrenia, meta-analyses have supported the existence of volumetric reductions in the hippocampus, thalamus, amygdala, and anterior cingulate cortex and a volumetric increase in the size of the

ventricles, particularly the lateral ventricles (Baiano et al., 2007; Baiano et al., 2008; Konick and Friedman, 2001; Nelson et al., 1998; Wright et al., 2000). A reduction in the area of the corpus callosum has also been identified (Arnone et al., 2008). A meta-analysis of bipolar disorder MRI studies revealed reductions in whole brain volume and an increase in the volume of the lateral ventricles and globus pallidus (Arnone et al., 2009). An increased incidence of white matter hyperintensities (WMHs), areas of high intensity visible on certain types of MRI scan, is a well-established finding in patients with bipolar disorder (Mahon et al., 2010). The relevance of WMHs to the pathophysiology of bipolar disorder remains unknown: WMHs can be caused by ischemia, gliosis, edema, and demyelination, and it is unclear which cause contributes to the presence of WMHs in bipolar disorder (Mahon et al., 2010). It has been suggested that WMHs disrupt brain tissue and thus neuronal connectivity (Schloesser et al., 2008). Further supporting the role of aberrant connectivity in both bipolar disorder and schizophrenia, abnormalities in white matter, which consists mostly of myelinated axons and glial cells, have been detected in both conditions (Heng et al., 2010; Thomason and Thompson, 2011). As the myelination of axons is necessary for the efficient transmission of action potentials, white matter abnormalities suggest reduced connectivity. Indeed, Harrison and Weinberger (2005) have termed schizophrenia a "disorder of connectivity". Importantly, some of these changes have been detected in first-episode patients and those genetically at high-risk of developing schizophrenia or bipolar disorder, indicating that they are not simply a result of drug treatment of disease progression (Jung et al., 2010; Sprooten et al., 2011a; Vita et al., 2009; Vita et al., 2006).

1.4.2 Abnormalities found in the post-mortem brain

Post-mortem studies offer the advantage over imaging studies of permitting a more detailed examination of abnormalities at the cellular level; however, they are, of course, subject to the possible confounds of long-term medication, circumstances surrounding death, and post-mortem factors. At a macroscopic level, a decrease in brain weight has been observed in several studies of schizophrenic patients (Brown et al., 1986; Bruton et al., 1990; Pakkenberg, 1987) consistent with the volumetric reductions reported by MRI studies.

Post-mortem studies also support the notion that white matter abnormalities contribute to the pathogenesis of schizophrenia and bipolar disorder. In the post-mortem brains of patients with schizophrenia, several studies have found subcortical white matter neurons to be increased in density and spatial distribution (Connor et al., 2011). An increase in white

matter neuron density has been reported in bipolar disorder too (Connor et al., 2009). This increase in density may result, in part, from altered neuronal migration: several studies have identified aberrantly positioned neurons in schizophrenia (Arnold et al., 1991; Falkai et al., 2000; Jakob and Beckmann, 1994; Kovalenko et al., 2003). Several other cytoarchitectural abnormalities have been identified in both schizophrenia and bipolar disorder. In the prefrontal cortex (PFC) of schizophrenic patients, pyramidal neurons have been found to be more densely packed, and it is thought that this difference arises from a reduction in neuropil and soma size (Rajkowska et al., 1998; Selemon and Goldman-Rakic, 1999). Abnormalities in GABAergic interneurons have also been detected in both schizophrenia and bipolar disorder (Benes and Berretta, 2001; Fatemi et al., 2005; Marin, 2012; Wang et al., 2011a; Woo et al., 2004). In both schizophrenia and bipolar disorder, a reduction in the expression of GAD67, the rate limiting enzyme in GABA synthesis has been observed (Akbarian et al., 1995; Guidotti et al., 2000; Hashimoto et al., 2003; Thompson Ray et al., 2011). GABAergic interneurons are involved in the inhibition of pyramidal neurons, which is necessary for the maintenance of normal gamma oscillations, the bursts of coordinated firing thought to play an important role in cognition (Bartos et al., 2007). Abnormal gamma oscillations have been detected in schizophrenic patients performing tests of working memory (Haenschel et al., 2009; Uhlhaas and Singer, 2010) and in patients with bipolar disorder during an implicit emotional task (Liu et al., 2012). Both schizophrenia and bipolar disorder are associated with aberrant myelination. The number of oligodendroglial cells, the glial cell type responsible for myelination in the central nervous system (CNS), is reduced in various brain regions in patients with schizophrenia or bipolar disorder (Uranova et al., 2004; Vostrikov et al., 2007), and down-regulation of myelin-related genes has been detected in both conditions (Hakak et al., 2001; Matthews et al., 2012; Tkachev et al., 2003). Deficits in myelination may contribute to the reduced connectivity mentioned in section 1.4.1.

1.4.3 Abnormalities in neurotransmitter systems

Multiple lines of evidence suggest that aberrant neurotransmission plays a fundamental role in the pathogenesis of both schizophrenia and bipolar disorder. In addition to the GABAergic deficits described in section 1.4.2, abnormalities have also been detected in glutamate, dopamine, and acetylcholine function, leading to difficulties in determining the primary pathology. As might be expected, given the observed abnormalities in neurotransmission, deficits in synaptic plasticity have been observed in both schizophrenia and bipolar disorder.

The dopamine hypothesis was one of the first theories of the aetiology of schizophrenia. In its initial incarnation, the dopamine hypothesis proposed hyperdopaminergia as playing a causal role in schizophrenia (Howes and Kapur, 2009). This conclusion was based on the discoveries that (i) the clinical effectiveness of antipsychotics equated with their affinity for dopamine D2 receptors where they exert an antagonistic effect and (ii) the symptoms of schizophrenia could be induced or exacerbated by amphetamine, which increases synaptic concentrations of dopamine and other monoamines (Lieberman et al., 1987; Matthysse, 1973; Seeman and Lee, 1975; Snyder, 1976). The discovery that there is an increase in the density of striatal dopamine D2 receptors in schizophrenic patients appeared to support the hypothesis of hyperdopaminergia (Seeman, 1992); however, these studies were unable to rule out the contribution of long-term antipsychotic use. Interestingly, subsequent studies using positron emission tomography (PET) or single-photon emission computed tomography (SPECT) have demonstrated an increase in dopamine synthesis and amphetamine-induced dopamine release in drug-free patients with schizophrenia (Laruelle et al., 1996; Lindstrom et al., 1999). Moreover, it has been shown that patients with treatment-resistant schizophrenia do not show the same increase in dopamine synthesis observed in patients whose symptoms are ameliorated by antipsychotics, suggesting differences in the underlying pathology and/or effects of antipsychotic medication (Demjaha et al., 2012).

The original dopamine hypothesis was limited by its ability to account only for the positive symptoms of schizophrenia; it provided no explanation for either the negative or cognitive aspects of the disorder. To remedy this deficit, the hypothesis was extended such that it described a hyperactive mesolimbic dopamine system, underlying the positive symptoms, and a hypoactive mesocortical dopamine system, underlying the cognitive symptoms (Davis et al., 1991; Weinberger, 1987). Hypoactivity of the mesocortical dopamine system was suggested by the discovery that dopamine D1 receptor availability is increased in schizophrenic patients, with this increase correlating positively with working memory impairments (Abi-Dargham et al., 2002).

Dopamine dysregulation has also been proposed to play a role in bipolar disorder (Berk et al., 2007). It is thought that during manic episodes there is an excess of dopaminergic signaling, while episodes of depression are characterised by reduced dopamine signaling. In support of this hypothesis, episodes of hypomania can be induced in patients with bipolar disorder by administering the dopamine precursor L-Dopa (Murphy et al., 1971) and dopamine agonists have been shown to be effective in alleviating depression in patients with

bipolar disorder (Berk et al., 2007). In common with schizophrenia, an increase in dopamine D2 receptor density has been reported in bipolar disorder patients, with increased density correlating positively with the severity of psychotic symptoms (Pearlson et al., 1995).

Subsequent to the dopamine hypothesis, a glutamate hypothesis of schizophrenia, which proposes glutamate hypofunction as a pathogenic mechanism, was proposed (Olney and Farber, 1995). Suggestive evidence for glutamate hypofunction came from the observation that the concentration of glutamate in the cerebrospinal fluid of patients with schizophrenia is reduced (Kim et al., 1980) and further support came from the finding that low doses of NMDA receptor antagonists, such as ketamine and phencyclidine, can mimic the symptoms of schizophrenia when administered to healthy individuals (Adler et al., 1999; Javitt and Zukin, 1991; Krystal et al., 1994). Importantly, the symptoms elicited by these drugs are reminiscent of the negative and cognitive symptoms of schizophrenia, as well as the positive symptoms, thus suggesting that glutamatergic hypofunction might have greater explanatory power than the dopamine hypothesis. Studies of glutamate receptor expression in the post-mortem brains of schizophrenic patients have provided support for the notion of glutamatergic hypofunction with a decrease in the expression of certain subclasses of glutamate receptor being reported by several studies (Meador-Woodruff and Healy, 2000). It is important to note that the dopamine and glutamate hypotheses of schizophrenia do not represent competing theories of schizophrenia, evidence for dysfunction in both neurotransmitter systems can be reconciled by circuit-based models. Lisman et al. (2008) describe a model whereby glutamatergic hypofunction results in a reduction in the activity of fast-spiking interneurons resulting in the disinhibition of pyramidal cells, which can produce hyperdopaminergia, particularly in the hippocampus. Moreover, glutamatergic neurons synapse directly onto mesocortical dopaminergic neurons in the ventral tegmental area, regulating their activity such that glutamatergic hypofunction would result in a decrease in the activity of mesocortical dopamine neurons (Takahata and Moghaddam, 2000; Westerink et al., 1998).

Abnormal glutamate function also appears to play a role in bipolar disorder. There is evidence for an increase in glutamatergic neurotransmission in patients with bipolar disorder (Eastwood and Harrison, 2010) and Lamotrigine, which reduces cortical glutamate, has been found to have mood-stabilising and antidepressant effects (Muzina et al., 2005). Similarly, the mood stabiliser lithium has been found to protect neurons in primary cultures from glutamate-induced excitotoxicity (Chuang et al., 2002). Perhaps paradoxically, decreased

expression of certain glutamatergic receptors has been detected in the post-mortem brains of patients with bipolar disorder (Beneyto et al., 2007; Nudmamud-Thanoi and Reynolds, 2004; Scarr et al., 2003); however, it is possible that this decrease represents a compensatory mechanism or medication effect.

Recently, interest has grown in the contribution of cholinergic dysfunction to schizophrenia. Compared to the general population, there is an increased incidence of cigarette smoking amongst patients with schizophrenia and it has been suggested that this might represent an attempt to compensate for deficits in nicotinic acetylcholine function (Aubin et al., 2012). A decrease in acetylcholine receptor expression has been reported in schizophrenic patients (Deng and Huang, 2005; Perl et al., 2003; Scarr et al., 2009) and the expression of nicotinic acetylcholine receptors (nAChRs) has been shown to be differentially modulated by smoking in schizophrenic smokers (Mexal et al., 2010). Interestingly, agonists of α7 nAChRs have been shown to alleviate cognitive deficits associated with schizophrenia and reverse the effects of phencyclidine (AhnAllen, 2012). This finding is of great interest given the lack of efficacy of many currently available antipsychotics in the amelioration of cognitive dysfunction (Money et al., 2010). Cholinergic abnormalities are also implicated in bipolar disorder: hippocampal α7 nAChRs have been found to be decreased in the post-mortem brains of patients with bipolar disorder (Thomsen et al., 2011). In common with schizophrenia, the incidence of cigarette smoking is increased amongst individuals with bipolar disorder (Heffner et al., 2011).

Other neurotransmitter systems identified as functioning abnormally in schizophrenia and bipolar disorder include the serotonergic system. The atypical antipsychotics clozapine and risperidone are believed to act, in part, via their antagonistic effects on serotonin receptors (Kapur and Remington, 1996) and selective serotonin reuptake inhibitors are often used to treat the depressive symptoms of bipolar disorder (Shastry, 2005). Furthermore, serotonin receptor expression has been reported to be abnormal in both schizophrenic and bipolar disorder patients (Castensson et al., 2003; Eastwood et al., 2001; Lopez-Figueroa et al., 2004).

Deficits in synaptic plasticity are believed to play a key role in both schizophrenia and bipolar disorder. Aberrant expression of several synapse-related genes has been detected for both disorders (Lopez de Lara et al., 2010; Maycox et al., 2009; Mirnics et al., 2000; Ryan et al., 2006), and mood stabilisers, antidepressants and antipsychotics are believed to exert their

therapeutic effects, at least in part, via an effect on synaptic function (D'Sa and Duman, 2002; Konradi and Heckers, 2001; Shaldubina et al., 2001). Moreover, abnormalities in synaptic plasticity have been detected in several animal models of schizophrenia and depression (Pittenger and Duman, 2008; Yin et al., 2011) and, in recent years, direct evidence has been obtained for abnormal plasticity in schizophrenic patients (Frantseva et al., 2008; Hasan et al., 2011).

Taken together, since the formulation of the dopamine and glutamate hypotheses of schizophrenia, subsequent evidence suggests a widespread disruption of normal neurotransmitter, and thus synaptic, function in both schizophrenia and bipolar disorder. Improved understanding of the inter-relationship between different neurotransmitter systems, and identification of relevant functional schizophrenia/bipolar disorder-associated genetic variants, will aid in establishing the primary pathology.

1.5 The genetic architecture of schizophrenia and bipolar disorder

Despite the clear evidence for a genetic contribution to the risk of developing schizophrenia or bipolar disorder, the underlying genetic architecture of the two conditions remains poorly understood. Two contrasting models have dominated the debate: the common disease-common variant (CDCV) model and the multiple rare variants model. The CDCV model proposes that, within an individual, multiple common variants in different genes, each of small effect size, confer susceptibility for the disorder (i.e. the disorder is polygenic). To apply the CDCV model to conditions which are binary in nature (i.e. an individual has the disorder or does not have it), a liability threshold has been invoked: the underlying distribution of liability for the disorder is assumed to be continuous but only those individuals who pass a minimum burden of common variants actually get the disease (Falconer, 1989). In contrast, the multiple rare variants model posits that the majority of the variance for complex diseases is due to the actions of rare variants (typically with a minor allele frequency of less than 1%). These variants may occur in different genes in different individuals, making the disorder genetically heterogeneous. A mixed model in which other loci in the genetic background act in a polygenic fashion together with environmental factors to modify the expressivity of rare alleles provides an alternative to either the CDCV or rare variant models interpreted in the most narrow of senses. Indeed, this model is congruent with experimental findings that have indicated that genetic background effects are common and can have a substantial effect on phenotype (Nadeau, 2001; Shao et al., 2008).

For various reasons the CDCV model has dominated the field of psychiatric genetics; however, in recent years, the premises of this hypothesis have been called into question and the plausibility of other hypotheses considered (McClellan et al., 2007; Mitchell, 2012; Mitchell and Porteous, 2011). An argument often used in favour of the polygenic CDCV model of schizophrenia and bipolar disorder is that neither disorder segregates in a manner consistent with classical Mendelian inheritance (Gottesman and Shields, 1967). The recurrence risk to relatives of schizophrenic patients decreases more quickly than would be expected if the disorder was caused by a single gene acting with a dominant mode of inheritance (Gottesman and Shields, 1967; Risch, 1990). The involvement of multiple common variants of small effect has also been used as an explanation for the fairly constant prevalence of schizophrenia over time despite the decrease in fertility associated with this disorder (Gottesman and Shields, 1967). Risch et al. (1990) have argued that should different rare alleles play a causal role in different families, one would expect the clinical presentation of the disorder to be distinct between families. Moreover, it has been argued that if these disorders were caused by rare variants of large effect, the variants involved would have been detected by genetic linkage studies (Risch, 1990). This latter argument can easily be refuted on the basis that many genetic linkage studies have been carried out across multiple unrelated families: risk alleles would have to be sufficiently common to contribute to disease pathogenesis in enough of the families studied to be detected using this approach.

Further doubt has been cast over the CDCV hypothesis of schizophrenia and bipolar disorder by recent findings that question many of the assumptions on which the acceptance of this hypothesis was based. One such assumption was that the rate of *de novo* mutations was too low to replenish highly penetrant deleterious alleles, which would be eliminated by purifying selection (Gottesman and Shields, 1967). A hint that this assumption was incorrect came from the observation that an increased incidence of both schizophrenia and bipolar disorder is associated with advanced paternal age (Frans et al., 2008; Matheson et al., 2011). Paternal age is associated with an increase in *de novo* germline mutations (Kong et al., 2012) and an increase in the incidence of hereditary disorders (Glaser and Jabs, 2004). The advent of whole-genome sequencing and its application to family pedigrees has facilitated direct assessment of the rate of spontaneous mutations and their contribution to disease. Based on the sequencing of a family quartet, (Roach et al., 2010) have estimated that there will be approximately 70 new mutations in each diploid genome. Following whole-genome sequencing in 78 parent-offspring trios, Kong et al. (2012) have predicted that *de novo*

mutations occur at a rate of roughly 2 per a year, when their data is fitted with a linear model, or double every 16.5 years under an exponential model. Under both models, paternal age explains at least 94% of the variation in the occurrence of *de novo* mutations. Furthermore, an increase in the rate of spontaneous mutations with predicted deleterious consequences has been identified in individuals with schizophrenia (Awadalla et al., 2010). Further supporting the role of spontaneous mutations in the pathogenesis in schizophrenia is the discovery that *de novo* copy number variants (CNVs) are associated with the disorder (Kirov et al., 2012; Malhotra et al., 2011; Xu et al., 2008). The involvement of CNVs in the pathogenesis of schizophrenia and bipolar disorder will be discussed further in section 1.6.2.2. *De novo* mutations may also represent a causal factor for bipolar disorder as the condition has been associated with an increase in somatic gene conversion and deletion (Ross, 2011), processes that occur when highly homologous but non-allelic regions of the genome exchange genetic information in a unidirectional fashion.

The involvement of *de novo* mutations in schizophrenia and bipolar disorder could also account for some instances of non-Mendelian inheritance. Furthermore, *de novo* pathogenic variants could, at least in part, explain the observed recurrence risks amongst relatives of schizophrenic patients. In fact, Mitchell and Porteous (2011) have demonstrated that the observed familial recurrence risks for schizophrenia are compatible with a mixed model in which varying proportions of cases of schizophrenia are attributable to dominant, recessive, or *de novo* modes of inheritance.

Risch's (1990) argument that distinct genetic causes of a disorder should be reflected by distinct phenotypic presentation is now outdated. Accumulating evidence highlights the fact that individual mutations are associated with a broad range of psychiatric conditions (reviewed in Mitchell and Porteous, 2011). An example of such a mutation, which will be discussed further in section 1.7.4.2, is a balanced translocation affecting the gene *Disrupted in Schizophrenia 1* (*DISC1*) (Millar et al., 2000b). This translocation was identified in a family with a high incidence of psychiatric illness; translocation carriers are variously affected by a wide range of psychiatric conditions including major depressive disorder, schizophrenia, and bipolar disorder (St Clair et al., 1990).

As alluded to previously, a mixed model, in which both common and rare variants confer susceptibility offers a compromise between the polarised CDCV and rare variant models. In a recent publication (Lee et al., 2012) used mathematical modelling to estimate the

proportion of variance in liability to schizophrenia attributable to SNPs of different minor allele frequencies. This analysis indicated that SNPs across the range of minor allele frequencies considered explained some proportion of the total variance. It is important to note that the genotypes of the SNPs in this study were ascertained by whole-genome genotyping chips and, therefore, rare variants would have been under-represented in this analysis. Indeed, a model that incorporated all SNPs (both genotyped and imputed) could only account for 23% of the variance in liability to schizophrenia. Nevertheless, these findings highlight the need to consider variants spanning the entire spectrum of allelic frequencies.

Taken together, the evidence calls for a re-evaluation of our understanding of the genetic architecture of schizophrenia and bipolar disorder. Ascertaining the genetic architecture of schizophrenia and bipolar disorder is more than a matter of academic debate: several methods are available for the detection of disease-associated variables, with the choice of method depending on the nature of the variant to be detected. These methods will be introduced in section 1.6. Until the underlying architecture of a condition is understood, it has been argued that "the genetic variants related to human disease that have been identified to date primarily reflect the methods used to detect them" (Campbell and Manolio, 2007).

1.6 Methods for the identification of genes involved in schizophrenia and bipolar disorder

Several methods for the identification of genetic variants involved in hereditary disorders exist. These will be introduced below, together with some examples of schizophrenia and bipolar disorder candidate genes that have been identified through their application.

1.6.1 Linkage analysis

Linkage analysis involves the comparison of closely related individuals to identify regions of the genome that segregate with a given phenotype. As genetic loci that are located in close proximity are less likely to be separated during meiotic recombination, they tend to be inherited together more often than distally located loci. Thus, by studying the segregation of genotyped markers with the phenotype of interest, the genomic region harbouring the causal variant can be narrowed down. The resolution of linkage analysis is usually fairly low: it is able to identify phenotype-associated chromosomal regions, which may span several megabases of DNA. Linkage analysis is most applicable for the detection of genomic regions

containing causal variants that are highly penetrant and of large effect; the method lacks power when applied to conditions resulting from the actions of many genes of small effect.

1.6.1.1 Schizophrenia

Several linkage studies have been carried out for both schizophrenia and bipolar disorder, which have between them implicated much of the genome. In a review of the literature, Harrison and Weinberger (2005) highlight four regions (6p22-24, 8p21-22, 1q21-22, 10q25.3-26.3) as having attained genome-wide significance for linkage to schizophrenia but point out that each of these regions has been studied in other studies in which significant linkage was not observed. In a meta-analysis of 32 independent genome-wide linkage studies of schizophrenia, Ng et al. (2009) detected genome-wide linkage to chromosome 2q and suggestive evidence for linkage to chromosome 5q and, in European subjects only, chromosome 8p. Interestingly, *Neuregulin 1* (*NRG1*), a well-supported schizophrenia candidate gene (section 1.7.4.1), is located within the linked region on chromosome 8p. Another gene in this region, *PPP3CC*, which encodes the calcineurin gamma subunit, has also been implicated in schizophrenia by association studies (Gerber et al., 2003; Horiuchi et al., 2007; Kyogoku et al., 2011; Liu et al., 2007) and the finding of altered expression in the brains of schizophrenic patients (Eastwood et al., 2005). On chromosome 2q, *zinc-finger protein 804A (ZNF804A*), which has been implicated in schizophrenia by genome-wide association studies (GWAS) (O'Donovan et al., 2008) and candidate gene association studies (Riley et al., 2010; Steinberg et al., 2011), and a meta-analysis of association studies (Williams et al., 2011), is located within a nominally significant region, rather than the region that showed genome-wide significance.

1.6.1.2 Bipolar disorder

In a recent review of whole-genome linkage studies for bipolar disorder, Alsabban et al. (2011) highlight a lack of congruence between the approximately 20 individual studies and combined analyses of these studies that had been carried out at that time. A large scale meta-analyses had failed to identify any linkage peaks at the level of genome-wide significance (Segurado et al., 2003), whilst a mega-analysis reported significant linkage to chromosome 6q for bipolar disorder 1 and 8q for bipolar disorder 1 and bipolar disorder 2 (McQueen et al., 2005). Similarly, in a review of linkage studies of bipolar disorder performed until December 2007, Serretti and Mandelli (2008) report the existence of linkage peaks on almost

every single chromosome. Some peaks have received support from multiple studies (e.g. 4p16.1, 6q21 and 11p15.5); however, the majority have been identified only by single studies. A few linkage peaks span genomic regions containing genes implicated by association studies, such as *dopamine receptor D4 (DRD4)* and the serotonin transporter gene *SLC6A4*.

1.6.1.3 Conclusions from linkage studies

Interpretations of the lack of consistency between individual linkage studies have varied. Some, such as Alsabban et al. (2011), have concluded that the lack of consistent major "hits" indicates that single-gene forms of the disorder in question (in this case, bipolar disorder) are exceedingly rare, if not non-existent. However, as mentioned in section 1.5, if many cases of schizophrenia and bipolar disorder are attributable to the effects of single mutations that are so rare that they only play a causal role in a small number of individuals, then the act of combining several unrelated families in a single linkage sample is likely to render the results of linkage analysis essentially meaningless.

1.6.2 Association analysis

Association analysis presents an alternative approach to linkage analysis, which is complimentary in terms of the kind of variants it can detect. A large number of cases selected for a phenotype of interest and controls (without the phenotype of interest) are genotyped for a set of markers and a comparison carried out to detect those variants that occur statistically significantly more frequently in the case group compared to the control group. Sometimes, the genotyped markers may contain the causal variant; however, more often, it will be the case that they are in linkage disequilibrium (LD) with the casual variant. LD describes the phenomenon whereby the rate at which markers at two or more genetic loci co-occur deviates from that expected based on random recombination during meiosis. The existence of regions of high LD can be exploited in the search for disease-associated variants by the use of "haplotype-tag" SNPs (Johnson et al., 2001). These are SNPs that capture the genetic variation within haplotype blocks (regions in which just a few common haplotypes are observed) in a non-redundant fashion. Association analysis has higher resolution than linkage analysis and is capable of detecting variants of smaller effect size. Prior to the availability of affordable high-throughput genotyping methods, association analysis would typically either be carried out to study genes implicated by linkage analysis (positional

candidates) or to investigate genes of interest based on their known functions (functional candidates). Nowadays, the advent of such methods has resulted in an increase in the number of GWASs being performed. These operate on the same principles as standard association analysis, except, instead of assessing association to a few markers, association is assessed to markers covering the entire genome. GWASs offer the advantage of, in theory, providing an unbiased method for identifying variants involved in the disease pathogenesis; however, as is true for all association analyses, their validity depends on both the genetic architecture of the phenotype under consideration and the degree of phenotypic heterogeneity present in the case group.

1.6.2.1 Schizophrenia

Numerous association studies have been performed for both schizophrenia and bipolar disorder, implicating many genes, of which a few have received replicated support. For schizophrenia, at the time of Sun et al.'s (2008) pre-GWAS era survey of association studies, over 500 genes had been tested for association with schizophrenia in more than 2000 association studies. In a review of the schizophrenia association study literature, Harrison and Weinberger (2005) identify the best supported schizophrenia candidate genes based on association and linkage data, as well as functional plausibility and evidence for altered expression in schizophrenia. Based on the evidence from association studies alone, *dystrobrevin binding protein 1* (*DTNBP1*) and *NRG1* are the best supported genes; however, when considering the other lines of evidence, *catechol-O-methyl transferase* (*COMT*) became one of the best supported candidate genes due to its functional plausibility. *COMT* degrades catecholamines, such as dopamine, adrenaline and noradrenaline, by catalysing their methylation. As abnormalities in dopamine function are believed to play a key role in the pathogenesis of schizophrenia (Howes and Kapur, 2009), altered *COMT* function is a plausible aetiological factor. A SNP in exon four of the COMT gene results in an amino acid substitution from valine (val) to methionine (met); the amino acid change affects the stability of COMT, such that the val form has significantly higher enzyme activity than met-COMT (Lachman et al., 1996). Several association studies have been carried out to assess the involvement of this amino acid change in schizophrenia with mixed results. Of those studies detecting a positive association, val-COMT has been associated with schizophrenia more frequently than met-COMT (Harrison and Weinberger, 2005). This finding was surprising: the classic dopamine hypothesis of schizophrenia, which postulates an increase in dopamine transmission at dopamine D2 receptors, would have predicted association to the lower

activity (i.e. higher dopamine) met-COMT. Subsequent studies have, to some extent, resolved this paradox by showing that COMT is particularly involved in the regulation of dopamine levels in the PFC (Weinberger et al., 2001), a region in which dopaminergic hypofunction has been detected in the brains of schizophrenic patients (Howes and Kapur, 2009). Indeed, the val158met polymorphism has been found to be associated with various aspects of prefrontal cognitive function relevant to schizophrenia (Bilder et al., 2002; Egan et al., 2001; Malhotra et al., 2002). Nevertheless, despite the strong biological plausibility of variation in *COMT* as a causal factor in schizophrenia, the most recent meta-analysis of *COMT* association studies found no evidence for the association of *COMT* SNPs, including the val158met polymorphism, and haplotypes with schizophrenia (Okochi et al., 2009).

Small-scale association studies investigating particular functional and positional candidate genes have now been largely superseded by GWASs. In a recent mega-analysis carried out by the Schizophrenia Psychiatric GWAS Consortium, samples of European ancestry were combined from 17 previous studies to give 9,394 cases and 12,462 controls and association assessed for a large number of SNPs across the genome (Ripke et al., 2011). Associations attaining genome-wide significance were observed for 136 SNPs, with the majority of these mapping to a region within chromosome 6p21.32-6p22.1 where the major histocompatibility complex (MHC) is located. Other significantly associated SNPs were located in regions harbouring genes implicated in schizophrenia such as *transcription factor 4* (*TCF4*) and *neurogranin (NRGN)*. Genetic variation in TCF4 has been associated with schizophrenia in an independent sample (Li et al., 2010), associated with sensorimotor gating in both schizophrenic patients and healthy controls (Quednow et al., 2012) and associated with cognitive performance, negative symptoms and age-of-onset in patients with psychosis spectrum disorders (Wirgenes et al., 2012). Reduced NRGN expression has been identified in the prefrontal cortex of schizophrenic patients (Broadbelt et al., 2006), genetic variation in the gene has been associated with schizophrenia in males (Ruano et al., 2008), and sequencing has identified rare variants present only in cases with schizophrenia (Shen et al., 2012).

An independent replication sample (8,442 cases and 21,397 controls) was assessed for a subset of the SNPs showing the strongest evidence for association in Ripke et al.'s (2011) initial sample. In this sample, a high level of agreement was observed with the results obtained in the initial sample. Combined analysis of the initial sample and the replication sample identified genome-wide significant association to seven loci, five of which were

novel (1p21.3, 2q32.3, 8p23.2, 8q21.3 and 10q24.32-q24.33), with the remaining two loci, located at 6p21.32-p22.1 and 18q21.2, containing the MHC and *TCF4*, respectively.

1.6.2.2 Bipolar disorder

A recent meta-analysis (i.e. the pooled analysis of published results) of bipolar disorder association studies failed to identify any variants attaining the threshold for significance following correction for multiple testing (Seifuddin et al., 2012). Four genes were found to contain variants that attained nominal significance ($p \leq 0.05$), including *BDNF*, which encodes brain derived neurotrophic factor. *BDNF* is a strong functional candidate for bipolar disorder: in rats, its expression shows a long-term reduction in response to neonatal stress (Russo-Neustadt et al., 2001), which is believed to play a role in the pathogenesis of bipolar disorder, and is increased following treatment with the mood-stabilising drugs lithium and valproate (Einat et al., 2003). Moreover, variants in BDNF have been implicated in bipolar disorder by several independent association studies (Serretti and Mandelli, 2008). Other genes receiving replicated support for association with bipolar disorder (in the pre-GWAS era) include *DAOA*, *DISC1*, *DRD4*, *DTNBP1*, *NRG1*, *SLC6A3*, *SLC6A4*, and *TPH2* (Serretti and Mandelli, 2008).

Several GWASs have been performed for bipolar disorder. Studies yielding genome-wide significant association are summarised in table 1.1. In a recent mega-analysis (i.e. the pooled analysis of raw data) carried out by The Psychiatric GWAS Consortium, variants within two genes, *CACNA1C* (encoding the α (1C) subunit of the voltage-dependent L-type calcium channel) and *ODZ4* (encoding Odz4/Ten-m4), attained genome-wide significance (Sklar et al., 2011). This mega-analysis represented the largest case-control association study of bipolar disorder carried out to date, including subjects of American, Canadian, and European ancestry who had previously been analysed in smaller GWASs. Seifuddin et al. (2012) carried out a comparison of their meta-analysis results for bipolar disorder association studies with the findings of the mega-analysis. Of the four genes identified by Seifuddin et al. (2012) as containing nominally significantly associated variants, none were implicated by Sklar et al.'s (2011) mega-analysis.

Study	Details	Genes implicated	Further details
(Baum et al., 2008)	A test/replication study design was employed. 550,000 SNPs were assessed for association in the initial sample (NIMH sample) using pooled DNA: 461 unrelated BPD 1 cases, 563 controls. SNPs meeting replication-testing criteria were assessed for association in the second sample (German sample): 772 BPD 1 patients and 876 control individuals. Both samples were of European ancestry.	*DGKH*	The strongest association was to a marker within the first intron of *DGKH* ($p = 1.5 \times 10^{-8}$).
(Sklar et al., 2008)	372,193 SNPs were assessed for association in a case-control group comprising 1,461 cases with BPD 1 and 2,008 controls. Subjects were of European and American ancestry.	*MYO5B*	Haplotype analysis identified association to SNPs within the *MYO5B* gene ($p = 2.04 \times 10^{-8}$). Nominally significant association was detected in the single-marker test of association ($p = 1.66 \times 10^{-7}$).
(Ferreira et al., 2008)	1,769,948 SNPs were assessed in 4,387 cases (81.2% BPD 1, 15.6% BPD 2, 2.21% SAB, and 0.935% manic disorder) and 6,209 controls. The sample included two previously assessed samples, the STEP-UCL sample (Sklar et al., 2008) and the WTCCC sample (WTCCC, 2007) and a third novel sample. Subjects were of European and American ancestry.	*ANK3*	Association was detected to an imputed intronic SNP within *ANK3* ($p = 9.1 \times 10^{-9}$).
(Cichon et al., 2011)	A test/replication study design was employed. 511,978 SNPs were assessed for association in the initial GWAS sample of 682 cases (99.6% BPD 1, 0.29% BPD 2, and 0.15% SAB) and 1,300 controls. The most significant 48 SNPs were assessed for association in a replication sample of 1729 cases (81.50% BPD 1, 10.99% BPD 2, and 7.51% SAB) and 2313 controls. Further evidence for replication was sought in a second replication sample comprising 6,030 cases (85.14% BPD 1, 9.04% BPD 2, 3.78% SAB, 1.97% BPD-NOS, and 0.0663% manic disorder) and 31,748 controls. Subjects were of European ancestry.	*NCAN*	Significant association to an exonic SNP in the *NCAN* gene was identified in meta-analyses of (i) the initial GWAS sample and the first replication sample ($p = 3.02 \times 10^{-8}$) and (ii) the initial GWAS sample and both replication samples ($p = 2.14 \times 10^{-9}$).
(Sklar et al., 2011)	Mega-analysis of previously reported GWAS samples (WTCCC, 2007; Ferreira et al., 2007; Sklar et al., 2008; 8; Scott et al., 2009; Smith et al., 2009; Djurovic et al., 2010; Cichon et al., 2012). Employed a test/replication design. The primary GWAS sample comprised 7,481 cases (84.1% BPD 1, 11.0% BPD 2, 3.52% SAB and 1.40% other BPD diagnoses) and 9,250 controls. 2,542,952 SNPs were assessed for association. Replication analysis was carried out for 34 SNPs showing association in the primary GWAS sample. Association to these SNPs was assessed in 4,496 cases (for the 4,111 cases for whom sub-type diagnoses were available: 81.6% BPD 1, 15.1% BPD 2, 1.92% SAB, and 1.39% BPD-NOS) and 42,422 controls. Subjects were of European, American and Canadian ancestry.	*ODZ4* and *CACNA1C,*	Meta-analysis of the primary GWAS sample and the replication sample identified intronic SNPs in *CACNA1C* ($p = 1.52 \times 10^{-8}$) and *ODZ4* ($p = 4.40 \times 10^{-8}$) as being significantly associated.

Table 1.1. Summary of positive bipolar disorder genome-wide association study (GWAS) findings. Only those studies in which association attained the threshold for genome-wide significance of $p \leq 5 \times 10^{-8}$ are mentioned. For each GWAS, details of the study such as study design, numbers of cases and controls, the number of single nucleotide polymorphisms (SNPs) assessed, and the ancestry of the subjects are provided. Genes containing associated SNPs are indicated together with further details about the association. Abbreviations not already specified: BPD 1 = bipolar 1 disorder; BPD 2 = bipolar 2 disorder; BPD-NOS = bipolar disorder not otherwise specified; NIMH = National Institute of Mental Health; SAB = schizoaffective disorder, bipolar type.

1.6.2.3 Conclusions from association analyses

Taken together, association studies of schizophrenia and bipolar disorder have implicated several variants, each predicted to confer a small increase in risk (odds ratios have typically been around 1.1 (Ripke et al., 2011). At the outset, it was widely expected that GWASs would reveal much of the genetic variation responsible for the heritable component of psychiatric disorders such as schizophrenia and bipolar disorder. More recently, it has, however, become apparent that, collectively, the associated variants can only explain a small percentage of the heritable variance: in recent studies, the variance explained has ranged from 3-6% (Purcell et al., 2009; Ripke et al., 2011; Ruderfer et al., 2011). This discovery has led to the notion of "missing heritability" and, subsequently, several theories as to where the missing heritability might be found. One approach to the missing heritability problem has been to consider marginally significant variants (i.e. those failing to meet the threshold for genome-wide significance but attaining a more liberal threshold, such as $p \leq 0.5$) with the hope that this might improve the amount of variance accounted for. Two strategies employed in the analysis of marginally associated variants are polygenic analysis and pathway analysis.

Purcell et al. (2009) have used a polygenic approach in the analysis of schizophrenia GWAS data. SNPs with *p*-values falling below a significance threshold (e.g. $p \leq 0.1$ or $p \leq 0.5$) in a discovery sample comprising schizophrenic patients and controls were used to calculate polygenic scores for each individual in a target sample, which also included schizophrenic patients and controls. To assess whether the polygenic score reflected schizophrenia risk, Purcell et al. (2009) assessed whether the cases in the target sample had a higher mean score than controls. Score alleles identified in the discovery sample were found to be significantly enriched amongst cases in the target sample, although they only explained around 3% of the variance. As the set of variants contributing to the polygenic score was likely to contain several false positives, Purcell et al. (2009) hypothesised that the amount of variation explained if only the true risk alleles were considered would be higher than 3%. To estimate the percentage of the variation explained by the true risk alleles a series of simulations were performed. These converged on an estimate of 34%; however, to reach this figure, assumptions were made about the genetic architecture of schizophrenia (such as the existence of a liability threshold), that may render the estimate invalid. Nevertheless, Purcell et al.'s (2009) demonstration that relaxing the significance threshold for the inclusion of SNPs in the polygene score resulted in an increase in the enrichment of score alleles in the

target sample cases suggests that the study of marginally significant *p*-values may yield useful insights.

The rationale behind pathway analysis of marginally associated variants is that by taking into account prior knowledge about the way genes operate in a biological context, the identification of true risk variants can be facilitated, even when these variants fail to attain genome-wide significance in single-marker tests of association (Wang et al., 2011a). Jia et al. (2010) applied pathway analysis to a schizophrenia case-control GWAS dataset and identified several pathways, including glutamate metabolism and the TGF-beta signalling pathway, as being overrepresented. Highlighting the potential for this approach to complement traditional GWAS analysis, several genes in the significantly overrepresented pathways had not been implicated by single-marker tests of association. By identifying biological pathways containing multiple implicated genes, this approach has the potential to offer insight, and thus generate testable hypotheses, regarding genetic interactions that may play an important role in disease pathogenesis.

The extent to which strategies designed to reveal the "hidden heritability" present in GWAS data will account for missing heritability in the context of psychiatric genetics remains to be seen. In light of the uncertainty surrounding the underlying genetic architecture of psychiatric conditions, discussed in section 1.5, it is, however, a viable prospect that the missing heritability problems stems, at least in part, from a mismatch between the type of variants best detected by GWAS and the nature of the risk variants involved in psychiatric illness.

The design and interpretation of GWASs has traditionally been predicated on the CDCV hypothesis. The majority of variants detected by GWAS chips are common variants (e.g., the Omni family of microarrays from Illumina can detect variants with minor allele frequencies greater than 2.5-5%, depending on the specific array; http://www.illumina.com/documents/products/brochures/brochure_omni_microarrays.pdf) and, while the genotyped SNPs are not themselves expected to be causal, association to a given marker has generally been interpreted as a reflection of the actions of another as-of-yet unidentified common variant with which the marker is in LD with. This line of reasoning has led some to interpret the results emerging from GWASs as supporting a polygenic CDCV genetic architecture (Purcell et al., 2009). Under this model, the failure of GWASs to account for anything more than a small proportion of the total heritable component of a

given disorder might be attributable to action of many common variants each conferring an increase in risk so infinitesimally small that they are not detectable using currently studied sample sizes. While this explanation receives some support from the analysis of marginally significant variants discussed above, a strong case has been made for the reconsideration of the involvement of rare variants (Dickson et al., 2010; Goldstein, 2009; Mitchell, 2012).

One possibility is that, due to the imperfect LD between the genotyped variant and the causal variant, the effect size of the causal variant is underestimated. Rare variants, which are likely to have evolved on a particular haplotype background, but, which, by definition will only co-occur with the genotyped variant in a small number of individuals, are likely to be most affected. Moreover, as rare variants may have developed on multiple haplotypes at any given locus, it is possible that their effects will cancel out when assessed in a large sample genotyped only for common variants. Furthermore, as a given rare variant is only likely to play a causal role in a very small number of cases within a sample, even though within these cases it may account for a large proportion of the genetic component of the disease, at the level of the whole sample, its estimated effect size will be diluted.

The issue of phenotypic heterogeneity is highly pertinent to the study of psychiatric disorders and has consequences for the interpretation of the missing heritability problem. The diagnostic categories of schizophrenia and bipolar disorder are broad (even when considering bipolar disorder 1 and 2 separately) and can result in patients diagnosed with the same disorder sharing few symptoms. In light of this phenotypic heterogeneity there is no reason to assume that schizophrenia or bipolar disorder represent unitary conditions in terms of their underlying aetiology. If schizophrenia and bipolar disorder are not unitary conditions, association analysis of cases selected purely by diagnostic status will produce results that are meaningless. A valid assessment of the extent of the missing heritability problem can only be made in the absence of the confounding effect of phenotypic heterogeneity; as such, it appears likely that the extent of the missing heritability problem for psychiatric conditions remains unknown. Moreover, the issue of phenotypic heterogeneity is relevant to arguments regarding the likely contribution of rare and common variants: if conditions such as schizophrenia comprise a sub-set of conditions that are individually rare, the theoretical basis of the CDCV hypothesis is thrown into question.

It is important to note that rare and common variants do not represent mutually exclusive mechanisms for genetic susceptibility. Indeed, as mentioned in section 1.4, it seems probable

that genetic background will play an important role in determining the effect of a given mutation. It has been suggested that genetic background may act in a polygenic fashion to modify the expressivity of rare variants (Mitchell and Porteous, 2011). As such, certain genetic backgrounds might render an individual less capable of overcoming the deleterious effects of a rare mutation than others. Under this model, some associated variants may represent modifiers of disease risk rather than directly causal variants.

As more genetic linkage and association studies for schizophrenia and bipolar disorder have been performed it has become increasingly apparent that efforts to detect causal variants are hampered by a lack of understanding of what kind of variants should be sought. Ultimately, if an unbiased exploration of the genetic basis of schizophrenia and bipolar disorder is to be achieved, it is likely that re-sequencing of large numbers of affected individuals will be necessary. Moreover to uncover the full-spectrum of regulatory and structural variants, whole genome-sequencing, rather than exome sequencing must be performed.

1.6.3 Cytogenetic studies

The detection of chromosomal abnormalities, such as translocations, deletions, or inversions, using cytogenetic methods has proven to be a useful approach in the search for genes involved in schizophrenia and bipolar disorder. One advantage to this approach is that, in contrast to association and linkage studies, cytogenetic methods permit the location of the variant to be accurately identified. Perhaps the best known gene implicated in the pathogenesis of psychiatric illness by a gross cytogenetic abnormality is the *DISC1* gene. *DISC1* was detected by virtue of its presence at the breakpoint of a balanced translocation, which segregates with various psychiatric diagnoses including schizophrenia, bipolar disorder, and major depressive disorder, in a large family (St Clair et al., 1990; Millar et al., 2000). The contribution of variation in the *DISC1* gene to psychiatric illness will be discussed further in section 1.7.4.2.2. Other candidate genes for schizophrenia and/or bipolar disorder implicated by their disruption by a chromosomal abnormality include *ABCA13* (encoding ATP-binding cassette subfamily A member 13) (Knight et al., 2009), *GRIK4* (encoding glutamate receptor, ionotropic, kainite 4) (Pickard et al., 2006), *NPAS3* (encoding neuronal PAS domain protein 3) (Kamnasaran et al., 2003; Pickard et al., 2005), and *PDE4B* (encoding phophodiesterase 4B) (Millar et al., 2005). The discovery that cytogenetic lesions in several genes appear to play a causal role in schizophrenia and bipolar disorder supports the role of rare variants in these conditions (Mitchell and Porteous, 2011).

1.7 Dysregulation of gene expression as a pathogenic mechanism for psychiatric illness

The role of altered gene expression in the pathogenesis of psychiatric illness is an area that has received increasing attention in recent years. In part, this interest has been fuelled by the relative paucity of risk variants with an effect on the encoded protein uncovered by approaches such as association and linkage analysis (Bray, 2008) and the fact that many of the most significant associations detected by GWAS are to variants located in non-coding regions (Duan et al., 2010). While it is possible that some non-coding and synonymous variants are in linkage disequilibrium with variants affecting the encoded protein, it is likely that at least some variants will confer risk via an effect on gene expression. Such variants might have an effect at the RNA level by altering transcription, stability or splicing or at the protein level via an effect on translation or stability; the ultimate outcome for any such variant would be to alter the level of the encoded protein. In relation to the rare vs. common variant hypotheses of the genetic architecture of psychiatric illness, one possible contribution of regulatory variants to disease pathogenesis might be in determining the degree to which a deleterious gene product is expressed.

1.7.1 Evidence for gene expression abnormalities in schizophrenia and bipolar disorder

Studies of gene expression in patients with schizophrenia have identified changes at multiple loci involved in processes such as synaptic function (Maycox et al., 2009; Mirnics et al., 2000; Vawter et al., 2006), mitochondrial and metabolic function (Altar et al., 2005; Prabakaran et al., 2004), myelination (Hakak et al., 2001; Sugai et al., 2004; Tkachev et al., 2003) and inflammation (Saetre et al., 2007). There is also evidence of altered gene expression in bipolar disorder, with aberrantly expressed genes contributing to pathways that overlap with those identified for schizophrenia. These include synaptic function (Cruceanu et al., 2012; Lopez de Lara et al., 2010; Ryan et al., 2006) and mitochondrial function (Lopez de Lara et al., 2010). Moreover, gene expression studies support the idea that there are pathogenic mechanisms common to schizophrenia and bipolar disorder: several genes have been found to be dysregulated in both conditions (Shao et al., 2008; Thomsen et al., 2011), and a study by Pedrosa et al. (2010) in which chromatin immunoprecipitation followed by microarray analysis (ChIP-chip) was performed to identify the transcriptional targets of β-catenin identified several genes with known involvement in schizophrenia and bipolar

disorder. Additionally, the discovery that glutamate receptor expression is abnormal in both bipolar disorder and schizophrenia (Beneyto et al., 2007; Woo et al., 2007; Woo et al., 2004) supports the role of glutamatergic dysfunction, a well-established pathophysiological mechanism for schizophrenia (Kantrowitz and Javitt, 2012), in the aetiology of bipolar disorder. As these studies did not select patients on the basis of genotype, it is likely that they reflect the combined influence of *cis*- and *trans*-acting effects, including environmental influences, converging on common molecular pathways.

Several well supported candidate genes for schizophrenia have been found to show dysregulated expression in at least some samples. These include *DTNBP1* (Talbot et al., 2004; Tang et al., 2009; Weickert et al., 2008; Weickert et al., 2004), *G72* (*DAOA*) (Korostishevsky et al., 2004), *RGS4* (Bowden et al., 2007; Erdely et al., 2006; Mirnics et al., 2001), *NRG1* (Hashimoto et al., 2004; Law et al., 2006; Nicodemus et al., 2009) and *DISC1* (Maeda et al., 2006; Nakata et al., 2009; Sawamura et al., 2005). The evidence implicating altered *NRG1* and *DISC1* expression in psychiatric illness will be discussed further in sections 1.7.4.1.5 and 1.7.4.2.5.

In addition to the evidence implicating dysregulated gene expression in the pathogenesis of schizophrenia and bipolar disorder, a study by Mar et al. (2011) highlights another manner in which abnormalities in gene expression might manifest. Using neural stem cells derived from patients with schizophrenia, Mar et al. (2011) performed genome-wide transcriptional profiling and assessed variance in gene expression. Compared to control subjects, patients with schizophrenia showed a marked reduction in gene expression variance, suggesting that gene regulation is more tightly constrained. Mar et al. (2011) suggest that overly constrained biological pathways would be less able to adapt to environmental stressors and, therefore, might precipitate a pathological state. Although replication of these findings would be necessary in order to draw firm conclusions, it seems that the study of gene expression variance might yield clues into the aetiology of psychiatric illness that would be missed by the traditional approach of simply studying differences in mean expression levels.

1.7.2 Genetic mechanisms contributing to dysregulated gene expression in psychiatric illness

As suggested above, there are several mechanisms by which abnormalities in gene expression might occur. Changes in gene expression can be brought about by the actions of

cis-acting regulatory variants located in the transcribed sequence, promoter region, or in distal regulatory elements located on the same chromosome, variation in *trans*-acting factors, such as proteins involved in transcriptional regulation or protein degradation, and copy number variation.

1.7.2.1 Cis- and trans-acting factors

Evidence for the role of *cis*-acting variants has come from studies in which disease-associated polymorphisms located in, or in linkage disequilibrium with, regulatory regions have been shown to be associated with gene expression level. Several variants in the *NRG1* gene have been shown to be associated with both schizophrenia and either endogenous gene expression level or transcriptional activity using the *in vitro* luciferase reporter assay (Law et al., 2006; Nicodemus et al., 2009; Pedrosa et al., 2009; Tan et al., 2007). Likewise, a schizophrenia-associated variant in the promoter region of the alpha7 neuronal nicotinic acetylcholine receptor subunit gene (*CHRNA7*) has been shown to result in decreased promoter activity in the luciferase reporter assay (Leonard et al., 2002).

If alterations in gene expression level mediated by *cis*-acting regulatory variants contribute to the pathogenesis of psychiatric illness, then there is no reason why changes in gene expression resulting from *trans*-acting factors should not also play a role. Using the expression level of the schizophrenia susceptibility gene *DTNBP1* as a quantitative trait, (Bray et al., 2008) performed genome-wide linkage analysis to detect genes involved in the regulation of *DTNBP1* expression, which might, therefore, also contribute to disease susceptibility. This approach highlighted the role of both *cis*- and *trans*-acting factors in the regulation of *DTNBP1* expression. Intriguingly, a linkage peak was observed in a region of chromosome 8p that encompasses *NRG1*, highlighting the possibility of a functional link between the two genes.

One *trans*-acting regulatory mechanism that has been the focus of increasing attention in recent years is microRNA (miRNA) mediated regulation. miRNAs are short (~22 nucleotide) single-stranded DNA molecules that mediate post-transcriptional regulation of gene expression by binding to partially complementary binding sites located at the 3' ends of mRNAs. In the vast majority of cases miRNA-mediated regulation results in repression of gene expression; however, there have been reports of miRNA-mediated increases in gene expression via upregulation of translation under some circumstances (Vasudevan et al.,

2007). miRNAs repress gene expression by two predominant mechanisms: destabilising the mRNA transcript or blocking translation initiation (Huntzinger and Izaurralde, 2011). miRNAs biogenesis involves several steps. In animals, miRNAs are initially transcribed as long primary miRNAs, which are then cut into hairpin-shaped pre-miRNAs of ~70 nucleotides by the nuclear RNase III Drosha (Lee et al., 2003). The pre-miRNA is exported out of the nucleus and cleaved by the cytoplasmic RNase III Dicer to form ~22 nucleotide miRNA duplexes (Bernstein et al., 2001). One strand of the duplex is then usually degraded and the remaining strand becomes the mature miRNA; although, in some cases, both strands are expressed at a high level (Guo and Lu, 2010).

There is now substantial evidence for the aberrant expression of miRNAs in both schizophrenia and bipolar disorder (Beveridge and Cairns, 2012; Forero et al., 2010). Genes encoding proteins involved in miRNA biogenesis, including *Drosha* and *Dicer*, have also been found to show increased mRNA expression in schizophrenia (Beveridge and Cairns, 2012). Furthermore, in a recent genome-wide association study, a variant within the primary transcript of miR-137 was identified as being associated with schizophrenia and four other variants attaining genome-wide significance were located in genes identified as predicted targets of miR-137 (Ripke et al., 2011). These genes have subsequently been confirmed, *in vitro*, as miR-137 targets (Kwon et al., 2011). miRNAs have been shown to play an important role in brain development and in structural plasticity (Forero et al., 2010), processes that, when disrupted, are known to contribute to the pathogenesis of schizophrenia (Harrison and Weinberger, 2005). Several miRNAs found to be disregulated in schizophrenia are known to regulate genes that have been implicated in the disorder, either by association of variants or gene expression analysis (Beveridge et al., 2010; Miller et al., 2012). One such example is miR-195, which shows increased expression in the brains of patients with schizophrenia (Beveridge et al., 2010), and has been found to regulate several genes relevant to schizophrenia, including *BDNF* (Mellios et al., 2008), *RELN* (Beveridge et al., 2010), and *GRIN3* (Beveridge et al., 2010). Interestingly, miRNAs have shown some promise as peripheral blood biomarkers: Lai et al. (2011) identified a seven-marker miRNA peripheral blood expression signature that could differentiate between patients with schizophrenia and controls with an area under the curve of receiver operating characteristics of 93%.

The potential pathogenic role of epigenetically mediated regulation of gene expression has already been alluded to in the context of monozygotic twins who are discordant for

psychiatric illness (section 1.3). Several studies have detected differences in epigenetic modifications in patients with schizophrenia or bipolar disorder; however, there are inconsistencies between individual studies (Gavin and Akbarian, 2012). DNA methylation of CpG islands within promoter regions is generally thought to mediate transcriptional inhibition (Klose and Bird, 2006). Increases in the expression of enzymes required for DNA methylation have been detected in the post-mortem brains of patients with schizophrenia (Guidotti et al., 2007; Veldic et al., 2004; Veldic et al., 2005), suggesting decreased gene expression via increased methylation as a pathogenic mechanism. The expression of both *RELN* and *GAD67*, genes expressed from CpG-island containing promoters, has been consistently found to be down-regulated in the post-mortem brains of patients with schizophrenia (Akbarian et al., 1995; Fatemi et al., 2005; Guidotti et al., 2000; Hashimoto et al., 2008), and both genes have also been shown to be down-regulated in the post-mortem brains of patients with bipolar disorder (Guidotti et al., 2000). In the post-mortem brains of patients with schizophrenia increased methylation of the *RELN* promoter has been reported (Abdolmaleky et al., 2005; Grayson et al., 2005) and GAD-67 expression has been found to correlate negatively with levels of the DNA methyltransferase DNMT1 (Veldic et al., 2005). Findings regarding the hypermethylation of the *RELN* promoter in schizophrenia have not been unanimous (Mill et al., 2008; Tochigi et al., 2008); however, such between-study variation is not surprising when considering an aetiologically heterogeneous disorder.

In addition to DNA methylation, epigenetic regulation of gene expression can occur via modification of histone proteins. DNA is organised into chromatin by wrapping around octamers of histone proteins to form nucleosomes. Covalent posttranslational modifications of histone proteins, such as acetylation, methylation, or phosphorylation, alter chromatin structure or function by changing the charge of the nucleosome and/or affecting the recruitment of non-histone proteins (Kouzarides, 2007). These modifications alter the interaction of the DNA sequence with molecules involved in transcription and, therefore, regulate gene expression (Berger, 2007). Histone acetylation promotes an open chromatin structure, which is accessible to the transcriptional machinery (Gorisch et al., 2005). The level of histone acetylation is maintained by the opposing actions of histone acetyltransferase (HAT) and histone deacetyltransferase (HDAC) enzymes (Bannister and Kouzarides, 2011).

In schizophrenia post-mortem brain samples, HDAC1 levels have been found to be elevated (Benes et al., 2008; Sharma et al., 2008), with HDAC1 level correlating negatively with GAD67 expression (Sharma et al., 2008). In keeping with these findings, reduced levels of

acetylated histone 3 have been detected in the peripheral blood mononuclear cells of patients with schizophrenia (Gavin et al., 2008). In contrast to acetylation, dimethylation of lysine 9 on histone 3 is a repressive histone mark. Levels of this mark have been found to be increased in patients with schizophrenia (Gavin et al., 2009). Based on these findings, it has been suggested that schizophrenia is characterised by a restrictive chromatin state, which results in reduced transcription and expression of certain genes (Gavin et al., 2009).

The role of aberrant epigenetic modification in bipolar disorder has been less well studied; however, a genome-wide epigenomic screen of patients with bipolar disorder and schizophrenia identified several loci showing altered methylation, some of which were common to both conditions (Mill et al., 2008). Altered expression of HDACs has also been reported in bipolar disorder (Hobara et al., 2010). Interestingly, the action of the mood stabilising drug valproate has been linked to HDAC inhibition (Phiel et al., 2001). Moreover, in a mouse model of hypermethylation, the antipsychotics clozapine and sulpiride have been found to demethylate hypermethylated *RELN* and *GAD67* promoters, with the action of both drugs being potentiated by valproate (Dong et al., 2008). These findings support the hypothesis that altered epigenetic regulation of gene expression might contribute to the pathogenesis of schizophrenia and bipolar disorder; however, they also highlight the importance of considering treatment effects when studying gene expression in medicated patients.

1.7.2.2 Copy number variation

In 2004, two studies demonstrated that submicroscopic (< 500kb) variations in DNA in copy number (CNVs) are common in normal human genomes (Iafrate et al., 2004; Sebat et al., 2004). Studies in rodents have shown that gene dosage differences resulting from CNVs show a weak but significant positive correlation with gene expression (Guryev et al., 2008; Henrichsen et al., 2009). Furthermore, differences in gene expression have been identified in genomic regions in the vicinity of CNVs: Stranger et al. (2007) found that while CNVs account for 18% of the variation in gene expression in human lymphoblastoid cell lines, more than half of the associations involved genes mapping outside of CNVs. This suggests that some gene expression changes associated with CNVs will result from disruption of *cis*- or *trans*-regulatory regions and/or alterations to the chromatin structure.

Accumulating evidence implicates rare CNVs in the pathogenesis of both schizophrenia and bipolar disorder (Malhotra and Sebat, 2012). Association studies have repeatedly identified an enrichment of CNVs in cases compared to controls (Buizer-Voskamp et al., 2011; ISC, 2008; Kirov et al., 2009; Walsh et al., 2008). Family studies have highlighted a high prevalence of de novo CNVs in schizophrenia; however, their contribution to sporadic compared to familial cases requires further clarification (Kirov et al., 2012; Malhotra et al., 2011; Xu et al., 2008). The genes affected by CNVs in patients have been found to contribute to functions known to be abnormal in schizophrenia, such as neurodevelopment and synaptic activity (Walsh et al., 2008; Malhotra et al., 2011).

In bipolar disorder, findings from association studies looking at the incidence of CNVs have been equivocal: two studies have reported an increased burden of CNVs in cases compared to controls (Priebe et al., 2011; Zhang et al., 2009), whilst two other studies failed to find any evidence for the involvement of CNVs (Grozeva et al., 2010; McQuillin et al., 2011), although McQuillin et al. (2011) do report the discovery of some rare CNVs found to occur only in cases. In common with schizophrenia, de novo CNVs appear to occur more frequently amongst patients with bipolar disorder, although, they have not been associated with sporadic incidences of the condition, as might be hypothesised (Malhotra et al., 2011). There is some evidence to suggest that de novo (Malhotra et al., 2011) and inherited CNVs (Zhang et al., 2009a; Priebe et al., 2012) might promote an earlier age of onset; however, evidence for the involvement of inherited CNVs in age of onset has not been replicated in some studies (Grozeva et al., 2010; McQuillin et al., 2011).

1.7.3 Issues surrounding the assessment of gene expression in psychiatric illness

The most fundamental issue surrounding the study of the contribution of gene expression to the pathogenesis of psychiatric illness is the ability to measure gene expression in a biologically relevant tissue. Clearly, when studying disorders such as bipolar disorder or schizophrenia, the tissue of choice would be the brain; however, for obvious reasons obtaining brain tissue from sufficient numbers of cases and controls is often impossible. Moreover, the use of post-mortem brain tissue is affected by several potential confounds that can be difficult to control for given the naturalistic nature of post-mortem studies. One such confound is the cause of death and the related issue of the state of the subject immediately prior to death. A reduction in brain pH has been observed following a prolonged agonal state

(Harrison and Kleinman, 2000) and altered brain pH has been shown to affect post-mortem gene expression levels (Harrison et al., 1995; Mexal et al., 2006; Tomita et al., 2004; Vawter et al., 2006). Time and season of death are also likely to affect gene expression and should be considered when using post-mortem brain tissue (Lewis, 2002). The evidence for the effect of post-mortem interval (PMI; the time between death and freezing of the tissue) on gene expression is not conclusive: some studies have detected an effect on gene expression (Birdsill et al., 2011), although this effect appears to be far less widespread than that of brain pH (Harrison et al., 1995); others have failed to detect a correlation between PMI and gene expression (Tomita et al., 2004).

Several studies have investigated the validity of using peripheral tissues, such as the blood, as a proxy for ascertaining gene expression in the brain. Sullivan et al. (2006) compared the expression levels of around 33,000 genes in the blood and the central nervous system and found a moderate positive correlation of 0.5 between the two. In a study that assessed a larger sample for the expression of a smaller number of genes in the blood and three brain regions, small but significant correlations were observed between expression levels in the blood and the brain($r = 0.24$-0.32); however, it should be noted that different platforms were used to measure blood and brain expression, potentially reducing the correlation between the tissue types (Cai et al., 2010). Moreover, when genes were considered as co-expression modules, 67% of genes in the cortex were part of a conserved gene expression module. Analysis of miRNA expression provides further support for the use of the blood as a surrogate for brain tissue in studies of gene expression: Liang et al. (2007) performed cluster analysis to assess miRNA co-expression in several different tissues and found the brain and peripheral blood mononuclear cells (together with thymus, adrenal gland, and testes) to form a unique cluster. Clearly, blood gene expression levels must be interpreted with caution when extrapolating to the brain. In support of the use of blood expression levels in the context of psychiatric illness, it is interesting to note that some schizophrenia-associated expression changes observed in the brain have been recapitulated in the blood (Glatt et al., 2005). Additionally, in patients with bipolar disorder or schizophrenia, BDNF has been found to show reduced expression in the hippocampus and prefrontal cortex (Thompson Ray et al., 2011; Weickert et al., 2003) and reduced serum expression, even in antipsychotic-naive schizophrenia patients (Grillo et al., 2007; Jindal et al., 2010). Further investigation of the comparability of blood and brain gene expression, ideally using a within-subject design, would be beneficial, as would further assessment of the effects of antipsychotics and mood-stabilising drugs on the relationship between blood- and brain-based gene expression.

1.7.4 *NRG1* and *DISC1*: two leading candidate genes for psychiatric illness susceptibility

As the work carried out in this thesis centres around *NRG1* and *DISC1*, two of the best-supported candidate genes for schizophrenia, which have also both been implicated in bipolar disorder, the remainder of this introduction will focus on these two genes.

1.7.4.1 NRG1

1.7.4.1.1 The NRG1 gene and its isoforms

The *NRG1* gene is located at chromosome 8p12-8p21. It is a large (~1.3Mb) gene, containing several transcription start sites and alternative splice sites that give rise to at least 31 isoforms (Mei and Xiong, 2008). All isoforms contain a core epidermal growth factor (EGF)-like domain (EGFc); however further sources of variability give rise to the distinct isoforms. These isoforms are classified into six types based on their 5' exon usage (Falls, 2003; Harrison and Law, 2006) (figure 1.1). Types IV-VI were identified more recently than types I-III (Steinthorsdottir et al., 2004) and less is known about their biological functions. Isoform-specific variation results from alternative splicing, resulting in different types of EGF domain: differential exon retention results in either EGFα or EGFβ, with the β variant predominating in the brain and being a more potent receptor activator than the α variant. After the EGF domain, most NRG1 isoforms contain a C-terminal transmembrane domain (TMc), which is sometimes preceded by a type 1 or type 4 stalk but can occur without a stalk (TMc-containing isoforms without a stalk are known as "2" isoforms). Some NRG1 isoforms contain a type 3 stalk: these isoforms are truncated prior to the TMc domain and are synthesised as soluble isoforms. All TMc-containing isoforms contain a carboxy-terminal region, which can take three forms (a, b, or c) depending on exon usage. NRG1 types I, II, IV, and V contain an immunoglobulin (Ig) domain and are sometimes referred to as Ig-NRG1s; NRG1 type III contains a cysteine rich domain (CRD), and is sometimes referred to as CRD-NRG1.The CRD domain contains an additional N-terminal transmembrane domain (TMn). No characteristic domain has yet been identified for NRG1 type VI.

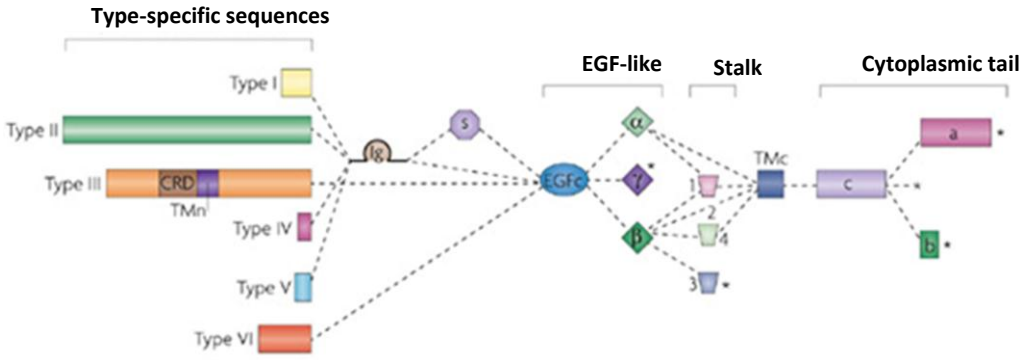


Figure 1.1. Types of Neuregulin 1 (NRG1). NRG1 isoforms can be classified into six types (I-VI) based on alternative 5' exon usage (type-specific sequences). Variation also arises from the existence of multiple epidermal growth factor (EGF)-like domain types, the type of stalk (1, 2, 3,or 4) preceding the C-terminal transmembrane domain (TMc), and alternative splicing of the C-terminal domain (a, b, and c). Types I, II, IV, and V have an immunoglobulin (Ig)-like domain between the N-terminal domain and the EGF domain. Some isoforms additionally contain a spacer (S) domain. All NRG1 types contain the core EGF domain (EGFc), with alternative exon retention resulting in the inclusion of α, β and γ variants. Type III NRG1 has a cysteine rich domain (CRD) in its N-terminal region, which includes an additional transmembrane domain (TMn). * indicates a stop codon. Adapted from Mei and Xiong, 2008.

1.7.4.1.2 NRG1 processing and signalling

NRG1 proteins interact with ErbB receptor tyrosine kinases, binding to ErbB3 and ErbB4 receptors via the EGF-like domain. NRG1 binding of ErbB receptors promotes ErbB dimerisation and autophosphorylation of tyrosine residues within the cytoplasmic domain of the receptor. This creates docking sites for various proteins, such as Grb2, Shc, and the regulatory subunit of phosphoinositide-3-kinase (PI3-kinase), that are involved in the activation of downstream signalling pathways. NRG1-ErbB4 signalling serves to mediate a variety of functions, including cell proliferation, migration, differentiation, adhesion, and apoptosis (Yarden and Sliwkowski, 2001). In the nervous system, NRG1 signalling has been implicated in synaptic formation, transmission, and plasticity (Mei and Xiong, 2008). Taken together, the cellular functions in which NRG1 is involved render NRG1 a strong functional candidate for schizophrenia and bipolar disorder.

With the exception of the glial growth factor isoform (GGF2), which is synthesised as a soluble fragment, NRG1 isoforms are synthesised as transmembrane proteins. The Ig-NRG1s are single-pass transmembrane proteins that undergo activity-dependent proteolytic cleavage between the EGF and TMc domains to release a soluble N-terminal fragment, which contains the EGF and Ig domains (Harrison and Law, 2006). The soluble N-terminal fragment activates ErbB receptors via both paracrine and autocrine signalling (Falls et al., 2003).

In contrast to the Ig-NRG1s, type III NRG1 (CRD-NRG1) contain two transmembrane domains, and are thus synthesised as two-pass transmembrane proteins. Proteolytic cleavage of type III NRG1s creates a transmembrane N-terminal fragment (Wang et al., 2001). This fragment engages with ErbB receptors in a juxtacrine fashion, a process which is required for the regulation of the myelination of axons in the peripheral nervous system by Schwann cells (Taveggia et al., 2005). In addition to participating in canonical forward signalling, the interaction of type III NRG1 with ErbB receptors also elicits back signalling: the intracellular domain of NRG1 is released following proteolysis and translocates to the nucleus where it regulates gene expression, including the repression of pro-apoptotic genes (Bao et al., 2003). Neuronal depolarisation can also stimulate NRG1 back signalling (Bao et al., 2003). In addition to its effects on cell survival, NRG1 back signalling has been shown to regulate the expression of postsynaptic density protein-95 (PSD-95; Bao et al., 2004) and α7 nicotinic acetylcholine receptors (nAChRs) via activation of the PI3K pathway (Hancock et al., 2008; Zhong et al., 2008), and regulate axon pathfinding of sensory neurons (Hancock et al., 2011).

1.7.4.1.3 NRG1 as a functional candidate for schizophrenia and bipolar disorder

Many of the known functions of NRG1 mirror functions known to be abnormal in schizophrenia and bipolar disorder. As might be expected for a gene with so many isoforms, the functions of NRG1 are diverse and are essential for the normal functioning of many organs including the brain. The necessity of NRG1 for normal development was revealed by the discovery that pan-Nrg1 knockout (KO) mice (binding of all Nrg1 isoforms with ErbB receptors was prohibited by disruption of the EGF-like domain) die at embryonic day (E) 10.5 due to a defect in cardiogenesis (Meyer and Birchmeier, 1995). Here, those functions of NRG1 relevant to our understanding of the pathophysiology of schizophrenia and bipolar disorder will be considered.

As mentioned in section 1.4, white matter abnormalities, which may be caused by aberrant myelination and/or neuronal migration, are a well-established finding in the brains of schizophrenic and bipolar disorder patients. NRG1 signalling through ErbB receptors has been implicated in both processes. In the CNS, axon myelination is performed by oligodendrocytes. NRG1 has been linked to many aspects of oligodendrocyte function, including proliferation, survival, maturation, and migration (Fernandez et al., 2000; Ortega et al., 2012; Sussman et al., 2005; Vartanian et al., 1999); however, it should be noted that normal myelination has been reported both in conditional null mutant mice that lack Nrg1 expression in projection neurons from various developmental stages and mice lacking oligodendrial expression of ErbB3 and ErbB4, receptors necessary for Nrg1 signalling (Brinkmann et al., 2008). Nevertheless, a role for NRG1 signalling in CNS myelination, in humans at least, is suggested by imaging studies showing an association between *NRG1* variants previously associated with risk of psychosis and white matter density and integrity (McIntosh et al., 2008; Sprooten et al., 2009).

NRG1 has been implicated in neuronal migration by the finding that NRG1-ERBB signalling promotes radial glia formation and facilitates the movement of neurons along radial glia, a process that is essential for normal cortical development (Anton et al., 1997; Rio et al., 1997). Interestingly, in the context of the abnormalities in GABAergic interneurons observed in the brains of schizophrenic and bipolar disorder patients (see section 1.4.2), Nrg1 signalling via ErbB4 receptors has recently been show to play a role in the guidance of migrating GABAergic interneurons from the ganglionic eminence to their cortical targets (Li et al., 2012). Sei et al. (2007) have demonstrated the potential involvement of aberrant NRG1 signalling in the neurodevelopmental abnormalities observed in the brains of schizophrenic patients: in an *in vitro* system, in which NRG1 is used to induce the migration of B lymphoblasts, cells from schizophrenic patients were found to show decreased NRG1-induced migration.

A role for Nrg1-ErbB4 signalling in axon guidance, another process key to neurodevelopment, has been demonstrated by López-Bendito et al. (2006). Following the tangential migration of a population of GABAergic interneurons from the lateral ganglionic eminence to the medial ganglionic eminence, a permissive corridor is created through which thalamocortical axons navigate through the telencephalon. The permissive nature of this corridor has been shown to rely, in part, on the interaction of CRD-Nrg1, expressed by cells

lining the corridor, and diffusible Ig-Nrg1, released from the pallium, with ErbB4 receptors expressed by thalamic neurons. These two isoforms of Nrg1 appear to contribute to axon pathfinding and outgrowth, respectively. A subsequent study has shown CRD-Nrg1 to regulate the pathfinding of sensory neurons in both the spinal cord and periphery (Hancock et al., 2011).

NRG1 has also been implicated in the regulation of several neurotransmitter systems that have been shown to function abnormally in schizophrenia and bipolar disorder. NRG1 appears to play a key role in regulating neurotransmitter receptor expression: to date, there is evidence to suggest that NRG1 is involved in regulating the expression of NMDA, $GABA_A$, and ACh receptors (Liu et al., 2001b; Okada and Corfas, 2004; Ozaki et al., 1997; Rieff et al., 1999). Furthering support for the notion that aberrant NRG1 signalling might contribute to the GABAergic deficits observed in schizophrenia and bipolar disorder, Abe et al. (2011) have demonstrated that Nrg1 regulates the expression and sensitivity of AMPA receptors in GABAergic interneurons, thus affecting their firing rate.

Studies demonstrating the role of NRG1 in regulating synaptic plasticity extend support for NRG1's functional candidacy. NRG1 has been shown to regulate LTP in the hippocampus (Kwon et al., 2008) and an impairment in LTP observed in heterozygous Nrg1 mutant mice can be rescued by exogenous Nrg1 (Bjarnadottir et al., 2007).

1.7.4.1.4 Overview of genetic studies implicating variation in NRG1 *in schizophrenia and bipolar disorder*

Several linkage studies have implicated the chromosomal locus, 8p21.1-22, where *NRG1* is situated in schizophrenia, although support for the involvement of this region has not been unanimous (reviewed in Harrison and Law, 2006). However, as mentioned previously (section 1.6.1.1), the existence of a risk variant within the chromosome 8p region, at least in individuals of European ancestry, has been supported by a meta-analysis of 32 independent genome-wide linkage studies of schizophrenia (Ng et al., 2009).

The first evidence for association of variants in *NRG1* with schizophrenia came from a study carried out by (Stefansson et al., 2002), who identified a haplotype (Hap_{ICE}) at the 5' end of the gene that doubled the risk of schizophrenia. Subsequent studies have replicated this finding (Stefansson et al., 2003; Williams et al., 2003) and others have reported association

to other markers or haplotypes either nearby to or overlapping HapICE (Corvin et al., 2004; Georgieva et al., 2008; Hall et al., 2004; Kim et al., 2006; Lachman et al., 2006; Prata et al., 2009; Tang et al., 2004; Thomson et al., 2007; Yang et al., 2003; Zhao et al., 2004). While the majority of association studies have focussed on markers at the 5' end of the gene, others that have considered SNPs further downstream have reported association to 3' haplotypes (Petryshen et al., 2005; Thomson et al., 2007; Yang et al., 2003). However, it should be noted that the evidence is not unequivocal, with some association studies failing to identify any risk variants within NRG1 (Ingason et al., 2006; Jonsson et al., 2009; Moon et al., 2011). Although the association of *NRG1* variants with bipolar disorder has been less well studied, positive association findings suggest that certain *NRG1* variants may also be risk factors for bipolar disorder (Green et al., 2005; Prata et al., 2009; Thomson et al., 2007).

1.7.4.1.5 Evidence implicating altered NRG1 expression in the pathophysiology of schizophrenia and bipolar disorder

As mentioned above (section 1.7.4.1.5), several association studies have been carried out implicating variation in *NRG1* in the pathophysiology of schizophrenia and accumulating evidence suggests that *NRG1* variants may also confer risk for bipolar disorder. The vast majority of the associated variants do not have an effect on the encoded protein, suggesting that they may confer risk via altered gene regulation. It is, of course, possible that some non-coding variants are in linkage disequilibrium with coding variants; however, this is unlikely to always be the case. Indeed, *NRG1* SNPs associated with schizophrenia have been shown to be associated with altered NRG1 expression (Law et al., 2006; Nicodemus et al., 2009). Moreover, some studies have detected increased levels of NRG1 in the brains of patients with schizophrenia (Hashimoto et al., 2004; Law et al., 2006), while Parlapani et al. (2010) found decreased expression of NRG1 type 1 but increased expression of NRG1 type 2 in the post-mortem brains of elderly schizophrenic patients. In a recent study, Marballi et al.(2012) assessed the expression of full-length NRG1 and NRG1 cleavage products in the post-mortem brains of patients with schizophrenia or bipolar disorder. No differences were observed in the amount of full-length NRG1 in either group; however, an increase in the NRG1 N-terminal cleavage product was detected in the schizophrenic patients and the hippocampal expression of a 50kDa product was decreased in both groups. Consistent with the potential role of altered NRG1-ERBB signalling in schizophrenia, antipsychotics have been found to down-regulate NRG1-ERBB signalling (Pan et al., 2011).

1.7.4.1.6 Background to the NRG1 *association study carried out in this thesis*

In this thesis, the analyses described in chapter 2 are based on a study carried out by Thomson et al. (2007). Here, Thomson et al.'s (2007) study will be described briefly to provide the background to the work presented in chapter 2.

Thomson et al. (2007) analysed the association of haplotype-tagging SNPs across the entire *NRG1* gene in a sample of patients with schizophrenia or bipolar disorder and healthy controls drawn from the Scottish population. Significant associations were detected to haplotypes in two regions (defined by linkage disequilibrium) of the *NRG1* gene: region A, located at the 5' end of the gene, which overlaps with the *NRG1* promoter region and extends into the first intron, and region B, which spans a largely intronic region but also overlaps with the sensory and motor neuron-derived factor (SMDF) and heregulin alpha (HRGα) isoforms, and the 3' exons of all other isoforms.

Given the evidence presented in section 1.7.4.1.5 for the likely involvement of altered *NRG1* expression in schizophrenia, a plausible hypothesis is that the associated haplotypes identified by Thomson et al. (2007) reflect the involvement of regulatory variants. In this thesis, an attempt will be made to replicate Thomson et al.'s (2007) findings for the two most significant haplotypes in regions A and B in two new case-control samples, comprising schizophrenic patients, bipolar disorder patients and healthy controls. It was hypothesised that this analysis might aid in the identification of a region containing regulatory variants that could be studied further *in vitro* to establish the effects of putative regulatory variants.

1.7.4.2 DISC1

1.7.4.2.1 Initial discovery of the DISC1 *gene*

The *DISC1* gene was first identified at the breakpoint of a balanced chromosomal translocation t(1;11) (q42.1;14.3) that segregates with psychiatric illness in a large Scottish family (StClair et al., 1990; Millar et al., 2000). Of the 87 family members assessed for the translocation, 37 were found to be carriers (Blackwood et al., 2001). In the most recent description of the family, 29 translocation carriers underwent psychiatric assessment and of these individuals, 18 were diagnosed with major mental illness (schizophrenia (7 cases),

bipolar disorder (1 case), or recurrent major depression (10 cases)) and an additional three individuals were diagnosed with anxiety, alcoholism, and minor depression, respectively. Of the thirty-eight non-carriers who underwent psychiatric assessment, none were diagnosed with a major psychiatric illness, whilst three were diagnosed with minor depression, one with alcoholism, and one with adolescent conduct disorder. Linkage analysis resulted in a logarithm of odds (LOD) score of 7.1 for the segregation of the translocation with the broad diagnosis of major mental illness, which included schizophrenia, bipolar disorder, and recurrent major depression, and a LOD score of 3.6 for the narrow diagnosis of schizophrenia (Blackwood et al., 2001). In addition to disrupting *DISC1*, the translocation disrupts a non-coding RNA gene located antisense to *DISC1* called *DISC2* (Millar et al., 2000), and, transcripts resulting from a gene on chromosome 11, thought to be a non-coding RNA gene, have also been identified that would also be disrupted by the translocation (Zhou et al., 2008). Analysis of lymphoblastoid cell lines derived from carriers of the t(1;11) translocation has revealed that the translocation results in the formation of fusion transcripts, which, if expressed as proteins, have been demonstrated to have a deleterious effect on mitochondrial function (Eykelenboom et al., 2012).

While not all carriers of the translocation were diagnosed with a psychiatric condition, there is evidence to suggest the existence of functional brain deficits even in diagnosis-free carriers. The P300 is an event related potential (ERP) thought to reflect the speed and efficiency of information processing. It is typically found to occur with reduced amplitude and increased latency in patients with schizophrenia (Bramon et al., 2004). Blackwood et al. (2001) found carriers of the translocation to have deficits in their P300 ERP characteristic of schizophrenia, with deficits existing in translocation carriers without any psychiatric diagnosis. While it should be noted that this study considered only 12 translocation carriers, of which only 3 were not diagnosed with a psychiatric condition, these results are consistent with other studies that have identified abnormal P300 ERPs in the relatives of patients with schizophrenia (Bramon et al., 2005). Moreover, Shaikh et al. (2011) have reported an association between *DISC1* variants and P300 amplitude and latency in a sample that included both patients with schizophrenia and control subjects.

1.7.4.2.2 Overview of genetic studies implicating variation in DISC1 *in schizophrenia and bipolar disorder*

Following the identification of *DISC1*, genetic linkage and association studies have identified variation in *DISC1* as playing a more general role in conferring risk for psychiatric illness. Positive association and linkage studies have provided further support for the role of variation in *DISC1* in schizophrenia, bipolar disorder, and recurrent major depression, and extended the range of *DISC1*-associated conditions to include autism and Asperger's syndrome (Bradshaw and Porteous, 2012; Chubb et al., 2008). Moreover, variation in DISC1 has been associated with aspects of cognitive function, including cognitive ageing (Burdick et al., 2005; Carless et al., 2011; Palo et al., 2007; Thomson et al., 2005); these results are highly pertinent in the context of the cognitive deficits central to the schizophrenia phenotype (Barch and Ceaser, 2012). While the accumulating evidence presents a strong case for the pathogenic role of certain *DISC1* variants, it is important to note that not all association studies have been positive (Arai et al., 2007; Devon et al., 2001; Hotta et al., 2011; Kockelkorn et al., 2004; Sanders et al., 2008; Zhang et al., 2005). Genome-wide association studies (GWASs) have not identified significant association between variants within the *DISC1* gene and schizophrenia (Athanasiu et al., 2010; Need et al., 2009; O'Donovan et al., 2008; Ripke et al., 2011; Sullivan et al., 2008), although Sullivan et al. (2008) identified several single nucleotide polymorphisms (SNPs) that showed a trend for association but did not remain significant after correction for multiple testing. GWAS findings from bipolar disorder samples echo those from schizophrenia: in the largest GWAS carried out for bipolar disorder to date, Sklar et al. (2011) did not detect association to any variants within the *DISC1* gene. The failure to detect association to *DISC1* by GWAS, however, does not negate the evidence supporting DISC1's involvement in psychiatric illness arising from other methods: allelic heterogeneity may render *DISC1* risk variants too individually rare to be detected by GWAS and phenotypic heterogeneity might make the grouping together of affected individuals as a unitary case group invalid.

1.7.4.2.3 Structure of the DISC1 *gene and the encoded protein*

In recent years it has become apparent that the *DISC1* gene gives rise to a huge variety of transcripts through alternative splicing: Nakata et al. (Nakata et al., 2009) have identified more than 50 DISC1 splice variants in the brain, some of which arise from splicing to the upstream gene *TSNAX*. Excluding those transcripts that use the TSNAX transcription start

site (TSS), all of the DISC1 transcripts have been shown to share the same TSS. The gene gives rise to 10 alternative termination codons.

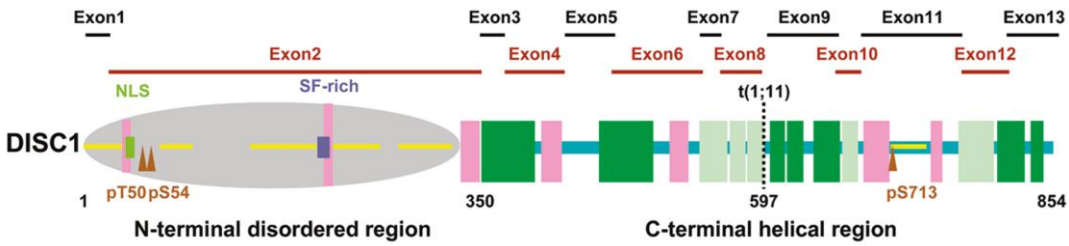


Figure 1.2. Schematic of the DISC1 structure. The DISC1 secondary structure has been predicted to comprise a predominantly disordered N-terminal region and an alpha helical coiled-coil containing C-terminal tail (Chubb et al., 2008). The N-terminal head domain contains only two conserved regions, which encode a nuclear localisation signal (NLS) and serine-phenylalanine- (SF) rich motif. Mapped secondary structure is indicated as follows: alpha helices are indicated in pink, coiled-coil helices are indicated in dark green, and ambiguous helices are indicated in light green. Yellow horizontal lines indicate predicted disordered regions. Known and putative phosphorylation sites are indicated by brown triangles and the location of the t(1;11) translocation breakpoint is indicated by a vertical dashed line. Numbers below the schematic indicate amino acid numbers and the thirteen exons encoding full-length DISC1 are shown above the schematic. Figure taken from Soares et al. (2011).

Understanding of the structure of encoded protein remains relatively poor. The full-length transcript encodes a protein 854 amino acids in length, which is believed to comprise an N-terminal "head" domain and an alpha helical coiled-coil containing C-terminal tail (Chubb et al., 2008; Soares et al., 2011) (figure 1.2). Traditionally, the N-terminal head domain has been thought to be formed of, approximately, the first 350 amino acids, and the C-terminal coiled-coil domain to comprise amino acids ~350-854. The N-terminal head domain has often been referred to as "globular" (Chubb et al., 2008); however, a recent study by Soares et al. (2011) suggests that this region may be predominantly disorganised and suggests a revised definition of the N-terminal region such that it is considered to include amino acids 1-326. The N-terminal domain shows little homology with known protein structures, containing only two recognisable motifs: a nuclear localisation signal (Ma et al., 2002) and serine-phenylalanine-rich motif (Taylor et al., 2003). The alpha helical coiled-coil C-terminal domain, in contrast, shows greater conservation amongst orthologs (Millar et al., 2000; Ma et al., 2002; Taylor et al., 2003).

1.7.4.2.4 DISC1: interactors, functions, and their relationship to schizophrenia and bipolar disorder

The discovery that DISC1 interacts with a large number of proteins has led to the idea that DISC1 acts as a molecular "hub" (Camargo et al., 2007). Understanding of the functions of DISC1 has, therefore, largely been shaped through the identification and investigation of its interactors.

In 2007, Camargo et al. carried out a yeast two hybrid screen to identify interactors of DISC1. This resulted in the identification of 289 interactors, several of which were already known, in screens with the full-length protein. Analysis of the functions of the identified interactors revealed enrichment for proteins involved in synaptic function and dendritic/axonal morphology. Subsequent studies have confirmed several of the interactors identified by Camargo et al. (2007) and supported the idea that DISC1 plays a key role at the synapse and in neurodevelopment.

Many of DISC1's interactors are involved in microtubule function: DISC1 is known to interact with several microtubule-associated proteins, such as nuclear distribution protein nudE1 (NDE1), NDE-like 1 (NDEL1), lissencephaly protein 1 (LIS1), microtubule-interacting protein associated with TRAF3 (MIPT3), microtubule-associated protein 1A (MAP1A), and dynactin (Brandon and Sawa, 2011). Further supporting DISC1's role in microtubule function, DISC1 has been shown to localise to the centrosome, where microtubules are organised (Morris et al., 2003). A recent study demonstrated that the phosphorylation of DISC1 underlies the developmental switch from neuronal proliferation to neuronal migration (Ishizuka et al., 2011). Unphosphorylated DISC1 interacts with GSK3β to regulate the canonical Wnt signalling pathway, a process that supports the proliferation of mitotic progenitor cells. Upon phosphorylation at serine 710, DISC1 recruits Bardet-Biedl syndrome proteins to the centromere, a process required for migration. Thus, altered DISC1 function could contribute to the cell migration abnormalities observed in schizophrenia and bipolar disorder via disruption of microtubule function. Consistent with notion, studies in animal models have revealed mutant versions of DISC1 to result in aberrant cell migration (Kamiya et al., 2005; Young-Pearse et al., 2010). Knockdown of *Disc1* has also been reported to affect migration: knockdown of *Disc1* in newly generated cells in the adult hippocampus has been shown to result in over-extended migration (Duan et al., 2007; Enomoto et al., 2009; Kim et al., 2009). Recently, Steinecke et al. (2012) showed, in mice,

that knockdown of *Disc1* in the medial ganglionic eminence, either *ex utero* or *in utero*, resulted in a reduction in the number of interneurons that reached their target destinations. This finding highlights a mechanism by which altered DISC1 function might contribute to GABAergic interneuron abnormalities observed in schizophrenia and bipolar disorder.

As well as the *Disc1* knockdown-induced deficits in neuronal migration, abnormalities in neuronal morphology have also been reported following knockdown. Duan et al. (2007) found neurons expressing a short hairpin construct used to knockdown *Disc1* expression to have enlarged soma and enhanced dendritic outgrowth. In migrating interneurons, Steinecke et al. (2012) found *Disc1* knockdown to result in an increase in the length of the leading process and a reduced number of side branches. Further supporting a role for DISC1 in the regulation of cell morphology, mice carrying mutant versions of *Disc1* in which point mutations have resulted in amino acid substitutions (L100P and Q31L) have been shown to have a decrease in dendritic length in the frontal cortex (Lee et al., 2011). These findings highlight the possibility that the reduction in neuropil seen in the brains of a proportion of schizophrenic patients might result, in some cases, from aberrant DISC1 function.

Evidence is emerging to suggest that, in common with NRG1, DISC1 is involved in axon myelination. In zebrafish in which *disc1* has been knocked down, oligodendrocyte development is impaired (Wood et al., 2009). Consistent with this finding, mice expressing an inducible C-terminal truncated version of *DISC1*, which is expressed predominantly in the forebrain, show premature oligodendrocyte differentiation and increased proliferation of oligodendrocyte progenitors (Katsel et al., 2011). Moreover, expression of the mutated protein affected the expression of Nrg1 and the Nrg1 receptors ErbB3 and ErbB4, suggesting that some of Disc1's effects on oligodendrocytes might be mediated through altered Nrg1 signalling. In humans, a *DISC1* variant previously associated with schizophrenia was found to affect white matter integrity, suggesting that *DISC1* variants that alter white matter development might confer risk for schizophrenia (Sprooten et al., 2011b).

Together with DISC1's involvement in dendritic outgrowth and myelination, studies showing that DISC1 plays a key role at the synapse highlight DISC1 as playing a central role in neuronal connectivity. DISC1 is known to be expressed at the synapse, particularly at excitatory synapses, in the human brain (Kirkpatrick et al., 2006) and subcellular fractionation has indicated that DISC1 is enriched in the post-synaptic density (Hayashi-Takagi et al., 2010). Consistent with its synaptic location, many of DISC1's protein

interactors are synaptic proteins, including Kalirin-7, PSD-95 (Hayashi-Takagi et al., 2010) and TNIK (Wang et al., 2011b). Wang et al. (2011) have shown DISC1 binding of TNIK to result in a reduction in TNIK activity. Inhibition of TNIK activity was found to result in a decrease in the expression of several post-synaptic proteins, including PSD-95 and the GluR1 AMPA receptor subunit, as well as decreasing AMPA receptor-mediated miniature excitatory post-synaptic currents (mEPSCs). Consistent with these findings, in neuronal cultures in which DISC1 has been knocked down, increases in the surface expression of GluR1 and the frequency of mEPSCs have been observed (Hayashi-Takagi et al., 2010). These findings suggest that DISC1 plays a key role, at least in part via its interaction with TNIK, in regulating synaptic formation and strength. Moreover, abnormalities in glutamatergic neurotransmission are believed to be central to the pathophysiology of schizophrenia, indicating a mechanism by which altered DISC1 function might play a causal role.

In addition to DISC1's role in regulating glutamatergic neurotransmission, there is evidence for the involvement of DISC1 in the function of two other neurotransmitter systems implicated in schizophrenia and bipolar disorder: the dopamine and GABA systems. Mice carrying L100P mutant *Disc1* have been found to have enhanced dopamine function: these mice show an enhanced response to amphetamine, have an increase in the number of striatal dopamine receptors, and show behavioural abnormalities that are ameliorated by the dopamine D2 receptor antagonist haloperidol (Lipina et al., 2010). Further evidence for DISC1's role in dopamine function comes from the finding that mice in which *Disc1* was knocked down transiently in the pre- or peri-natal stages show abnormalities in postnatal mesocortical dopaminergic maturation (Niwa et al., 2010). Niwa et al. (2010) also detected a reduction in parvalbumin, a marker of fast-spiking GABAergic interneurons following *Disc1* knockdown. Similarly, mice expressing a dominant negative C-terminal truncated form of *DISC1* have been shown to have reduced parvalbumin reactivity in the PFC (Hikida et al., 2007).

The interaction of DISC1 with GSK3β was mentioned previously in the context of the role of DISC1 in neuronal proliferation and migration. As GSK3β is a key target of lithium (Klein and Melton, 1996) and is involved in several neurodevelopmental processes (Hur and Zhou, 2010) this interaction has been the subject of much interest. In zebrafish, Disc1 has been shown to regulate both canonical, β-catenin-mediated, and non-canonical Wnt signalling during embryogenesis (De Rienzo et al., 2011) and Singh et al. (2011) have demonstrated

that *DISC1* polymorphisms can disrupt Wnt/GSK3β signalling and cortical development (Singh et al., 2011). It should, however, be noted that two of the polymorphisms assessed by Singh et al. (2011) were common variants, suggesting the possibility for functional compensation in the majority of individuals. GSK3β binding sites on DISC1 overlap with binding sites for the enzyme phosphodiesterase 4B (PDE4B) (Mao et al., 2009; Murdoch et al., 2007). As phosphodiesterases are the only enzymes involved in the hydrolysis of cyclic adenosine 3',5'-monophosphate (cAMP) , a second messenger involved in learning, memory and mood (Benito and Barco, 2010), they are of great relevance to schizophrenia and bipolar disorder, as well as other psychiatric disorders. Indeed, *PDE4B* polymorphisms and altered expression have been associated with schizophrenia (Fatemi et al., 2008b; Guan et al., 2012; Numata et al., 2008) and a translocation affecting the gene has been identified in an individual with schizophrenia and a relative with a psychotic illness (Millar et al., 2005). Recent evidence supports the existence of functional convergence between GSK3β and PDE4B: Carlyle et al. (2011) demonstrated the coordinated modulation of cAMP signalling by DISC1, GSK3β and PDE4B, and Lipina et al. (2012) have shown combined subthreshold doses of rolipram (a PDE4 inhibitor) and TDZD-8 (a GSK3 inhibitor) to be effective in ameliorating behavioural abnormalities in the L100P *Disc1* mouse model.

1.7.4.2.5 Evidence for the role of altered DISC1 expression in schizophrenia and bipolar disorder

Initial evidence for the role of altered DISC1 expression in the pathogenesis of psychiatric disorders came from the study of lymphoblastoid cell lines derived from carriers of the t(1;11) translocation. An observed reduction in DISC1 protein expression of 50% in these cell lines highlighted reduced DISC1 expression as a risk factor for the onset of illness in translocation carriers (Millar et al., 2005). The findings from subsequent studies investigating the role of DISC1 expression in psychiatric illness have, however, been equivocal. Some studies (Maeda et al., 2006; Nakata et al., 2009; Sawamura et al., 2005) have identified altered DISC1 expression in individuals diagnosed with psychiatric illnesses, whereas others (Dean et al., 2007; Rastogi et al., 2009) have not identified any such differences. One study (Lipska et al., 2006) reported no change in DISC1 mRNA expression in the dorsolateral prefrontal cortex or the hippocampus in patients with schizophrenia, but identified a modest (~20%) but significant increase in hippocampal DISC1 protein expression in patients. Furthermore, this study reported a positive correlation between DISC1 mRNA and protein expression in the hippocampus. Factors such as differences in the

brain regions assessed, differences in the DISC1 transcripts quantified, and phenotypic heterogeneity might have contributed to the lack of consensus between these studies. Moreover, if altered DISC1 expression is a risk factor for psychiatric illness, it is likely that it will not (i) always result in the onset of psychiatric illness or (ii) be present in affected individuals diagnosed, thus reducing statistical power to detect a significant association. Furthermore, it is important to note that the demonstration of altered mRNA expression does not necessarily indicate a change in protein expression. These issues are discussed further in section 3.5.

Whilst not a direct assessment of gene expression, the presence of copy number variation in a region is at least suggestive of altered expression of the affected genes. Two studies have reported the presence of CNVs involving the *DISC1* locus in individuals diagnosed with autism spectrum disorders: Williams et al. (2009) reported the case of a boy with autism spectrum disorder who was found to have a deletion involving *DISC1*, *DISC2*, and *TSNAX*, the gene located immediately 5' of *DISC1*, and Crepel et al. (2010) identified a microduplication involving seven genes, including DISC1, in two brothers with autism and mild mental retardation. In a genome-wide association analysis of CNVs with schizophrenia, Glessner et al. (2010) identified an association between large CNVs (≥100kb) affecting *DISC1* and schizophrenia. The key question, however, when discussing CNVs in relation to gene expression is to what extent do changes at the DNA level resulting from CNVs result in altered gene expression? Assessment of the relationship between CNVs and gene expression in cell lines from individuals in the HapMap project revealed that CNVs could account for almost 18% of the variation in mRNA expression (Stranger et al., 2007). The relationship between CNVs and protein expression may, however, be less direct: a study in cancer cells revealed that in most cases there is no correspondence between gene copy number and the level of protein expression, although the protein expression of some oncogenes and tumour suppressor genes was found to be affected by gene copy number (Geiger et al., 2010). In light of these findings, it would be of interest to investigate the relationship between CNVs and gene expression in individuals diagnosed with psychiatric illnesses.

1.8 Summary

To summarise, psychiatric disorders, such as schizophrenia and bipolar disorder, are severe conditions, often poorly treated by currently available medications, which represent a considerable economic burden. While a genetic basis for both schizophrenia and bipolar

disorder is well-established, and several candidate genes have been identified by classical genetic techniques, few causative variants have been uncovered. The frequent implication of variants in non-coding regions of the genome by association studies, observations of altered gene expression in affected individuals and the growing evidence for the aberrant function of mechanisms involved in the regulation of gene expression, such as miRNA-mediated regulation, in affected individuals suggest the investigation of gene expression to be a worthwhile endeavour in furthering understanding of the aetiology of these conditions.

1.9 Aims of this thesis

The aims of this thesis were as follows:

1. To assess association of six intronic SNPs previously found to form two three-SNP haplotypes associated with either schizophrenia or schizophrenia and bipolar disorder (Thomson et al., 2007) in two novel case-control samples and to perform combined analyses of the new samples with the original sample assessed in Thomson et al. (2007).

2. To use a combination of bioinformatics and information available in the published literature to define a candidate promoter region for *DISC1* and to design a series of nested promoter fragments for assessment in the dual luciferase assay.

3. To determine the transcriptional activity of the *DISC1* promoter fragments and, on the basis, of this information to design further fragment(s) for empirical assessment.

4. To use publicly available data to identify transcription factors potentially involved in the regulation of DISC1 expression and to experimentally assess their effects on *DISC1* promoter activity using the dual luciferase assay.

Chapter 2

Association analysis of *Neuregulin 1* candidate regions in schizophrenia and bipolar disorder

Chapter 2: Association analysis of *Neuregulin 1* candidate regions in schizophrenia and bipolar disorder

2.1 Introduction

Neuregulin 1 (*NRG1*) is one of the best supported candidate susceptibility genes for schizophrenia. Evidence for its involvement in the disorder comes both from genetic linkage and association studies (see section 1.7.4.1.4) and from the known functions of *NRG1*, which mirror many aspects of schizophrenia (see section 1.7.4.1.3). In 2007, (Thomson et al., 2007) carried out association analysis to assess the role of variation in the *NRG1* gene in conferring risk for schizophrenia and bipolar disorder in a sample from the Scottish population. Prior to this study, one study had identified *NRG1* as a susceptibility gene for bipolar disorder (Green et al., 2005), thus adding *NRG1* to a growing list of genes implicated in both schizophrenia and bipolar disorder (Owen et al., 2007).

In order to assess the role of *NRG1* in schizophrenia and bipolar disorder, Thomson et al. (2007) analysed the association of haplotype-tagging SNPs across the entire *NRG1* gene in a sample of patients with schizophrenia and bipolar disorder and healthy controls drawn from the Scottish population. This represented the first study to select markers systematically, based on linkage disequilibrium (LD) across the whole of *NRG1*, thus permitting an unbiased assessment of the existence and location of susceptibility variants. Thomson et al. (2007) selected haplotype-tagging SNPs to tag haplotypes of greater than 10% frequency in the HapMap Data Release 7 using data from the CEPH trios of Utah residents with ancestry from northern and western Europe (CEU).

Using a Nyholt-adjusted preliminary significance threshold of $p = 0.0016$, Thomson et al. (2007) identified haplotypes that showed significant association with schizophrenia and/or bipolar disorder in two regions of the *NRG1* gene (figure 2.1). The first region, region A, was located at the 5' end of the gene, overlapping the majority of SNPs and haplotypes identified as being associated with schizophrenia in previous studies, including HapICE, the *NRG1* haplotype identified as being associated with schizophrenia in the first association study to identify variation in *NRG1* as a risk factor for schizophrenia (Stefansson et al., 2002). The second region, region B, was located at the less well studied 3' end of the gene.

Region A was found to contain multiple haplotypes that were nominally associated with schizophrenia and/or bipolar disorder. The most significant of these, which shall be henceforth referred to as "the region A haplotype", was a three-SNP haplotype that was associated with schizophrenia in both global (pg = 0.0043) and individual (pi = 0.00032) tests of haplotype association. This haplotype was not associated with bipolar disorder (pg = 0.40, pi = 0.084), and association in the combined case group fell just short of the Nyholt-adjusted significance threshold (pg = 0.0080, pi = 0.0017). This haplotype comprised SNPs rs1503491, rs553950, and rs327329 (T-T-T), and was present with an estimated frequency of 1.3% in schizophrenia, and was not predicted to occur in individuals with bipolar disorder or control individuals. Following correction for multiple testing by permutation analysis, this haplotype no longer showed significant association to schizophrenia.

Region B harboured the most significant haplotype in Thomson et al.'s (2007) study. This haplotype was significantly associated with schizophrenia (pg = 0.0078, pi = 0.00014), nominally associated with bipolar disorder (pg = 0.13, pi = 0.0022), and showed the most significant association in the combined case group (pg = 0.0034, pi = 0.000062). Association to this haplotype survived permutation correction in both the schizophrenia and the combined case group (pi = 0.024 and pi = 0.016).

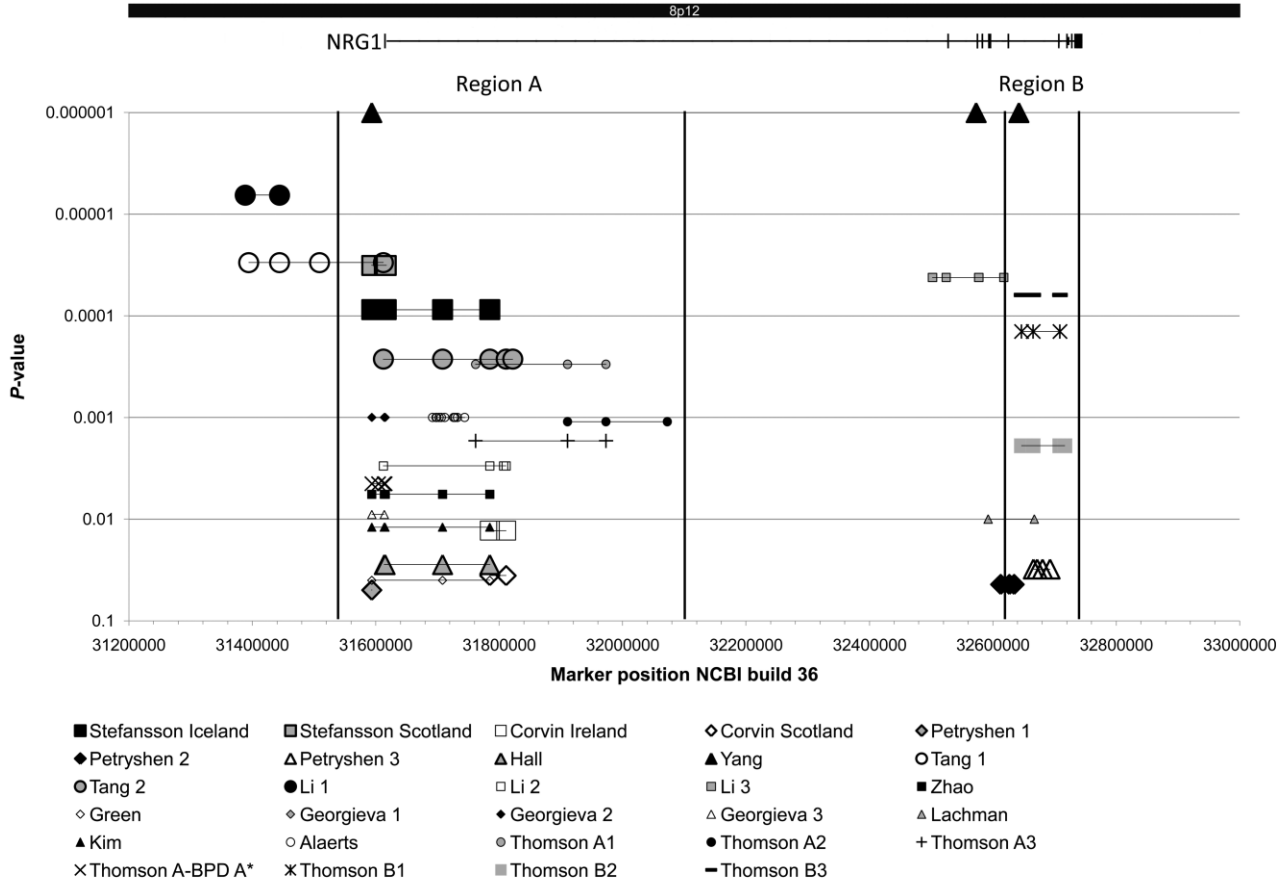


Figure 2.1. Schematic showing the most significant individual haplotype *p*-values from previous association studies of *NRG1* and schizophrenia and/or bipolar disorder. The significance of each haplotype is indicated by its location on the y-axis and the genomic position of the haplotype, according to NCBI build 36, is indicated by position on the x-axis. Vertical lines indicate the boundaries of regions A and B defined by LD blocks. All SNPs and microsatellites are within chromosome 8p12. This figure has been published previously in Walker et al. (2010).

The region A haplotype spans a region in intron 1, and the 5' boundary of region A, defined by LD, extends into the putative *NRG1* promoter region. Region B also spans a largely intronic region, although it does contain the coding sequence of the sensory and motor neuron-derived factor (SMDF) isoform, and the 3' exons of all other isoforms. As discussed in the introduction (section 1.7.4.1.5), there is evidence that the dysregulation of *NRG1* expression might play a pathogenic role in schizophrenia: association studies have identified a paucity of SNPs with an effect on the encoded protein; altered *NRG1* expression has been identified in the post-mortem brains of patients with schizophrenia; and some schizophrenia associated variants show association with *NRG1* expression level (Harrison and Law, 2006). In this context, it was hypothesised that the region A and region B haplotypes might confer risk via a variant that has an effect on *NRG1* expression by altering transcription, translation, splicing, or RNA and/or protein stability.

The work presented in this chapter has previously been published (Walker et al., 2010).

2.2 Aims

The aims of this chapter were to confirm whether a variant contained within, or in LD with, (i) the region A and (ii) the region B haplotypes confers susceptibility to (i) schizophrenia, (ii) bipolar disorder, or (iii) the combined schizophrenia and bipolar disorder case group with a view to determining whether haplotypes showing replicated association might confer susceptibility to schizophrenia and/or bipolar disorder by tagging variants that disrupt NRG1 regulatory mechanisms. This was achieved by assessing association in two independent case-control samples from the Scottish (n = 307 control subjects, 303 schizophrenic patients, and 239 bipolar disorder patients) and German populations (n = 397 control subjects, 396 schizophrenic patients, and 400 bipolar disorder patients). In addition, association was assessed in combined samples comprising (i) the Scottish sample assessed by Thomson et al. (2007), and the novel Scottish sample analysed in this chapter (n = 765 control subjects, 682 schizophrenic patients and 601 bipolar disorder patients), and (ii) the two Scottish samples and the German sample (n = 1162 control subjects, 1078 schizophrenic patients and 1001 bipolar disorder patients). The analysis of the combined samples was carried out in order to assess association in a larger sample, which would, in theory, have greater power to detect common variants of small effect size.

2.3 Materials and methods

2.3.1.1 Scottish case-control sample (Scottish 2)

Individuals with bipolar disorder or schizophrenia (table 2.1), according to Diagnostic and Statistical Manual of Mental Disorders (4th Edition) (DSM–IV) and/or International Classification of Diseases 10th edition (ICD-10) criteria were recruited from inpatient and outpatient services at psychiatric hospitals in South East and South Central Scotland, including the Royal Edinburgh Hospital, and the Ravenscraig Hospital, near Inverclyde. Diagnoses were reached by consensus between two psychiatrists (Professor Douglas Blackwood and Dr. Walter Muir) and were based on performance on the Schedule for Affective Disorder and Schizophrenia - Lifetime version (SADS-L) or the Structured Clinical Interview for DSM-IV (SCID), interviews, and a detailed review of medical case notes.

Control subjects were recruited from the same population, with the majority (>80%) being recruited through the Scottish National Blood Transfusion Service, which only accepts unmedicated individuals who do not have a chronic illness. The remaining control subjects were recruited from hospital staff or the general population, and were screened to exclude individuals currently taking medication, or with a family history of psychiatric illness.

2.3.1.2 German case-control sample

The German case-control sample (table 2.1) was obtained through collaboration with Dr. Sven Cichon and colleagues (Department of Genomics, Life and Brain Center, University of Bonn, Germany). Individuals with a lifetime diagnosis of bipolar disorder 1 or schizophrenia, according to DSM-IV criteria, were recruited from consecutive admissions to the inpatient unit of the Department of Psychiatry and Psychotherapy of the University of Bonn and of the Central Institute of Mental Health in Mannheim. Final diagnoses were reached using a consensus best estimate procedure (Leckman et al., 1982), based on medical records, family history, and information obtained through a structured clinical interview for and Statistical Manual of Mental Disorders (3rd Edition, Revised) (DSM-III-R) (SCID-1) (Spitzer et al., 1992).

Control subjects were recruited from the Bonn region of Germany by Dr. Marcella Rietschel. Both the patient and control samples were of German ancestry, extending back at least three generations.

Sample	Control	SCZ	BPD	Total
Scottish 1	458*	386	368	1212
Scottish 2	307	303	239	849
German	397	396	400	1193
Merged Scottish**	765	682	601	2048
Merged All	1162	1078	1001	3241

Table 2.1. Numbers of participants in the Scottish 1, Scottish 2, and German case-control samples, and in the two combined samples, Merged Scottish (Scottish 1 and Scottish 2) and Merged All (Scottish 1, Scottish 2, and German). The number of control subjects, subjects diagnosed with schizophrenia (SCZ) and subjects diagnosed with bipolar disorder (BPD) in each study is shown, together with the total. *corrected from Thomson et al. (2007), ** the Merged Scottish totals for schizophrenia (SCZ) and bipolar disorder (BPD) are not equal to the sum of the Scottish 1 and Scottish 2 samples as 13 subjects from the Scottish 1 sample were excluded from the merged analyses due to a subsequent change in their diagnostic status.

2.3.1.3 Combined samples

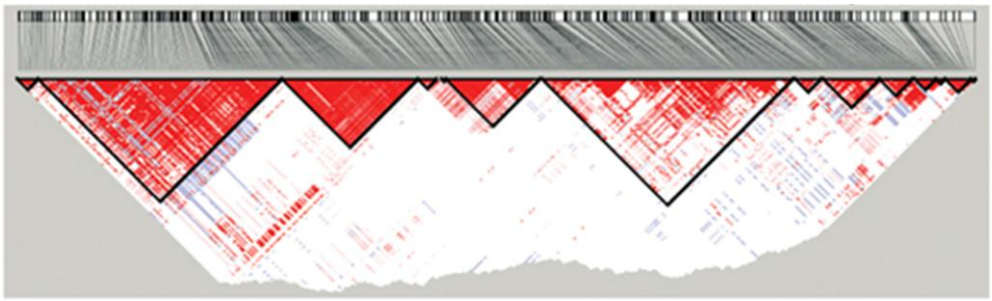
Two combined sample sets were created: firstly, the Scottish 1 sample was combined with the Scottish 2 sample to create a combined Scottish dataset (Merged Scottish); and secondly, the two Scottish samples were combined with the German sample (Merged All). The statistical program SPSS 14.0 was used to combine the datasets.

2.3.2 Selection of SNPs

SNPs were selected for replication analysis on the basis that they had previously been found to form the two three-SNP haplotypes showing the most significant association with schizophrenia and/or bipolar disorder in a Scottish sample (Thomson et al., 2007). This sample shall henceforth be referred to as the Scottish 1 sample.

These six SNPs were initially selected for genotyping in the Scottish 1 population as they are haplotype-tagging SNPs; that is, together with the other 34 SNPs genotyped by Thomson et al. (2007), they efficiently capture genetic variation across the *NRG1* locus. Thomson et al. (2007) defined the LD structure of *NRG1* using HapMap Data Release 7 using data from the CEPH trios of Utah residents with ancestry from northern and western Europe (CEU). Haplotypes were defined using the solid spine of LD algorithm (*D*'>0.8) and adjacent blocks were joined where the multiallelic *D*' exceeded 0.85. Haplotype-tagging SNPs (MAF > 10%) were then selected using Haploview. Of the six SNPs genotyped in the present study, the three region A SNPs are located in haplotype blocks two and three, and the three region B SNPs are located in haplotype blocks nine and ten (figure 2.2).

A



B

Block	1	2	3	4	5	6	7	8	9	10	11	12	13	14
SNP	1-6	7-9	10-11	12	13	14-16	17-31	32	33-35	36-37	NT	NT	38-39	40

Figure 2.2. LD map of the *NRG1* locus. A. The region chr8: 31,52, 000–32,854,000 (hg17) was downloaded from HapMap into Haploview (CEPH trios from Utah with western or northern European ancestry (CEU), Data Release 16c, June 2005) and the LD map generated from 409 SNPs of >10% minor allele frequency. LD was calculated by Haploview version 2.5 using solid spine of LD>0.8 and adjacent blocks were joined where the multiallelic D'>0.8, resulting in a total of 14 haplotype blocks across the *NRG1* locus. **B**. Table indicating the haplotype block location of the 40 haplotype-tagging SNPs genotyped by Thomson et al. (2007). These SNPs were selected to tag haplotypes of greater than 10% frequency in the HapMap data release 7). The SNPs genotyped in the present study are located in haplotype blocks two and three (region A SNPs) and nine and ten (region B SNPs). Figure adapted from Thomson et al. (2007).

2.3.3 Genotyping and quality control

2.3.3.1 Scottish 2 sample

Genotyping was performed at the Genetics Core of the Wellcome Trust Clinical Research Facility, at the Western General Hospital, Edinburgh, using the Illumina BeadArray™ platform. Probes and primers were designed by Illumina. In addition to the individuals in the Scottish 2 sample, some individuals who had previously been genotyped as part of the Scottish 1 sample were included in this genotyping set. As the initial genotyping of the Scottish 1 sample was carried out using either the Illumina BeadArray™ platform or TaqMan assays on an ABI 7900HT Fast Real-Time PCR System (Applied Biosystems), this provided an opportunity to assess the replicability of genotyping on two different occasions using partially overlapping methods. Andrea Christoforou (AC) used the Illumina BeadStudio software to make genotype calls, discarding data points with a GenCall score of less than 0.25. AC then clustered the SNPs, excluding samples with a genotype call rate of less than 80%. SNPs with a locus success rate of less than 90% and samples with a genotyping success rate less than 90% were excluded from the analysis. AC then reformatted the genotype data for input into Dr. Naomi Wray's Fortran program *BasicAS (Basic Association Study*; Wray, unpublished*)*, such that it included sample ID, gender, diagnosis, and genotype.

After AC had formatted the data for input to *BasicAS*, I identified individuals who had previously been genotyped as part of the Scottish 1 sample, checked for discrepancies in their gender, genotype, and diagnostic data and when no discrepancies were identified removed duplicated samples from the dataset. No discrepancies were identified for either gender or genotype; however where diagnostic status had changed from schizophrenia to bipolar disorder, or vice versa, individuals were included in the Scottish 2 sample using their most recent information, and excluded from the Scottish 1 sample on forming the combined (Merged Scottish and Merged All) samples. When diagnostic status was either uncertain or had changed to unipolar depression, individuals were not included in the Scottish 2 sample, and were excluded from the Scottish 1 sample when the combined samples were formed. I then used *BasicAS* to perform a test of Hardy-Weinberg Equlibrium (HWE), as described below (section 2.3.4.1).

2.3.3.2 German sample

The three region B SNPs were genotyped in the German sample in 1202 unrelated individuals using the high-throughput Illumina BeadArray™ platform at the Department of Genomics, Life, and Brain Centre at the University of Bonn, Germany. Initial quality control measures were implemented at the University of Bonn by Drs. Sven Cichon (SC), Axel Hillmer (AH), and Per Hoffman (PH): genotypes between duplicate samples included in the dataset were compared to assess the consistency of genotype calls; samples with a genotype call rate of less than or equal to 50% were excluded (nine samples); and SNPs that failed to genotype were removed (five SNPs). None of the three region B SNPs were affected by these measures. AC reformatted the data, using Progeny (www.progenygenetics.com), for entry into *BasicAS*. I then used *BasicAS* to assess HWE, as described below (section 2.3.4.1).

2.3.4 Statistical analysis

Assessments of HWE and the single-marker analysis were performed using *BasicAS*. *BasicAS* was also used to create linkage format files to be used as input for the haplotype analysis program Cocaphase version 2.4 (Dudbridge, 2003) (www.mrc-bsu.cam.ac.uk/personal/frank/software/unphased).

2.3.4.1 Assessment of Hardy Weinberg Equilibrium

To assess whether genotype frequencies at any of the six markers deviated from HWE in control individuals, a χ^2 test of independence with one degree of freedom was implemented in *BasicAS*. Deviation from HWE was considered to have occurred when $p \leq 0.01$.

2.3.4.2 Comparison of control subjects

The χ^2 test of independence, as implemented in *BasicAs* was used to test for allele frequency differences between control individuals in each of the three samples. To account for the testing of multiple markers, *p*-values were subjected to permutation analysis, and the threshold for significance set at 0.05 for corrected *p*-values.

2.3.4.3 Single-marker analysis

Differences in allele and genotype frequencies between cases and controls were assessed using the χ^2 test of independence with one and two degrees of freedom, respectively, as implemented in *BasicAS*. Odds ratios (ORs) and their 95% confidence intervals (CI) were calculated in Microsoft Excel. The significance of ORs was assessed using the Z-statistic.

2.3.4.4 Haplotype analysis

Haplotype frequency estimation and comparisons between cases and controls were carried out using Cocaphase 2.404 (Dudbridge, 2003). Haplotypes were assessed in both global and individual tests of haplotype frequency. The global test assesses the significance of the overall difference in haplotype frequencies between cases and controls at a given haplotype locus, whereas the individual test compares the case-control frequencies of a particular haplotype compared to a reference haplotype at that locus. To avoid the potential confounding effects of rare haplotypes, haplotypes with a frequency of less than or equal to 1% in both the cases and controls were grouped together for the global test of significance.

ORs and 95% CIs were calculated using the most common haplotype in the control sample as the reference. When the most common haplotype was the haplotype of interest, the second most common haplotype was used as the reference haplotype.

2.3.4.5 Multiple testing correction

In the separate analyses of the Scottish 2 and German samples, associations directly replicating those identified in the analysis of the Scottish 1 sample (Thomson et al., 2007) were accepted as significant when $p \leq 0.05$, due to the *a priori* evidence for the involvement of these SNPs. Novel associations attaining nominal significance ($p \leq 0.05$) were corrected by permutation analysis. In the merged analyses, all nominally significant associations were corrected by permutation analysis, as the merged samples were considered to be novel samples.

As permutation analysis is a resampling-based method, it maintains the correlation structure present in a dataset. This is important when correcting for multiple testing in a dataset that contains non-independent variables, such as SNPs in linkage disequilibrium (LD). For such

data, the use of methods that assume independence, such as the Bonferroni correction, would be overly conservative, resulting in inflation of the type II error rate (Nyholt, 2004).

Permutation analysis was performed using Cocaphase version 2.404 (Dudbridge, 2003). The case-control status of individuals was randomly shuffled and the test statistic re-calculated under these new conditions. This re-shuffling process was performed 1000 times, creating a random distribution of the test-statistic from the null hypothesis of no association between case-control status and genotype. Churchill and Doerge (1994) recommend the use of at least 1000 permutations for an alpha level of 0.05.

As Cocaphase only corrects for the most significant result within a certain window (i.e. single-marker, two-SNP haplotype, or three-SNP haplotype), an application developed by Omer Jilani (OJ) was used to correct for other significant results in a given window. This application compiles the 1000 permuted *p*-values produced by Cocaphase and these are used to produce a distribution to which empirically obtained *p*-values are compared and corrected. Corrections were made at two levels: the single-test level (p_{st}) and the experiment-wise level (p_{ew}). The single-test level corrects for multiple comparisons within a particular subgroup (e.g. the schizophrenic group in the German sample) and region, at the single-marker or x-SNP haplotype (i.e. two-SNP or three-SNP) level. The experiment-wise level combines *p*-values across diagnostic groups (schizophrenia, bipolar disorder, and the combined case group), and single-marker and haplotype tests to give a regionally corrected *p*-value within each sample. This was achieved by permuting each test individually and then selecting the best *p*-value from each of the 1000 permutations across all tests to create a new distribution of 1000 permuted *p*-values. No correction was made for the use of both allelic and genotypic tests, as these tests are highly correlated and should not, therefore, greatly increase the number of independent tests performed.

2.3.5 Imputation of the missing three-SNP region B haplotype

As genotyping of one of the region B SNPs, rs2919390, was unsuccessful in the Scottish 2 sample, the ability of the two-SNP region B haplotype to predict the three-SNP region B haplotype was assessed in the Scottish 1 sample. Genotype data was uploaded to Haploview v4.2, which uses the expectation-maximisation algorithm to estimate haplotypes from unphased genotype data. Haplotypes were defined using the "custom" option for defining

haplotype blocks. r^2 between the associated two-SNP region B haplotype (GC) and the associated three-SNP region B haplotype (AGC) was calculated by treating each haplotype as an "allele" and using the following formula for multi-allelic r^2, where h_{ij} is the observed frequency of a two loci haplotype comprising allele i (haplotype GC), which has frequency p_i, at the first locus, and allele j (haplotype AGC), which has frequency p_j, at the second locus:

$$r^2 = \frac{(hij - (pi \times pj))^2}{\sqrt{p_i \times (1-p_i) \times q_j \times (1-q_j)}}$$

2.4 Results

2.4.1 Selection of SNPs

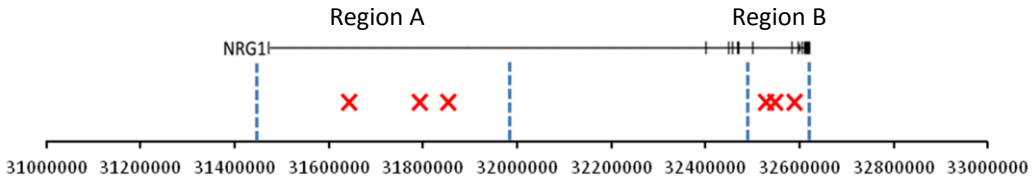


Figure 2.3. Schematic showing the location of the three region A and three region B SNPs selected for replication analysis (UCSC human genome browser, hg 19). All six SNPs were genotyped in the Scottish 2 sample; only the three region B SNPs were genotyped in the German sample. The locations of the six SNPs assessed in this study relative to the *NRG1* gene (chr 8: 31496911-32622558) are indicated by red crosses. Regions A (chr 8: 31469479-31991227) and B (32499740-32624387) are delimited by the boundaries of linkage disequilibrium blocks (dashed blue lines).

Six intronic SNPs were selected for replication analysis. These SNPs formed the two three-SNP haplotypes that were found to show the most significant association with schizophrenia and/or bipolar disorder in the Scottish 1 sample. Three SNPs formed the 5', region A, haplotype and three formed the 3', region B, haplotype (figure 2.3). All six SNPs were genotyped in the Scottish 2 sample, whereas only the three region B SNPs were genotyped in the German sample (table 2.2).

Region	SNP number	SNP name	Alleles	Location (chr. 8)	Genotyping success rate (%)	
					Scottish 2	German
A	1	rs1503491	C/T	31642106	99.88	N/A
	2	rs553950	G/T	31791362	100.0	N/A
	3	rs327329	C/T	31853498	100.0	N/A
B	4	rs2919390	A/G	32526955	N/A*	99.33
	5	rs6988339	C/T	32545916	99.76	99.83
	6	rs3757930	C/T	32589118	99.76	99.66

Table 2.2. Details of the three region A and three region B SNPs selected for replication in the Scottish 2 sample and the German sample (region B only). SNPs are listed together with their rs numbers and their genomic coordinates (UCSC human genome browser, hg 19). N/A = not available. *data discarded due to poor GenCall score (< 0.25).

2.4.2 Genotyping and quality control

2.4.2.1 Scottish 2 sample

The six SNPs (table 2.2) were genotyped in the Scottish sample using the Illumina BeadArray™ platform at the Genetics Core of the Wellcome Trust Clinical Research Facility, as described in section 2.3.3.1. Following the quality control measures described in section 2.3.3 (performed by AC), one of the region B SNPs (rs2919390) was excluded from the dataset due to a poor GenCall score. This left five SNPs to be included in the analyses, three in region A, and two in region B.

At this point, I received the data for the five SNPs. The dataset contained genotype information from both new individuals (n = 847) and individuals previously genotyped as part of the Scottish 1 dataset (n = 842). I searched for samples duplicated from the Scottish 1 sample and checked that the gender, genotype, and diagnostic information for these individuals was consistent with the information obtained on initial genotyping of the Scottish 1 sample in 2004. If an exact match was detected, the sample was removed from the current dataset. Mismatches were detected for thirteen duplicated samples; in all cases, the mismatch was for diagnostic status. These individuals were excluded from the merged analyses.

Markers in control individuals were examined for deviation from Hardy-Weinberg Equilibrium (HWE) using the χ^2 test of independence, as implemented in *BasicAS*. It is assumed that in a large population, under conditions of random mating, and free from the influence of selection, mutation, or migration, genotype frequencies for individual markers will be in HWE (Sham, 1998). Deviations from HWE can be indicative of inaccurate genotyping (Hosking et al., 2004). None of the five markers genotyped in the Scottish 2 sample were found to deviate from HWE (all $p > 0.05$; table 2.3).

Marker (region)	Genotype			HWE *p*-value
	11 (%)	12 (%)	22 (%)	
rs1503491 (A)	161 (52.4)	125 (40.7)	21 (6.8)	0.622
rs553950 (A)	244 (79.5)	61 (19.9)	2 (0.7)	0.385
rs327329 (A)	19 (6.2)	133 (43.3)	155 (50.5)	0.172
rs6988339 (B)	102 (33.3)	149 (48.7)	55 (18.0)	0.964
rs3757930 (B)	158 (51.5)	122 (39.7)	27 (8.8)	0.620

Table 2.3. Assessment of Hardy-Weinberg Equilibrium in control individuals of the Scottish 2 sample. Hardy-Weinberg Equilibrium (HWE) was assessed for each of the five genotyped markers using the χ^2 test of independence, as implemented in *BasicAS*. For each marker, the number and percentage of control individuals identified as carrying each genotype are shown, together with the associated *p*-value. Genotypes are indicated using a "1" to refer to the allele that occurs first alphabetically and a "2" to represent the allele that occurs second. Deviation from HWE was defined as $p \leq 0.05$.

2.4.2.2 German sample

The three region B SNPs were genotyped in the German sample using the high-throughput Illumina BeadArray platform at the Department of Genomics, Life, and Brain Centre at the University of Bonn, Germany, as described in section 2.3.3.2. All three SNPs survived the quality control measures implemented SC, AH, and PH, described in section 2.3.3.2.

At this point, I received the data for these three SNPs. Deviations from HWE were assessed in the control sample using the χ^2 test of independence, as implemented in *BasicAs*. None of the SNPs were found to deviate from HWE (all $p > 0.05$; table 2.4).

Marker (region)	Genotype			HWE *p*-value
	11 (%)	12 (%)	22 (%)	
rs2919390 (B)	139 (35.0)	182 (45.8)	76 (19.1)	0.236
rs6988339 (B)	152 (38.4)	187 (47.2)	57 (14.4)	0.966
rs3757930 (B)	199 (50.1)	155 (39.0)	43 (10.8)	0.402

Table 2.4. Assessment of Hardy-Weinberg Equilibrium in the control individuals of the German sample. Hardy-Weinberg Equilibrium (HWE) was assessed for each of the three genotyped markers using the χ^2 test of independence, as implemented in *BasicAS*. For each marker, the number and percentage of control individuals identified as carrying each genotype are shown, together with the associated *p*-value. Genotypes are indicated using a "1" to refer to the allele that occurs first alphabetically and a "2" to represent the allele that occurs second. Deviation from HWE was defined as $p \leq 0.05$.

2.4.2.3 Comparison of control groups

As the individual case-control datasets were to be combined for the merged analyses, it was important to assess whether there were differences between the two datasets. Therefore, allele frequency differences between the control subjects of each sample were assessed using the χ^2 test of independence in *BasicAS* (table 2.5). The two sets of Scottish controls (Scottish 1 and Scottish 2) were found to differ significantly on two SNPs: a region A SNP, rs553950, (uncorrected $p = 0.029$), and a region B SNP, rs6988339, (uncorrected $p = 0.011$). However, neither of these differences remained significant after correction for multiple testing by permutation analysis (rs553950 corrected $p = 0.129$; rs6988339: corrected $p = 0.053$).

Comparison/ marker (region)	Allele frequency (%)		Uncorrected *p*-value (corrected *p* – value)
Scottish 1 vs. Scottish 2	**Scottish 1**	**Scottish 2**	
rs1503491 (A)	669 (74.2)	447 (72.8)	0.553
rs553950 (A)	841 (92.6)	549 (89.4)	0.029 (0.129)
rs327329 (A)	236 (26.5)	171 (27.9)	0.550
rs6988339 (B)	580 (64.2)	353 (57.7)	0.011 (0.053)
rs3757930 (B)	625 (69.4)	438 (71.3)	0.430
Scottish 1 vs. German	**Scottish 1**	**German**	
rs2919390 (B)	489 (54.5)	460 (57.9)	0.150
rs6988339 (B)	580 (64.2)	491 (62.0)	0.357
rs3757930 (B)	625 (69.4)	553 (69.6)	0.928
Scottish 2 vs. German	**Scottish 2**	**German**	
rs6988339 (B)	353 (57.7)	491 (62.0)	0.101
rs3757930 (B)	438 (71.3)	553 (69.6)	0.491

Table 2.5. Comparison of controls groups. The allele frequencies of the control groups of each sample were compared to assess the existence of systematic differences between the samples. Comparisons were carried out using the χ^2 test of independence in *BasicAS*. For each comparison, the list of markers assessed is listed, together with the number and percentage of individuals in each sample carrying the allele that occurs first alphabetically. Uncorrected p-values for each comparison are shown. Where an uncorrected *p*-value attained the threshold for nominal significance ($p \leq 0.05$), correction for multiple testing was performed by permutation analysis and a corrected *p*-value is shown in parentheses.

To further compare the two Scottish samples, the correlation between the allele frequencies for each marker was calculated (figure 2.4). This revealed a high level of correlation (r^2 = 0.98). Based on these findings, the merged analyses were performed, although the potential confound of population differences will be considered when interpreting the results.

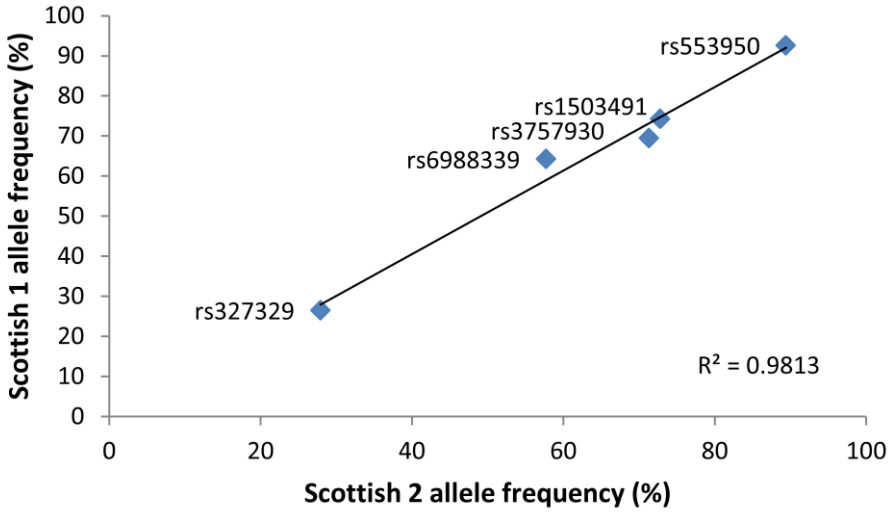


Figure 2.4. Correlation between the allele frequencies of the control individuals in the Scottish 1 and Scottish 2 samples. The correlation between allele frequencies for the five genotyped markers was assessed by calculation of the coefficient of determination, R^2, which is displayed above. Allele frequencies for the Scottish 1 sample are displayed as percentages on the y-axis and allele frequencies for the Scottish 2 sample are displayed as percentages on the x-axis. For each marker, the displayed allele frequency relates to the allele appearing first alphabetically.

2.4.3 Association analyses

Association analyses were performed to assess the association of the three region A SNPs and the region B SNPs with schizophrenia, bipolar disorder, and the combined case group (schizophrenia and bipolar disorder). Single-marker and haplotype analyses were carried out, with the two regions being analysed separately. Both regions were assessed in the Scottish 2 and the Merged Scottish samples, whereas only region B was assessed in the German and Merged All samples.

2.4.3.1 Single-marker analysis

Differences in allele and genotype frequencies between cases and controls were assessed using the χ^2 test of independence, as implemented in *BasicAS*. The results of this analysis are

displayed in table 2.6. The results of the single-marker analysis for the Scottish 1 sample, which has been analysed previously (Thomson et al., 2007), are shown for comparison.

Marker (region)	Allele frequency		*p*-value	Genotype frequency			*p*-value	HWE *p*-value
Group	**Allele 1 (%)**	**Allele 2 (%)**		**11 (%)**	**12 (%)**	**22 (%)**		
rs1503491 (A)	**C**	**T**		**CC**	**CT**	**TT**		
Scottish 1								
BPD	518 (71.5)	206 (28.5)	0.237	183 (50.6)	152 (42.0)	27 (7.5)	0.486	0.551
SCZ	578 (75.3)	190 (24.7)	0.609	220 (57.3)	138 (35.9)	26 (6.8)	0.612	0.494
BPD+SCZ	1097 (73.5)	395 (26.5)	0.702	403 (54.0)	290 (38.9)	53 (7.1)	0.836	0.933
Controls	669 (74.2)	233 (25.8)		246 (54.5)	177 (39.2)	28 (6.2)		0.607
Scottish 2								
BPD	353 (74.2)	123 (25.8)	0.615	131 (55.0)	91 (38.2)	16 (6.7)	0.827	0.970
SCZ	440 (72.6)	126 (27.4)	0.939	162 (53.5)	116 (38.3)	25 (8.3)	0.719	0.513
BPD+SCZ	793 (73.3)	289 (26.7)	0.827	293 (54.2)	207 (38.3)	41 (7.6)	0.758	0.597
Controls	447 (72.8)	167 (27.2)		161 (52.4)	125 (40.7)	21 (6.8)		0.622
Merged Scottish								
BPD	861 (72.5)	327 (27.5)	0.507	310 (52.2)	241 (40.6)	43 (7.2)	0.787	0.680
SCZ	1005(73.9)	355 (26.1)	0.790	376 (55.3)	253 (37.2)	51 (7.5)	0.505	0.353
BPD+SCZ	1865 (73.2)	683 (26.8)	0.864	686 (53.8)	494 (38.8)	94 (7.4)	0.703	0.697
Controls	1116 (73.6)	400 (26.4)		407 (53.7)	302 (39.8)	49 (6.5)		0.481
rs553950 (A)	**G**	**T**		**GG**	**GT**	**TT**		
Scottish 1								
BPD	650 (89.5)	76 (10.5)	**0.028**	291 (80.2)	68 (18.7)	4 (1.1)	**0.023**	0.990
SCZ	688 (89.6)	80 (10.4)	**0.028**	311 (81.0)	66 (17.2)	7 (1.8)	**0.009**	0.121
BPD+SCZ	1339 (89.6)	155 (10.4)	**0.012**	602 (80.6)	134 (17.9)	11 (1.5)	**0.012**	0.264
Controls	841 (92.6)	67 (7.4)		387 (85.2)	67 (14.8)	0 (0.0)		0.204
Scottish 2								
BPD	428 (89.5)	50 (10.5)	0.946	191 (79.9)	46 (19.2)	2 (0.8)	0.955	0.671
SCZ	542 (89.4)	64 (10.6)	0.989	244 (80.5)	54 (17.8)	5(1.7)	0.430	0.324
BPD+SCZ	970 (89.5)	114 (10.5)	0.964	435 (80.3)	100 (18.5)	7(1.3)	0.612	0.646
Controls	549 (89.4)	65 (10.6)		244 (79.5)	61 (19.9)	2 (0.7)		0.385
Merged Scottish								
BPD	1067 (89.5)	128 (10.5)	0.109	477 (80.0)	113 (19.0)	6 (1.0)	0.114	0.809
SCZ	1217 (89.5)	143 (10.5)	0.093	549 (80.7)	119 (17.5)	12 (1.8)	**0.013**	0.068
BPD+SCZ	2284 (89.5)	268 (10.5)	0.058	1026 (80.4)	232 (18.2)	18 (1.4)	**0.026**	0.242
Controls	1390 (91.3)	132 (8.7)		631 (82.9)	128 (16.8)	2 (0.3)		0.088
rs327329 (A)	**C**	**T**		**CC**	**CT**	**TT**		
Scottish 1								
BPD	198 (26.9)	538 (73.1)	0.840	20 (5.4)	184 (41.3)	236 (51.6)	0.881	0.077
SCZ	191 (24.8)	579 (75.2)	0.442	30 (7.8)	131 (34.0)	224 (58.2)	0.079	0.085
BPD+SCZ	389 (25.8)	1117 (74.2)	0.735	50 (6.6)	289 (38.4)	414 (55.0)	0.578	0.964
Controls	236 (26.5)	656 (73.5)		26 (5.8)	184 (41.3)	236 (52.9)		0.204
Scottish 2								
BPD	133 (27.8)	345 (72.2)	0.583	16 (6.7)	94 (39.3)	129 (54.0)	0.643	0.840
SCZ	169 (27.9)	437 (72.1)	0.988	24 (7.9)	121 (39.9)	158 (52.1)	0.562	0.901
BPD+SCZ	295 (27.2)	793 (72.8)	0.778	40 (7.4)	215 (39.7)	287 (53.0)	0.531	0.976
Controls	171 (27.9)	443 (72.1)		19 (6.2)	133 (43.3)	155 (50.5)		0.172
Merged Scottish								
BPD	321 (26.7)	881 (73.3)	0.852	36 (6.0)	249 (41.4)	316 (52.6)	0.969	0.153
SCZ	357 (26.2)	1005 (73.8)	0.561	54 (7.9)	247 (36.3)	380 (55.8)	**0.048**	0.124
BPD+SCZ	677 (26.4)	1887 (73.6)	0.645	90 (7.0)	496 (38.7)	696 (54.3)	0.265	0.899
Controls	407 (27.0)	1099 (73.0)		45 (6.0)	317 (42.1)	391 (51.9)		0.065

Table 2.6. Results of the single-marker association analysis. See page 74 for legend.

Marker (region)	Allele frequency		*p*-value	Genotype frequency			*p*-value	HWE *p*-value
Group	**Allele 1 (%)**	**Allele 2 (%)**		**11 (%)**	**12 (%)**	**22 (%)**		
rs2919390 (B)	**A**	**C**		**AA**	**AC**	**CC**		
Scottish 1								
BPD	428 (58.8)	300 (41.2)	0.079	130 (35.7)	168 (46.2)	66 (18.1)	0.189	0.365
SCZ	463 (60.8)	299 (39.2)	**0.010**	143 (37.5)	177 (46.5)	61 (16.0)	**0.036**	0.615
BPD+SCZ	891 (59.8)	599 (40.2)	**0.010**	273 (36.6)	345 (46.3)	127 (17.0)	**0.037**	0.315
Controls	489 (54.5)	409 (45.5)		134 (29.8)	221 (49.2)	94 (20.9)		0.870
German								
BPD	495 (62.3)	299 (37.7)	0.073	154 (38.8)	187 (47.1)	56 (14.1)	0.145	0.949
SCZ	451 (57.7)	331 (42.3)	0.916	124 (31.7)	203 (51.9)	64 (16.4)	0.225	0.210
BPD+SCZ	476 (60.0)	318 (40.0)	0.328	278 (35.3)	390 (49.5)	120 (15.2)	0.204	0.379
Controls	460 (57.9)	334 (42.1)		139 (35.0)	182 (45.8)	76 (19.1)		0.236
rs6988339 (B)	**A**	**G**		**AA**	**AG**	**GG**		
Scottish 1								
BPD	433 (59.0)	301 (41.0)	**0.032**	132 (36.0)	169 (46.0)	66 (18.0)	0.113	0.355
SCZ	448 (58.0)	324 (42.0)	**0.010**	133 (34.5)	182 (47.2)	71 (18.4)	**0.041**	0.529
BPD+SCZ	881 (58.5)	625 (41.5)	**0.006**	265 (35.2)	351 (46.6)	137 (18.2)	**0.026**	0.272
Controls	580 (64.2)	324 (35.8)		191 (42.3)	198 (43.8)	63 (13.9)		0.312
Scottish 2								
BPD	297 (62.1)	181 (37.9)	0.137	90 (37.7)	117 (49.0)	32 (13.4)	0.289	0.533
SCZ	338 (56.0)	266 (46.0)	0.545	97 (32.1)	144 (47.7)	61 (20.2)	0.781	0.571
BPD+SCZ	635 (58.7)	447 (41.3)	0.686	187 (34.6)	261 (48.2)	93 (17.2)	0.921	0.906
Controls	353 (57.7)	259 (42.3)		102 (33.3)	149 (48.7)	55 (18.0)		0.964
German								
BPD	469 (58.8)	229 (41.2)	0.189	141 (35.3)	187 (46.9)	57 (17.8)	0.380	0.511
SCZ	508 (64.1)	284 (35.9)	0.376	165 (41.7)	178 (44.9)	53 (13.4)	0.637	0.649
BPD+SCZ	976 (61.4)	614 (38.6)	0.795	306 (38.5)	365 (45.9)	124 (15.6)	0.839	0.382
Controls	491 (62.0)	301 (38.0)		152 (38.4)	187 (47.2)	57 (14.4)		0.966
Merged Scottish								
BPD	725 (60.4)	475 (39.6)	0.550	221 (36.8)	283 (47.2)	96 (16.0)	0.790	0.734
SCZ	782 (57.4)	580 (42.6)	**0.024**	229 (33.6)	324 (47.6)	128 (18.8)	0.085	0.480
BPD+SCZ	1506 (58.8)	1056 (41.2)	0.087	450 (35.1)	607 (47.4)	224 (17.5)	0.231	0.434
Controls	932 (61.5)	584 (38.5)		293 (38.7)	347 (45.8)	118 (15.6)		0.365
Merged All								
BPD	1195 (59.8)	803 (40.2)	0.194	362 (36.2)	470 (47.0)	167 (16.7)	0.438	0.491
SCZ	1290 (59.9)	864 (40.1)	0.216	394 (36.6)	502 (46.6)	181 (16.8)	0.464	0.328
BPD+SCZ	2483 (59.8)	1669 (40.2)	0.140	756 (36.4)	972 (46.8)	348 (16.8)	0.344	0.237
Controls	1424 (61.7)	884 (38.3)		445 (38.6)	534 (46.3)	175 (15.2)		0.477

Table 2.6. Results of the single-marker association analysis (continued). See next page for legend.

Marker (region)	Allele frequency		*p*-value	Genotype frequency			*p*-value	HWE *p*-value
Group	**Allele 1 (%)**	**Allele 2 (%)**		**11 (%)**	**12 (%)**	**22 (%)**		
rs3757930 (B)	**C**	**T**		**CC**	**CT**	**TT**		
Scottish 1								
BPD	533 (72.6)	201 (27.4)	0.160	193 (52.6)	147 (40.1)	27 (7.4)	0.332	0.891
SCZ	558 (72.3)	214 (27.7)	0.204	197 (51.0)	164 (42.5)	24 (6.5)	0.186	0.236
BPD+SCZ	1090 (72.4)	416 (27.6)	0.115	390 (51.8)	311 (41.3)	52 (6.9)	0.148	0.344
Controls	625 (69.4)	275 (30.6)		220 (48.9)	185 (41.1)	45 (10.0)		0.507
Scottish 2								
BPD	347 (72.9)	129 (27.1)	0.568	124 (52.1)	99 (41.6)	15 (6.3)	0.547	0.416
SCZ	437 (72.4)	167 (27.6)	0.694	158 (52.3)	121 (40.1)	23 (7.6)	0.868	0.980
BPD+SCZ	784 (72.6)	296 (27.4)	0.579	282 (52.2)	220 (40.7)	38 (7.0)	0.651	0.579
Controls	438 (71.3)	176 (28.7)		158 (51.5)	122 (39.7)	27 (8.8)		0.620
German								
BPD	537 (67.5)	259 (32.5)	0.348	176 (44.2)	185 (46.5)	37 (9.3)	0.105	0.241
SCZ	523 (66.4)	265 (33.6)	0.162	173 (43.9)	177 (44.9)	44 (11.2)	0.194	0.900
BPD+SCZ	1060 (66.9)	524 (33.1)	0.179	349 (44.1)	362 (45.7)	81 (10.2)	0.086	0.362
Controls	553 (69.6)	241 (30.4)		199 (50.1)	155 (39.0)	43 (10.8)		0.966
Merged Scottish								
BPD	870 (72.6)	328 (27.4)	0.168	313 (52.3)	244 (40.7)	42 (7.0)	0.242	0.551
SCZ	987 (72.5)	375 (27.5)	0.168	353 (51.8)	282 (41.4)	46 (6.8)	0.162	0.303
BPD+SCZ	1859 (72.6)	701 (27.4)	0.105	666 (52.0)	526 (41.1)	88 (6.9)	0.098	0.247
Controls	1063 (70.2)	451 (29.8)		378 (49.9)	307 (40.6)	72 (9.5)		0.402
Merged All								
BPD	1408 (70.6)	586 (29.4)	0.697	489 (49.0)	429 (43.0)	79 (7.9)	0.155	0.259
SCZ	1511 (70.3)	639 (29.7)	0.849	526 (48.9)	459 (42.7)	90 (8.4)	0.270	0.469
BPD+SCZ	2917 (70.4)	1227 (29.6)	0.738	1015 (49.0)	888 (42.9)	169 (8.2)	0.115	0.193
Controls	1616 (70.0)	692 (30.0)		577 (50.0)	462 (40.0)	115 (10.0)		0.114

Table 2.6. Results of the single-marker association analysis (continued). The association between intronic *NRG1* single nucleotide polymorphisms (SNPs) and bipolar disorder (BPD), schizophrenia (SCZ), and bipolar disorder and schizophrenia (BPD+SCZ) was assessed in case-control samples from Scotland (Scottish 2) and Germany. Association had previously been reported in the Scottish 1 sample (Thomson et al., 2007) and the results of this analysis are shown here for comparison. In addition to the single case-control samples, two combined samples were analysed, these were: (i) the Merged Scottish sample, which comprised the Scottish 1 and the Scottish 2 samples; and (ii) the Merged All sample, which comprised the Scottish 1, Scottish 2, and German samples. The SNPs analysed here were previously found to form two three-SNP haplotypes in the Scottish 1 sample: one haplotype located at the 5' end of the gene (region A) and the other located at the 3' end of the gene (region B). Five SNPs were analysed in the Scottish 2 sample, as one region B SNP (rs2919390) did not genotype successfully. Only the three region B SNPs were assessed in the German sample. Results are shown for allelic and genotypic tests of association. For both tests, *p*-values attaining the nominal significance level of $p \leq 0.05$ are highlighted in yellow. *P*-values are also displayed for the test of Hardy-Weinberg Equilibrium.

2.4.3.1.1 Scottish 2 and German case-control samples

No individual SNP or genotype was found to be associated with bipolar disorder, schizophrenia, or the combined case group in either the Scottish 2 or the German samples.

2.4.3.1.2 Merged Scottish case-control sample

Association was observed to three SNPs in the Merged Scottish sample, which comprises the Scottish 1 and the Scottish 2 samples, in the allelic or the genotypic tests of association (table 2.6).

The region B SNP rs6988339, was nominally associated with schizophrenia (uncorrected p = 0.024); however, this association did not withstand correction for multiple testing (corrected p_{st} = 0.059). The G allele was more prevalent in the schizophrenia group than the control group, conferring an increase in disease risk ($OR_{G/A}$ = 1.18, 95% CI: 1.10-1.37, table 2.7). Only a trend towards significance was observed for this allele at the genotype level ($p < 0.1$). This finding is consistent with the association to this SNP observed in the Scottish 1 sample (Thomson et al., 2007), although at a lower level of significance.

In keeping with initial findings in the Scottish 1 sample (Thomson et al., 2007), the region A SNP rs553950 was found to show nominally significant association in the schizophrenia group (uncorrected p = 0.013) and the combined case group (uncorrected p = 0.026) in the test of genotype significance, with this association remaining significant after correction for multiple testing in both groups (schizophrenia: corrected p_{st} = 0.030; combined cases: corrected p_{st} = 0.047). In both groups, genotypes containing the T allele were found to occur more frequently amongst the cases than the controls. Pair-wise comparisons between genotypic odds ratios (OR; table 2.7) suggest a recessive mode of inheritance as the TT genotype confers an increase in risk for both schizophrenia and the combined case group when compared to the GG genotype (schizophrenia: $OR_{TT/GG}$ = 6.90, 95% CI: 1.54-30.95, Z-test p = 0.006; combined cases: $OR_{TT/GG}$ = 5.54, 95% CI = 1.28-23.94, Z-test p = 0.011), whilst the GT genotype does not pose an increased risk compared to the GG genotype (schizophrenia: $OR_{GT/GG}$ = 1.07, 95% CI: 0.81-1.41, Z-test p = 0.318; combined cases: $OR_{GT/GG}$ = 1.11, 95% CI: 0.88-1.41, Z-test p = 0.185). A recessive mode of inheritance is further corroborated by the finding that the strength of the association to the TT genotype in

both the schizophrenia and the combined case groups is increased when tested under a recessive model (TT vs. GG+GT) using a χ^2 test (schizophrenia: $p = 0.004$; combined cases: $p = 0.011$).

Another region A SNP, rs327329, showed nominally significant association with schizophrenia at the genotype level (uncorrected $p = 0.048$); however, this association was no longer significant after correction for multiple testing (corrected $p_{st} = 0.129$). In keeping with Thomson et al., (2007), homozygous CC and TT genotypes are more common amongst the cases than the controls (table 2.7). However, pair-wise comparison of ORs does not reveal an obvious mode of inheritance (table 2.7).

Sample	SNP	Allele		*P*-value	OR (95% CI)	Genotype			*P*-value (corrected *p*-value)	OR (95% CI)			OR *p*-value		
		1 (%)	2 (%)			11 (%)	12 (%)	22 (%)							
	rs553950	G	T		T/G	GG	GT	TT		TT/GG	GT/GG	TT/GT	TT/GG	GT/GG	TT/GT
Scottish 1	SCZ	688 (89.6)	80 (10.4)	**0.032**	**1.46 (1.04-2.05)**	311 (81.0)	66 (17.2)	7 (1.8)	**0.023**	N/A	N/A	N/A	N/A	N/A	N/A
	BPD+SCZ	1097 (73.5)	395 (26.5)	**0.010**	**4.52 (3.44-5.95)**	602 (80.6)	134 (17.9)	11 (1.5)	**0.009**	N/A	N/A	N/A	N/A	N/A	N/A
	Controls	841 (92.6)	67 (7.4)	**0.006**		387 (85.2)	67 (14.8)	0 (0.0)							
Merged Scottish	SCZ	1217 (89.5)	143 (10.5)	0.093	1.23 (0.96-1.59)	549 (80.7)	119 (17.5)	12 (1.8)	**0.013 (0.030)**	**6.90 (1.54-30.95)**	1.07 (0.81-1.41)	**6.45 (1.41-29.44)**	**0.006**	0.318	**0.008**
	BPD+SCZ	2284 (89.5)	268 (10.5)	0.058	1.24 (0.99-1.54)	1026 (80.4)	232 (18.2)	18 (1.4)	**0.026 (0.047)**	**5.54 (1.28-23.94)**	1.11 (0.88-1.41)	**4.96 (1.13-21.74)**	**0.011**	0.185	**0.017**
	Controls	1390 (91.3)	132 (8.7)			631 (82.9)	128 (16.8)	2 (0.3)							

Table 2.7. Odds ratio analysis for significant single markers and genotypes. See page 79 for legend.

Sample	SNP	Allele		*P*-value	OR 95% CI)	Genotype			*P*-value (corrected *p*-value)	OR (95% CI)			OR *p*-value		
		1 (%)	2 (%)			11 (%)	12 (%)	22 (%)							
	rs327329	**C**	**T**		**T/C**	**CC**	**CT**	**TT**		**CC/TT**	**TT/CT**	**CC/CT**	**CC/TT**	**TT/CT**	**CC/CT**
Scottish 1	SCZ	191 (24.8)	579 (85.2)	0.442	1.09 (0.87-1.36)	30 (7.8)	131 (34.0)	224 (58.2)	0.079	1.22 (0.70-2.12)	**1.33 (1.00-1.78)**	1.62 (0.92-2.87)	0.246	**0.026**	**0.049**
	Controls	236 (26.5)	656 (73.5)			26 (5.8)	184 (41.3)	236 (52.9)							
Merged Scottish	SCZ	357 (26.2)	1005 (73.8)	0.561	1.04 (0.88-1.23)	54 (7.9)	247 (36.3)	380 (55.8)	**0.048 (0.129)**	1.23 (0.81-1.88)	**1.25 (1.00-1.55)**	**1.54 (1.00-2.37**	0.163	**0.024**	**0.024**
	Controls	407 (27.0)	1099 (73.0)			45 (6.0)	317 (42.1)	391 (51.9)							
	rs6988339	**A**	**G**		**G/A**	**AA**	**AG**	**GG**		**GG/AA**	**AG/AA**	**GG/AG**	**GG/AA**	**AG/AA**	**GG/AG**
Scottish 1	SCZ	448 (58.0)	324 (42.0)	**0.010**	**1.29 (1.06-1.58)**	133 (34.5)	182 (47.2)	71 (18.4)	**0.041**	**1.62 (1.08-2.43)**	1.32 (0.98-1.78)	**1.62 (1.08-2.43)**	**0.010**	**0.034**	0.156
	Controls	580 (64.2)	324 (35.8)			191 (42.3)	198 (43.8)	63 (13.9)							
Merged Scottish	SCZ	782 (57.4)	580 (42.6)	**0.024 (0.059)**	**1.18 (1.10-1.37)**	229 (33.6)	324 (47.6)	128 (18.8)	0.085	**1.39 (1.00-2.37)**	1.19 (0.95-1.50)	1.16 (0.87-1.56)	**0.017**	0.064	0.157
	Controls	932 (61.5)	584 (38.5)			293 (38.7)	347 (45.8)	118 (15.6)							

Table 2.7. Odds ratio analysis for significant single markers and genotypes (continued). See next page for legend.

Table 2.7. Odds ratio analysis for nominally significant single markers and genotypes in the Merged Scottish sample, which comprised two Scottish case-control samples: the Scottish 1 sample, previously analysed by Thomson et al. (2007), and the Scottish 2 sample. Nominally significant associations in the allelic and genotypic tests of association ($p \leq 0.05$) were subjected to permutation analysis to correct for multiple testing. *P*-values corrected at the single-test level are displayed in parentheses. Associations attaining nominal significance are highlighted in yellow, while those remaining significant after permutation analysis (carried out in the Merged Scottish sample only) are highlighted in green. Odds ratios (OR) were calculated relative to the non-risk allele or genotype and considered significant when the lower boundary of their 95% confidence interval (CI) was not less than one, and their Z-test *p*-value was less than or equal to $p \leq 0.05$. The *p*-values for ORs satisfying these requirements are highlighted in blue and their corresponding ORs and 95% CIs are indicated in bold. Results from the analysis of the Scottish 1 sample are shown for comparison. N/A-not available (due to genotype count of 0 in control sample).

2.4.3.1.3 Merged all case-control sample

No individual SNP or genotype was found to be significantly associated with bipolar disorder, schizophrenia, or the combined case group in the Merged all case-control sample.

2.4.3.2 Haplotype analysis

Single-marker analysis is most useful when the aetiological variant has been genotyped. As SNPs in the current study were initially selected by Thomson et al. (2007) to tag blocks of LD, it is highly unlikely that they are aetiological variants. It is more probable that association to these SNPs indicates that the SNP is in LD with a disease risk variant. Haplotype analysis is likely, therefore, be more powerful than single-marker analysis as aetiological variants might be in stronger LD with the haplotype than with individual SNPs within the haplotype (Akey et al., 2001; Zaykin et al., 2002) or could be present on a specific genetic background (Cordell and Clayton, 2002). Moreover, haplotypes are more robust than single markers, as the LD between a haplotype and an aetiological variant is less likely to be affected by random drift and mutations at the marker locus (Akey et al., 2001).

Haplotype frequencies were estimated, using the expectation-maximisation (EM) algorithm, and analysed in Cocaphase. Haplotypes were analysed separately for the region A and the region B SNPs, as these regions are separated by ~0.5Mb. All possible two- and three-marker haplotypes formed by consecutive SNPs were assessed in each region in both global (*p*g) and individual (*p*i) tests of significance.

Haplotype analysis yielded three nominally significant ($p \leq 0.05$) haplotypes: two that replicate those identified in the Scottish 1 dataset, and one novel haplotype (table 2.8). Haplotypes attaining significance in the Scottish 1 dataset are shown for comparison.

Region		A			B						
SNP ID	**rs1503491**	**rs553950**	**rs327329**	**rs2919390**	**rs6988339**	**rs3757930**	**Case frequency**	**Control frequency**	***Pi* (corrected *Pi*)**	***Pg***	**OR (95% CI)**
Block no.	**2**	**3**	**3**	**9**	**9**	**10**					
Scottish 1											
SCZ	T	T	T				0.013	4.13×10^{-13}	0.000320	0.00434	N/A
SCZ				A	G		0.401	0.321	0.00120	0.00545	1.42 (1.13-1.77)
SCZ					G	C	0.331	0.249	0.00140	0.0110	1.51 (1.20-1.91)
SCZ				A	G	C	0.334	0.240	0.000100	0.00780	1.59 (1.23-2.05)
BPD+SCZ				A	G		0.392	0.321	0.000800	0.00502	1.36 (1.12-1.65)
BPD+SCZ					G	C	0.327	0.249	0.000400	0.00343	1.47 (1.20-1.80)
BPD+SCZ				A	G	C	0.326	0.240	0.000100	0.00553	1.52 (1.22-1.89)
German											
BPD				*C*	*A*	*C*	0.237	0.294	**0.020 (st: 0.128, ew:0.453)**	0.353	0.726 (0.556-0.949)
Merged Scottish											
SCZ	T	T	T				0.010	0.00300	**0.0280 (st: 0.208, ew: 0.581)**	0.168	3.44 (1.16-10.3)
SCZ					G	C	0.329	0.273	**0.00370 (st: 0.0220, ew: 0.0720)**	**0.0303**	1.31 (1.10-1.56)
BPD+SCZ					G	C	0.320	0.273	**0.008000 (st: 0.0440, ew: 0.139)**	**0.0287**	1.24 (1.06-1.45)
Merged All											
BPD+SCZ					G	C	0.3004	0.270	**0.0452 (st: 0.139, ew: 0.427**	0.108	1.18 (1.05-1.34)

Table 2.8. Haplotypes demonstrating nominally significant association to schizophrenia, bipolar disorder, or the combined case group. See next page for legend.

Table 2.8. Nominally significant estimated haplotypes showing association to bipolar disorder (BPD), schizophrenia (SCZ), or the combined case group (BPD+SCZ). Haplotypes showing nominally significant association were identified in the Scottish 1, German, Merged Scottish (Scottish 1 and Scottish 2) and Merged All (Scottish 1, Scottish 2, and German) samples. Alleles forming the significant haplotype are indicated in the second and third columns. Block number indicates the LD block in which each SNP is located according to Thomson et al. (2007). For the German, Merged Scottish and Merged All samples, *p*-values attaining nominal significance ($p \leq 0.05$) in the individual or global tests of significance are highlighted in bold. *P*-values attaining nominal significance in the individual test of significance were corrected by permutation analysis at the single-test (st) and experiment-wise (ew) levels. Corrected *p*-values are displayed in parentheses. Alleles forming the novel haplotype identified in the German sample are indicated in italics. Haplotypes attaining nominal significance at the Nyholt-corrected significance threshold of $p \leq 0.0016$ in the Scottish 1 sample (Thomson et al., 2007) are presented for comparison. Matching colours indicate corresponding haplotypes in the original Scottish 1 sample and current analysis.

2.4.3.2.1 Scottish 2 and German case-control samples

None of the haplotypes identified by Thomson et al. (2007) were found to be significant in either the Scottish 2 or the German case-control sample. A novel three-SNP region B haplotype (C-A-C) was identified as significant in bipolar disorder cases in the German sample at the individual (uncorrected *p*i = 0.020) but not global (uncorrected *p*g =0.353) level of significance. However, following correction by permutation analysis the individual *p*-value was not statistically significant (corrected p_{st}= 0.128).

2.4.3.2.2 Merged Scottish and Merged All case-control samples

The three-SNP region A haplotype (T-T-T), which was nominally associated with schizophrenia in the Scottish 1 sample (Thomson et al., 2007), was significantly associated with schizophrenia in the Merged Scottish sample in the individual (uncorrected *p*i = 0.028) but not the global test (uncorrected *p*g = 0.168) of significance. However, this association did not survive multiple testing correction (corrected *p* = 0.208).

In keeping with the Scottish 1 sample, the two-SNP region B haplotype involving SNPs rs6988339 and rs3757930 (G-C) was significantly associated with schizophrenia and the combined case group in the Merged Scottish sample in both the individual (schizophrenia: uncorrected p_i = 0.0037; combined cases: uncorrected p_i = 0.0080) and global (schizophrenia: uncorrected p_g = 0.030; combined cases: uncorrected p_g =0.029) haplotype tests. These associations remained significant after correction of individual *p*-values by permutation analysis (schizophrenia: corrected p_{st} = 0.022; combined cases: corrected p_{st} =

0.044). This haplotype was more common in cases (schizophrenia: 32.9%, combined cases: 32.0%) than controls (27.3%), conferring an increased disease risk (schizophrenia: OR = 1.3, 95% CI: 1.1-1.6; combined cases: OR = 1.2, 95% CI: 1.1-1.5). This haplotype was also associated with the combined case group in the Merged All group in the individual (uncorrected $p_i = 0.045$) but not the global (uncorrected $p_g = 0.108$) test of significance. The association in the Merged All group did not remain significant after permutation analysis (corrected $p_{st} = 0.139$).

2.4.3.2.3 Imputation of the missing region B haplotype

In an attempt to compensate for the genotyping failure of one of the region B SNPs (rs2919390) in the Scottish 2 sample, the ability of the two-SNP (rs6988339 and rs3757930) region B haplotype to predict the three-SNP region B haplotype was assessed in the Scottish 1 sample. Genotype data was uploaded to Haploview v4.2 and r^2 calculated as a measure of LD between the two- and three-SNP region B haplotypes. The two SNP G-C haplotype, which was significant in the Merged Scottish sample, was found to predict the three-SNP A-G-C haplotype , which was significantly associated with schizophrenia and the combined case group in the Scottish 1 sample (Thomson et al., 2007), with an r^2 of 0.97 (figure 2.5). It therefore seems likely that the same three-SNP haplotype would show significant association in the Merged Scottish sample.

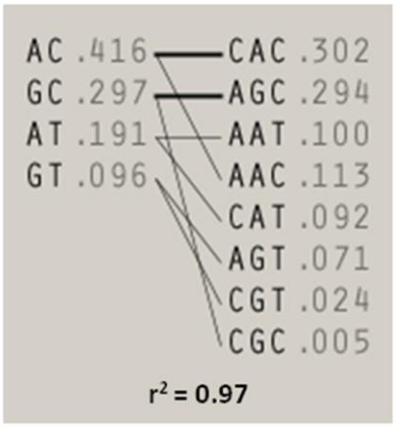


Figure 2.5. Prediction of the three-SNP region B haplotype from the two-SNP (rs6988339 and rs3757930) haplotype. The ability of the two-SNP region B haplotype (GC), which was significantly associated with schizophrenia and the combined case group in the Merged Scottish sample, to predict the three-SNP region B haplotype (AGC), which was significantly associated with schizophrenia and the combined case group, in the Scottish 1 sample (Thomson et al., 2007), was assessed. Genotype data from the Scottish 1 sample for the markers composing the two- and three-SNP region B haplotypes was uploaded to Haploview v4.2 and r^2 between the GC and AGC haplotypes calculated as a measure of LD. All possible allele combinations (individual haplotypes) for the two-SNP haplotype (left-hand side) and the three-SNP haplotype (right-hand side) are displayed. The frequency of each individual haplotype is shown to the right of the haplotype. Lines connecting the two- and three- SNP haplotypes indicate the co-occurrence of these haplotypes in greater than (i) 0% (thin lines) or (ii) 20% (thick lines) of chromosomes. The two-SNP G-C haplotype was found to predict the three-SNP A-G-C haplotype with $r^2 = 0.97$.

2.5 Summary and discussion

In this chapter, association analysis was carried out to assess the involvement of six haplotype-tagging SNPs in *NRG1*, a leading candidate gene for schizophrenia, in schizophrenia and bipolar disorder in two independent case-control samples from Scotland and Germany. These SNPs were previously found to be significantly associated with schizophrenia and a combined schizophrenia and bipolar disorder case group in a Scottish sample (Scottish 1; Thomson et al., 2007). In this sample, the SNPs were found to form two significantly associated three-SNP haplotypes, one located at the 5' end of the gene (region A) and the other located at the 3' end of the gene (region B). In addition to testing association in the two new case-control samples, two merged samples were assessed: the Merged Scottish sample comprised the Scottish 1 sample and the new Scottish sample (Scottish 2), and the Merged All sample comprised the two Scottish samples and the German sample.

Individual analyses of the two novel case-control samples failed to replicate any of Thomson et al.'s (2007) previously reported associations. On combining the two Scottish samples, the two-SNP region B haplotype (rs6988339 and rs3757930) was found to be significantly associated with schizophrenia and the combined case group, with these associations withstanding correction for multiple testing by permutation analysis. Association was always to the G-C haplotype, consistent with Thomson et al.'s (2007) findings. No associations in the Merged All sample remained significant after permutation correction.

Unfortunately, one of the region B SNPs (rs2919390) failed to genotype in the Scottish 2 sample. Thus, only two haplotype-tagging SNPs in region B could be analysed in this sample. By considering the LD between the two- and three-SNP region B haplotypes in the Scottish 1 sample, the ability of the two-SNP haplotype to predict the three-SNP haplotype was assessed. This revealed a high level of predictability, with the two-SNP haplotype that showed significant association in both the Scottish 1 and the Merged Scottish samples predicting the three-SNP haplotype that was significantly associated in the Scottish 1 sample on 98.9% of occasions. It is therefore likely that very little information was lost by the failure to genotype the region B SNP, rs2919390, and it seems fair to conclude that had this SNP not failed to genotype, the three-SNP region B haplotype would have been likely to

show significant association in the schizophrenia and combined case groups in the Merged Scottish sample.

The interpretation of association findings in studies of psychiatric illness is complicated by both heterogeneity at the level of observable phenotypes and a lack of understanding of the underlying genetic architecture. Both issues call into question the rationale of performing association studies to detect genetic loci that contribute to disease susceptibility. This uncertainty opens both positive and negative association findings up to several equally plausible explanations, which can often only be differentiated between by further studies.

A fundamental assumption behind the use of association analysis is that the genetic contribution to the disorder in question involves several common variants whose effects are detectable at the population level. This assumption essentially re-states the polygenic common disease common variant (CDCV) hypothesis of psychiatric illness, which describes a situation in which, within an individual, the combined effect of several common variants of modest effect size in different genes contributes to disease predisposition. To accommodate the fact that diagnostic status is a discrete variable, a liability threshold was invoked: although the underlying distribution of liability is continuous, only those individuals who pass a minimum burden of common variants actually get the disease (Falconer, 1989). Although the CDCV hypothesis is just one of many hypotheses of the potential genetic architecture of psychiatric disorders, is has, historically, been the most widely accepted. In recent years, however, the premises of this hypothesis have been called into question, and the plausibility of other hypotheses considered (Mitchell, 2012; Mitchell and Porteous, 2011). These authors have argued that epidemiological observations of schizophrenia and bipolar disorder are compatible with the effects of multiple rare mutations occurring in a large number of genes, with mutations in different genes contributing to disease pathogenesis in different individuals (i.e. genetic heterogeneity). It is suggested that whilst, within an individual, the presence of a certain rare mutation may be necessary for disease onset, the resultant phenotype will be dependent on the genetic background (Mitchell and Porteous, 2010).

In light of this uncertainty, the results presented in this chapter must be interpreted with caution. Assuming a CDCV model, it can be predicted that an increase in sample size would confer an increase in statistical power, and thus greater likelihood of detecting association to

variants of small effect size. A corollary of this argument is that failure to replicate findings between samples is likely to be attributable to random sampling effects rendering (at least some of) the samples unrepresentative of the populations from which they were drawn, assuming, of course, that the samples are not drawn from genetically divergent populations. As such, the positive findings in the Scottish 1 sample (Thomson et al., 2007) may have been products of the "winner's curse" (Trikalinos et al., 2004), the well-documented phenomenon whereby initial estimates of effect size are often inflated compared to subsequent estimates. The reduced effect sizes detected in the Merged Scottish sample would be likely to represent a more accurate depiction of effect size in the population; these effect sizes are in-line with results from genome wide association studies (Cichon et al., 2009). The absence of significant associations in the two new samples could reflect insufficient statistical power to detect effects of the magnitude estimated by the Merged Scottish sample. Under the CDCV model, the lack of significant associations in the Merged All sample is more difficult to interpret. One explanation is that the further increase in sample size following the addition of the German sample to the two Scottish samples resulted in an even more accurate estimate of effect size, and thus that variants in the *NRG1* regions assessed do not confer susceptibility to schizophrenia and/or bipolar disorder in either of these populations. Alternatively, systematic differences may have existed between the Scottish and German cases, rendering invalid their combination in one sample.

One potential source of systematic difference between the Scottish and German samples is LD structure. Differences in LD structure could result in a haplotype tagging a susceptibility variant in one sample but not another sample. This difference could either reflect a true difference in the LD structure of the populations that the samples are drawn from, or be the result of random sampling effects. Too few markers were genotyped in the present study to examine the LD structure of the different samples. However, previous comparison of LD structure between different European populations has found that whilst similar patterns of LD are found in different populations, there are some genomic regions where shifts in the bounds of LD blocks can be observed (Mueller et al., 2005). The authors concluded that "The observed population differences in haplotype frequencies and LD structure may affect the power to detect phenotype-genotype associations". Whilst Mueller et al., (2005) did not include any samples drawn specifically from Scotland, or even the United Kingdom, in their analysis, their findings highlight the possibility of differences in the LD structure of the Scottish and German populations.

The SNPs genotyped in the present study had previously been identified as haplotype-tagging SNPs in the initial *NRG1* association study in the Scottish 1 sample carried out by Thomson et al. (2007). As such, these SNPs were specifically selected to efficiently capture genetic variation within their haplotype blocks. Nevertheless, rare variants and haplotypes present in this region would not necessarily have been captured by the genotyped variants (when Thomson et al. (2007) defined the *NRG1* haplotype structure, haplotype-tagging SNPs were selected from SNPs with MAF at least 10% to tag haplotypes of greater than 10% frequency). As such, it is possible that functional variants were not effectively captured by the current analyses. One approach to further investigating the genetic variation present at this locus would be to genotype a larger number of variants in a larger number of subjects (i.e. deep genotyping); however, this approach will always be limited by its inability to capture as-of-yet unknown sequence variants. Re-sequencing of the *NRG1* locus, in contrast, would permit an unbiased assessment of the variants present in this region.

It is possible that certain *NRG1* variants only confer risk for psychiatric illness on a particular genetic background. As such, between-samples differences in genetic background might have influenced the power of each sample to detect risk variants in LD with the markers/haplotypes analysed. It is possible that any differences between samples reflect true population differences between the Scottish and German populations in genetic background. Such heterogeneity amongst different European populations has been demonstrated for another schizophrenia- and bipolar disorder-risk gene, *DISC1* (Hennah et al., 2009).To investigate this heterogeneity, Hennah et al. (2009) used conditional association analysis and found a *DISC1* variant that confers risk only on certain genetic backgrounds. It is already known that variants within *NRG1* interact epistatically with each other and with variants in *ERBB4*, which encodes an NRG1 receptor, and the schizophrenia-risk gene *AKT1* (Nicodemus et al., 2009). Furthermore, a functional interaction has been detected between NRG1 and DISC1 that is mediated by ErbB2 and ErbB3 receptors and PI3/Akt signalling (Seshadri et al., 2010). Together, these findings suggest that genotype at many functionally relevant loci will ultimately determine the phenotypic effect of variants within *NRG1*.

Phenotypic differences between the cases included in the Scottish and German samples represents a further potential confound. A potential source of between-samples variation is the difference in inclusion criteria for bipolar disorder used in the Scottish and German samples: the Scottish samples included individuals diagnosed with both bipolar disorder 1

and bipolar disorder 2, whereas the German sample only included individuals with bipolar disorder 1. Furthermore, it is possible that systematic differences were introduced between the Scottish and German cases if the diagnoses made by the Scottish and German clinical teams were affected by different biases. In the study of conditions, such as psychiatric illnesses, where there is no established diagnostic marker, this latter concern is particularly pertinent.

The issue of phenotypic heterogeneity is of serious concern when considering the validity of using association analysis to study psychiatric illnesses. The diagnostic categories of schizophrenia and bipolar disorder are broad (even when considering bipolar disorder 1 and 2 separately) and can result in patients diagnosed with the same disorder sharing few symptoms. In light of this phenotypic heterogeneity there is no reason to assume that schizophrenia or bipolar disorder represent unitary conditions in terms of their underlying aetiology. As such, it is not surprising that association studies in which the only inclusion criterion for cases is diagnostic status often yield either no significant findings, or findings that are not replicable. Even if a CDCV model operates within specific disease subtypes, the pooling together of all cases with a given diagnosis will significantly diminish statistical power to detect associated variants.

Phenotypic heterogeneity also has significant implications for the predicted genetic architecture of psychiatric illness. If conditions such as schizophrenia and bipolar disorder are no longer considered as unitary, then it follows that each condition may in fact consist of several genetically distinct conditions that happen to share varying degrees of phenotypic overlap. This possibility calls into question the dogma that these conditions are common, and thus the rationale of the CDCV model. A strong argument against the involvement of rare variants of large effect size in psychiatric illness is that such variants would have been detected with ease by genetic linkage and association studies. As both of these approaches have been limited in their success in identifying variants of large effect size, the CDCV model gained in favour. However, the results obtained from association analyses are also compatible with a scenario in which the case groups consist of several different sub-conditions, each individually rare, caused by different rare mutations. When considered at the sample level, the estimated effect size of a rare mutation would be diluted compared with its effect within a particular sub-condition. Furthermore, it is possible that genotyped variants tag several rare variants of large effect size. Association to the genotyped variant

may, therefore, represent different rare variants acting in different individuals. As such, a rare variant model could explain the pattern of positive and negative associations detected in the current study.

There is evidence to suggest that phenotypic heterogeneity in schizophrenia and bipolar disorder can be related to particular *NRG1* variants. Several studies have demonstrated association of *NRG1* variants with particular aspects of psychiatric illness (Bramon et al., 2008; Goes et al., 2009; Hall et al., 2006; Prata et al., 2009; Sei et al., 2007). The assessment of association in phenotypically homogenous samples selected for relevant endophenotypes may, therefore, represent an improvement over the study of cases selected solely on diagnostic status. The validity of this approach will, however, depend upon the rarity of the mutations involved and the specificity of the relationship between genotype and endophenotype. Nevertheless, such an approach has the potential to detect pathways whose disruption contributes to a given endophenotype, and thus has the potential to identify novel therapeutic targets. It is, however, important to note that even if genetic variants are detected that play a causal role in determining variation in an endophenotype, these variants may represent modifiers of disease-risk rather than causal variants *per se*. It has been suggested that endophenotypes of psychiatric illness, such as cognitive deficits, may be indicative of a brain that is less well-adapted for dealing with the effects of disease-causing rare mutations; thus, such endophenotypes may frequently co-occur with, but not play a directly causal role in, psychiatric illness (Barnett et al., 2006).

2.6 Conclusions

To conclude, this study found no evidence for association to region A and produced evidence consistent with, but not indubitably supportive of, the presence of a risk variant for schizophrenia and/or bipolar disorder within, or in LD with, region B in an enlarged Scottish case-control sample.

The observations of (i) failure to replicate Thomson et al.'s (2007) initial association of *NRG1* region B haplotype with schizophrenia and/or bipolar disorder in the second Scottish sample and (ii) continued support, albeit with reduced effect size, for the involvement of a variant in this region in the Merged Scottish sample suggests two possible conclusions. Firstly, a variant in LD with region B haplotype might confer risk for schizophrenia and/or

bipolar disorder in the Scottish population, with the increased sample size of the Merged Scottish sample permitting a more accurate assessment of the effect size of this variant. Alternatively, the initial association of the region B haplotype in the first Scottish sample (Thomson et al., 2007) may, due to random sampling error, not be representative of the involvement of this region in susceptibility to psychiatric illness in the Scottish population. It will only be possible to distinguish between these two interpretations with further studies, which should be designed so as to minimise the potential confounds of genetic and phenotypic heterogeneity.

These findings should be interpreted in the context of a substantial body of genetic and functional evidence supporting the candidacy of *NRG1* as a schizophrenia- and bipolar disorder-susceptibility gene. Failure to find evidence for association in the two new samples when analysed individually raises questions regarding the impact of allelic, genotypic, and phenotypic heterogeneity on the comparability and representativeness of different case-control samples. These issues highlight limitations in our understanding of the genetic architecture of psychiatric illnesses and, therefore, the potential limitations of association analysis. Our lack of understanding of the genetic architecture of psychiatric illness makes it difficult to predict how the role of candidate risk genes, such as *NRG1*, would be most usefully assessed. Association analysis of more phenotypically homogenous samples may be more likely to identify risk variants; however, as discussed previously, the validity of this approach is subject to caveats. While deep genotyping may improve the validity of association analysis, if very rare variants within *NRG1* do play a role in the pathogenesis of psychiatric illness, then they are only likely to be detected through the sequencing of large numbers of affected individuals. A recent study in which the regions ~3kb upstream of the transcription start sites of the *NRG1* type I, type II/IV, and type III isoforms were resequenced in individuals with schizophrenia and control subjects found an increased load of rare novel variants in cases (Weickert et al., 2012), highlighting the potential of this approach to elucidate the contribution of variation in *NRG1* to psychiatric illness.

Chapter 3

Characterisation of the *DISC1* promoter

Chapter 3: Characterisation of the *DISC1* promoter

3.1 Introduction

Genetic and functional evidence highlight *DISC1* as one of the leading candidate genes for psychiatric illness. Nevertheless, relatively little is known about how variation in the gene might confer risk for major mental illness. One potential mechanism is via the dysregulation of *DISC1* expression. This chapter describes the first characterisation of the *DISC1* promoter region. Improved understanding of the mechanisms governing wildtype DISC1 regulation will permit predictions to be made about the effects of variants. Some of the work presented in this chapter has previously been published (Walker et al., 2012).

3.1.1 *DISC1* expression and susceptibility to psychiatric illness

3.1.1.1 Evidence for the role of altered DISC1 expression in conferring susceptibility to psychiatric illness

DISC1 was first identified in a large Scottish family in which a balanced chromosomal translocation t(1;11) (q42.1;14.3) disrupting the *DISC1* locus segregates with schizophrenia, bipolar disorder, and recurrent major depression (Millar et al., 2000b; St Clair et al., 1990). Lymphoblastoid cell lines derived from carriers of the t(1;11) translocation have been found to have a 50% reduction in the expression of DISC1 protein (Millar et al., 2005). The findings from subsequent studies investigating DISC1 expression in psychiatric illness have, however, been equivocal. Some studies (Maeda et al., 2006; Nakata et al., 2009; Olincy et al., 2012; Sawamura et al., 2005) have identified altered DISC1 expression in individuals diagnosed with psychiatric illnesses, whereas others (Dean et al., 2007; Rastogi et al., 2009) have not identified any such differences. One study (Lipska et al., 2006) reported no change in DISC1 mRNA expression in the dorsolateral prefrontal cortex or the hippocampus in patients with schizophrenia, but identified a significant increase of ~20% in hippocampal DISC1 protein expression in patients. Furthermore, this study reported a positive correlation between DISC1 mRNA and protein expression in the hippocampus.

Several factors could underlie the lack of consistency between the above studies, including differences in the composition of the patient groups: dysregulated DISC1 expression might contribute to the pathogenesis of the disease only in some cases. Other factors such as differences in the brain regions assessed and differences in the DISC1 transcripts quantified

might also have contributed to the lack of consensus between these studies. Furthermore, it is important to note that the demonstration of altered mRNA expression does not necessarily indicate a change in protein expression. These factors are discussed further in section 3.5.

Whilst not a direct assessment of gene expression, the presence of copy number variation in a region is at least suggestive of altered expression of the affected genes. Two studies have reported the presence of copy number variants (CNVs) involving the *DISC1* locus in individuals diagnosed with autism spectrum disorders: Williams et al. (2009) reported the case of a boy with autism spectrum disorder who was found to have a deletion involving *DISC1*, *DISC2*, and *TSNAX*, and Crepel et al. (2010) identified a duplication involving seven genes, including DISC1, in two brothers with autism and mild mental retardation. In a genome-wide association analysis of CNVs with schizophrenia, Glessner et al. (2010) identified an association between large CNVs (≥100kb) affecting *DISC1* and schizophrenia. The key question, however, when discussing CNVs in relation to gene expression is to what extent do changes at the DNA level resulting from CNVs result in altered gene expression? Assessment of the relationship between CNVs and gene expression in cell lines from individuals in the HapMap project revealed that CNVs could account for almost 18% of the variation in mRNA expression (Stranger et al., 2007). The relationship between CNVs and protein expression may, however, be less direct: a study in cancer cells revealed that in most cases there is no correspondence between gene copy number and the level of protein expression, although the protein expression of some oncogenes and tumour suppressor genes was found to be affected by gene copy number (Geiger et al., 2010). In light of these findings, it would be of interest to investigate the relationship between CNVs and gene expression in individuals diagnosed with psychiatric illnesses.

3.1.1.2 Evidence for the role of sequence variation in the DISC1 upstream region in the regulation of DISC1 expression

Studies examining the effect of polymorphisms in the upstream region of *DISC1* on *DISC1* gene expression are somewhat limited: to date, only three studies have been carried out (Carless et al., 2011; Hayesmoore et al., 2008; Hennah and Porteous, 2009). Of these three studies, two looking at the effect of single nucleotide polymorphisms (SNPs; Hennah and Porteous, 2009; Carless et al., 2011) have identified SNPs in the DISC1 upstream region that are associated with DISC1 expression level.

Using data available from publicly available databases, Hennah and Porteous (2009) assessed 754 SNPs located in the region from 10kb upstream of *TSNAX* to 10kb downstream of *DISC1* for association with DISC1 expression values drawn from 210 lymphoblastoid cell lines from the four HapMap cohorts (CEU: Utah residents with Northern and Western European ancestry from the CEPH collection; YRI: Yoruba in Ibadan, Nigeria; CHB: Han Chinese in Beijing, China ; and JPT: Japanese in Tokyo, Japan). Only one SNP, rs1765778, which is located 30kb upstream of DISC1 demonstrated association in all four populations, and a further five SNPs were found to be associated in three of the four populations, with the common exception being the Japanese population. In all cases, the minor allele (G), as defined by allele frequencies in the CEU population was found to be associated with decreased DISC1 expression.

Many more SNPs were found to be significantly associated with DISC1 expression level in a study carried out by Carless et al. (2011). The greater number of associated SNPs could be attributable to the fact that this study was carried out in a larger sample (1232 individuals). Fifteen SNPs in the *DISC1* promoter region (defined as the region ~2kb upstream of the gene) were identified as being associated with DISC1 expression level; of these fifteen SNPs, only one, rs3738398, which is located 221 bp upstream of the *DISC1* transcription start site (TSS), overlapped with the SNPs identified by Hennah and Porteous (2009). The direction of effect for this SNP was consistent between the two studies.

A different approach to the identification of SNPs involved in the *cis*-regulation of DISC1 expression was taken by Hayesmoore et al. (2008). In this study individuals heterozygous for a SNP, rs3738401, located in *DISC1* exon 2, were assessed for expression imbalance between the maternally and paternally inherited transcripts. One individual was found with such an expression imbalance. Sequencing of the region extending approximately 1kb upstream of the *DISC1* TSS revealed a novel insertion polymorphism, a tandem duplication of the 22 nucleotides located -168 to -147 relative to the *DISC1* TSS. The duplication affected a dinucleotide repeat region, such that chromosomes carrying the duplicated sequence carried one $(TG)_8$ repeat and one $(TG)_{10}$ repeat. Variation in the numbers of repeats in dinucleotide repeat regions has previously been associated with regulation of promoter activity (Borrmann et al., 2003; Wang et al., 2005). In the case of the *DISC1* promoter, however, Hayesmoore et al. (2008) concluded that that the repeat was not *per se* functional, as individuals who did not show allelic expression imbalance were also heterozygous for the number of TG repeats in the repeat region.

It is worth noting that *cis*-regulatory mechanisms involved in the control of *DISC1* expression will, to varying extents, operate in a tissue-specific and developmental period-specific manner. This will, in part, reflect between-tissue and developmental variation in the expression of transcription factors. As such, the three studies carried out to date, which have been carried out in adult subjects and have investigated expression in lymphoblastoid cell lines (Hennah and Porteous, 2009), lymphocytes (Carless et al., 2011) and the post-mortem brain (Hayesmoore et al., 2008), are likely to have identified variants affecting the function of only a small number of the *cis*-regulatory modules present upstream of *DISC1*. In order to establish a more comprehensive understanding of the *cis*-regulatory control of DISC1 expression the effect of sequence variants in the *DISC1* putative promoter region should be studied in a range of tissues across development. Moreover, it will be of particular interest, in the context of psychiatric illness, to identify variants affecting region- and cell-specific expression of *DISC1* in the brain.

3.1.1.3 Evidence for the involvement of sequence variation in the DISC1 upstream region in conferring susceptibility to psychiatric illness

A key question when relating the experiments carried out in this chapter to the field of psychiatric genetics is whether variation in, or in linkage disequilibrium with, the *DISC1* promoter region has been shown to be associated with the diagnosis of psychiatric illness or related endophenotypes. Perhaps one of the most obvious variants to consider in relation to this question is rs3738398, the SNP identified by both Hennah and Porteous (2009) and Carless et al. (2011) as being associated with DISC1 expression level. Hennah et al. (2003) identified the C allele of this SNP as participating in a haplotype that was understransmitted from parents to female offspring diagnosed with schizophrenia, and, in an initial study, Kockelkorn et al. (2004) identified the association of this variant with schizophrenia, with the G allele being more prevalent amongst cases (figure 3.1). However, upon analysis of a larger sample, Kockelhorn et al. (2004) found that the association was no longer significant. This latter result is in keeping with subsequent association studies that also found no evidence for association between this SNP and schizophrenia and mood disorders (Arai et al., 2007; Saetre et al., 2008; Schumacher et al., 2009). Similarly, Carless et al. (2011) found no association between this SNP and performance on several neurocognitive measures relevant to psychiatric illness. It is, of course, possible that this SNP only confers risk in the

context of other variants; association analysis conditioned on the presence of genotype at other variant(s) could be performed to assess this hypothesis.

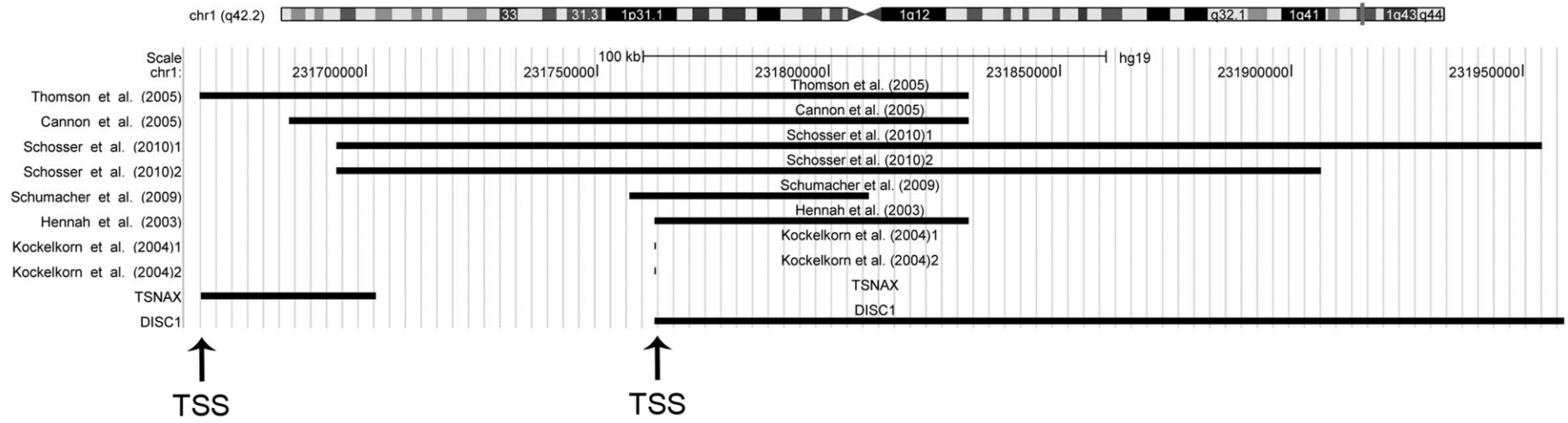


Figure 3.1. Schematic showing the genomic locations of variants and haplotypes in the *DISC1* upstream region that have been associated with psychiatric illness. The genomic locations (UCSC human genome browser, hg19) of associated haplotypes and variants are indicated by black rectangles. Track names indicate the study in which significant association was detected. The locations of *DISC1* (long variant) and the upstream gene *TSNAX* are shown for reference, and the transcription start site (TSS) of each gene is indicated by an arrow. Thomson et al. (2005) detected association between an eight-marker haplotype and bipolar disorder in males. Cannon et al. (2005) found a four-marker haplotype to be associated with schizophrenia. Schosser et al. (2010) identified a three-marker haplotype as being associated with both bipolar disorder and major depressive disorder (1) and a two-marker haplotype as being associated with bipolar disorder (2). Schumacher et al. (2011) identified an eight-marker haplotype as being associated with schizophrenia. Hennah et al. (2003) identified a three-marker haplotype as being under-transmitted to females with schizophrenia. Kockelkorn et al. (2004) found both the TG dinucleotide repeat polymorphism (1) and the SNP rs3738398 (2) to be associated with schizophrenia in an initial sample, but failed to replicate these associations in a larger sample.

Several haplotypes spanning the *DISC1* upstream region have been found to show significant association to psychiatric illnesses, supporting the notion that variation in this region might contribute to illness susceptibility (Cannon et al., 2005; Hennah et al., 2003; Schosser et al., 2010; Schumacher et al., 2009; Thomson et al., 2005) (figure 3.1). One potential mechanism for this risk is via the disturbance of neural systems required for normal cognitive function: the four-marker haplotype identified by Cannon et al. (2005) as being associated with schizophrenia was also associated with visuospatial working memory, verbal learning and memory, and reaction time to visual targets, independent of diagnostic status. Identification of the functional variants responsible for these positive association and linkage findings is, however, necessary before speculating too much on the possible mechanisms of their action. It should also be noted that, to date, no GWAS study has identified variation in the *DISC1* promoter region as conferring risk for schizophrenia or bipolar disorder at the level of genome-wide significance. Nevertheless, as discussed in section (1.6.2.3), failure to identify association by GWAS does not negate the potential pathogenic contribution of variation in this region.

The polymorphic TG dinucleotide repeat region located upstream of *DISC1* investigated by Hayesmoore et al. (2008; discussed in section 3.1.1.2) occurs in tandem with a polymorphic CG dinucleotide repeat region. This region, which has the sequence $(TG)_8(CG)_8$ in the UCSC human genome browser (chr1: 231762399-231762430, hg19), was assessed for association with schizophrenia by Kockelkorn et al. (2004; figure 3.1). In keeping with the findings for rs3738398, the number of TG repeats was found to be associated with schizophrenia in the initial sample, but not the larger sample. One explanation for these results is that the findings in the initial, smaller, sample were due to random variation resulting from the small sample size; alternatively, it is possible that the compositions of the two samples differed, and that variation in these repeats is associated with a phenotype that was more prevalent in the small sample. A lack of evidence for association between the length of the TG repeat region and schizophrenia and mood disorders was also reported by Devon et al. (2001) and Arai et al. (2007); however, it should be noted that the assay implemented by Devon et al. (2001) did not distinguish between the $(TG)_n$ polymorphism and a $(CG)_n$ polymorphism located immediately downstream.

3.1.2 Sequence features of promoter regions

3.1.2.1 Core promoter architecture

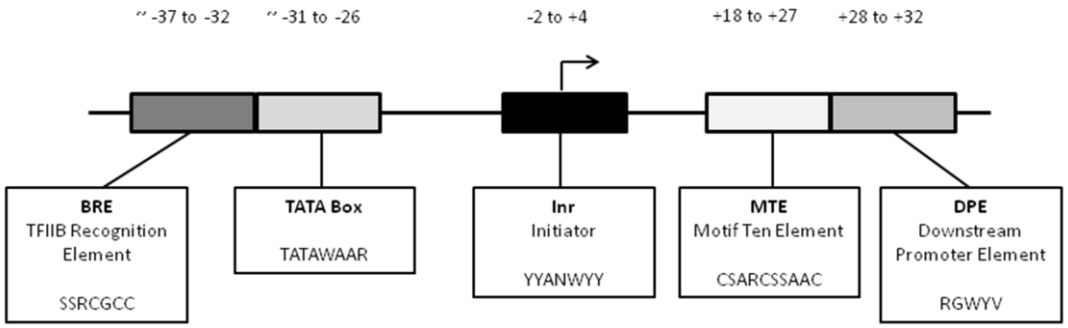


Figure 3.2. Schematic of core promoter elements that can contribute to basal transcription. Motifs are depicted by filled rectangular boxes and their usual location relative to the transcription start site (where the transcription start size is +1) indicated above. The name and the consensus sequence of each core promoter element are given in the boxes below. Consensus sequences are denoted in IUPAC notation for degenerate nucleic acids (R= A or G; S = C or G; V = A or C or G; W = A or T; Y = C or T). The direction of transcription is indicated by the arrow above the Initiator.

The eukaryotic core promoter comprises the genomic region extending approximately 35 bp upstream and downstream of the TSS. In this region, the basal transcription factors, the factors minimally essential for transcription, interact with the DNA sequence to mediate transcription. Initial studies of core promoter regions highlighted the TATA box as a key feature required for transcription (Breathnach and Chambon, 1981). The TATA box, which has the consensus sequence TATAWAAR, is present approximately 31 bp to 26 bp upstream of the TSS (figure 3.2) (Smale and Kadonaga, 2003). Transcription from TATA box-containing promoters is initiated by the binding of the TATA box by TATA box binding protein (TBP), a component of the multiprotein complex, transcription factor IID (TFIID). Other basal transcription factors (TFIIA-H), including RNA polymerase II, then bind to form the preinitiation complex and transcription is initiated at a site approximately 30 bp downstream. However, as an increasing number of core promoter regions have been identified, it has become apparent that only a minority (10-20%) include a TATA box (Cooper et al., 2006; Gershenzon and Ioshikhes, 2005). Those genes transcribed from TATA

box-containing promoters tend to encode proteins expressed only in specific tissues or contexts (Schug et al., 2005).

The initiator (Inr) is a core promoter element, located -2 to +4 bp relative to the TSS, which can occur both in TATA box-containing and TATA-less promoters. The Inr consensus sequence, YYANWYY, comprises an adenosine at the +1 position, surrounded, predominantly, by pyrimidines (figure 3.2; Smale and Kadonaga, 2003). The Inr can bind TFIID via components known as TBP-associated factors (TAFs), and it does not appear to be essential for the TBP component to interact with DNA sequence in the -30 bp region in TATA-less promoters for transcription to occur (Martinez et al., 1995). Experiments in which the Inr has been deleted, or inserted into promoters that naturally do not contain an Inr have demonstrated that the presence of an Inr results in high levels of transcription from a specific start site (Smale and Kadonaga, 2003).

The downstream promoter element (DPE) is a core promoter element located +28 to +32 bp relative to the TSS, usually found in promoters lacking a TATA box but containing an Inr (Smale and Kadonaga, 2003). The DPE is a TFIID recognition site with the consensus sequence RGWYV, which is additionally often associated with the presence of a guanine nucleotide at position +24 (figure 3.2). The Inr and the DPE appear to function as a unified core promoter element: the binding of TFIID binding and basal transcriptional activity is eliminated by mutation of either the Inr or the DPE (Burke and Kadonaga, 1996), and decreased by alteration of the spacing between the Inr and the DPE (Kutach and Kadonaga, 2000).

Another core promoter element located downstream of the TSS is the motif ten element (MTE) (Lim et al., 2004). This element has the consensus sequence CSARCSSAAC, is located at +18 to +27 bp relative to the TSS (figure 3.2), and, in common with the DPE, acts together with the Inr to bind TFIID (Lim et al., 2004).

The transcription factor II B recognition element (BRE) is marked out by being one of the only core promoter elements not to recognise TFIID. The BRE is situated upstream of the TATA box, at approximately -37 to -32 bp relative to the TSS, and has the consensus sequence SSRCGCC (figure 3.2; Smale and Kadonaga, 2003). The function of the BRE remains to be fully understood: in archaeal promoters, the interaction of the BRE with transcription factor B serves to enhance the assembly of the preinitiation complex, and thus

transcription, whereas studies in humans suggest that binding of TFIIB to the BRE can sometimes exert a repressive effect on transcription (Smale and Kadonaga, 2003).

3.1.2.2 Transcription factor binding sites

The core promoter motifs described above are sufficient for the assembly of the preinitiation complex and can, therefore, support a basal level of transcription. To fine-tune this transcription additional gene-specific transcription factors (henceforth referred to as transcription factors) are required. The binding of these transcription factors to *cis*-regulatory sites can support or repress transcription from a promoter, thus permitting control over where and when a gene is expressed. Modulation of transcriptional activity by transcription factors is achieved via the recruitment of co-activators and co-repressors. Co-activators include histone modifying enzymes and chromatin remodelling enzymes that act to promote an open chromatin structure required for transcription (Spiegelman and Heinrich, 2004). Other co-activators operate by recruiting RNA polymerase II and interacting with the basal transcription factors (Spiegelman and Heinrich, 2004). Co-repressors are frequently associated with histone deacetylase activity (Spiegelman and Heinrich, 2004), which promotes a closed chromatin structure (Cress and Seto, 2000). Co-repressors have also been identified that block transcription via direct interaction with basal transcription factors (Burke and Baniahmad, 2000).

The identification of transcription factor binding sites (TFBSs) presents a challenge: TFBSs consist of short, degenerate sequences that occur frequently in the genome (Bulyk, 2003); as such, not all binding sites are functional (Whitington et al., 2009). In fact, between-promoter variability in the binding site for a given transcription factor, and thus variability in the affinity of the site for the transcription factor, is thought to be one mechanism by which diversity in gene expression levels can be generated (Stormo, 2000). Computational programs can, therefore, aim only to infer the binding potential of a site but cannot determine functionality (Cartharius et al., 2005). The computational identification of putative TFBSs has been greatly aided by the development of positional weight matrix (PWM) representations of known binding sites. PWMs represent the likelihood of finding a particular nucleotide at each location in a binding site, thus permitting assessment of the similarity between a given transcription factor binding site and a DNA sequence of interest (Cartharius et al., 2005; Stormo, 2000). Additional tools to identify functional TFBSs include inter-species comparison, on the basis that functional regulatory regions are more

likely to be conserved between species (see section 3.1.2.5 for further discussion of this issue), and the comparison of the upstream sequences of co-regulated genes within a species, as such genes are hypothesised to be under the regulatory control of shared pathways (Pennacchio and Rubin, 2001).

The advent of high-throughput chromatin immunoprecipitation (ChIP) has made a significant contribution to the identification of TFBSs. In a standard ChIP experiment, transcription factors are cross-linked to their binding sites, the chromatin is sheared into short fragments, and the DNA fragments binding to a transcription factor of interest isolated by immunoprecipitation. The identity of these DNA fragments is then discerned using one of several possible techniques. One possibility is microarray hybridisation (ChIP-chip). The main limitation of microarrays is their poor spatial resolution; in this regard, high-throughput sequencing of the immunoprecipitated DNA fragments is superior (Ho et al., 2011). In comparison to purely computational approaches to TFBS identification, ChIP has the advantage of reflecting tissue-, or cell-, specific chromatin architecture. The study of ChIP-ed regions of DNA is, therefore, a useful approach to narrowing down putative functional TFBSs. Further experimental assessment is, however, still required to determine both the functionality of a site and the effect of the binding of a transcription factor on promoter activity (i.e. whether it represses or promotes transcription) (Tijssen et al., 2011). The study of ChIP-identified binding sites is likely to offer particular advantages for the identification of TFBSs in promoters that are poorly conserved and/or that drive the expression of genes with few known co-expressed genes. In this regard, the use of ChIP data may be particularly beneficial for the identification of putative binding sites in the *DISC1* promoter: The *DISC1* gene has been shown to have undergone a high level of divergence in recent evolutionary history (Bord et al., 2006). Although Bord et al. (2006) only examined the open reading frame, as evolutionary divergence was found to be particularly marked at the N-terminal of the gene, it seems plausible that a high level of divergence might also be evident in the region upstream of the TSS.

3.1.2.3 CpG islands

CpG islands are regions of DNA enriched for CpG dinucleotides. The CpG dinucleotide occurs relatively rarely: as it is a DNA methyltransferase substrate, methylation of the cytosine can result in deamination to form thymine, a mutation which is inefficiently repaired (Bird, 2002). The promoters of approximately 60% of human genes fall near a CpG

island (Venter et al., 2001). CpG island promoters are associated with genes that are constitutively activated, such as housekeeping genes, but are also found in tissue-specific genes (Gardiner-Garden and Frommer, 1987). Interestingly, CpG-rich promoters have been reported as being enriched in the brain (Gustincich et al., 2006). Investigation of the relationship between the presence of CpG islands and core promoter elements has revealed that TATA boxes, Inr and DPE elements are more common in promoters that do not contain CpG islands, whereas BREs are more common in CpG island-containing promoters (Blake et al., 1990; Gershenzon and Ioshikhes, 2005). A key difference between CpG island-containing promoters and TATA-box containing promoters is that the former often have multiple TSSs whereas the latter usually only contain one TSS (Sandelin et al., 2007; Smale and Kadonaga, 2003). One mechanism by which CpG islands are thought to contribute to constitutive expression is by destabilising the chromatin structure (Ramirez-Carrozzi et al., 2009), and thus making the DNA sequence accessible to transcription factors (see section 3.1.2.3, below, for an introduction to the role of chromatin structure in transcription).

3.1.2.4 Chromatin structure in the promoter region

In order for transcription to take place the promoter region of a gene must be accessible. Chromatin structure plays a large part in determining the accessibility of DNA, with a lightly packed chromatin structure, euchromatin, promoting transcriptional activity. Initial studies of nucleosome occupancy in promoter regions revealed the region around the TSS to be depleted of nucleosomes, leading to the term "nucleosome free region" (NFR) (Ozsolak et al., 2007; Yuan et al., 2005). Subsequent studies have shown, however, that these NFRs do, in fact, contain nucleosomes that bind DNA unstably. This unstable binding of DNA was found to be due to the substitution of canonical histone proteins with variant histone proteins (Henikoff et al., 2009; Jin and Felsenfeld, 2007).

Epigenetic modification of histones is another mechanism by which chromatin structure can be made accessible to the transcriptional machinery. Histone acetylation is a key mechanism for creating an open chromatin structure (Kadonaga, 1998). It has been suggested that acetylation reduces the ability of histones to bind DNA by neutralising positive charges at the lysine and arginine residues (Hong et al., 1993). In contrast to histone acetylation, the effects of histone methylation are more variable, with methylated histones marking areas of both active and repressed transcription: for example, the trimethylation of lysine 4 on histone H3 (H3K4me3) is frequently associated with actively transcribed genes (Barski et al., 2007;

Guenther et al., 2007), whereas the trimethylation of lysine 27 on histone H3 (H3K27me3) is generally a repressive histone modification (Cao et al., 2002; Kirmizis et al., 2004). The promoters of some genes have been found to carry both the activating H3K4me3 mark and the repressive H3K27me3 mark. These regions, termed "bivalent domains", have been identified in embryonic stem cells and are thought to "poise" key developmental genes for lineage-specific activation or repression (Mikkelsen et al., 2007).

In regions enriched for GC or TG repeats, the formation of left-handed Z-DNA, in contrast to the conventional right-handed B-DNA, may contribute to transcriptional activation (Liu et al., 2006). The Z-DNA conformation has been associated with transcriptional activity by the finding that transcription can induce Z-DNA formation and that Z-DNA formation near the promoter regions of genes can regulate their transcription *in vivo* (Liu et al., 2006; Liu et al., 2001a; Wittig et al., 1992). Typically, Z-DNA is less stable than B-DNA; however, under certain conditions the Z-DNA conformation can be stabilised. One such condition is in reaction to the torsional stress created by negative superhelicity; Liu et al. (2006) propose that sequences rich in GC or TG repeats release negative superhelical torsion by converting to a Z-DNA formation. Superhelical torsion can result from the actions of ATP-utilising chromatin remodelling enzymes (Havas et al., 2000), and actively transcribing RNA polymerase (Liu and Wang, 1987; Wittig et al., 1992). It has been suggested that the zigzag structure of Z-DNA may promote the disruption, or even ejection, of the nucleosome (Liu et al., 2006), thus permitting the access of transcriptional machinery to the DNA sequence.

3.1.2.5 Conservation of promoter elements

As regions of functional DNA are likely to have undergone fewer substitutions than neutral DNA, the cross-species comparison of promoter regions can aid in the identification of regulatory motifs. Cross-species comparison has had mixed success in the identification of functional *cis*-regulatory elements: some studies have found a high percentage of the conserved non-coding regions assessed to have regulatory function in *in vitro* assays (Grice et al., 2005; Nobrega et al., 2003), while others have been less successful (Martin et al., 2004). Key to the success of the identification of functional regulatory motifs by cross-species comparison is the selection of the species to compare. When making this choice a trade-off must be made between the ability to detect functional promoter elements with high power and specificity, which is facilitated by the comparison of evolutionarily distant species, and the ability to detect the majority of functional promoter elements, facilitated by

the comparison of evolutionarily close species. To address this issue, Prabhakar et al. (2006) carried out a systematic assessment of the ability of cross-species comparison with various combinations of species to identify functional human *cis*-regulatory elements. Comparison with five eutherian mammals (mouse, rat, cat, cow, and pig) or six simians (human, baboon, colobus, squirrel monkey, owl monkey, marmoset, and dusky titi) resulted in the ability to detect cis-regulatory elements with a sensitivity of 53-80% and a true positive rate of 27-67%. In contrast, comparison with more distant species failed to identify many empirically defined functional non-coding elements.

The combination of conservation data with knowledge of functional regulatory motifs has been shown to be successful in identifying functional regulatory elements. ESPERR (evolutionary and sequence pattern extraction through reduced representation) is a computational approach to functional element prediction that implements this approach. ESPERR designates regulatory potential scores to conserved regions of the genome based on their similarity to known regulatory elements (Taylor et al., 2006). For the detection of *cis*-regulatory modules, the use of regulatory potential scores has been shown to confer an advantage over methods only considering conservation (King et al., 2005).

3.2 Aims

The aims of this chapter were as follows:

- To use bioinformatic analysis to assess the *DISC1* upstream region for sequence features associated with promoter activity, and to use this information to select a series of *DISC1* promoter fragments to assess experimentally for promoter activity.
- To use dual luciferase reporter assays to assess the transcriptional activity of the *DISC1* promoter constructs.
- To use publicly available ChIP-seq data to identify transcription factors potentially involved in the regulation of DISC1 expression.

3.3 Materials and Methods

3.3.1 *In silico* analysis of the *DISC1* promoter region

DISC1 has previously been shown by 5' rapid amplification of cDNA ends (5' RACE) to have one TSS (Nakata et al., 2009). This TSS, located at chr1: 231762561 (February 2009, GRCh build 37; http://genome.ucsc.edu/), was, therefore, considered as the *DISC1* TSS for subsequent definition of the promoter region. An initial candidate promoter region extending from 1000bp upstream of the TSS to the translation start site (chr1: 231761561-231762613) was downloaded from the UCSC human genome browser (http://genome.ucsc.edu/). This region was used for subsequent analyses of promoter-related sequence features.

3.3.1.1 Detection of canonical core promoter motifs

The candidate DISC1 promoter region was assessed for canonical core promoter motifs using the eukaryotic core promoter predictor program YAPP, (http://www.bioinformatics.org/yapp/cgi-bin/yapp.cgi), which permits the user to search for core promoter elements, and synergistic combinations of these elements, in their expected location relative to a known TSS (table 3.1). Following a previously described method (Cartharius et al., 2005), the test sequence was compared against positional weight matrix representations (table 3.2) of core promoter elements to calculate a matrix similarity score, which only reaches one when the test sequence corresponds to the most conserved nucleotide at every position of the matrix. Searches were carried out using similarity score thresholds of 0.8, 0.7, and 0.6.

CpG islands were identified using the "CpG Islands" track of the UCSC human genome browser. CpG islands represented by this track have been defined according to the following criteria: GC content of 50% or greater, length greater than 200 bp, and a ratio of greater than 0.6 of the observed number of CG dinucleotides to the expected number of CG dinucleotides according to the number of Cs and Gs present in the segment.

Core promoter element	**Location searched (base pair relative to TSS)**	
	Start	**Stop**
Transcription factor IIB recognition element	-43	-31
TATA box	-35	-24
Initiator	-3	-1
Motif ten element	> 16	> 18
Downstream promoter element	> 27	> 29

Table 3.1. Core promoter elements searched for in the *DISC1* putative promoter region. Core promoter elements were searched for using the YAPP Eukaryotic Core Promoter Predictor program (http://www.bioinformatics.org/yapp/cgi-bin/yapp.cgi). This program permits the user to specify the location of the transcription start site (TSS) and then searches for core promoter elements in their expected location relative to the TSS, as specified above. The inputted sequence is compared, using a sliding windows algorithm, to position weight matrix representations (supplementary table 1) of the core promoter elements and a matrix similarity score calculated. Core promoter elements are identified when the matrix similarity score exceeds a pre-defined threshold. The *DISC1* candidate promoter region (chr1: 231761561-231762613) was searched using three thresholds: the default threshold of 0.8, and 0.7, and 0.6.

Core promoter element | **Position weight matrix**

Transcription factor IIB recognition element

Pos	A	C	G	T
1	0	68.9	31.1	0
2	0	67.6	32.4	0
3	35.1	0	64.9	0
4	0	100	0	0
5	0	0	100	0
6	0	100	0	0
7	0	100	0	0

TATA box

Pos	A	C	G	T
1	17.7	21.1	29	32.2
2	19.3	36.1	36.4	8.2
3	6.6	14.8	6.8	71.8
4	83.4	0	0	16.6
5	0	0	0	100
6	95	0	0	5
7	72.3	0	0	27.7
8	94.2	0	5.8	0
9	53.3	0	20.1	26.6
10	29.3	9	51.2	10.5
11	17.7	32.5	37.7	12.1
12	22.7	33	33.2	11.1

Initiator

Pos	A	C	G	T
1	0	55.4	0	44.6
2	0	75.0	0	25.0
3	100	0	0	0
4	23.2	28.6	26.8	21.4
5	28.6	0	0	71.4
6	16.1	42.8	0	41.1
7	0	51.8	16.1	32.1

Motif ten element

Pos	A	C	G	T
1	3.4	34.5	60.3	1.7
2	24.1	41.4	31	3.4
3	87.9	3.4	8.6	0
4	8.6	5.2	74.1	12.1
5	1.7	94.8	0	3.4
6	1.7	41.4	53.4	3.4
7	10.3	44.8	44.8	0
8	43.1	0	56.9	0
9	12.1	8.6	67.2	12.1
10	5.2	86.2	3.4	5.2
11	1.7	5.2	89.7	3.4
12	17.2	34.5	46.6	1.7

Downstream promoter element

Pos	A	C	G	T
1	51.7	0	48.3	0
2	0	0	100	0
3	58.8	0	0	41.2
4	0	55.2	0	44.8
5	21.5	30.5	48	0

Table 3.2. Core promoter position weight matrix representations used by the YAPP Eukaryotic Core Promoter Predictor program to identify core promoter elements. See next page for legend.

Table 3.2. Core promoter position weight matrix representations used by the YAPP Eukaryotic Core Promoter Predictor program to identify core promoter elements. The inputted sequence is compared, using a sliding windows algorithm, to the core promoter element position weight matrix representations and conservation index and matrix similarity scores calculated, following the methods of Cartharius et al. (2005). For each nucleotide position of the core promoter element (Pos), the position weight matrix indicates the probability (in percentage terms) of observing an A, C, G, or T.

3.3.1.2 Assessment of epigenetic modification

Regions of the *DISC1* candidate promoter region carrying epigenetic marks relevant to transcriptional activity were identified using the "ENCODE Regulation" super-track on the UCSC human genome browser (hg19) and the "GIS ChIP-PET" track on the UCSC human genome browser (hg18). The ENCODE Regulation super-track contains three sub-tracks detailing regions identified as being enriched for three histone marks: H3K4me1, H3K4me3, and H3K27ac. These regions were identified by chromatin immunoprecipitation followed by sequencing (ChIP-seq) in several human cell lines (table 3.3). The GIS ChIP-PET track displays regions enriched for H3K4me3 and H3K27me3 identified by ChIP followed by paired-end di-tag (PET) sequencing in the human embryonic stem cell line, hES3.

Cell line name	Description	Chromatin modification assessed		
		H3K4me1	H3K4me3	H3K27Ac
Gm12878	Lymphoblastoid	✓	✓	✓
H1-hESC	Embryonic stem cell	✓	✓	
HepG2	Liver carcinoma		✓	✓
HMEC	Mammary epithelial	✓	✓	✓
HSMM	Skeletal muscle myoblast	✓	✓	✓
HUVEC	Umbilical vein endothelial	✓	✓	✓
K562	Myelogenous leukaemia	✓	✓	✓
NHEK	Epidermal keratinocyte	✓	✓	✓
NHLF	Lung fibroblast	✓	✓	✓

Table 3.3. Cell lines assessed for epigenetic chromatin modifications for the ENCODE Regulation super-track on the UCSC human genome browser (hg18). ChIP followed by sequencing (ChIP-seq) was used to assess the cell lines for three modifications: H3K4me1, which is often found near regulatory elements; H34me3, which is associated with the promoters of actively transcribed or poised genes; and H3K27ac, which is associated with active regulatory elements. The name of the cell line, a description of its origin, and the chromatin modifications assessed in each cell line (indicated by a tick) are shown.

3.3.1.3 Assessment of potential for Z-DNA formation

The online program Zhunt (http://gac-web.cgrb.oregonstate.edu/zDNA/) (Ho et al., 1986) was used to assess the potential of the *DISC1* candidate promoter region for Z-DNA formation. The potential for Z-DNA formation in response to negative superhelicity is calculated by consideration of the energy required to stabilise dinucleotides in the Z-DNA conformation compared to the B-DNA confirmation. These energy requirements were either determined empirically or predicted based on empirically defined values using an algorithm that took into account the following parameters: base identity (A, C, G, or T) and base conformation (*syn* or *anti*). As the base conformation cannot be defined from the input sequence, Zhunt implements an algorithm to maximise the alternation of *syn* and *anti* base pairs in the sequence, a pattern characteristic of Z-DNA. The Z-DNA forming potential of a sequence was then calculated using an algorithm that takes into account sequence effects on the energy required for Z-DNA stabilisation by including nearest neighbour interactions. The resultant score reflects the superhelical density required for the onset of Z-DNA formation. By comparison with the scores obtained for random nucleotide sequences, a Z-score is derived that reflects the number of random nucleotide bases that must be searched in order to find a sequence that is as likely or more likely to form Z-DNA than the sequence in question. A Z-score of above 700 is considered to identify regions likely to form Z-DNA.

3.3.1.4 Identification of evolutionarily conserved regions with regulatory potential

As mentioned previously (section 3.1.2.5), the identification of functional non-coding elements can be aided by the combination of conservation data with knowledge of the sequence characteristics of known regulatory elements. The ESPERR algorithm introduced in section 3.1.2.5 has been applied to the identification *cis*-regulatory elements and the results of this analysis are available in the "ESPERR Regulatory Potential (7 Species)" track of the UCSC human genome browser. This track was, therefore, chosen to detect regions with conserved regulatory potential in the *DISC1* candidate promoter region. The track displays regulatory potential scores calculated from alignments of human, chimpanzee, macaque, mouse, rat, dog, and cow genetic sequences, thus striking a balance between the use of evolutionary convergent and divergent organisms, as suggested by Prabhakar et al. (2006). Aligned regions for which regulatory potential scores are calculated are defined as regions of the reference genome in which no region of more than 100 bases lacks alignment in at least three non-human species. For these regions, regulatory potential scores are

calculated based on the similarity of the sequence to the sequences of known regulatory elements versus regions of neutral DNA contained in a training set. Scores below 0 are deemed to reflect similarity with neutral DNA and scores above 0.1 to indicate very marked resemblance to alignment patterns typical of regulatory elements present in the training set. As sequence data for the macaque is missing in all but the most 5' region of the *DISC1* candidate promoter region (macaque sequence is available for the region aligned to human chr1: 231761561-231761764), it is worth noting that for the majority of the *DISC1* candidate promoter region the only non-human species assessed for alignment were chimpanzee, mouse, rat, dog, and cow.

Regulatory potential scores for the candidate *DISC1* promoter region were downloaded from the UCSC human genome browser, using the tables function, and evolutionarily conserved regions likely to be of regulatory importance identified by filtering regulatory potential scores using a threshold of ≥ 0.1.

3.3.1.5 Identification of putative transcription factor binding loci

Transcription factors potentially involved in the regulation of DISC1 expression were identified using the "Transcription Factor ChIP" track of the "ENCODE Regulation" super-track of the UCSC human genome browser (hg18 and hg19). Genomic regions bound by a particular transcription factor were identified by ChIP-seq in several cell lines; however, not every transcription factor was assessed in every cell line.

Potential FOXP2 binding sites were identified in the region upstream of *DISC1* corresponding to the long promoter fragment. This region (chr1: 231761579-231762608, hg19) was downloaded from the UCSC human genome browser and analysed using the DNA Pattern Find program of the Sequence Manipulation Suite (Stothard, 2000) at http://www.bioinformatics.org/sms2/dna_pattern.html. The region was searched for the consensus binding sites of the general FOX family of transcription factors, the FOXP family, and FOXP2, as well as sites differing from the consensus sequences by one base pair, as this level of deviation has previously been identified as being tolerated by FOXP2 (Vernes et al., 2007). The FOX family bind to sites with the consequence sequence TRTTKRY (Overdier et al., 1994; where R = A or G, K = G or T, and Y = C or T). FOXP binds to sites with the consensus sequence TATTTRT (Wang et al., 2003), and FOXP2 binds to sites with the consensus sequence AATTTG or ATTTGT (Stroud et al., 2006; Wu et al., 2006), as well as

the core binding site, ATTT (Konopka et al., 2009). As experimental assessment of FOXP2 binding has revealed that it can bind to the non-consensus sequence AAAGSAAA (S = G or C) (Vernes et al., 2011), this site was also searched for.

3.3.2 Luciferase reporter vectors

Three luciferase reporter vectors purchased from Promega were used in this study: the promoterless vector, pGL4.10, which encodes firefly luciferase reporter gene *luc2*; pGL4.13, which expresses *luc2* under the control of an SV40 promoter; and pRL-TK, which expresses Renilla luciferase (*Rluc*) under the control of a thymidine kinase promoter.

3.3.3 Cloning of *DISC1* promoter fragments

3.3.3.1 DNA isolation from a bacterial artificial chromosome

A bacterial artificial chromosome (BAC), RP11-17H4, covering the region chr1: 231,697,040- 231,766,242 (hg19), which includes the *DISC1* putative promoter region, was obtained from the BACPAC Resources Centre as a stab culture. The stab culture was streaked out onto a lysogeny broth (LB) agar plate containing 20µg/ml chloramphenicol and left to grow overnight at 37°C. The following day, a single colony was picked and placed in 3ml LB medium supplemented with 25µg/ml chloramphenicol and left to grow for 16 hours at 37°C in a shaking incubator. DNA was then isolated using the peqGOLD plasmid mini kit II (PEQLAB). The culture was centrifuged at 5000 x *g* for 10 minutes at room temperature, the supernatant discarded, and the pellet resuspended in 500µl solution I (resuspension buffer) containing RNase A (100µg/ml). The cell suspension was transferred to a 2ml microcentrifuge tube, and 500µl solution II (lysis buffer) added. The tube was inverted and then incubated at room temperature for five minutes to obtain a clear lysate. DNA was then precipitated by adding 700µl solution III (neutralisation buffer), inverting the tube, and centrifuging at 10000 x *g* for 10 minutes at room temperature. The supernatant (750µl) was then loaded onto a PerfectBind DNA column in a 2ml collection tube. The column was centrifuged at 10000 x *g* for 1 minute at room temperature, the flow-through discarded, and a further 750µl of supernatant loaded onto the column. The column was centrifuged again at 10000 x *g* for 1 minute at room temperature and the flow-through discarded. These steps were repeated until all the supernatant had been processed. The column was then washed using 500µl PW buffer to remove protein contamination. PW buffer was added to the column, centrifuged at 10000 x *g* for 1 minute, and the flow-through discarded. A second

wash step was then carried out using 750µl DNA wash buffer, which removes salt residues from the membrane, the column was centrifuged at 10000 x *g* for 1 minute, and the flow-through discarded. The second wash step was repeated before the column was dried to remove ethanol by centrifuging at 10000 x *g* for 2 minutes. DNA was then eluted by placing the column into a clean 1.5ml microcentrifuge tube, adding 75µl elution buffer and centrifuging at 5000 x *g* for 1 minute. Isolated DNA was stored at -20°C until required.

3.3.3.2 Primer design

Primers were designed using the online program Primer 3 (http://frodo.wi.mit.edu/) to amplify a series of four nested fragments covering different extents of the *DISC1* candidate promoter region (figure 3.3). Each promoter fragment was amplified using a unique forward primer with a common reverse primer. Each primer contains a region of genomic sequence, a restriction site (NheI for the forward primers, HindIII for the reverse primer), and an additional 4bp to ensure efficient cutting by the restriction enzyme.

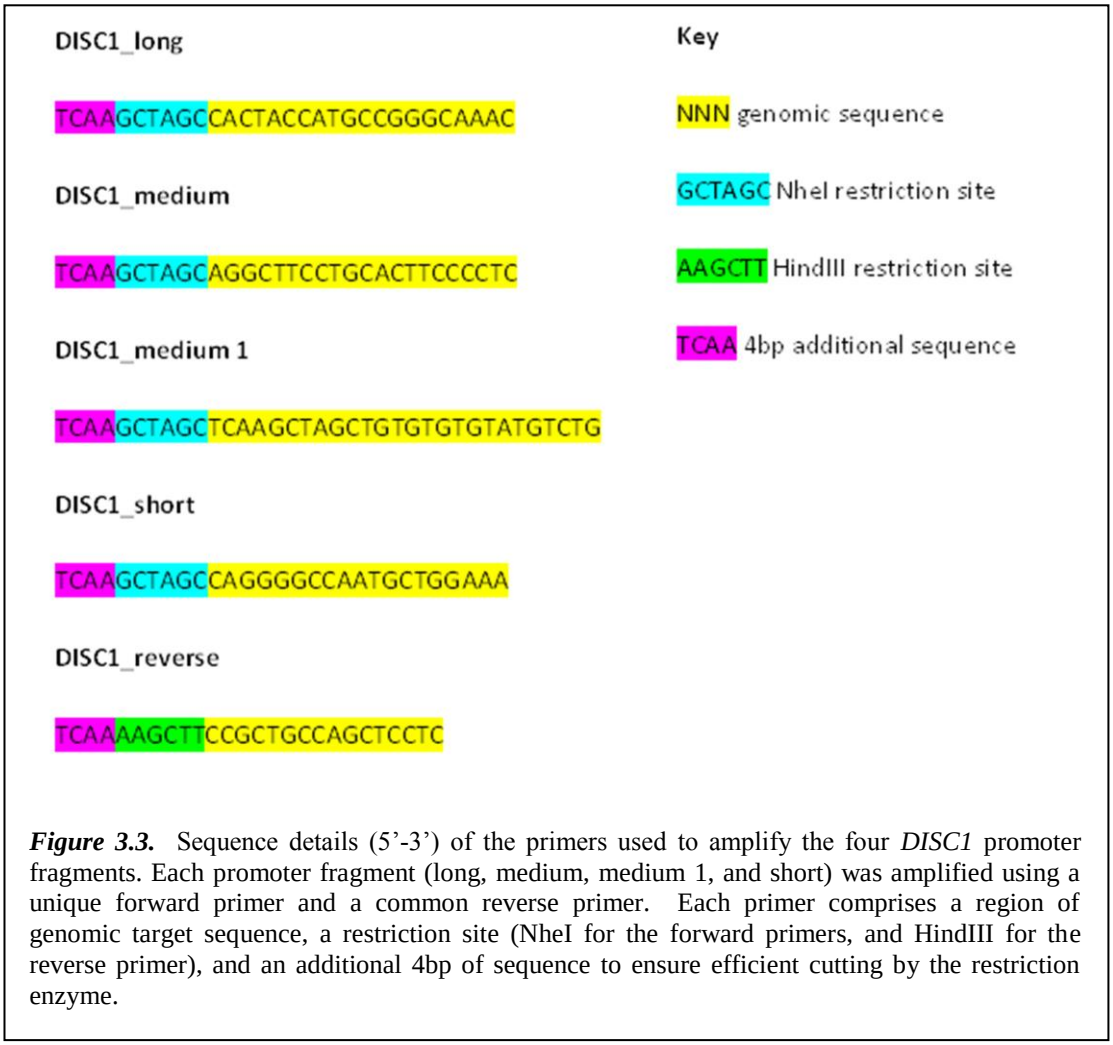


Figure 3.3. Sequence details (5'-3') of the primers used to amplify the four *DISC1* promoter fragments. Each promoter fragment (long, medium, medium 1, and short) was amplified using a unique forward primer and a common reverse primer. Each primer comprises a region of genomic target sequence, a restriction site (NheI for the forward primers, and HindIII for the reverse primer), and an additional 4bp of sequence to ensure efficient cutting by the restriction enzyme.

3.3.3.3 Polymerase chain reaction (PCR) amplification of DISC1 *promoter fragments*

The four promoter fragments were PCR-amplified from a BAC (RP11-17H4) using the Expand High Fidelity Plus PCR System (Roche), which uses a thermostable DNA polymerase with proofreading ability. One microlitre (2.5ng) of template DNA was added to a reaction containing 10µl 5x Expand High Fidelity Plus buffer, 10mM dNTPS, 4µM forward and reverse primers, 0.5µl (2.5U) Expand High Fidelity Plus Enzyme Mix, and 32.5µl double distilled water (ddH_2O). Polymerase amplification was carried out using a thermal cycler (MJ Thermocycler) using the parameters detailed in figure 3.4 for touch-down PCR. The promoter fragments were then PCR purified to remove any contaminants that might interfere with subsequent stages using the QIAquick PCR Purification Kit (Qiagen), according to the manufacturer's instructions, and eluted in 32µl elution buffer

Short and medium fragments

Initial denaturation: 94°C for 2 minutes
Denaturation: 94°C for 15 seconds
Annealing: 70°C for 30 seconds (- 1°C per cycle) } x 10 cycles
Extension: 72°C for 45 seconds
Denaturation: 94°C for 15 seconds
Annealing: 60°C for 30 seconds } x 30 cycles
Extension: 72°C for 45 seconds
Final extension: 72°C for 7 minutes

Long fragment

Initial denaturation: 94°C for 2 minutes
Denaturation: 94°C for 15 seconds
Annealing: 70°C for 30 seconds (- 1°C per cycle) } x 10 cycles
Extension: 72°C for 1 minute 15 seconds
Denaturation: 94°C for 15 seconds
Annealing: 60°C for 30 seconds } x 30 cycles
Extension: 72°C for 1 minute 15 seconds
Final extension: 72°C for 7 minutes

Figure 3.4. Thermocycling parameters for touch-down PCR amplification of the long, medium, and short *DISC1* promoter fragments.

3.3.3.4 Restriction enzyme digest

Enzymatic digestion of the inserts (DISC1 promoter fragments) and the pGL4.10 vector with the restriction enzymes *NheI* and *HindIII* (Roche). Digests were carried out sequentially for each enzyme. For the inserts, 28µl of each PCR product was digested using 0.5µl (5U) enzyme, 3.5µl 10x restriction enzyme buffer (*NheI*: Roche buffer M; *HindIIII*: Roche buffer B), and 3µl ddH_2O. For the vector, 3µg pGL4.10 was digested using 0.5µl (5U) enzyme, 10µl 10x restriction enzyme buffer, and 87.5µl ddH_2O. Digests were carried out overnight at 37°C.

3.3.3.5 Product purification

Digested products were purified using the QIAquick PCR Purification Kit according to the manufacturer's instructions. Buffer PB, which binds small single- or double-stranded PCR products (~100 bp) and primers up to 40 nucleotides in length, was added to each sample in a 5:1 volumetric ratio and mixed. The sample was then applied to a QIAquick spin column, which was placed in a 2ml collection tube and centrifuged at 17900 x *g* for 1 minute. Flow-through was discarded and the spin column replaced in the collection tube. The sample was then washed with 750µl buffer PE by centrifuging at 17900 x *g* for 1 minute, the flow-through discarded, and the column centrifuged for a further 1 minute to remove residual ethanol. Purified DNA was then eluted by placing the spin column in a clean 1.5ml collection tube, adding 32µl elution buffer, resting for 1 minute, and then centrifuging at 17900 x *g* for 1 minute. DNA was stored at -20°C until required.

3.3.3.6 Spectrophotometric analysis of DNA using the NanoDrop

The concentration and purity of DNA samples was measured using a Thermo Scientific NanoDrop 1000 spectrophotometer, a full-spectrum spectrophotometer. Accurate determination of sample concentration and purity can be made from as little as 1-2µl of sample, thus reducing the amount of sample required compared to traditional spectrophotometric methods. Samples are pipetted onto a pedestal onto the end of a fibre optic cable. An arm, containing a second fibre optic cable, is the lowered onto the sample, such that the liquid bridges the gap between the two fibre optic ends. A pulsed xenon flash lamp is used as the light source, and a CCD array is used to analyse the light after passing through the sample.

Samples were measured by selecting the "nucleic acids" program on start-up and setting the measurement option to "DNA-50". The machine was then "blanked" by taking a reading from a reference sample (the elution buffer used to re-suspend the purified DNA in section 3.3.3.5). Each sample (1.5µl) was then pipetted onto the pedestal (which was cleaned between samples), and the absorption of the sample calculated by comparison of the intensity of light that transmitted through the sample compared to that which transmitted through the blank reference sample, using the equation:

$$\text{Absorbance} = -\log\left(\text{Intensity}_{\text{sample}}/\text{Intensity}_{\text{blank}}\right)$$

Concentration was calculated from absorbance of light at 260nm using the Beer-Lambert equation, which predicts a linear change in absorbance with concentration:

$$A = e*b*c$$

Where A is the sample absorbance measurement, e is the wavelength-dependent extinction coefficient (L mol^{-1} cm^{-1}), b is the path length in centimetres, and c is the analyte molarity (M).

Sample purity was determined by calculation of the ratio of absorbance at 260nm and 280nm (260/280) and 260nm and 230nm (260/230). Pure DNA should have a 260/280 ratio of ~1.8. Lower ratios can indicate the presence of contaminating agents absorbing at around 280nm, such as protein or phenol. The 260/230 ratio should be in the range of 1.8-2.2. Lower ratios can be caused by contaminating salts, phenol, or protein.

3.3.3.7 Treatment with shrimp alkaline phosphatase

Digested pGL4.10 was treated with shrimp alkaline phosphatase (SAP) to remove 5' phosphate groups in order to prevent re-ligation. One microgram of DNA was added to a reaction containing 1µl SAP and 7µl dilution buffer (USB) and left for 4 hours at 37°C before being heated for 10 minutes at 65°C to inactivate the phosphatase.

3.3.3.8 Ligation of DISC1 *promoter fragments and the pGL4.10 promoterless vector*

The inserts were then ligated into the pGL4.10 vector upstream of the coding sequence of *luc2*. Ligations were carried out using a 5:1 molar ratio of insert to vector using the Rapid DNA Ligation Kit (Roche), according to the manufacturer's instructions. Thirty-six femtomoles (fmoles) vector and 180 fmoles insert were added to 1x buffer 2 to make a total volume of 10µl. This mixture was added to 10µl 2x buffer 1 and 1µl T4 DNA ligase and left for 1 hour at room temperature.

3.3.3.9 Endotoxin free purification of plasmid DNA

Two microlitres of the ligated vector and insert were electroporated into 20µl ElectroMax DH10B electrocompetent cells (Invitrogen), using a MicroPulser (Bio-Rad). The electroporated cells were then suspended in 400µl SOC (Super Optimal Broth with Catabolite repression) medium (Invitrogen), and left to shake for 1 hour at 37°C, before being diluted 1/5 with SOC medium (20µl cell suspension and 80µl SOC medium). The diluted mixture was then applied to an LB-agar plate containing ampicillin (100µg/ml) and left to grow overnight at 37°C.

A single colony was selected for each promoter construct and used to inoculate a starter culture of 2ml LB-broth with ampicillin (100µg/ml). The starter culture was left to grow for 8 hours in a shaking incubator at 37°C. After 8 hours, the starter culture was poured into a flask containing 200ml LB-broth with ampicillin (100µg/ml) and grown for 16 hours in a shaking incubator at 37°C. The culture was then centrifuged at 6000 x *g* for 15 minutes at 4°C in a Beckman centrifuge. DNA was purified using an EndoFree Plasmid Purification kit (Qiagen), according to the manufacturer's instructions. An EndoFree kit was used to improve transfection efficiency. The supernatant was discarded and the cell pellet resuspended in 10ml buffer P1. The sample was then lysed by adding 10ml buffer P2, mixing, and incubating the sample for 5 minutes at room temperature. DNA was then precipitated by adding 10ml buffer P3 and mixing. The sample was then applied to a QIAfilter cartridge and incubated for 10 minutes at room temperature, in order to let a precipitate containing various contaminants, such as genomic DNA, protein, and detergent to form. Following the incubation, the sample was filtered through the cartridge, 2.5ml of the endotoxin removal buffer (buffer ER) added to the lysate, the sample mixed, and incubated

on ice for 30 minutes. A QIAGEN-tip 500 was equilibrated by adding 10ml buffer QBT. The sample was then added to the equilibrated QIAGEN-tip and allowed to enter the resin in the QIAGEN-tip by gravity flow. Contaminants were removed by washing the QIAGEN-tip twice with 30ml buffer QC, and the DNA eluted by placing the QIAGEN-tip into a 30ml collection tube and applying 15ml buffer QN. Precipitation of the DNA was carried out by adding 10.5ml isopropanol, mixing, and centrifuging at 15000 x *g* for 30 minutes at 4°C. After decanting the supernatant, the DNA pellet was washed with 5ml endotoxin-free 70% ethanol and centrifuged at 15000 x *g* for 10 minutes. The supernatant was removed, the pellet air-dried, and then re-suspended in 250µl endotoxin-free TE buffer. Purified DNA was stored at -20°C until required.

3.3.3.10 Sequencing reaction

Prior to use, the promoter constructs were sequenced to confirm integrity using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). Each construct was sequenced using a forward primer and a reverse primer located in the multiple cloning site of pGL4.10, either side of where the *DISC1* promoter fragment was inserted. The primer sequences used were as follows: forward: 5'-CTAGCAAAATAGGCTGTCCC-3'; reverse: 5'-CCGTCTTCGAGTGGGTAGAA-3'. Reactions were carried out in 96-well plates. Two hundred nanograms of plasmid DNA was added to a sequencing reaction containing 1µl Terminator Ready Reaction Mix, 1µl BigDye Sequencing Buffer, 1µl 3.2 pmol primer, and 3µl ddH_2O. The reaction was carried using the following thermal cycling conditions: 96°C for 1 minute (initial denaturation), followed by 30 cycles of 96°C for 10 seconds (denaturation), 50°C for 5 seconds (annealing), and 60°C for 4 minutes (extension).

3.3.3.11 Ethanol/EDTA precipitation of sequencing reaction

The sequencing reactions were precipitated by adding 2.5µl 125mM Ethylenediaminetetraacetic acid (EDTA) and 30µl 100% EtOH to each well. The plate was then sealed and inverted four times to mix the reaction. Following a 15 minute incubation at room temperature, the plate was spun in a Jouan centrifuge at 2500 x *g* for 30 minutes. The plate was unsealed and inverted over a paper towel and spun briefly to remove the EtOH. Forty microlitres 70% EtOH was then added to each well, the plate resealed and mixed by inverting 4 times. The plate was then centrifuged at 2500 x *g* for 15 minutes, the seal removed, the plate inverted over a paper towel again and centrifuged briefly to remove the

EtOH. Wells were air dried and the plate sealed with adhesive film before being sent to the Medical Research Council Human Genetic Unit to be sequenced by Agnes Gallacher.

3.3.3.12 Assessment of the DNA conformation of the DISC1 promoter constructs

As the conformation of plasmid DNA has been shown to affect the efficiency with which it is transfected (Cherng et al., 1999), the DISC1 promoter constructs were visually assessed for the proportion of DNA in the supercoiled form, which transfects most efficiently (Cherng et al., 1999), the open circle form, and the linear form. These three forms can be differentiated by gel electrophoresis as they migrate at different rates: The supercoiled form migrates fastest through the gel; the next fastest is the linear; and the open circle form migrates the slowest. One hundred nanograms of plasmid DNA was combined with 4 x Orange G loading buffer (1g Ficoll-400 (Sigma) and 0.2ml 50mM EDTA made up to a total volume of 10ml with ddH_2O. Orange G added to colour) and made up to a total volume of 8µl with ddH_2O. Samples were electrophoresed on a 1% tris-borate-EDTA (TBE) gel containing 5µl SYBRSafe (Invitrogen). The gel was then visualised under ultraviolet illumination.

3.3.4 Dual luciferase reporter assay

3.3.4.1 Cell culture and transient transfection

The neuroblastoma cell lines, SH-SY5Y and LAN-5, and the human embryonic kidney cell line, HEK293, were maintained in Dulbecco's Minimal Essential Medium (DMEM; Invitrogen) supplemented with 10% foetal bovine serum (FBS) at 37°C in a humidified 5% CO_2 atmosphere. Cells were grown to ~80% confluency in tissue culture flasks before being transfected using Nucleofector (Lonza Biologics), according to the manufacturer's instructions. Cell culture medium was aspirated off, the cells were washed in PBS, and dissociated using TrypLE Express (Invitrogen). Cells were then counted and, for each promoter construct, 1×10^6 cells were harvested by centrifuging at 90 x *g* for 10 minutes at room temperature. After removing the supernatant, the cell pellet was resuspended in 100µl Nucleofector Solution V. The 100µl cell suspension was then combined with a total of 2µg DNA, comprising (i) 1961ng of the appropriate reporter construct (promoterless pGL4.10 for the negative control, or pGL4.13 vector for the positive control, or pGL4.10-*DISC1* promoter construct (long, medium, medium 1, or short)), and (ii) 39ng of the transfection

efficiency control pRL-TK. The solution was transferred to a cuvette and inserted into the Nucleofector Cuvette Holder and transfected using the program A-023, which is recommended for transfecting SH-SY5Y and HEK293 cells with high viability (there is no established Nucleofector program recommended for LAN-5 cells). The solution was then mixed immediately with 500µl culture medium, and 31.25µl was pipetted into each of 6 wells of a 96-well plate, which had previously been filled with 62.5µl culture medium and warmed to 37°C. These six wells were treated as technical replicates. The activity of each promoter construct was assessed in three experimental replicates.

3.3.4.2 Measurement of luciferase activity

Following transfection with the promoter constructs, cells were incubated at 37°C for 24 hours, after which time the cells had reached approximately 80% confluency. At this point, the cell culture medium was removed and the cells washed with PBS. Lysis was achieved by adding 20µl 1 x Passive Lysis Buffer (Promega) to each well and shaking at room temperature for 1 hour.

The dual luciferase reporter (DLR) assay was carried out using the Dual-Luciferase Reporter Assay System (Promega). Five microlitres of cell lysate was transferred from each well of the 96-well cell culture plate to a well of a 96-well half-well opaque white plate. An opaque plate was used to reduce cross-talk between adjacent wells. The plate was then inserted into an Omega FLUOstar luminometer (BMG Labtech). The first well was then injected with 12.5µl luciferase assay reagent II (LARII), which activates firefly luciferase, and a reading of firefly luciferase activity taken over a period of 10 seconds following a 2 second post-injection delay, during which time the plate was shaken at 550rpm. Once the firefly luciferase activity had been measured, the well was injected with 12.5µl Stop & Glo reagent, which quenches the firefly luciferase signal and activates the *Renilla* luciferase signal. Again, *Renilla* luciferase activity was measured over a period of 10 seconds following a 2 second post-injection delay whilst the plate was shaken. This process was then repeated for the remaining wells on the plate.

3.3.4.3 Calculation of normalised relative luciferase activity

Assessment of the activity of each *DISC1* promoter construct was carried out in three experimental replicates. Within each experimental replicate, six technical replicates were

performed. The stages involved in the calculation of the normalised relative luciferase values for each promoter construct are summarised in figure 3.5. Transfection efficiency was controlled for by normalising the firefly luciferase value for each well (obtained from the empty pGL4.10 vector, pGL4.10-*DISC1* promoter construct, or pGL4.13) to the corresponding *Renilla* luciferase value. For each plate, subtraction of the mean normalised luciferase value for the empty pGL4.10 vector controlled for the contribution of background luminescence. The pGL4.13 vector, which expresses the firefly luciferase gene *luc2* under the control of the constitutively active SV40 promoter, acted as a between-plate calibrator.

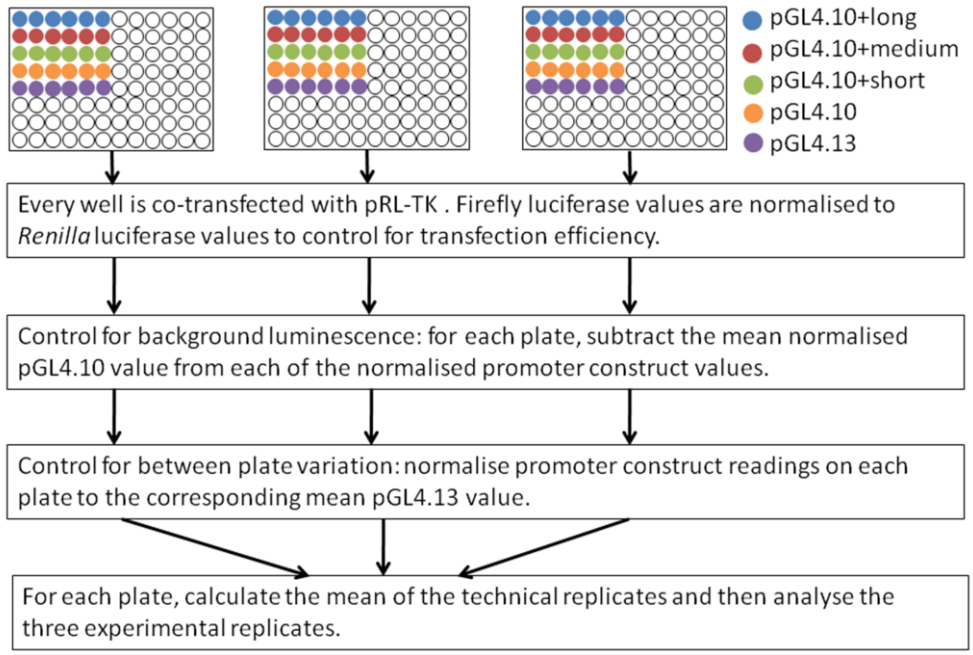


Figure 3.5. Schematic summarising the stages involved in obtaining normalised relative luciferase activity values for each of the *DISC1* promoter constructs assessed using the dual luciferase reporter assay. Cells were co-transfected with a firefly luciferase (*luc2*) containing vector (either the promoterless vector pGL4.10 or pGL4.13, which expresses *luc2* under the control of an SV40 promoter), together with pRL-TK, which expresses *Renilla* luciferase under the control of a thymidine kinase promoter. Transfections were performed in six technical replicates (indicated by the six wells of the same colour on each 96-well plate) and repeated in three experimental replicates.(represented by the three 96-well plates).

3.3.4.4 Statistical analysis of DLR assay data

Statistical analysis was carried out using SPSS (Statistical Package for the Social Sciences) 17.0 (Apache Software Foundation). Significant main effects were identified by one-way

analysis of variance (ANOVA) and post-hoc analysis was carried out using Tukey's Honestly Significant Difference (HSD) test. When analysing data with equal sample sizes, the *F*-test has been shown to be robust to deviations from the normality and homogeneity of variances assumptions (Boneau, 1960); therefore, formal assessment of these assumptions was not performed.

3.4 Results

3.4.1 *In silico* analysis of the *DISC1* candidate promoter region

3.4.1.1 Selection of a DISC1 candidate promoter region

The region spanning from 1000 bp upstream of the *DISC1* TSS to the translation start site (chr1: 231761561-231762613) was selected as the *DISC1* candidate promoter region. This region was chosen based upon the findings of the ENCODE Pilot Project, which characterised 45 promoters and found that, on average, the sequence -300 bp to -50 bp relative to the TSS contributes positively to core promoter activity, whereas, in 55% of the promoters studied, the region further upstream, -1000 bp to -500 bp relative to the TSS, confers a repressive effect (Cooper et al., 2006) (figure 3.6). The region between the TSS and the translation start site was included as this region can influence promoter activity and sometimes contains canonical promoter motifs, such as the downstream promoter element (DPE) (Kadonaga, 2002).

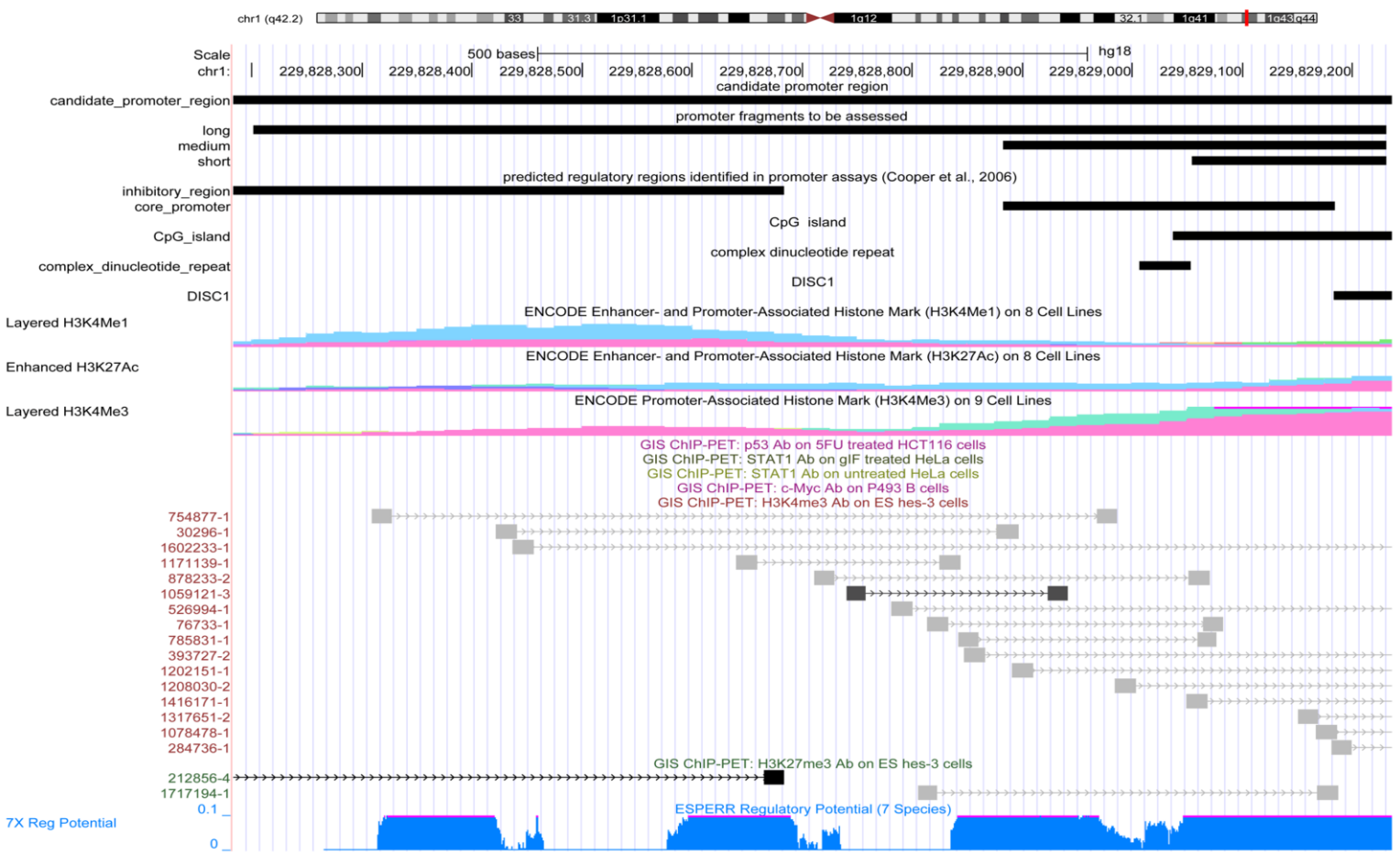


Figure 3.6. Schematic showing the information used from the UCSC human genome browser (hg18) in the design of the *DISC1* promoter fragments. See next page for full legend.

Figure 3.6. Schematic showing the information used from the UCSC human genome browser (hg18) and other publicly available sources in the design of the *DISC1* promoter fragments. The region designated as the *DISC1* candidate promoter region (chr1: 229828184-229829236) is shown together with the promoter fragments to be assessed in the dual luciferase reporter assay (long, medium, and short). The track marked "DISC1" indicates the location of the DISC1 transcript. Regions previously identified as generally contributing positively or negatively to promoter activity in the promoters investigated as part of the ENCODE Pilot Project (Cooper et al., 2006) are indicated. A CpG island identified from the UCSC human genome browser CpG Island track is depicted together with the complex dinucleotide repeat region. Regions identified as carrying histone modifications associated with promoter activity are denoted by three sub-tracks of the ENCODE Regulation super-track and the GIS-ChIP-PET track. The ENCODE Regulation sub-tracks indicate regions identified by ChIP-chip as carrying three histone marks: H3K4me1, H3K4me3, and H3K27ac. Data from eight (H3K4me1 and H3K27ac) or nine (H3K4me3) cell lines is presented in overlay format, such that each cell line is represented by a different colour and the height of the bars represents the extent of enrichment at each position in the genome. The GIS-ChIP-PET track shows regions identified by ChIP-PET as carrying the H3K4me3 modification and the repressive H3K27me3 modification in human embryonic stem cells (hES3). Regions enriched for a histone modification are indicated by two blocks, representing the ends of the di-tag pair, connected by a thin arrowed line. The direction of the arrows indicates the orientation of the ChIP-ed sequence and the colour of the line indicates the number of sequences identified overlapping that particular location (light grey: one or two sequences, dark grey: three sequences, and black: four or more sequences). The ESPERR Regulatory Potential (7 Species) track displays regulatory potential scores calculated from alignments of human, chimpanzee, macaque, mouse, rat, dog, and cow genetic sequences. Regulatory potential scores below 0 are deemed to reflect similarity with neutral DNA and scores above 0.1 to indicate very marked resemblance to alignment patterns typical of regulatory elements present in the training set. NB. The overlay format of the ENCODE Regulation sub-tracks is best viewed online (http://genome.ucsc.edu/); here the track displays in transparent overlay method rather than the solid overlay format available in the downloaded image.

3.4.1.2 Detection of canonical promoter motifs

Analysis of the *DISC1* candidate promoter region using the eukaryotic core promoter prediction program, YAPP, revealed a lack of canonical core promoter elements, including the TATA box, Transcription factor IIB Recognition Element (BRE), Initiator, and DPE, in their expected location relative to the TSS when the default matrix similarity score threshold of 0.8 was used. To ensure that this threshold was not overly conservative, the analysis was repeated with the lower thresholds of 0.7 and 0.6. This still yielded no matches. The region does contain a CpG island (Chr1: 231762415-231763115), which is a common feature of constitutively active genes. Interestingly, at the 5' end of the CpG island, there is a complex dinucleotide repeat region ($(TG)_4TATGTC(TG)_8(CG)_8$).

3.4.1.3 Identification of epigenetic modifications associated with transcriptional activity in the DISC1 *candidate promoter region*

The "ENCODE Regulation" super-track on the UCSC human genome browser (hg19) and the "GIS ChIP-PET" track on the UCSC human genome browser (hg18), which indicate regions of ChIP-identified histone modification, were used to assess the *DISC1* candidate promoter region for H3K4me3, H3K4me1, H3K27ac, and H3K27me3 modifications (figure 3.5). H3K4me1 is often found near regulatory elements, H3K4me3 is associated with promoters that are actively transcribed or poised for transcription, and H3K27ac is associated with active regulatory elements. H3K27me3 is a repressive histone modification.

H3K4me3, H3K4me1, and H3K27ac histone marks were found to be present in all the cell lines assessed by the ENCODE regulation track, although to different extents in each cell line. The H3K4me3 and H3K27ac modifications were most enriched nearest the *DISC1* transcription start site, whereas the H3K4me1 modification was present further upstream. The GIS ChiP-PET track identifies both the H3K4me3mark and the H3K27me3 mark in the *DISC1* promoter region in embryonic stem cells, suggesting that *DISC1* might fall into the category of genes under the control of a bivalent promoter.

3.4.1.4 Assessment of the Z-DNA forming potential of the DISC1 *candidate promoter region*

Assessment of the *DISC1* candidate promoter region for its potential to form Z-DNA in response to negative superhelicity was assessed using the program Zhunt. A region of 56 base pairs with the genomic coordinates chr1: 231762380-231762435 was identified as being highly likely to form Z-DNA with a Z-score of 31396934.60, compared to the threshold Z-score for Z-DNA formation of 700 (Ho et al., 1986). This region spans the complex dinucleotide repeat region.

3.4.1.5 Identification of evolutionarily conserved regions with regulatory potential in the DISC1 *candidate promoter region*

Regions of the *DISC1* candidate promoter region likely to contain conserved regulatory elements were identified using the "ESPERR Regulatory Potential (7 Species)" track of the UCSC human genome browser. As sequence data for the macaque is missing in all but the most 5' region of the *DISC1* candidate promoter region (macaque sequence is available for the region aligned to human chr1: 231761561-231761764), the majority of comparisons within this region were between human, chimp, cow, dog, rat, and mouse. Using a regulatory potential score of 0.1 as the threshold for the detection of putative regulatory importance, the identified loci visually clustered into four distinct regions (figure 3.6).

3.4.2 A series of *DISC1* promoter constructs for assessment in the dual luciferase reporter assay

A series of three nested promoter constructs (pGL4.10-short, -medium, and -long) were designed to characterise the candidate promoter region (figure 3.6). The long construct extends from -982 bp to +47 bp relative to the TSS, and, therefore, extends almost to the translation start site, which is located at +52 bp relative to the TSS. This construct was designed to include the region from -1000 bp to -500 bp identified by Cooper et al. (2006) as typically repressing promoter activity in the 45 promoters studied as part of the ENCODE Pilot Project. In the *DISC1* candidate promoter region, this typically repressive region coincides with the area showing the greatest enrichment for the repressive H3K27me3 histone modification. Two regulatory potential peaks are included in this promoter fragment, possibly suggesting two clusters of regulatory motifs.

The medium construct, which spans the region from -300 bp to +47 bp relative to the TSS, was designed to correspond to the region identified by Cooper et al. (2006) as usually contributing positively to core promoter activity. This construct contains the complex dinucleotide repeat region, and an additional 124 bp upstream. Consistent with the possibility that this region might contribute positively to promoter activity, the medium fragment coincides with regions identified as showing the greatest enrichment for the H3K4me3 and the H3K27ac histone modifications, which are often associated with actively transcribed genes. Two regions of high regulatory potential are contained within the medium fragment.

The short construct, which includes the region -129 bp to +47 bp relative to the TSS, was designed to assess the contribution of the sequence from the translation start site to the 3' end of the complex dinucleotide repeat region. Like the medium fragment, the short fragment includes sequence identified as carrying the typically activating histone modifications, H3K4me3 and H3K27ac. The short construct contains only the more 3' region of the two regions of high regulatory potential included in the medium fragment.

3.4.3 Cloning of the *DISC1* promoter constructs and confirmation of sequence

The *DISC1* promoter constructs were cloned upstream of the luciferase reporter gene, *luc2*, in the promoterless pGL4.10 vector (figure 3.7). The constructs were sequenced to confirm their integrity using primers located in the pGL4.10 multiple cloning region, either side of the *DISC1* promoter fragment insert. All constructs were found to contain an insert with the correct sequence in the correct orientation. A sequencing trace for the *DISC1* medium promoter construct is shown as an example (figure 3.8).

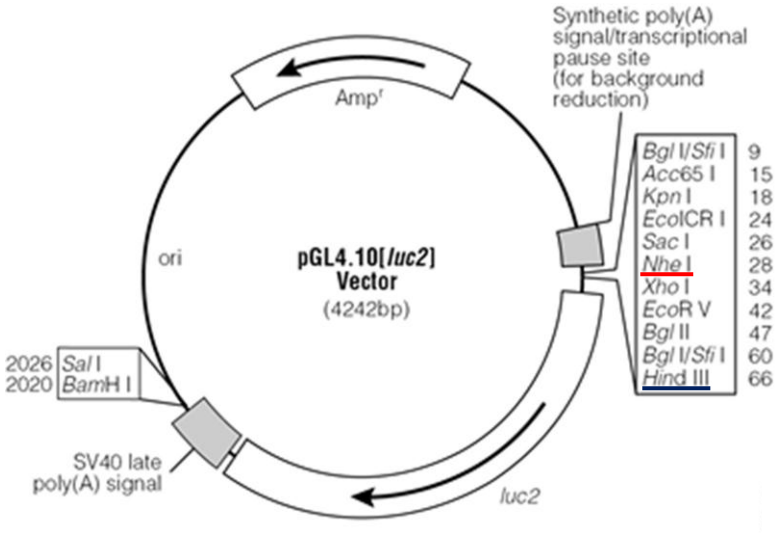


Figure 3.7. Schematic showing the features of pGL4.10, the promoterless vector used to assess the promoter activity of the *DISC1* promoter constructs. *DISC1* promoter constructs were cloned between the Nhe1 (underlined in red) and HindIII (underlined in blue) restriction sites in the multiple cloning region, upstream of the firefly luciferase (*luc2*) gene. The vector additionally contains a synthetic poly(A) signal/transcriptional pause site to reduce background expression, an SV40 late poly(A) signal to enhance RNA stability and translation efficiency, a ColEI-derived origin of replication (ori), and an ampicillin resistance gene (Amp^r).

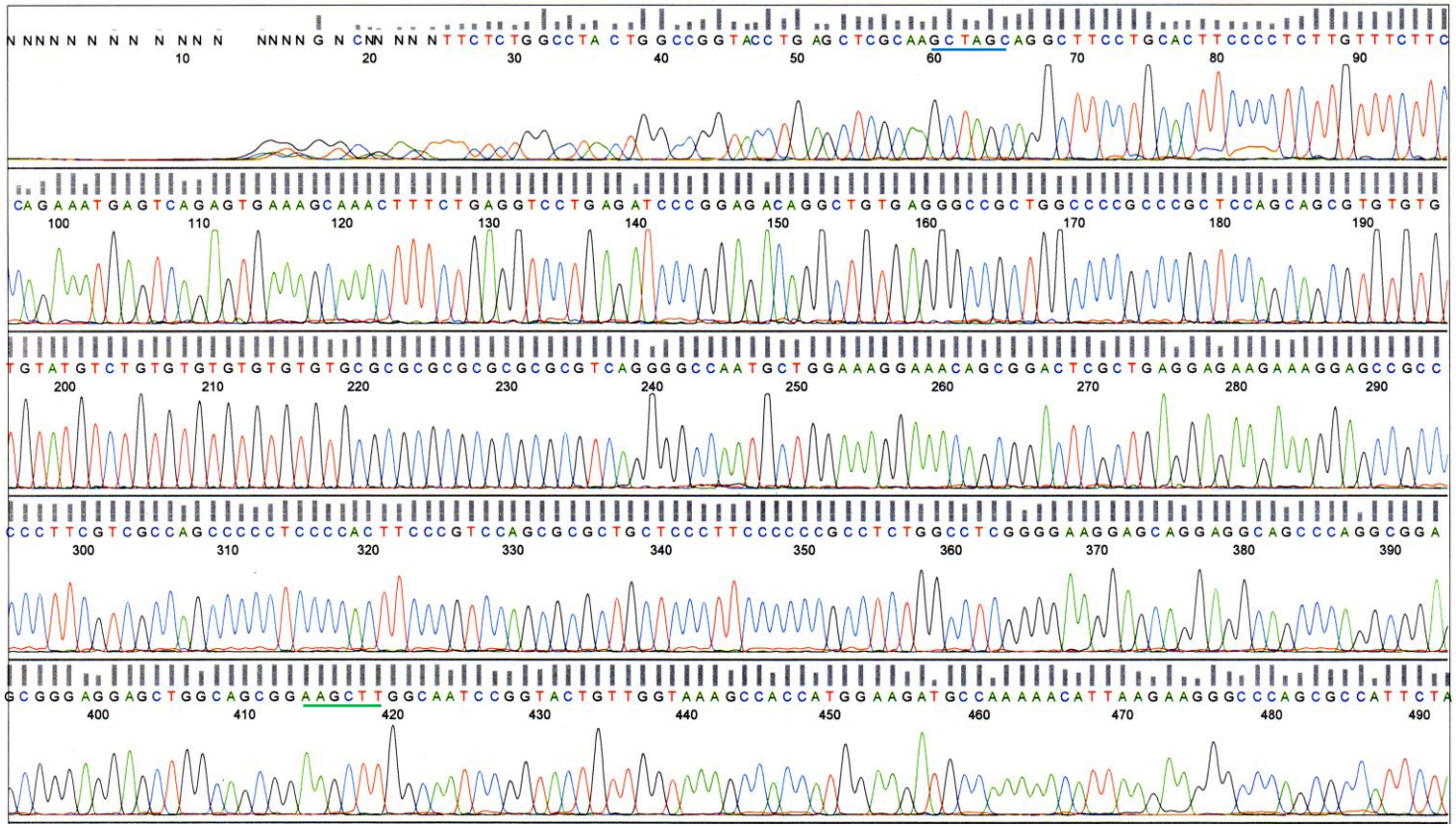


Figure 3.8. Sequencing trace showing the *DISC1* medium promoter fragment cloned into the promoterless pGL4.10 vector. Insert sequence and orientation were assessed using a forward primer located upstream of the intended insert location in the pGL4.10 multiple cloning region. Restriction sites are indicated by underlining: the blue line indicates the NheI site and the green line indicates the HindIII site. The *DISC1* medium promoter fragment is present as a 348 bp insert between the two restriction sites.

3.4.4 Assessment of *DISC1* promoter constructs in the dual luciferase reporter assay

3.4.4.1 Cell line selection

Three cell lines were chosen for the assessment of the *DISC1* promoter constructs: two neuroblastoma cells lines, SH-SY5Y and LAN-5, and the human embryonic kidney cell line, HEK293. The rationale for assessing the promoter constructs in multiple cell lines was that this would reveal the activity of the constructs in a range of cellular environments, thus reducing the likelihood that any findings would be cell line specific. All three cells lines express DISC1 endogenously (James et al., 2004; Murdoch et al., 2007).

3.4.4.2 Assessment of DNA conformation

Plasmid DNA can exist in supercoiled, open circle, and linear forms. As the supercoiled form has been shown to transfect with the greatest efficiency (Cherng et al., 1999), the proportion of supercoiled DNA in each of the *DISC1* promoter construct preparations was assessed. Gel electrophoresis revealed the majority of DNA to be in supercoiled form and did not suggest any obvious differences in the proportions of supercoiled DNA between the three constructs (figure 3.9).

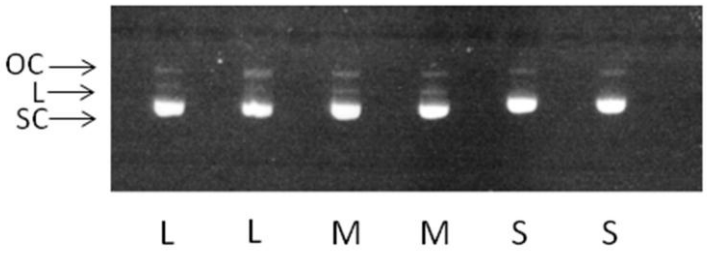


Figure 3.9. Assessment of plasmid DNA conformation. The conformation of the plasmid DNA preparations for the pGL4.10-short (S), -medium (M), and –long (L) *DISC1* promoter constructs was assessed by gel electrophoresis. Two endo-free maxi-prep preparations of each construct were assessed for proportions of open circle (OC), linear (LC), and supercoiled (SC) DNA by electrophoresing 100ng of each preparation on a 1% tris-borate-EDTA gel.

3.4.4.3 Assessment of the short, medium, and long DISC1 *promoter constructs using the dual luciferase reporter assay*

The short, medium, and long *DISC1* promoter fragments were assessed for their ability to drive gene expression using the dual luciferase reporter (DLR) assay. SH-SY5Y, LAN-5, and HEK293 cells were co-transfected with (i) either a *DISC1* promoter construct (pGL4.10-long, -medium, or -short), the promoterless pGL4.10 vector, or pGL4.13, which expresses *luc2* under the control of an SV40 promoter, and (ii) the transfection efficiency control, pRL-TK. Transfections were performed in six technical replicates and the experiment repeated in three experimental replicates.

As described in section 3.3.4.3, normalisation of the firefly luciferase values (arising from the pGL4.10 or pGL4.13 vectors) to the *Renilla* luciferase value of the co-transfected pRL-TK vector provided a control for transfection efficiency. Promoterless pGL4.10 was assessed as an indicator of background luminescence. Within each experimental replicate, following normalisation for transfection efficiency, the mean value of the six pGL4.10 technical replicates was subtracted from the normalised luciferase values of the remaining wells on the plate. For each well, the resultant value was divided by the mean value of the six technical replicates for pGL4.13, which acted as an inter-plate (and thus inter-experimental replicate) calibrator.

One-way ANOVA revealed significant differences between the promoter activity of the three *DISC1* promoter constructs in both SH-SY5Y and HEK293 cells (SH-SY5Y: $p \leq 0.001$, figure 3.10A; HEK293: $p = 0.006$, figure 3.10B). In both cases, Tukey's HSD post-hoc test revealed that the promoter activity of the medium construct was significantly greater than that of the long (SH-SY5Y: $p \leq 0.001$; HEK293: $p = 0.016$), and the short (SH-SY5Y: $p \leq 0.001$; HEK293: $p = 0.006$) constructs. No significant difference was observed between the constructs in LAN-5 cells ($p = 0.206$; figure 3.10C); however, the trend for greater expression from the medium construct compared to either the long or the short constructs was maintained.

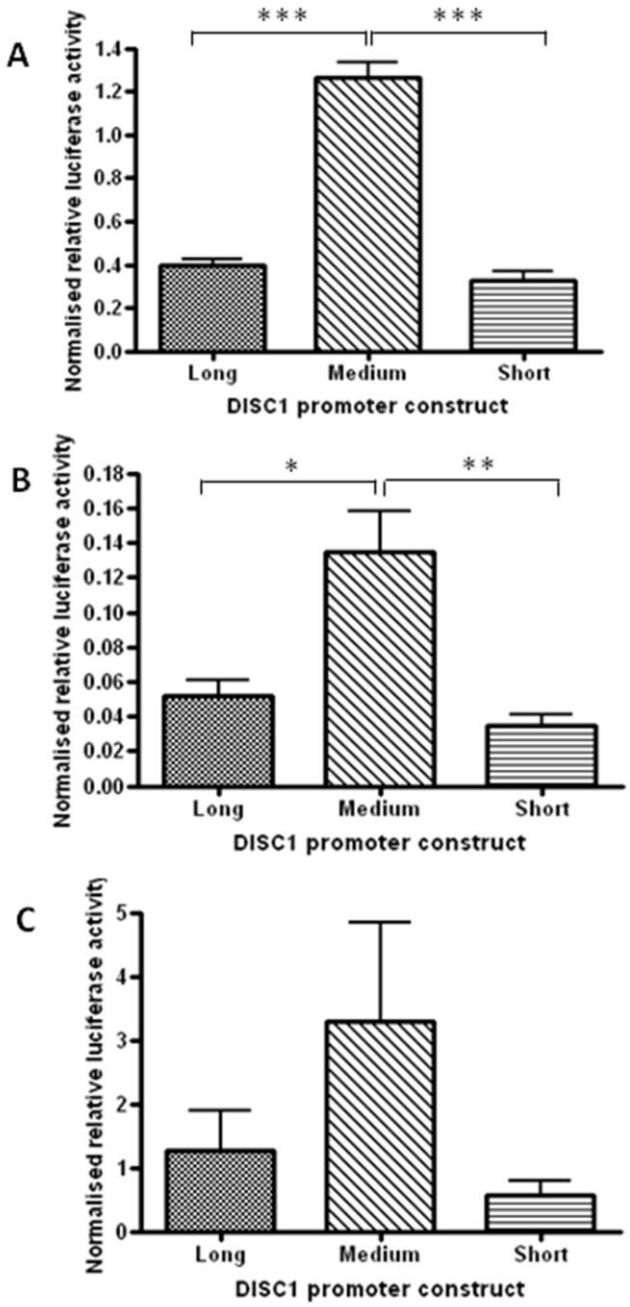


Figure 3.10. Characterisation of the *DISC1* candidate promoter region using the dual luciferase reporter assay. **A**., **B**., and **C**. Dual luciferase reporter assays comparing the promoter activity of the short, medium, and long *DISC1* promoter constructs reveal that the medium construct yields the highest level of promoter activity in SH-SY5Y (A), HEK293 (B), and LAN-5 (C) cells (n = 3). Statistical significance was assessed by one-way ANOVA followed by Tukey's HSD post-hoc test. *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$. Error bars represent standard error of the mean. This figure is adapted from a figure published previously in Walker et al. (2012)

3.4.4.4 Assessment of the role of the complex dinucleotide repeat region

To assess whether the complex dinucleotide repeat region underlies the enhanced promoter activity of the medium construct a further construct (pGL4.10-medium-1) was designed. This construct spans the region -176 bp to +47 bp relative to the TSS, thus comprising the short construct plus the dinucleotide repeat region. This construct does not contain the 124 bp 5' of the dinucleotide repeat region included in the medium construct, thus permitting a more accurate assessment of the contribution of the dinucleotide repeat region (figure 3.11A). Having established similar promoter activity profiles in all cell lines for the long, medium, and short constructs, assessment of the medium-1 construct was carried out only in SH-SY5Y cells. Tukey's HSD revealed that the medium-1 construct had significantly lower promoter activity than the medium construct ($p = 0.001$) and did not differ significantly from either the short, or the long construct (both $p \geq 0.4$, figure 3.11B).

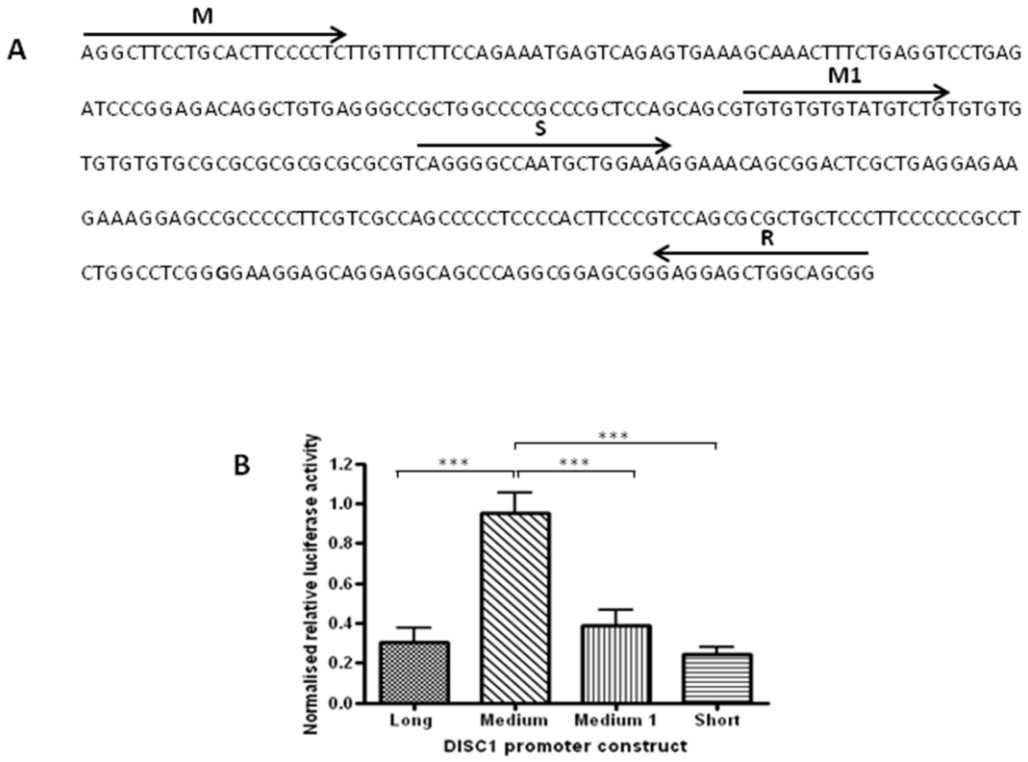


Figure 3.11. Assessment of the contribution of the dinucleotide repeat region to *DISC1* promoter activity. **A.** Design of the pGL4.10-medium-1 (M1) construct to assess the contribution of the complex dinucleotide repeat region. The genomic locations of the primers used to amplify the M1 construct (forward: M1, and reverse: R), are shown in relation to the primers used to amplify the medium (forward: M) and short (forward: S) constructs, which were amplified using a common reverse primer (R). **B.** Dual luciferase reporter assays suggest that the complex dinucleotide repeat region does not underlie the enhanced promoter activity of the medium promoter construct (n = 3). Statistical significance was assessed by one-way ANOVA followed by Tukey's HSD post-hoc test. $*p \leq 0.05$, $**p \leq 0.01$, $***p \leq 0.001$. Error bars represent standard error of the mean. This figure is adapted from a figure published in Walker et al. (2012).

3.4.5 Identification of transcription factors potentially involved in the regulation of *DISC1* expression

Having assessed the transcriptional activity of a series of *DISC1* promoter fragments, it was of interest to identify transcription factors that might be involved in regulating *DISC1* promoter activity. As discussed previously (section 3.1.1.2), for the identification of putative TFBSs, there are several advantages to using databases of ChIP-identified transcription factor-bound DNA regions over a purely computational approach. Hence, the ENCODE Regulation super-track on the UCSC human genome browser (hg 18 and hg 19) was used to search the *DISC1* candidate promoter region. Two questions were of interest: firstly, does the data within this track yield any clues to the mechanisms contributing to different levels of activity of the four DISC1 promoter fragments? Secondly, do any of the transcription factors identified as potentially regulating *DISC1* link the dysregulation of *DISC1* to any known risk factors/mechanisms for psychiatric illness?

With regards to the first question, it was difficult to determine which transcription factors in the ENCODE Regulation track, if any, might contribute to the activating properties of the medium construct compared to the medium 1 construct and the short construct as the resolution of the ChIP-seq performed by ENCODE was such that no transcription factor was found to bind to a region of DNA unique to the medium construct (figure 3.11; table 3.4). Sequencing of the ChIP-ed fragments with additional primers to increase the spatial resolution of the data would be necessary to determine more precisely the regions of DNA required for transcription factor binding. Looking at the data from the hg18 assembly, there is one transcription factor, Max, predicted to bind in a region only contained within the long construct (figure 3.11A). Max is a basic/helix-loop-helix/leucine zipper transcription factor that either forms heterodimers or homodimers to exert an effect on transcription. Depending on the nature of the dimer formed and the DNA sequence bound, Max can exert either an activating effect, a repressive effect, or no effect on transcription (Solomon et al., 1993). As such, further investigation of the nature of Max binding in the DISC1 promoter region would be a worthwhile pursuit in investigating the mechanisms behind the low transcriptional activity of the long fragment relative to the medium fragment.

Regarding the second question, the presence of a FOXP2-bound region immediately upstream of DISC1 in the hg18 assembly (figure 3.12A) was of particular interest. Moreover, the FOXP2-bound region was identified in a neuronal cell line, PFSK-1 (table

3.4), thus increasing the likelihood that this finding might be of relevance to psychiatric illness. It is important to note that FOXP2 was not amongst the transcription factors assessed for the hg19 version of the ENCODE Regulation track (i.e. its absence in this assembly does not indicate a negative result). Genetic variation in the *FOXP2* gene has previously been implicated in speech and language function (Feuk et al., 2006; Lai et al., 2001; Lennon et al., 2007; MacDermot et al., 2005; Shriberg et al., 2006; Tomblin et al., 2009; Zeesman et al., 2006), and FOXP2 has been shown to regulate a network of genes involved in brain development and neurite outgrowth (Vernes et al., 2011). Furthermore, some of FOXP2's transcriptional targets include genes previously implicated in autism (Mukamel et al., 2011; Vernes et al., 2008). Thus, a regulatory relationship between FOXP2 and DISC1 would represent a potential mechanism linking dysregulated *DISC1* expression to neurodevelopmental disorders. It would, therefore, be of great interest to establish experimentally whether *DISC1* is regulated by FOXP2.

Other transcription factors identified as potentially playing in the regulation of *DISC1* also have relevance to current understanding of psychiatric illness. Genetic variation inTCF4 (also known as TCF7L2) has been associated with schizophrenia (Alkelai et al., 2012; Hansen et al., 2011). Furthermore, TCF4 is a member of the Wnt signaling pathway, dysfunction of which has been implicated in psychiatric illness (Okerlund and Cheyette, 2011). HDAC2, a histone deacetylase, is one of the targets of histone deactylase inhibitors, which have attracted attention as potential drug therapies for psychiatric illness (Grayson et al., 2010). c-fos, FOSL2, c-jun, JunD and BATF are members of the activator protein 1 (AP-1), family of transcription factors that modulate gene expression in response to cytokines, growth factors, stress, and bacterial and viral infections (Betz et al., 2010; Hess et al., 2004). NF-κB, IRF4, and PU.1 are also regulators of the immune response (Carotta et al., 2010; Hayden et al., 2006; Shaffer et al., 2009), which, as introduced in section 1.3, is believed to play an important role in the aetiology of schizophrenia. Furthermore, NF-κB and PU.1 have been found to show altered expression in schizophrenia (Song et al., 2009; Weigelt et al., 2011). Taken together, the known functions of transcription factors identified by ChIP as potential regulators of DISC1 expression suggest that DISC1 might be involved in a regulatory network that when perturbed confers risk for psychiatric illness.

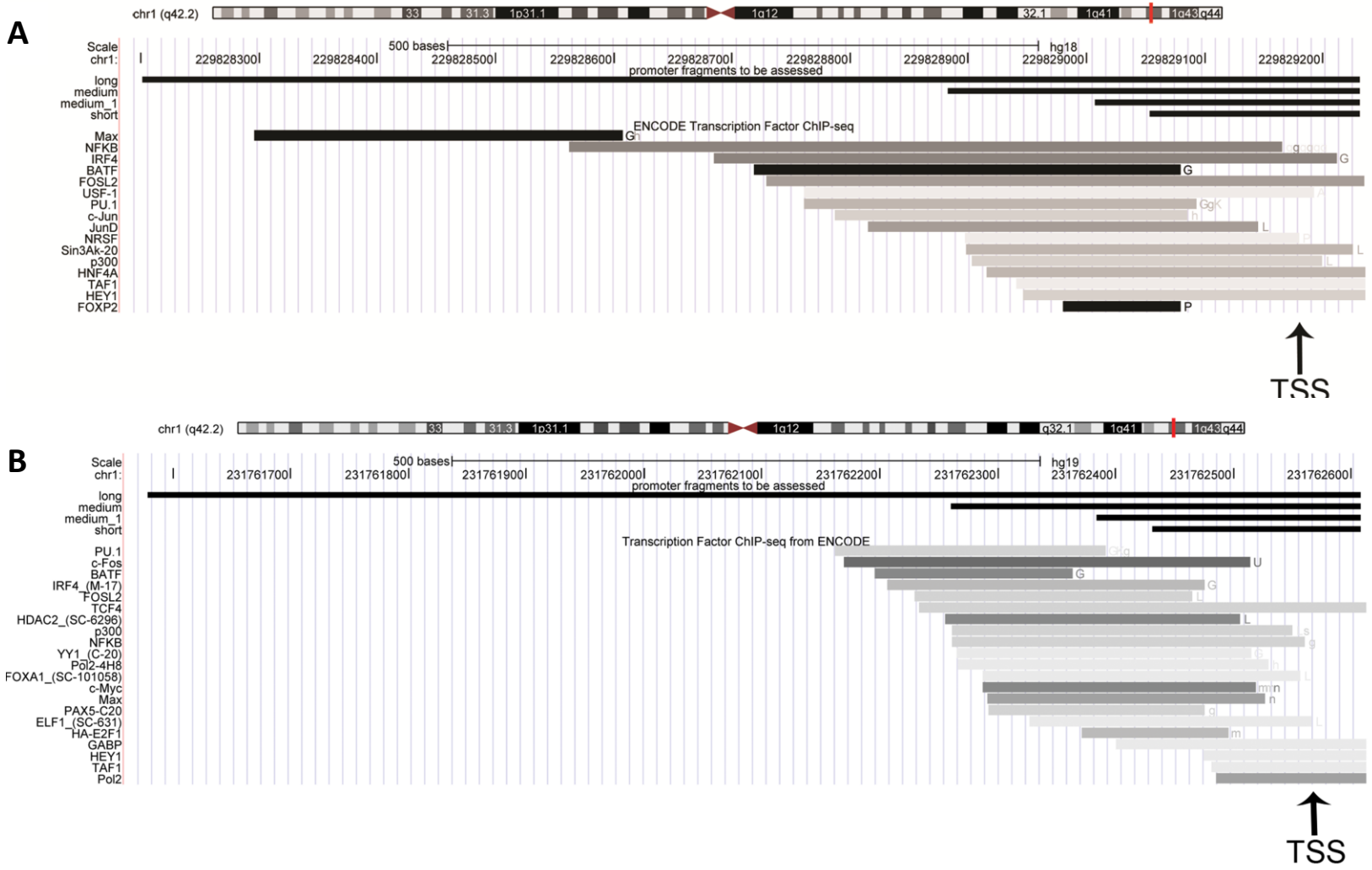


Figure 3.12. Schematic showing the location of regions of DNA identified by ChIP as being bound by a transcription factor. See next page for legend.

Figure 3.12. Schematic showing the location of regions of DNA identified by ChIP as being bound by a transcription factor. The *DISC1* upstream region was assessed for putative transcription factor binding regions using the ENCODE Regulation super-track of the UCSC human genome browser, which contains ChIP data for several transcription factors in several cell lines. As the cell lines and transcription factors assessed differed between the hg18 and hg19 assemblies of the genome browser, findings for both assemblies are displayed (A: hg18; B: hg19). The genomic region assessed was identical for both assemblies and has the co-ordinates chr1:229828184-229828231 for hg18 and chr1: 231761561-231762608 for hg19. For **A** and **B**, the location of the DISC1 transcription start site (TSS) is indicated by an arrow, and the locations of the four DISC1 promoter fragments assessed in the dual luciferase reporter assay are shown for reference. Regions identified as being bound by a transcription factor are indicated by a grey rectangular box, with the darkness of the box indicating the greatest observed level of enrichment for the transcription factor in any of the cell lines assayed (darker = greater enrichment). The identity of the transcription factor is indicated on the left-hand side of the image. At the right-hand side of each bar are letters representing the cell lines in which binding was detected, as these are not all visible here, this information is summarised in table 3.4.

Transcription factor	Assembly	Cell line(s)
Max	hg18 and hg19	Acute promyelocytic leukemia (NB4), lymphoblastoid (GM12878), and umbilical vein endothelial cells (HUVEC)
NFKB	hg18 and hg19	Lymphoblastoid (GM10847, GM12891, GM12892, GM15510, GM18951, GM19099, GM19193)
IRF4	hg18 and hg19	Lymphoblastoid (GM12878)
BATF	hg18 and hg19	Lymphoblastoid (GM12878)
FOSL2	hg18 and hg19	Hepatocellular carcinoma (HepG2)
PU.1	hg18 and hg19	Lymphoblastoid (GM12878 and GM12891) and myelogenous leukemia (K562)
USF-1	hg18 and hg19	Lymphoblastoid (GM12878), hepatocellular carcinoma (HepG2), and myelogenous leukemia (K562)
c-Jun	hg18 and hg19	Umbilical vein endothelial cells (HUVEC)
c-Fos	hg19	Umbilical vein endothelial cells (HUVEC)
JunD	hg18 and hg19	Hepatocellular carcinoma (HepG2),
TCF4	hg19	Colorectal carcinoma (HCT-116), and hepatocellular carcinoma (HepG2)
HDAC2	hg19	Hepatocellular carcinoma (HepG2)
YY1	hg19	Lymphoblastoid (GM12878)
Pol2-4H8	hg19	Colorectal carcinoma (HCT-116)
NRSF	hg18	Neuroectodermal (PFSK-1)
Sin3Ak-20	hg18	Hepatocellular carcinoma (HepG2)
p300	hg18 and hg19	Hepatocellular carcinoma (HepG2) and neuroblastoma cell line (SK-N-SH)
c-Myc	hg19	Human breast adenocarcinoma (MCF7) and acute promyelocytic leukemia (NB4)
FOXA1	hg19	Hepatocellular carcinoma (HepG2)
PAX5	hg19	Lymphoblastoid (GM12892)
HNF4A	hg18	Lymphoblastoid
TAF1	hg18 and hg19	Hepatocellular carcinoma (HepG2) and embryonic stem cells (H1-hESC)
HEY1	hg18 and hg19	Hepatocellular carcinoma (HepG2)
ELF1	hg19	Hepatocellular carcinoma (HepG2)
FOXP2	hg18	Neuroectodermal (PFSK-1)
HA-E2F1	hg19	Human breast adenocarcinoma (MCF7)
GABP	hg19	Hepatocellular carcinoma (HepG2)
Pol2	hg19	Fibroblast (ProgFib) and promyelocytic leukemia (NB4)

Table 3.4. Cell lines in which transcription factors have been ChIP-ed in the *DISC1* candidate promoter region. See next page for legend.

Table 3.4. Cell lines in which transcription factors have been ChIP-ed in the *DISC1* candidate promoter region. Transcription factors identified as potential regulators of DISC1 expression by ChIP are listed together with the assembly of the UCSC genome browser (hg18 or hg19) in which their interaction with the *DISC1* candidate promoter region (chr1:229828184-229828231 for hg18 and chr1: 231761561-231762608 for hg19) is indicated. Cell line(s) in which the interaction was identified are listed for each transcription factor.

3.4.5.1 In silico identification of putative FOXP2 binding sites in the DISC1 promoter

As the putative role of FOXP2 in the regulation of *DISC1* transcription was of particular interest, the *DISC1* upstream region was searched for potential FOXP2 binding sites. Consensus binding sites of increasing levels of specificity have previously been identified for members of the general FOX family of transcription factors, the FOXP family, and FOXP2. The FOX family bind to sites with the consequence sequence TRTTKRY (where R = A or G, K = G or T, and Y = C or T) (Overdier et al., 1994). FOXP binds to sites with the consensus sequence TATTTRT (Wang et al., 2003), and FOXP2 binds to sites with the consensus sequence AATTTG or ATTTGT (Stroud et al., 2006; Wu et al., 2006), as well as the core binding site, ATTT (Konopka et al., 2009). Experimental assessment of FOXP2 binding has revealed that it can bind to sites that deviate from the consensus binding site by one base pair (Vernes et al., 2007), or have the non-consensus sequence AAAGSAAA (S = G or C) (Vernes et al., 2011).

Searching the region upstream of *DISC1* that corresponds to the long promoter fragment (chr1: 231761579-231762608, hg19) for these sequences revealed 11 putative FOXP2 binding sites. As can be seen in table 3.5, of these sites, eight were unique to the long construct, two were located in sequence common to the long and the medium constructs, and one was present in all three constructs.

Sequence[a]	Class of binding site[b]					Position (nucleotide)[c]	No. occurrences in each construct[d]		
	FOX	FOXP	FOXP2	Core	NC		Long	Medium	Short
GCAAACA	+	–	–	–	–	-968	1	0	0
ATTT	–	–	–	+	–	-956, -733, -476, -394, -342, -266	5	1	0
TTTTGT	–	–	+	–	–	-780	1	0	0
CTTTGT	–	–	+	–	–	-403	1	0	0
AAAGCAAA	–	–	–	–	+	-251	0	1	0
AAAGGAAA	–	–	–	–	+	-113	0	0	1

Table 3.5. *In silico* screening of the promoter region of the *DISC1* gene (chr1: 231761579-231762608) for the presence of putative FOXP2 binding sites. Sites corresponding to the consequence binding site of the general FOX family (TRTTKRY), the FOXP family (TATTTRT), FOXP2 (AATTTG/ATTTGT), and the core FOXP2 binding site (ATTT) were identified. In addition, sites deviating from the FOXP2 consensus binding site by one base pair and non-consensus (NC) sites previously demonstrated to bind FOXP2 (Vernes et al., 2007, 2011) were searched for. [a] Deviations from the consensus binding site are indicated in red. Both of these sequences have previously been shown to bind FOXP2 (Vernes et al., 2007) [b] A plus sign (+) indicates the presence and a minus sign (−) indicates the absence of the indicated class of binding site. Where a putative binding site is compatible with more than one class of consensus binding site, the most specific class is indicated. [c] Position of the first nucleotide of the putative binding site relative to the transcription start site. [d] The number of times each putative binding site sequence is observed in each promoter construct is indicated. Where a sequence is present in a region of DNA common to more than one construct, only the shortest construct is indicated.

3.5 Summary and discussion

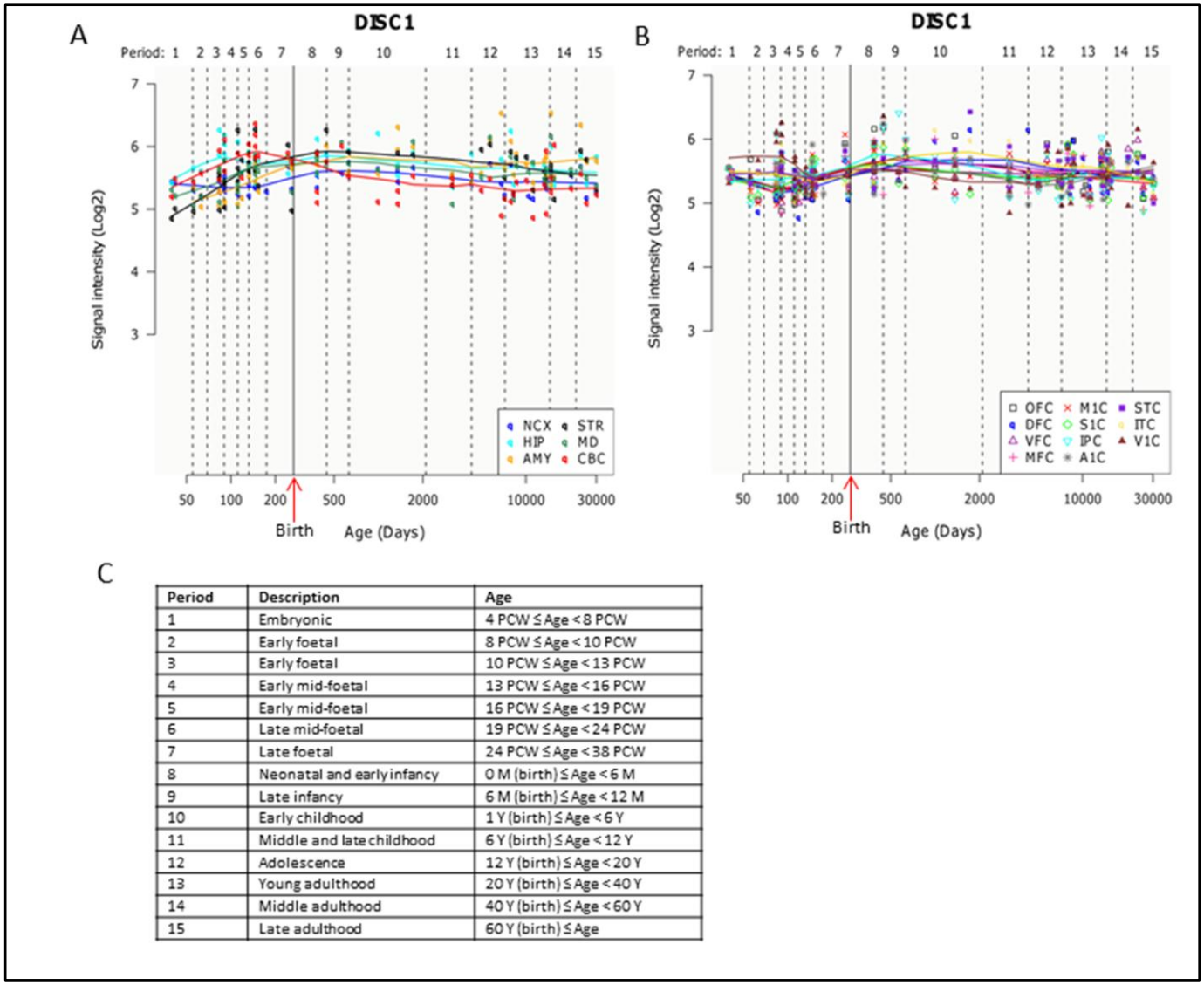
This chapter describes the characterisation of the promoter of *DISC1*, a leading candidate gene for schizophrenia and bipolar disorder, which has also been implicated in autism. The region was first analysed *in silico* to identify genomic features and epigenetic modifications commonly associated with promoter regions. These findings were used to inform the design of a series of constructs encoding nested fragments of the candidate *DISC1* promoter region for assessment in the dual luciferase assay.

The *DISC1* candidate promoter region was found to be devoid of canonical core promoter elements in the expected location relative to the previously identified transcription start site (TSS) (Nakata et al., 2009). The region was found to contain a CpG island, a feature of many constitutively expressed genes, which is in keeping with *DISC1*'s ubiquitous pattern of expression (Millar et al., 2000b). Assessment of ChIP-identified epigenetic modifications in the cell lines depicted by the ENCODE Regulation track revealed a pattern of enrichment consistent with actively transcribed genes. The presence of both the activating histone modification, H3K4me3, and the repressive modification, H3K27me3, in human embryonic stem cells (as indicated by the GIS ChIP-PET track) suggests that the *DISC1* promoter might fall into the bivalent class of promoters. The bivalent promoter is a hallmark of developmentally regulated stem cell genes (Johnson et al., 2011). Promoters carrying this modification are frequently transcriptionally silenced in embryonic stem cells but are ready to be activated upon differentiation (Bernstein et al., 2006). One putative mechanism for activation is the binding of RNA polymerase II in the bivalent state permitting swift transcriptional activation upon the loss of the H3K27me3 modification throughout development (Johnson et al., 2011).

The H3K27me3 modification is particularly prevalent at CpG island promoters and, in a genome-wide scan of chromatin state, 22% of promoters with a high CpG content were identified as carrying both the H3K4me3 and H3K27me3 histone modifications in mouse embryonic stem cells (Mikkelsen et al., 2007). Such promoters were associated with reduced transcriptional activity. Upon differentiation to neural progenitor cells, the majority of these bivalent promoters were found to resolve to H3K4me3 alone and show an increase in expression (Mikkelsen et al., 2007).

In light of these findings, it is somewhat surprising that DISC1 has previously been reported to be expressed at a modest level in human embryonic stem cells and to show reduced expression upon neural differentiation (Sun et al., 2011), leading to the conclusion that DISC1 might play an important role in stem cells.

Profiling of DISC1 mRNA expression in human post-mortem brain from a range of developmental stages has been carried out as part of the Human Brain Transcriptome project (http://hbatlas.org/; (Kang et al., 2011); figure 3.13). DISC1 mRNA expression was measured in six brain regions (neocortex, hippocampus, amygdala, striatum, mediodorsal nucleus of the thalamus and cerebellar cortex), with the neocortex being divided into eleven sub-regions for more detailed investigation. In the six brain regions, DISC1 mRNA expression was found to remain relatively constant across development, although some region-specific regulation was observed (figure 3.13A). While expression in the neocortex, striatum and mediodorsal nucleus peaks around during the neonatal/early infancy stage, cerebellar expression peaks during the late mid-foetal stage and hippocampal expression reaches its maximum during early foetal development and then remains fairly stable. In the 11 neocortical sub-regions, DISC1 expression shows little between-region variability and remains reasonably constant throughout development (figure 3.13B). Although these findings indicate relatively limited variation in DISC1 mRNA expression across development and between-brain regions, it is important to note that the assessment of expression in each region at each developmental time point would have encompassed multiple cell types, thus diluting the contribution of individual cell types, which may have shown greater temporal and/or spatial regulation. Studies in which individual cell types are isolated using techniques such as laser capture microdissection would facilitate a more accurate understanding of the spatial and temporal regulation of DISC1 expression.



C

Period	Description	Age
1	Embryonic	4 PCW ≤ Age < 8 PCW
2	Early foetal	8 PCW ≤ Age < 10 PCW
3	Early foetal	10 PCW ≤ Age < 13 PCW
4	Early mid-foetal	13 PCW ≤ Age < 16 PCW
5	Early mid-foetal	16 PCW ≤ Age < 19 PCW
6	Late mid-foetal	19 PCW ≤ Age < 24 PCW
7	Late foetal	24 PCW ≤ Age < 38 PCW
8	Neonatal and early infancy	0 M (birth) ≤ Age < 6 M
9	Late infancy	6 M (birth) ≤ Age < 12 M
10	Early childhood	1 Y (birth) ≤ Age < 6 Y
11	Middle and late childhood	6 Y (birth) ≤ Age < 12 Y
12	Adolescence	12 Y (birth) ≤ Age < 20 Y
13	Young adulthood	20 Y (birth) ≤ Age < 40 Y
14	Middle adulthood	40 Y (birth) ≤ Age < 60 Y
15	Late adulthood	60 Y (birth) ≤ Age

Figure 3.13. Developmental profiling of DISC1 mRNA expression in human post-mortem brain. Developmental profiling of DISC1 mRNA expression in human post-mortem brain samples was carried out by microarray as part of the Human Brain Transcriptome project (http://hbatlas.org/). A. DISC1 expression was measured in six brain regions (NCX = neocortex; HIP = hippocampus, AMY = amygdala, STR = striatum, MD = mediodorsal nucleus of the thalamus; CBC = cerebellar cortex) across 15 developmental periods. Birth is indicated by a red arrow. B. DISC1 expression was measured in eleven sub-regions of the neocortex (OFC = orbital prefrontal cortex; DFC = dorsolateral prefrontal cortex; VFC = ventrolateral prefrontal cortex; MFC = medial prefrontal cortex; M1C = primary motor cortex; S1C = primary somatosensory cortex; IPC = posterior inferior parietal cortex; A1C = primary auditory cortex; STC = superior temporal cortex; ITC = inferior temporal cortex; V1C = primary visual cortex) across 15 developmental periods. Birth is indicated by a red arrow. C. Table showing a description and range of ages for each of the 15 developmental periods assessed.

In the mouse brain, Sun et al. (2011) showed Disc1 mRNA to be expressed at a very low level in the cortices of mice at E18 and P7 and in the whole brain in adult mice. These observations are somewhat at odds with a previous study that reports the detection of Disc1 in the mouse cerebral cortex throughout development, with peaks in expression at E14, E16, and E18 (Austin et al., 2004). Furthermore, Disc1 has been found to show strong expression in the hippocampus at E18 (Meyer and Morris, 2008), and unpublished findings from our lab (Brown et al., manuscript in preparation) indicate that hippocampal expression of Disc1 peaks at E18. As such, further studies assessing both the expression of Disc1 in embryonic stem cells and later in development, and the contribution of epigenetic histone modification to the observed developmental expression pattern are necessary.

DISC1 transcription may also be regulated by disruption of the chromatin structure via the formation of Z-DNA. Computational assessment, using Zhunt, of the Z-DNA forming potential of the *DISC1* promoter region suggests that this region is more likely than many genomic regions to undergo a transition to Z-DNA under conditions of negative superhelicity, which could be invoked by the actions of chromatin remodelling enzymes or RNA polymerase II (Havas et al., 2000; Liu and Wang, 1987; Wittig et al., 1992). The Z-DNA forming region identified by Zhunt encompassed the complex dinucleotide repeat region. This region, with its composition of alternating purine and pyrimidine nucleotides matches the canonical Z-DNA forming sequence (Rich et al., 1983). In a previous study in which a region of DNA that acted as a transcriptional repressor was identified as having Z-DNA forming potential with a Z-score similar to that of the DISC1 candidate promoter region, mutation of the dinucleotides to non-Z-DNA forming dinucleotides reduced the repressor activity of the region in some cell lines (Ray et al., 2010). Clearly, however, multiple factors are likely to contribute to the transition of B-DNA to Z-DNA and empirical assessment would be necessary to determine whether Z-DNA formation does contribute to the transcriptional activity of the *DISC1* promoter.

In the dual luciferase assay, the highest level of transcription was obtained from the medium promoter construct, which spans the region from -300 bp to +47 bp relative to the TSS, while a reduced level of transcription was obtained from the long construct, which comprises the sequence of -982 bp to +47 bp relative to the TSS. This pattern of results was evident in all three cell lines assessed, suggesting that the observed differences are not specific to a particular cell type. However, whilst statistically significant differences were observed in SH-SY5Y cells and HEK293 cells, the differences did not attain significance in LAN-5 cells.

As the magnitude of the differences between the mean expression levels obtained from each construct are fairly similar in all three cell lines and the same number of experimental replicates was performed for each cell line, a likely explanation for the lack of statistical significance in LAN-5 cells is the greater level of variability observed. One potential explanation for this variability is that LAN-5 cells showed greater cell death following transfection than the other two cell lines. This may be because the protocol used for the Nucleofection of LAN-5 cells had not been optimised. Nucleofector Kit V was used with the A-023 electroporation program as this kit is recommended for several other neuronal cell lines (e.g. NG108-15, SH-SY5Y and SK-N-SH), and the program is recommended for the high viability transfection of SH-SY5Y cells, amongst others. A study published after I completed this work has used the T-020 program to electroporate LAN-5 cells (Petroni et al., 2011); however, as Lonza do not provide any details about the differences about the custom programs installed on the Nucleofector device, it is difficult to ascertain why T-020 might be more suitable than A-023. Nevertheless, although the results in LAN-5 cells were not statistically significant, the evidence does seem to support the conclusion that the medium promoter construct has greater transcriptional activity than either the short or the long constructs. The pattern of these findings is consistent with the ENCODE Pilot Project, which found that, in general, the region -300 bp to -50 bp relative to the TSS contributes positively to core promoter activity, whereas the region located-1000 bp to -500 bp relative to the TSS confers a repressive effect (Cooper et al., 2006).

Before considering the factors that might contribute to the differing promoter activities of the three promoter constructs, it is necessary to acknowledge the limitations of this work. An obvious limitation is that the promoter activity of the *DISC1* promoter fragments was assessed in isolation and the findings will, therefore, not necessarily reflect how the sequence contributes to promoter activity in the context of the genome. The effects of both structural properties influenced by the genomic context and interactions with regulatory elements outside of the region of interest would have been missed. Furthermore, as the promoter constructs were transfected transiently, the sequence of interest was assessed in a chromatin conformation different to that which occurs in the context of the genome; chromatin context is known to play an important role in transcriptional regulation (Smith and Hager, 1997). Results from the dual luciferase reporter assays performed here can, therefore, only be taken to indicate how a particular region might contribute to promoter activity. These findings can be used as a guide for the identification of regulatory elements and sequence variants that might be of interest; however, further experiments to assess *cis*-regions or *trans*-acting

factors of interest using approaches that take into account the genomic context of the promoter region should be performed.

In addition to being limited by their inherently artificial nature, dual luciferase reporter assays are also subject to the influence of many potentially confounding extraneous variables. In an assessment of the factors affecting reporter assays, Karimi et al. (2009) identified several factors that affected the relative reporter activities of two constructs that differed by one SNP. A SNP would only be expected to have a relatively small effect on promoter activity; therefore, extraneous variables would be likely to exert a larger effect on the outcome of the experiment than in situations, such as this one, where constructs of markedly different promoter activity are compared. Nevertheless, it is worth considering the extraneous variables that might have affected the findings presented in this chapter. The confounding variables identified by Karimi et al. (2009) included the transfection method used, the concentration of the DNA transfected, DNA conformation, the growth history of the cells, and cell cycle phase.

The transfection method used and the amount of DNA transfected were identical in all of the experiments performed in this chapter and, prior to use the long, medium, and short promoter constructs were assessed for DNA conformation and found to contain similar proportions of supercoiled, open circle, and linear DNA. It is possible, however, that between-experiment differences in the growth history of the cells and cell cycle phase may have contributed to between-experiment variability. Whilst in future experiments, it would be a good idea to attempt to control for these extraneous variables, the magnitude and the consistency of the differences between the promoter activities of the medium construct and the other two constructs suggests that extraneous variables did not exert a particularly large effect on the conclusions drawn from these experiments.

Whilst cognisant of the potential limitations of the experiments performed in this chapter, it is of interest to consider the mechanisms that might underlie the differences in the activity of the three promoter constructs. The medium promoter construct, which gave the highest level of transcriptional activity, contains the complex dinucleotide repeat region. Dinucleotide repeats have previously been identified as contributing to the transcriptional activity of some promoters (Borrmann et al., 2003; Wang et al., 2005); however; in the case of the *DISC1* promoter, the complex dinucleotide repeat region did not appear to underlie the enhanced transcriptional activity of the medium construct. The promoter activity of the medium 1

construct, which lacked the 124 bp 5' of the complex dinucleotide repeat region present in the medium construct, did not differ significantly from the short or the long constructs. It is, of course, the case that this analysis would not have detected any contribution of the complex dinucleotide repeat region that was dependent on the genomic context. The relatively small region of DNA considered in luciferase assays is a limitation of the technique. The involvement of distal regulatory elements in the control of gene transcription is well documented: regulatory elements located hundreds of kilobases away from the transcription start site have been reported previously (DiLeone et al., 2000; Spilianakis et al., 2005).

The results obtained in the present study can be compared with those of Hayesmoore et al. (2008). As mentioned in the introduction (section 3.1.1.2), Hayesmoore et al. (2008) concluded that the polymorphic TG repeat region, which forms part of the complex dinucleotide repeat region, does not influence DISC1 expression in the brain. Whilst this conclusion appears to fit with the present findings, limitations in Hayesmoore et al.'s (2008) approach render their findings compatible with several conclusions. Perhaps the most obvious limitation of Hayesmoore et al.'s (2008) approach was the lack of consideration of potential interaction effects between other sequence variants and the TG repeat polymorphism. The fact that some individuals heterozygous for polymorphisms affecting the number of TG repeats did not show allelic expression imbalance was interpreted to mean that the TG repeat polymorphism does not influence the expression of DISC in the brain. In fact, it is possible that the effect of the TG repeat polymorphism on DISC1 expression is modified by another sequence variant. Furthermore, the contribution of TG repeat number to the allelic expression imbalance shown by the individual affected by the duplication of the TG repeat region was dismissed somewhat prematurely. The affected individual carried one $(TG)_8$ and one $(TG)_{10}$ on the chromosome affected by the duplication and a $(TG)_9$ on the unaffected chromosome. Hayesmoore et al. (2008) concluded that, as $(TG)_9$ is similar in size to both $(TG)_8$ and $(TG)_{10}$, TG repeat number was unlikely to play a part in the allelic expression imbalance. This conclusion appears unwarranted: the duplication results in a TG repeat region with the sequence $(TG)_8TATGTC(TG)_{10}$, as the two sets of TG repeats are in such close proximity to each other it seems feasible that they might have an additive or interactive effect.

The assessment of the role of the complex dinucleotide region in *DISC1* promoter activity carried out in this chapter together with Hayesmoore et al.'s (2008) assessment of the TG

repeat region represent only a preliminary investigation of the function of this region. Before any firm conclusions can be drawn, further characterisation is required. Assessment of the association between polymorphisms in this region and DISC1 expression in a large sample, including the consideration of interaction effects with other genetic variants, would be informative. Further *in vitro* assessment could be carried out using reporter assays to assess the promoter activity of constructs carrying varying numbers of repeats in the complex dinucleotide repeat region in the context of upstream sequence. Using such assays, the length of the dinucleotide repeat region and the genetic background could be varied systematically permitting a more direct assessment of the effect of dinucleotide repeat region length on promoter activity than is possible in a naturalistic study, such as that carried out by Hayesmoore et al. (2008). Furthermore, the cloning of a longer section of the *DISC1* upstream region than studied in this chapter would improve the validity of the *in vitro* assessment.

ChIP data obtained as part of the ENCODE project was used to identify transcription factors that might be involved in the regulation of DISC1. Both the spatial resolution of ChIP-seq and the large number of transcription factors identified as binding the *DISC1* candidate promoter region, made it impossible to draw any conclusions about any individual transcription factors that might contribute to the relatively high level of promoter activity obtained from the medium fragment. The observation that the medium promoter fragment contains a region of high regulatory potential, according to the ESPERR Regulatory Potential (7 Species) track, not included in the short fragment or the medium 1 fragment suggests that conserved regulatory elements might contribute to the activating effect of the 124 bp unique to the medium fragment. As such, further investigation of the aligned sequences could yield clues to transcription factors that confer a positive effect on promoter activity.

In contrast to the medium, medium 1 and short fragments, there was a paucity of transcription factors predicted to bind to the genomic region unique to the long fragment. As diminished gene expression was observed from this fragment compared to the medium fragment, repressive elements may be located in the region located -982 bp to -301 bp relative to the TSS. As such, it was interesting to find that the transcription factor Max, which under some circumstances can exert a repressive effect (Solomon et al., 1993), had been identified by ChIP as binding in the within this region. Experimental assessment of (i) the binding of Max to this region, perhaps using electromobility shift assays, and (ii) the

effect of Max on the transcriptional activity of the long construct, using the dual luciferase reporter assay, would help to establish whether the repressive effect of Max binding contributes to the low activity of the long fragment.

Assessment of the ENCODE Regulation super-track for transcription factors identified by ChIP as binding in *DISC1* upstream region assessed in this chapter revealed several transcription factors with links to psychiatric illness. These include FOXP2, TCF4 and transcription factors with links to immune function (NF-κB, IRF4, and members of the AP-1 family).Variation in the *FOXP2* gene has been linked to speech and language function (Feuk et al., 2006; Lai et al., 2001; Lennon et al., 2007; MacDermot et al., 2005; Shriberg et al., 2006; Tomblin et al., 2009; Zeesman et al., 2006), and transcriptional targets of FOXP2 include genes, such as *CNTNAP2* and *MET*, implicated in autism (Mukamel et al., 2011; Vernes et al., 2008). TCF4 is a Wnt signalling responsive transcription factor. The Wnt signalling pathway is of great interest in the study of the pathogenesis of psychiatric illness due to its role in neurodevelopment and amenability to pharmacological targeting (Okerlund and Cheyette, 2011). As mentioned in the introduction (section 1.7.2.4.2), DISC1 has previously been implicated in Wnt signalling (De Rienzo et al., 2011; Mao et al., 2009); therefore, the regulation of *DISC1* expression by a Wnt signalling responsive transcription factor might represent a feedback mechanism. The finding that *DISC1* is potentially regulated by transcription factors implicated in immune function is interesting in the context of the growing evidence for abnormal immune function in psychiatric illness (see section 1.3) and the evidence, albeit not unequivocal, for the therapeutic effect of adjunctive treatments that target the immune system (Mansur et al., 2012). The role of these transcription factors in regulating *DISC1* warrants experimental assessment as the identification of regulatory links would assist in the synthesis of our understanding of the mechanisms involved in conferring risk for psychiatric illness.

Given the evidence for the involvement of FOXP2 in neurodevelopmental disorders, its putative role in regulating *DISC1* transcription was of particular interest. As such, the *DISC1* promoter fragments assessed in this chapter were assessed for potential FOXP2 binding sites. Eleven predicted FOXP2 binding sites were identified, the majority of which were unique to the long construct. As FOXP2 generally confers a repressive effect on transcription (Shu et al., 2001), the effects of FOXP2 may have contributed to the decreased promoter activity of the long construct relative to the medium construct; although, it should be noted that the ENCODE ChIP-seq assay did not detect FOXP2 binding in the sequence unique to the long

promoter fragment. As discussed previously (section 3.1.2.2), however, the computational prediction of TFBSs based on consensus sequences is plagued by false positives, and, therefore, the role of FOXP2 in regulating DISC1 must be assessed experimentally and the functionality of predicted binding sites determined using techniques such as electrophoretic mobility shift assays.

3.6 Conclusions

To conclude, the *DISC1* promoter region is devoid of canonical promoter motifs but does contain a CpG island, in common with many ubiquitously expressed genes. Histone modifications identified by ChIP by previous studies indicate that the *DISC1* promoter may exist in a bivalent state in embryonic stem cells, although this possibility requires experimental assessment. Dual luciferase reporter assays suggested the presence of positive regulatory elements in the region unique to the medium promoter fragment (-300 bp to -177 bp relative to the TSS) and repressive regulatory elements in the region unique to the long promoter fragment (-982 bp to -301 bp relative to the TSS). The presence of peaks in the ESPERR Regulatory Potential (7 Species) track in regions unique to both the medium and the long fragments suggest that certain features controlling the promoter activity of these fragments might show cross-species conservation. Several transcription factors identified by previous studies as potentially binding to the *DISC1* promoter region are themselves linked to psychiatric illness. This finding suggests that further investigation of the mechanisms regulating *DISC1* expression might help to integrate currently disparate strands of understanding regarding the pathogenesis of psychiatric illness. The assessment of the regulation of *DISC1* by one of these transcription factors, FOXP2, is the subject of the next chapter.

Chapter 4

Assessment of the DISC1-FOXP2 Regulatory Relationship

Chapter 4: Assessment of the FOXP2-DISC1 Regulatory Relationship

4.1 Introduction

Understanding how a gene is regulated, and, in turn, how it regulates other genes, permits the gene to be considered in the context of a regulatory network. Identifying the other members of this network can provide important clues to both normal gene function and the mechanisms that might be disturbed in a pathological state. Here, the role of one transcription factor, forkhead box P2 (FOXP2), in regulating DISC1 expression was assessed. In addition, a putative bi-directional regulatory relationship between DISC1 and FOXP2 expression was explored in a mouse model carrying an N-ethyl-N-nitrosourea (ENU)-induced point mutation, L100P, in the *Disc1* gene. Some of the work presented in this chapter has previously been published (Walker et al., 2012).

4.1.1 FOXP2

FOXP2 is a member of a large, *evolutionarily* conserved family of transcription factors, characterised by the presence of a forkhead box (FOX) DNA-binding domain (Benayoun et al., 2011). Members of the FOXP subfamily have, in addition to the FOX domain, a zinc finger domain and a leucine zipper motif, which permit regulation of their transcriptional activities by facilitating homo- and hetero-dimerisation (Takahashi et al., 2009). In addition to FOXP2, two other members of the FOXP subfamily, FOXP1 and FOXP4, are abundantly expressed in the brain (Takahashi et al., 2009).

4.1.1.1 The FOXP2 *gene*

The *FOXP2* gene displays a complex pattern of splicing resulting in multiple isoforms. To date, 26 *FOXP2* exons have been defined (Bruce and Margolis, 2002; MacDermot et al., 2005; Schroeder and Myers, 2008) (figure 4.1), which, according to the UCSC genome browser, are arranged to form 14 distinct transcripts. The FOXP2 isoform, isoform 1, investigated in this chapter is the major protein-coding transcript: it comprises 17 exons, utilises a start codon in exon 2, and encodes a protein of 715 amino acids (Lai et al., 2001). The majority of FOXP2 isoforms contain the same functional domains: exons 5-6 encode a polyglutamine tract, exons 8-10 encode the zinc finger domain and the leucine zipper motif, exons 12-14 encode the FOX domain, and exon 17 encodes an acidic C-terminal tail. A

notable exception is isoform FOXP2.10+, an alternatively spliced isoform that includes an elongated version of exon 10 followed by a polyadenylation site, which lacks the FOX domain and thus does not display transactivational activity (Vernes et al., 2006).

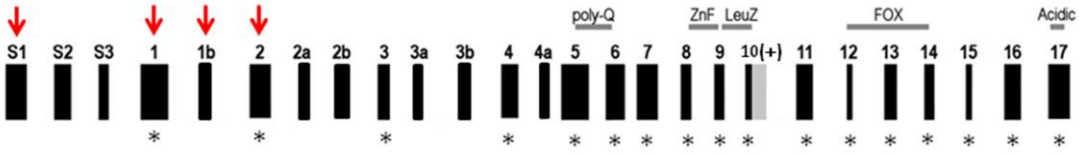


Figure 4.1. Exon structure of *FOXP* (not to scale). Exons are represented by horizontal bars with the exon number indicated above. Red arrows indicate transcription start sites detected by Schroeder and Myers (2008). Asterisks indicate exons included in FOXP2 isoform 1, the major protein coding transcript. Exon 10+ (black and grey bars) is an elongated version of exon 10 (black bar) formed by alternative splicing. Most FOXP2 isoforms include a poly-Q domain, a zinc finger domain (ZnF), a leucine zipper motif (LeuZ), a FOX domain, and an acidic C-terminal tail (Acidic). An exception is FOXP2.10+, an alternatively spliced isoform that includes an elongated version of exon 10 (10+) followed by a polyadenylation site. This transcript, therefore, lacks the FOX domain and the acidic C-terminal tail. Adapted from Schroeder and Myers (2008).

4.1.1.2 Mechanisms controlling the transcriptional regulation of FOXP2

In an initial attempt to understand the regulation of *FOXP2*, Schroeder and Myers (2008) mapped the gene's transcription start sites (TSSs). They identified four distinct TSSs located in exons S1, 1, 1b, and 2 (figure 4.1). The TSSs in exons S1 and 2 appeared to contribute to basal expression, whereas the TSSs in exons 1 and 1b appeared to be more cell line specific. Interestingly, the expression of the transcripts derived from the exon 1 and exon 1b TSSs was found to be highest in the brain, lung, and digestive tract, organs where the expression of FOXP2 is particularly important during development (Schroeder and Myers, 2008). Intriguingly, Schroeder and Myers (2008) failed to demonstrate transcriptional activity using putative promoter fragments derived from the sequence upstream of the exon 1 and exon 1b TSSs in transient transfection dual luciferase assays. Failure to detect activity is unlikely to be attributable to the choice of experimental cell line (PFSK-1), as endogenous expression of FOXP2 transcripts derived from these start sites was detected in PFSK-1 cells. This led to the suggestion that distal regulatory elements might be required to initiate transcription from these TSSs.

Analysis of the Encyclopaedia of DNA Elements (ENCODE) transcription factor binding site (TFBS) track on the UCSC genome browser reveals several binding sites in the FOXP2 promoter regions predicted by chromatin immunoprecipitation (ChIP) (Rosenbloom et al., 2010). The presence of two FOXP2-bound regions upstream of the exon S1 TSS and several FOXA1-bound regions upstream of both the exon S1 and exon 1 TSSs suggests that FOXP2 may regulate itself. A putative TCF4 (TCF7L2) binding site present upstream of exon S1 is of particular interest as it links the regulation of FOXP2 with the Wnt signaling pathway, dysfunction of which is believed to be a risk factor for psychiatric illness (Okerlund and Cheyette, 2011). Members of the Wnt-activated LEF/TCF family can bind to a common binding site (Hallikas et al., 2006), therefore the ChIP-identified TCF4 site is consistent with a previous study that found FoxP2 to be regulated by Lef1 during zebrafish central nervous system (CNS) development (Bonkowsky et al., 2008). In the context of the known role of DISC1 in regulating the Wnt signalling pathway (section 1.7.4.2.4), these observations raise the question of whether DISC1 function might affect the transcriptional regulation of FOXP2.

Whilst the bioinformatic prediction of TFBSs suffers from many limitations (section 3.1.2.2), it is interesting that a search for TFBSs in the 5' flanking region of exon S1 (-1500bp to +100bp relative to the TSS) using a TFBS prediction program identified a putative cAMP response element-binding (CREB) binding site (Bruce and Margolis, 2002). As mentioned in section 1.7.4.2.4, PDE4B, an enzyme responsible for the hydrolysis of cAMP is a DISC1-interating protein, thus this finding identifies another potential link between DISC1 function and FOXP2 regulation. This study also identifies EGR1 as a potential regulator of FOXP2. EGR1 has been found to be downregulated in the prefrontal cortex in patients with schizophrenia (Yamada et al., 2007), and the closely related transcription factor, Egr4, has been identified as being disregulated in a genetic mouse model of schizophrenia, which carries a mutation in the *Disc1* gene (section 4.1.2.1). Both EGR1 and EGR4 can bind to a common EGR consensus motif (Zipfel et al., 1997). Although preliminary, these observations suggest several potential links between the regulation of FOXP2 and genetic mechanisms with suspected involvement in the aetiology of schizophrenia and other psychiatric illnesses.

4.1.1.3 The pathogenic consequences of point mutations in the FOXP2 *gene*

The *FOXP2* gene was originally identified through mapping studies of a large multi-generational family (the KE family) with a monogenic speech and language disorder (Fisher et al., 1998). All affected members of the family have a heterozygous non-synonymous *FOXP2* mutation, yielding a subsitution (R553H) within the highly-conserved DNA-binding domain of the encoded protein (Lai et al., 2001). Affected individuals are profoundly impaired in the selection and sequencing of coordinated orofacial movements required for speech, resulting in the characterisation of their disorder as a developmental verbal dyspraxia (DVD) (Hurst et al., 1990; Vargha-Khadem et al., 1995; Vargha-Khadem et al., 1998; Watkins et al., 2002a). In addition, affected individuals display deficits in linguistic and grammatical processing, as well a decrease in IQ in both verbal and, to a lesser extent, non-verbal domains (Vargha-Khadem et al., 1995; Watkins et al., 2002a). Following extensive behavioural testing, Vargha-Khadem et al. (1995) concluded that the core deficit in affected family members is orofacial dyspraxia; however, what remains unclear is whether the additional deficits are secondary to the orofacial dyspraxia or whether they represent additional core deficits.

Another heterozygous *FOXP2* point mutation, R328X, which results in a prematurely truncated protein lacking the DNA-binding domain, was identified in a second family segregating DVD (MacDermot et al., 2005). Functional characterisation has revealed that both mutations can disturb nuclear localisation of FOXP2, as well as interfere with its capacity to act as a transcription factor (Vernes et al., 2006). Moreover, chromosomal rearrangements (including translocations and deletions) that disrupt *FOXP2* have been reported in other individuals and families with speech and language impairments (Feuk et al., 2006; Lennon et al., 2007; Shriberg et al., 2006; Tomblin et al., 2009; Zeesman et al., 2006).

4.1.1.4 Evidence for the involvement of FOXP2 in the transcriptional regulation of DISC1

DISC1 was first suggested as a potential direct target of FOXP2 by ChIP followed by microarray (ChIP-chip) in human foetal basal ganglia (Spiteri et al., 2007). Subsequently, a second ChIP screen carried out as part of the ENCODE project identified FOXP2 binding in the *DISC1* upstream region in the neuroectodermal cell line, PFSK-1 (Rosenbloom et al., 2010) (figure 4.2).

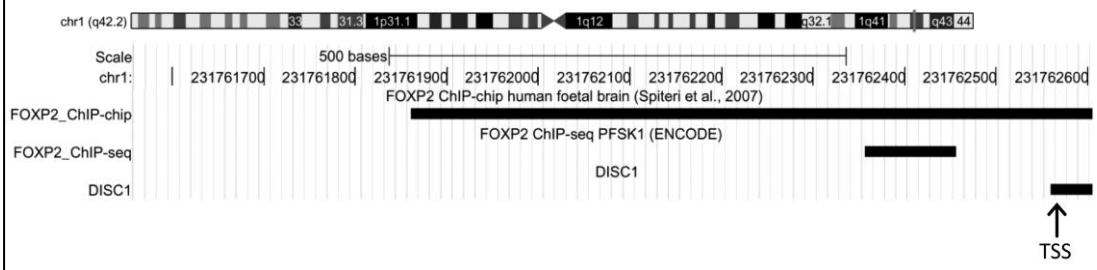


Figure 4.2. Schematic showing the genomic location of putative FOXP2 binding sites identified by chromatin immunoprecipitation (ChIP) in the *DISC1* upstream region. Potential FOXP2 binding sites were identified by either ChIP followed by sequencing (ChIP-seq) in human foetal brain (Spiteri et al., 2007), or by ChIP followed by microarray (ChIP-chip) in the PFSK-1 neuroectodermal cell line ("ENCODE Integrated Regulation" track on the UCSC human genome browser, Rosenbloom et al., 2010). The *DISC1* transcription start site (TSS; chr1: 231762561, UCSC genome browser, February 2009, GRCh build 37; http://genome.ucsc.edu/) is shown for reference.

The potential involvement of FOXP2 in the regulation of *DISC1* was of great interest as several lines of evidence suggest overlap in the phenotypic consequences of variation in the two genes. FOXP2 has been shown to regulate genes involved in susceptibility to autism spectrum disorder (ASD) (Mukamel et al., 2011; Vernes et al., 2008), a condition associated with genetic variation in *DISC1* (Crepel et al., 2010; Kilpinen et al., 2008; Williams et al., 2009; Zheng et al., 2011). In addition, there is tentative evidence to suggest that variation in *FOXP2* might be a risk factor for both schizophrenia and major depressive disorder: Li et al. (2012) identified association between a rare *FOXP2* variant and both conditions in a sample drawn from the Chinese Han population and Sanjuán et al. (2006) found a SNP and a haplotype in *FOXP2* to be associated with schizophrenia with auditory hallucinations. Association between gray matter concentration in patients with schizophrenia and a *FOXP2* variant has also been reported (Spaniel et al., 2011). One study has been published in which no association was identified between FOXP2 variants and schizophrenia (Sanjuan et al., 2005); however, only two variants were assessed.

Abnormalities in cognitive function represent an area of phenotypic overlap in conditions resulting from variation in *FOXP2* and/or *DISC1*. The reductions in cognitive function observed in individuals carrying the R553H FOXP2 mutation were described in the previous section and, as mentioned in the introduction, deficits in cognitive function are a core feature

of the schizophrenia phenotype that have been associated with variation in *DISC1* (section 1.1). Of particular interest, given the putative FOXP2-DISC1 regulatory relationship, is the association of *DISC1* variants with verbal working memory (Burdick et al., 2005). Characterisation of cognitive functioning in autism has resulted in less clear cut findings: some studies support the existence of an uneven pattern of cognitive functioning (Joseph et al., 2002; Kuschner et al., 2007); however, another has reported children with autism to have a wide range of cognitive abilities without the existence of a characteristic pattern of deficits (Siegel et al., 1996). It should, however, be noted that studies of cognitive function in autism have typically been limited to high-functioning individuals. Linguistic function is another area of apparent overlap in the conditions associated with variation in *FOXP2* and/or *DISC1*. Impaired language function is a core symptom of autism (Groen et al., 2008), while linguistic abnormalities are a well-established finding in schizophrenia (Barrett et al., 2009; Bull and Venables, 1974; Condray et al., 2002; John et al., 2011). Variation in DISC1 has been associated with verbal fluency (Palo et al., 2007) and the fMRI-assessed activity of multiple brain regions, including the pre/post-central gyrus, inferior frontal gyrus, and the cuneus during a language task (Chakirova et al., 2011; Whalley et al., 2012). The effects of variants on brain activation do, however, appear to differ between control subjects and those diagnosed with, or at high genetic risk of, bipolar disorder or schizophrenia. Despite the evidence for schizophrenia- and DISC1-related effects on language function, it is important to note that the linguistic abnormalities displayed by individuals carrying the R553H and R328X FOXP2 mutations, and those diagnosed with autism or schizophrenia show only partial overlap. For example, the contribution of orofacial praxia to linguistic dysfunction appears to be predominantly a feature of the conditions arising from the R553H and R328X FOXP2 mutations.

Consistent with the neurodevelopmental nature of the conditions associated with variation in FOXP2 and DISC1, both genes are known to play important roles in brain development: genome-wide ChIP-chip analysis has revealed FOXP2 to regulate genes involved in neurite outgrowth in the developing brain (Vernes et al., 2011) and the involvement of DISC1 in key neurodevelopmental processes, including neuronal migration, neurite outgrowth, axon myelination, and synaptic function has previously been mentioned in section 1.7.4.2.4. The available evidence suggests, therefore, that FOXP2 and DISC1 might contribute to partially overlapping neurodevelopmental pathways, the dysfunction of which contributes to the development of disorders that share some phenotypic similarities.

In this chapter, the ChIP-generated hypothesis that FOXP2 is involved in the regulation of DISC1 was tested, and the effect of the two pathogenic point mutations, R553H and R328X, assessed.

4.1.2 The *Disc1* L100P mutant mouse

A line of mutant mice carrying an ENU-induced point mutation in *Disc1* was developed by Clapcote et al. (2007). The mutation occurs in exon 2 of the gene and results in an amino acid substitution from leucine to proline at the 100th amino acid residue. First generation carriers of the mutation were backcrossed for six generations to generate heterozygotes with a predominantly C57BL/6J (98.4%) genetic background. At this point, heterozygotes were intercrossed to generate homozygous and heterozygous carriers of the point mutation.

Behavioural phenotyping of the *Disc1* L100P mice revealed a "schizophrenia-like" phenotype (Clapcote et al., 2007). The mice show deficits in prepulse inhibition (PPI) and latent inhibition (LI), which were reversed by treatment with the antipsychotics haloperidol and clozapine, and the PDE4 inhibitor rolipram. PPI describes the neurobehavioural phenomenon whereby the presentation of a low-intensity stimulus prior to a high-intensity stimulus reduces the startle reaction to the high-intensity stimulus (Hoffman and Searle, 1965). LI refers to the finding that pre-exposure to a non-reinforced stimulus decreases learning to the stimulus if it is subsequently rewarded (Lubow and Moore, 1959). Deficits in PPI and LI are found in individuals diagnosed with schizophrenia and are thought to reflect a deficit in the ability to filter out non-salient information, and thus in information processing (Braff and Geyer, 1990; Lubow and Gewirtz, 1995; Perry and Braff, 1994). The mice additionally demonstrated hyperactivity in the open field maze, which was reduced by clozapine, and impaired working memory, as assessed by the T maze. Deficits in working memory are frequently observed in individuals with schizophrenia (Barch, 2005).

Subsequent studies have provided mixed support for the behavioural deficits observed by Clapcote et al. (2007): the PPI deficit has been observed by Lipina et al. (2010), and Walsh et al. (2012) found the Disc1 L100P mice to be hyperactive in the open field. In contrast, Shoji et al. (2012) failed to replicate Clapcote et al.'s (2007) findings of impaired PPI and working memory but did observe the *Disc1* L100P mice to be hyperactive. Shoji et al. (2012) backcrossed the mice for a further two generations compared to Clapcote et al. (2007), so it is possible that their findings reflect the removal of additional ENU-induced mutations or

residual DBA/2J genetic material that remained in the genetic background of the mice assessed by Clapcote et al. (2007). Shoji et al. (2010) did, however, backcross their mice to C57BL/6JJcl mice rather than C57BL/6J mice, as used by Clapcote et al. (2007), so it is possible that their results have been influenced by between-substrain behavioural differences. It is also possible that differences in the experimental variables or the laboratory environment might have contributed to the differences between the two studies. Indeed, whilst Clapcote et al., (2007) assessed both male and female mice, Shoji et al. (2012) assessed only male mice. The gender of the mice cannot, however, fully explain the differences between these two studies, as Lipina et al. (2010) also only assessed male mice. Additional differences between the studies were the age at which the mice were assessed and their previous experience of behavioural tests: Clapcote et al. (2007) used experimentally naive mice aged between 12-16 weeks of age, while Shoji et al. (2012) used mice that were not experimentally naive and were aged between 16-19 weeks for assessment of PPI and 18-22 weeks for the assessment of working memory. Further studies are required to investigate the reasons behind Shoji et al.'s (2012) failure to replicate the behavioural phenotype of the *Disc1* L100P mice. Furthermore, in light of the neuroanatomical and neurochemical abnormalities identified in the L100P mice, which are described in the following paragraphs, it would be of interest to assess Shoji et al.'s (2012) mice at these levels.

Neuroanatomical investigation of the *Disc1* L100P mice has revealed a 13% reduction in adult brain volume (Clapcote et al., 2007), reduced neuron number, reduced neurogenesis, and altered neuronal distribution (Lee et al., 2011a). Neurons in the frontal cortex were found to have shorter dendrites and decreased surface area and spine density (Lee et al., 2011a). These abnormalities correspond to some of the neuroanatomical abnormalities present in patients with schizophrenia (Akbarian et al., 1995; Lawrie and Abukmeil, 1998; Ross et al., 2006; Wright et al., 2000).

A biochemical mechanism for at least some of the behavioural and neuroanatomical phenotypes displayed by the Disc1 L100P mice is suggested by the observation that the mutation reduces Disc1-Pde4b (Clapcote et al., 2007) and Disc1-Gsk3α and –Gsk3β binding (Lipina et al., 2011), which occur under normal circumstances in a region at the N-terminal of the gene. Binding by DISC1 serves to inhibit GSKβ activity (Mao et al., 2009); however, somewhat counterintuitively, no difference has been found in the kinase activities of either Gsk3α or β in the Disc1 L100P mice (Lipina et al., 2010). Similarly, the L100P mutation has been shown to have no effect on Pde4b activity (Clapcote et al., 2007). Nevertheless, genetic

and pharmacological inactivation of Gsk3 reverses the PPI and LI deficits and reduces the hyperactivity of the Disc1 L100P mice (Lipina et al., 2011). Furthermore, inactivation of Gsk3α rescues the spine density phenotype (Lee et al., 2011b). A synergistic interaction between Gsk3 and Pde4b has been highlighted by the finding that combined treatment of the Disc1 L100P mice with a GSK3 inhibitor, TDZD-8, and the PDE4 inhibitor rolipram, using each drug at a sub-threshold dose, corrects the PPI deficit and reduces hyperactivity (Lipina et al., 2012). In light of the paradoxical nature of these observations further work is required to understand how the altered interactions between Disc1 and Gsk3 and Pde4b affect the functioning of these enzymes and thus contribute to the behavioural and anatomical phenotypes of the mice.

Abnormalities in dopamine function have also been observed in the Disc1 L100P mouse: the mice are hypersensitive to the dopamine agonist amphetamine and show increased striatal expression of high-affinity D2 receptors (Lipina et al., 2010). Treatment with the dopamine antagonist haloperidol ameliorates the LI and PPI deficits, reduces hyperactivity and sensitivity to amphetamine (Lipina et al., 2010). Abnormalities in dopamine function may be a consequence of disrupted Pde4b function: the binding of Disc1 to Pde4b modulates cyclic adenosine monophosphate (cAMP) function, and thus dopamine synthesis (Nishi et al., 2008; Yamashita et al., 1997).

4.1.2.1 Altered gene expression in the Disc1 *L100P mouse*

During her PhD, Sarah Brown (SB) carried out microarray analysis to identify genes that are differentially expressed in the Disc1 L100P mouse. A subset of the genes identified as being differentially expressed by the microarray was followed up using qRT-PCR to measure gene expression throughout development. In addition to assessing the expression of these genes, SB characterised the developmental profile of *Disc1* in the Disc1 L100P mice and showed altered expression at E13 and P1 but not in the adult mouse, consistent with a previous report of Disc1 protein expression in the adult mouse (Clapcote et al., 2007). Developmental expression of *Disc1* in the Disc1 L100P mouse appears to be shifted such that peak expression occurs later in the mutant mice than their wildtype counterparts. A similar shift in peak expression was observed for *neurexins* (*NRXNs*) *1* and *3* (Brown et al., 2011), and *Pak3* (Brown et al., manuscript in preparation). Altered developmental expression was also confirmed for *Sort1*, *Cdh11*, and *Egr4*.

In this chapter the effect of the *Disc1* L100P mutation on the developmental expression of *Foxp2* was assessed using qRT-PCR. The rationale for carrying out this analysis is that, as described in section 4.1.1.2, the expression of FOXP2 may be regulated by several pathways with a known link to DISC1 function (section 1.7.4.2.4), including Wnt signalling, and cAMP- and EGR-dependent transcription. The expression of Foxp2 in the *Disc1* L100P mice has not been studied previously as expression of Foxp2 was not detectable above background on the microarray.

4.1.3 Introduction to the quantitative real-time polymerase chain reaction (qRT-PCR) technique

In this chapter, qRT-PCR is used twice: (i) to assess the effect of FOXP2 overexpression on DISC1 mRNA expression in HEK293 cells, and (ii) to characterise developmental Foxp2 expression in mice carrying the Disc1 L100P mutation and their wildtype counterparts. qRT-PCR is the most sensitive method for the accurate detection of low abundance mRNA (Bustin et al., 2000); however its validity depends on accurate normalisation of the expression values for the gene(s) of interest (GOI) to account for the amount of cDNA present in the reaction. The need for normalisation is borne from the fact that: (i) the efficiency of the reverse transcriptase reaction is variable, rendering the quantification of RNA insufficient as an estimate of cDNA quantity (Bustin and Nolan, 2004), and (ii) the quantification of cDNA by most methods (including nano-dropping) is not sufficiently accurate for direct normalisation to cDNA quantity.

The normalisation of GOI expression values to the expression values of internal control genes, "reference genes", measured in the same sample has become the most widely accepted approach to normalisation. The advantage of this approach is that the reference genes are exposed to the same conditions as the GOI at all stages of the experimental workflow, and therefore should normalise for variation introduced at any step. Unfortunately, no known gene meets the requirement for an "ideal reference gene"; that is, a gene is yet to be identified that shows invariant mRNA expression under all possible experimental conditions and at all developmental stages (Thellin et al., 1999; Vandesompele et al., 2002). The effect of variation in reference gene expression on study outcome has been explored by studies that have compared the results obtained when normalising gene-of-interest data to different reference genes. These studies have revealed reference gene choice to have a large effect on the outcome of a study, sometimes changing a finding from being

statistically significant to non-significant (Dheda et al., 2005; Ferguson et al., 2010). Therefore, it is important to identify reference genes whose expression is affected the least by the variables present in a certain experiment. One way of improving the chances accurately quantifying cDNA is to use multiple reference genes that participate in diverse cellular processes, and are, therefore, unlikely to be co-regulated (Vandesompele et al., 2002). This minimises the bias that any one non-ideal reference gene might introduce during normalisation.

The identification of suitable reference genes is plagued by the circular problem of having to assess the expression stability of a candidate reference gene when there is no reliable method to normalise the candidate. Several computer programs have been developed to address this problem, each implementing a slightly different algorithm. In this chapter, two different programs are used: geNorm and NormFinder. The different approaches used by the two programs are described below.

4.1.3.1 geNorm

The geNorm software developed by Vandesompele et al. (2002) works on the assumption that the expression ratio of two genuine reference genes should be constant across samples. For each candidate reference gene, the program calculates the pairwise variation value of the gene with each other candidate reference gene and assigns a gene stability value (*M* value), which represents the average pairwise variation of the gene with all other tested genes. Through a stepwise process of excluding the worst performing reference genes, the software ranks the candidate reference genes in order of stability. Identification of the minimum number of reference genes for accurate normalisation is permitted through the calculation of "V". V represents the variation in the sequential normalisation factors derived from the n and n+1 most stable genes. When $Vn/n+1 \leq 0.15$, there is no benefit to using an additional gene to calculate the normalisation factor. Normalisation factors are calculated by taking the geometric mean of the n most stable reference genes. The geometric mean is favoured over the arithmetic mean as it is more robust to outlying values and differences in abundance between the candidate reference genes (Vandesompele et al., 2002). It is recommended that geNorm is used with eight candidate reference genes measured in ten samples; however, the minimum requirements are that it is run on three candidate reference genes measured in two samples.

4.1.3.2 NormFinder

NormFinder employs a statistical model to assess the expression stability of candidate reference genes (Andersen et al., 2004). This model, which is based on a two-way ANOVA, assesses stability by comparing the expression of each candidate reference gene to a norm that is calculated by taking the mean of the expression values of all candidate reference genes for each sample. The program permits the user to indicate whether the samples belong to different groups (i.e. treated and control), and the most stable reference gene is then defined as the gene with the lowest inter- and intra-group variation. The most stable pair of reference genes is also calculated: this is the pair of reference genes that, when used in combination, result in the lowest combined standard deviation. An obvious assumption of this model is that the mean expression level of all the candidate genes does not differ between groups. Additionally, it is recommended that the model is used with a minimum of three candidate reference genes measured in a minimum of eight samples per a group (Andersen et al., 2004).

4.1.3.3 A comparison of the geNorm and Normfinder approaches

It has been shown that, in general, geNorm and NormFinder tend to agree on the most stable reference genes (Perez et al., 2008; Uddin et al., 2011); however, this is not always the case (Andersen et al., 2004). Furthermore, it has been demonstrated that the pairwise method used by geNorm has the tendency to rate co-regulated genes highly, independently of their expression stabilities; NormFinder is less susceptible to this confound (Andersen et al., 2004; Beekman et al., 2011). However, the NormFinder approach is limited by the assumption that the mean expression of the candidate reference genes does not vary between groups. This can be difficult both to predict and to assess, due to the circular problem of attempting to quantify reference genes. An advantage of NormFinder over geNorm is the ability to directly assess the inter-group variation of a gene. One advantage of geNorm over NormFinder is that it calculates the optimum number of reference genes to use. In contrast, NormFinder identifies the best pair of reference genes.

As there are different advantages and disadvantages associated with the use of both NormFinder and geNorm, both approaches were used in this chapter to identify the best reference genes. It was hoped that the use of two programs would (i) maximise the chance of identifying the most stable reference genes, and (ii) permit a comparison between the results

obtained using the reference genes identified by the two approaches, thus increasing the chance of identifying truly significant differences in gene expression.

4.2 Aims

The aims of this chapter were as follows:

- To assess the role of FOXP2 in regulating DISC1 expression
- To establish whether the pathogenic FOXP2 point mutations, R553H and R328X, alter any regulatory relationship observed
- To assess Foxp2 expression in the *Disc1* L100P mouse, as a preliminary step in assessing whether DISC1 might be involved in the regulation of FOXP2 expression

4.3 Materials and Methods

4.3.1 Expression constructs

4.3.1.1 FOXP2 expression constructs

The three FOXP2 constructs used in this study, pcDNA4/HisMax-FOXP2, pcDNA4/HisMax-FOXP2.R553H, and pcDNA4/HisMax-FOXP2.R328X were kind gifts from Dr. Simon Fisher at the University of Oxford.

4.3.1.2 DISC1 promoter expression constructs

The generation of these constructs has been described in section 3.3.3.

4.3.2 Cell culture and transfection

The neuroblastoma cell line, SH-SY5Y, and the human embryonic kidney cell line, HEK293, were maintained in Dulbecco's Minimal Essential Medium (DMEM; Invitrogen) supplemented with 10% foetal bovine serum (FBS) at 37°C in a humidified 5% CO_2 atmosphere. Cells were grown to ~80% confluency in T175 tissue culture flasks (Greiner) before seeding onto either 96-well plates (Sigma Aldrich) at 5 x 10^4 cells per well (dual luciferase reporter assays), or 6-well plates (Sigma Aldrich) at 5 x 10^5 cells per well (western blots and qRT-PCR), in order to reach ~80% confluency following overnight incubation. Transient transfections were carried out using either Fugene HD (Roche) or X-tremeGENE HP (Roche) according to the manufacturer's instructions. Transfections for the dual luciferase reporter assays and western blots were carried out using Fugene HD, while transfections performed for qRT-PCR were carried out using X-tremeGENE HP. The change from Fugene HD to X-tremeGENE HP was made because Fugene HD was discontinued by Roche. Both Fugene HD and X-tremeGENE HP are non-liposomal transfection reagents recommended for the transfection of high-density cells. Fugene HD was used at a ratio of 4:1 (transfection reagent to DNA), a ratio found by Elise Malavasi to result in optimal transfection efficiency of HEK293 cells with luciferase constructs (data not shown). As the ratio of X-tremeGENE HP to DNA had not previously been optimised it was used at two different ratios, 3:1 and 4:1, in all of the experiments presented in this chapter. These ratios were selected based on information from Roche who advised that X-tremeGENE HP

operated very similarly to Fugene HD (and could, therefore, probably be directly substituted) but that assessing a lower ratio of transfection reagent to DNA would be advisable in case of toxicity.

4.3.3 Western blotting

Western blotting was used to assess the effect of (i) the pcDNA3.1/HA (the control plasmid) and (ii) full-length FOXP2 on DISC1 protein expression.

4.3.3.1 Antibody selection

DISC1 protein expression was detected using a C-terminal antibody kindly supplied by Dr. Tetsu Akiyama from the University of Tokyo. The generation and validation of the antibody has been described previously (Ogawa et al., 2005). This antibody has been reported to detect the 100 kDa full-length isoform of DISC1 by immunoblotting (Murdoch et al., 2007; Ogawa et al., 2005).

FOXP2 protein expression was detected using a commercially available rabbit polyclonal antibody from Abcam (ab16046). This antibody binds to the C-terminal of FOXP2 and detects an 80 kDa product by immunoblotting (manufacturer's datasheet; http://www.abcam.com/FOXP2-antibody-ab16046.html) (Fatemi et al., 2008a; Hu et al., 2009). This antibody has been validated in PFSK-1 cells, a human neuroectodermal cell line, by ENCODE (http://www.epigenomebrowser.org/ENCODE/antibodies.html), who used it to perform the ChIP-seq experiment that identified *DISC1* as a putative target of FOXP2 (Rosenbloom et al., 2010).

The reference genes α-tubulin, β-actin, and GAPDH were detected using the following commercially available antibodies: α-tubulin (Abcam, ab18251); β-actin (Sigma), and GAPDH (Milipore, MAB374).

4.3.3.2 Transfection and protein extraction

For the assessment of the effect of the control plasmid on DISC1 expression level, cells were transfected with either 2μg of the control plasmid (pcDNA3.1/HA) or exposed to the

transfection reagent alone (mock transfection). Each condition was performed in three experimental replicates.

To assess the effect of FOXP2 on DISC1 expression level, cells were transfected with either 2µg of pcDNA4/HisMax-FOXP2, or 2µg of the control plasmid. Each condition was performed in five experimental replicates.

Twenty-four hours post-transfection each well was supplemented with 2000µl fresh DMEM with 10% FBS. Forty-eight hours post-transfection, cells were rinsed with PBS and lysed with 400µl PBS-1% Triton-X 100-10% glycerol protein extraction buffer, containing Complete Protease Inhibitor Cocktail (Roche). Protein lysates were gently mixed for 30 minutes at 4˚C, before being centrifuged at 16 060 x *g* for 20 minutes at 4˚C to remove cell debris.

4.3.3.3 Quantification of protein concentration

Protein concentrations were determined using the colorimetric Bio-Rad DC Protein Assay (Bio-Rad Laboratories). This assay is a modified version of the Lowry protein assay, which exploits the formation of copper-protein complexes under alkaline conditions. On the addition of Folin reagent an unstable product is formed, which is reduced, producing a blue colour.

A set of eleven standards of known protein concentration (0.0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, and 2.0 mg/ml) were prepared by diluting bovine serum albumin in dH_20. Five microlitres of each standard was pipetted onto a 96-well plate in duplicate. Experimental samples were diluted two-fold in dH_20 and 5µl of each sample was loaded onto the 96-well plate in triplicate. To each well was added a total volume of 25µl of Reagent A (alkaline copper tartrate solution) and Reagent S (surfactant solution), combined in a 50:1 ratio, and 200µl Reagent B (Folin reagent).

Following an incubation of fifteen minutes, the absorbance of each sample at 750nm was read using a Synergy HT plate reader (Bio Tek). A protein concentration-absorbance curve was plotted for the standards and the protein concentration of the experimental samples interpolated from this curve.

4.3.3.4 Protein samples from the L100P Disc1 mice

EG provided protein samples from the L100P *Disc1* mice and wildtype controls. Whole-brain protein was available from embryonic mice at embryonic days (E) 12.5, 15.5, and 17.5. Brain region-specific protein samples were available from postnatal mice at postnatal days (P) 1, 7, 21, and 90. Wildtype controls were littermates of the L100P mice at P21 and P90; for the other developmental stages non-littermate controls (cousins) were used.

4.3.3.5 Gel electrophoresis of protein samples

Immediately before electrophoresis, protein samples (10ug brain lysates, or 40ug cell lysates) were mixed with 5x loading buffer (final concentration 0.16M Tris-chloride (pH 6.8), 5% sodium dodecyl sulphate (SDS), 25% glycerol, 6% β-mercaptoethanol, 0.02% Bromophenol blue) and heated at 100˚C for 2 minutes. Protein samples and the Precision Plus Protein All Blue Standards molecular weight marker (Bio-Rad Laboratories) were fractionated on a 7% SDS-polyacrylamide-tris-acetate gel (Invitrogen) and electrophoretically transferred onto a polyvinyl difluoride (PVDF) membrane (Invitrogen). Gel electrophoresis and transfer efficiency were assessed by Ponceau S staining.

4.3.3.6 Immunolabelling of PVDF membrane and chemiluminescent detection of immunoreactive bands

The membrane was immersed in blocking buffer (phosphate buffered saline (PBS) containing 0.2% Tween (PBS-T) and 5% non-fat milk) for one hour, followed by incubation (overnight at 4˚C followed by four hours at room temperature) with the primary antibody (DISC1, 1:1000; FOXP2, 1:500; GAPDH, 1:100,000; Millipore) in blocking buffer. Membranes were then washed in PBS-T, and incubated with a horseradish peroxidase (HRP)-conjugated secondary antibody (For both DISC1 and FOXP2: swine-anti-rabbit, 1:2000, Dako; for GAPDH: goat-anti-mouse, 1:1000, Dako), in blocking buffer for one hour. Membranes were washed in PBS-T as before. Immunoreactive bands were visualised using the ‘ECL’ or ‘ECL Plus’ kit (GE Healthcare) on light-sensitive film.

4.3.3.7 Quantification and statistical analysis of immunoreactive bands

Immunoreactive bands were quantified by optical densitometry OD using GeneTools image analysis software (Syngene). Gene tools permits the user to manually quantify bands by drawing a rectangular box around each band of interest and selecting either automatic or manual background correction. Manual background correction was used for the experiments described in this chapter. A corrected optical density (OD) value is then returned for each band. For each sample lane on a given blot, the optical density value for the GOI immunoreactive band was normalised to the optical density value for a reference gene immunoreactive band.

Statistical analysis of normalised OD values was carried out using Statistical Package for the Social Sciences (SPSS) version 17.0. Between-genotype differences were assessed using the independent samples two-tailed Student's *t*-test and differences were deemed to be statistically significant when $p \leq 0.05$.

4.3.4 Reverse Transcription-Polymerase Chain Reaction (RT-PCR)

RT-PCR was carried out to assess the endogenous expression of FOXP2 by HEK293 and SH-SY5Y cells. One of the SH-SY5Y samples had previously been extracted and reverse transcribed into cDNA by SA. Niamh Ryan (NR) carried out the polymerase amplification and the gel electrophoresis of samples. I carried out all the other steps.

4.3.4.1 Primer design

RT-PCR primers were designed using the online program Primer 3 (http://frodo.wi.mit.edu/) to amplify a region within the first exon of FOXP2. According to the UCSC genome browser (hg19), this region is common to all eight FOXP2 RefSeq transcripts and is included in four non-RefSeq transcripts (identified by the "UCSC Genes" track of the UCSC genome browser). Two predicted short transcripts, which have not been validated by RefSeq, do not contain the amplified region. Primers were designed to comply with the following requirements: melting temperatures in the range of 50-63°C and a maximum difference in melting temperature of 10°C, a GC content of 40-60%, and length of around 20 nucleotides. The sequences of the designed primers are: forward: 5'-

AGGAATCTGCGACAGAGACAA-3'; and reverse: 5'-GTTGCAGATGCAGCAGTTCT-3'.

4.3.4.2 mRNA extraction

HEK293 and SH-SY5Y cells were grown to ~80% confluency in six-well tissue culture plates before being washed with PBS and dissociated by incubation in TrypLE^{TM} Express (500µl/well), a trypsin substitute, for five minutes at room temperature. A cell pellet was formed by centrifuging the lysates for five minutes at 300 x *g*. The supernatant was removed and the cell pellet washed in PBS.

Total RNA was then extracted using an RNeasy Mini Kit (Qiagen), according to the manufacturer's instructions. Briefly, cell pellets were lysed and homogenised in 600µl buffer RLT, which inactivates RNases. The lysates were then added to a Qiashredder column and centrifuged for 2 minutes at 17,000 x *g* and the supernatant transferred to a fresh microcentrifuge tube. Following the addition of 70% ethanol (600µl), 600µl of each lysate was applied to an RNeasy Mini spin column. RNeasy spin columns contain a silica membrane that binds RNA molecules longer than 200 nucleotides. After centrifuging for 15 seconds at 9,500 x *g*, the flow-through was discarded and the remaining 600µl of each lysate was added to the spin column and centrifuged for 15 seconds at 9,500 x *g*. The column was then washed to remove contaminants, such as carbohydrates, proteins, and fatty acids, by adding 350µl wash buffer, RW1, and centrifuging for 15 seconds at 9,500 x *g*. An on-column DNase digest was then performed: for each sample, 10µl DNase 1 was added to 70µl buffer RDD, the mixture was then pipetted directly onto the spin column silica membrane and incubated for 15 minutes at room temperature. The column was then washed by adding 350µl buffer RW1 and centrifuging for 15 seconds at 9,500 x *g*, before being transferred to a new collection tube. A second wash buffer, RPE (500µl), was then applied to the column, which was centrifuged for 15 seconds at 9,500 x *g*. The flow-through was discarded and another 500µl buffer RPE added to the column. This time the column was centrifuged for two minutes at 9,500 x *g* to dry the silica membrane. To eliminate carryover of buffer RPE, the column was transferred to a fresh collection tube and centrifuged for a further 1 minute at 9,500 x g. RNA was then eluted by placing the column in a fresh collection tube, adding 50µl RNase-free water, and centrifuging for one minute at 9,500 x g. To obtain the optimum RNA yield, a further 30µl RNase-free H_2O was applied to the column, which was centrifuged again for one minute at 9,500 x g.

The concentration and purity of total RNA samples was determined using the Thermo Scientific NanoDrop 1000 spectrophotometer (see section 4.2.4.3) and RNA was stored at -80°C until required for cDNA synthesis.

4.3.4.3 Spectrophotometric analysis of RNA using the NanoDrop

The concentration and purity of RNA samples was measured using a Thermo Scientific NanoDrop 1000 spectrophotometer (see section 3.3.3.6). Samples were measured by selecting the "nucleic acids" program on start-up and setting the measurement option to "RNA-40". The machine was then blanked by taking a reading from a reference sample (RNase-free H_2O). Each sample (1.5µl) was then pipetted onto the pedestal (which was cleaned between samples), and the absorption of the sample calculated as described in section 3.3.3.6.

Sample purity was determined by calculation of the ratio of absorbance at 260nm and 280nm (260/280) and 260nm and 230nm (260/230). Pure RNA should have a 260/280 ratio of ~2.0. Lower ratios can indicate the presence of contaminating agents absorbing at around 280nm, such as protein or phenol. The 260/230 ratio should be in the range of 1.8-2.2. Lower ratios can be caused by contaminating salts, phenol, or protein.

4.3.4.4 cDNA synthesis

cDNA was synthesised from DNase-treated RNA using the Transcriptor First Strand cDNA Synthesis kit (Roche). Two microlitres of random hexamer primer was added to 1µg RNA diluted to a total volume of 11.5µl with RNase-free H_2O. The RNA was denatured in a thermal cycler (Peltier Thermal Cycler 225, MJ Research) using the following program: 65°C, 10 minutes; 4°C, 5 minutes. The denatured RNA was then added to a reaction containing 5X Transcriptor Reverse Transcriptase reaction buffer (8µl), 10nM dNTP mix (4µl), RNase inhibitor (1µl), and Transcriptor Reverse Transcriptase enzyme (1µl). The reaction was incubated in a thermal cycler at 25°C for 10 minutes (primer annealing), followed by 50°C for 60 minutes (cDNA synthesis), and 85°C for 5 minutes (enzyme inactivation). Single-stranded DNA was stored at -20°C until required for PCR amplification.

4.3.4.5 PCR reaction and cycling conditions

One microlitre of cDNA was added to an PCR reaction comprising 2µl 10x PCR buffer (Qiagen), 4µl Q solution (Qiagen), 0.8µl dNTPs (5mM), 1µl combined forward and reverse primers (10µM), 0.1µl HotStarTaq DNA Polymerase (Qiagen), and 11.1µl RNase-free H_2O.

Polymerase amplification was carried out using a thermal cycler (Peltier Thermal Cycler 225, MJ Thermocycler) using the following parameters for touch-down PCR: 95°C for 5 minutes (initial Taq activation), followed by 10 touch-down cycles of 94°C for 30 seconds, 68°C for 30 seconds (-1°C each cycle), and 72°C for 1 minute, and then 30 cycles of 94°C for 30 seconds, 58°C for 30 seconds, and 72°C for 1 minute, followed by a final extension step of 72°C for 10 minutes.

4.3.4.6 Agarose gel electrophoresis

Agarose gels were prepared by adding 0.7-1.4g UltraPure™ agarose (Invitrogen) to 70ml tris-borate-EDTA (TBE) in a flask and heating until a clear solution was formed. The solution was allowed to cool for a few minutes before adding 7µl of the DNA gel stain SYBR Safe (Invitrogen). Once the SYBR safe had been mixed into the solution by swirling the flask, the solution was poured into a gel casting tray, gel combs inserted and the gel left to set. Once the gel had set, the combs were removed and the gel placed into an electrophoresis chamber containing TBE buffer.

Prior to loading, samples were mixed with 4 x Orange G loading buffer (1g Ficoll-400 (Sigma) and 0.2ml 50mM EDTA made up to a total volume of 10ml with ddH_2O. Orange G added to colour). A 1 kb or 1 kb Plus DNA ladder (Invitrogen) and the samples were then pipetted into the wells of the gel and electrophoresed at 100V until the samples were sufficiently fractionated. Samples were then visualised under UV light in a UVIdoc gel documentation system (UVITEC).

4.3.5 Quantitative Real-Time Polymerase Chain Reaction (qRT-PCR)

4.3.5.1 Selection of gene-of-interest (GOI) assays

Assays for human *DISC1* and mouse *Foxp2* were ordered from Applied Biosystems. Criteria for probe selection were: i) that the probe covered as many isoforms of the gene as possible, thus permitting a representative depiction of gene expression; and ii) that the probe was intron spanning, thus minimising the effect of gDNA contamination. Details of these assays are available in table 4.1.

Target	Species	Assay ID[a]	Amplicon length	Probe context sequence[b]	Sequence accession	Primer location (exon)	
						F	R
DISC1 and DISC1-TSNAX	Human	Hs00962133_m1	140	tatgataaagctg agacgttacaac	NM_001164546.1	3	4
					NM_001012959.1	3	4
					NM_001164538.1	3	4
					NM_001164539.1	3	4
					NM_001012957.1	3	4
					NM_001164540.1	3	4
					NM_001164541.1	3	4
					NM_001164542.1	3	4
					NM_001164544.1	3	4
					NM_001164545.1	3	4
					NM_001164547.1	3	4
					NM_001164548.1	3	4
					NM_001164549.1	3	4
					NM_001164551.1	3	4
					NM_018662.2	3	4
					NR_028393.1	7	8
					NR_028394.1	8	9
					NR_028395.1	8	9
					NR_028396.1	7	8
					NR_028397.1	7	8
					NR_028398.1	5	6
DISC1	Human	Hs00962131_m1	119	accgcgcaggcag ccgggattgctt	NM_001012957.1	1	2
					NM_001012958.1	1	2
					NM_001012959.1	1	2
					NM_001164537.1	1	2
					NM_001164538.1	1	2
					NM_001164539.1	1	2
					NM_001164540.1	1	2
					NM_001164541.1	1	2
					NM_001164542.1	1	2
					NM_001164544.1	1	2
					NM_001164545.1	1	2
					NM_001164546.1	1	2
					NM_001164547.1	1	2
					NM_001164548.1	1	2
					NM_001164549.1	1	2
					NM_001164550.1	1	2
					NM_001164551.1	1	2
					NM_001164552.1	1	2
					NM_001164553.1	1	2
					NM_001164554.1	1	2
					NM_001164555.1	1	2
					NM_018662.2	1	2
Foxp2	Mouse	Mm00475030_m1	85	ctgcctcaagctgg cttaagtcctg	NM_053242.4	8	9
					NM_212435.1	6	7

Table 4.1. Details of Applied Biosystems TaqMan gene expression assays used to measure genes of interest in this chapter. See next page for legend.

Table 4.1. Details of Applied Biosystems TaqMan gene expression assays used in this chapter. The table shows the target of the TaqMan assay, the species, the assay ID, the length of the PCR amplicon generated by the assay, the probe context sequence, the accession number of each transcript detected by each assay, and the location of the forward (F) and reverse (R) primers (denoted by exon number for each targeted transcript). [a] The m1suffix at the end of each assay ID indicates that the probe spans an exon junction and the assay will not detect any genomic DNA. [b] The probe context sequence is a 25 bp sequence that contains the probe sequence, which is usually 15-18 nucleotides long.

4.3.5.2 Selection of reference gene assays

4.3.5.2.1 Reference gene assays for the normalisation of DISC1 expression

Reference gene assays were ordered from PrimerDesign as part of their geNorm reference gene kits. Assays were ordered to target four reference genes: *glyceraldehyde 3-phosphate dehydrogenase* (*GAPDH*), s*uccinate dehydrogenase complex, subunit A* (*SDHA*), *splicing factor 3 subunit 1* (*SF3A1*), and *ubiquitin C* (*UBC*). All four assays are intron-spanning. Details of these assays are summarised in table 4.2.

Target	Species	Sequence accession	Primer location (exon)		Probe location (exon)
			Forward	Reverse	
GAPDH	Human	NM_002046	8	9	8
SDHA	Human	NM_004168	7	8	7
SF3A1	Human	NM_001005409.1	11	12	11
		NM_005877.4	11	12	11
UBC	Human	NM_021009	1	2	1-2

Table 4.2. Details of the PerfectProbe (PrimerDesign) reference gene assays used for the normalisation of DISC1 expression. Shown are the name of the probe target, the species the probe is designed to work in, the accession number of the target, the location of the forward and reverse primers and the probe (denoted by exon number for each targeted transcript).

4.3.5.2.2 Reference gene assays for the normalisation of Foxp2 expression

An assay for the reference gene *hypoxanthine phosphoribosyltransferase 1* (*Hprt1*) was ordered from Applied Biosystems (Mm01318743_m1). To permit comparison with previous gene expression profiling experiments carried on the Disc1 L100P RNA samples by Sarah Brown (SB) during her PhD, the same *Hprt1* TaqMan assay (Applied Biosystems) was used.

This assay spans an intron (between exons 3 and 4) and has been designated an "m1" suffix by Applied Biosystems, indicating that it has been shown to not detect a product in gDNA.

4.3.5.2.3 Use of geNorm and NormFinder reference gene selection algorithms

geNorm was used within the Biogazelle qbase^{PLUS} qRT-PCR software. When using geNorm, the optimum number of reference genes suggested by geNorm (as described in section 4.1.3.1) was used to normalise the GOI.

NormFinder was used within the GenEx qRT-PCR analysis software. On the basis that the use of more than one reference gene is likely to increase the likelihood of the reference genes giving an accurate depiction of cDNA quantity, the most stable pair of reference genes, as chosen by NormFinder, was used for normalisation of the GOI (section 4.1.3.2).

4.3.5.3 Sample preparation

4.3.5.3.1 Disc1 *L100P mice*

cDNA was synthesised from the brains of mice carrying the L100P *Disc1* mutation, and their wildtype littermates by SB during her PhD and by Helen Torrance (HT). SB and HT synthesised cDNA from mice at embryonic days (E) 13, 15, and 18, postnatal days (P) 1, 7, and 20, and adult mice (12-13 weeks of age).

4.3.5.3.2 HEK293 cells

Cells were transfected with either 2µg of i) pcDNA4/HisMax-FOXP2, or ii) the control plasmid pcDNA3.1/HA, or exposed to the transfection reagent alone (mock transfection), or left untransfected. Transfections with pcDNA4/HisMax-FOXP2 and the control plasmid were performed in four biological replicates. The mock and untransfected conditions were performed in three biological replicates. Twenty-four hours post-transfection each well was supplemented with 2000µl fresh DMEM with 10%FBS. Forty-eight hours post-transfection, the cells were washed with PBS and dissociated by incubation in TrypLE^{TM} Express (500µl/well) for five minutes at room temperature. A cell pellet was formed by centrifuging the lysates for five minutes at 300 x *g*. The supernatant was removed and the cell pellet washed in PBS.

Extraction of mRNA was carried as described in section 4.2.4.2, quantified as described in section 4.2.4.3, and cDNA synthesised as described in section 4.2.4.4.

4.3.5.4 Assessment of cDNA samples for gDNA contamination

cDNA samples were assessed for contamination by carrying out RT-PCR using primers that span an intron-exon boundary (table 4.3).

Tissue/cell type	Target gene	Primer sequences	Expected product size (bp)	
			cDNA	gDNA
HEK293 cells	*WDR1*	F: cttgctggggttgtttctgt R: ccacgacggtgggattta	238	590
Mouse brain	*Sorsc2*	F: cctactgcaaagagcccaat R: ggcgctgaactctcatagaa	156	625

Table 4.3. Details of the primers used in the RT-PCR carried out to detect gDNA contamination in cDNAs synthesised from mRNA extracted from either mouse brain or HEK293 cells. The sequences of the forward (F) and reverse (R) primers are given, together with the expected product size, in base pairs (bp), if the primers were to amplify their target from either cDNA or gDNA.

One microlitre of cDNA was added to an RT-PCR reaction comprising 2µl 10x PCR buffer (Applied Biosystems), 0.6µl dNTPs (5mM), 0.5µl combined forward and reverse primers (20µM), 0.3µl Taq DNA Polymerase (made in house), and 14.4µl RNase-free H_2O.

The reaction was carried out using a thermal cycler (Peltier Thermal Cycler 225, MJ Research) using cycling conditions for touch-down PCR: 93°C for 1 minute (initial denaturation), followed by 10 touch-down cycles of 93°C for 20 seconds, 65°C for 30 seconds (-1°C each cycle), and 72°C for 1 minute, and then 30 cycles of 93°C for 20 seconds, 55°C for 30 seconds, and 72°C for 1 minute, followed by a final extension step of 72°C for 10 minutes. PCR amplimers were fractionated by agarose gel electrophoresis, as described in section 4.3.5.4, and gDNA contamination assessed.

4.3.5.5 Quantitative Real-Time Polymerase Chain Reaction (qRT-PCR) analysis of mRNA expression

qRT-PCR was carried out using an Applied Biosystems 7900HT Fast Real-Time PCR System at the Wellcome Trust Clinical Research Facility (WTCRF), at the Western General Hospital, Edinburgh. A hydrolysis probe-based chemistry was chosen over the use of an intercalating dye, as hydrolysis probes permit greater target recognition specificity. Hydrolysis probes are fluorescently labeled DNA oligonucleotides that are designed to bind downstream of one of the primers. The 5' end of the probe is labeled with a fluorescent reporter molecule and at the 3' end of the probe there is a quencher molecule. When the probe is intact, the quencher suppresses fluorescence from the 5' end of the probe; however, during the PCR reaction, the probe is degraded by the polymerase enzyme, the quencher is physically separated from the fluorescent reporter molecule, and fluorescence is emitted. It is only possible for a probe to fluoresce if it has bound to the correct target; in contrast, the intercalating dye approach relies on the primers amplifying the correct target, as the dye will intercalate with any double-stranded DNA molecule that is formed during the reaction. Two different types of probe-based assay were used in this chapter: TaqMan assays (Applied Biosystems) and PerfectProbes assays (Primer Design). The main difference between these assays is the chemistry of probes: TaqMan probes are linear, whereas PerfectProbes create a stem-loop secondary structure (figure 4.3). The stem-loop structure brings the fluorescent reporter molecule and the quencher into greater spatial proximity, reducing background fluorescence and thus the number of qRT-PCR cycles required for target detection.

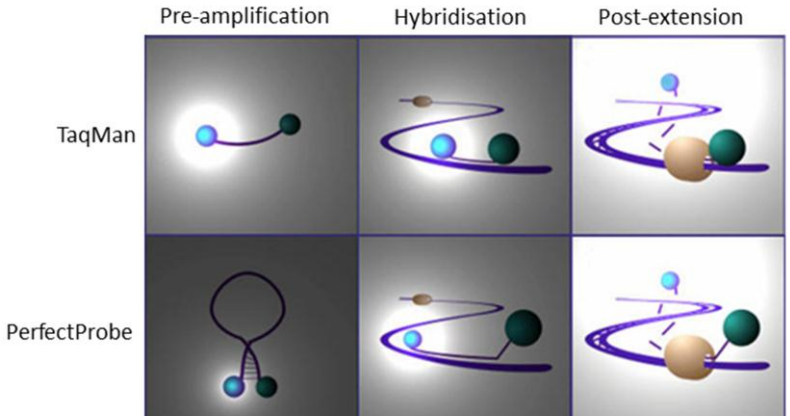


Figure 4.3. Schematic depicting the chemistries of TaqMan probes and Perfect Probes. Before amplification, TaqMan probes are linear while PerfectProbes form a stem-loop structure. The stem-loop structure brings the fluorescent reporter molecule (blue) into greater spatial proximity with the quencher molecule (green), resulting in less background fluorescence than linear TaqMan probes. Both probes types anneal to the DNA during the hybridisation step and are then cleaved by the 5'-3' exonuclease activity of the polymerase (beige), thus separating the fluorescent reporter molecule from the quencher and permitting fluorescence. Figure adapted from http://www.primerdesign.co.uk.

Samples were assayed for a GOI (either *DISC1* or *Foxp2*), together with either six (DISC1) or three (Foxp2) reference genes. Each sample was measured in three technical replicates. Reactions were carried out in 384-well plates, with samples arranged according to the sample maximisation approach recommended by Hellemans et al. (2007). This approach ensures that if the number of wells required for all sample and gene combinations exceeds one plate, as many samples as possible are analysed on the same plate and different genes are analysed on different plates, thus minimising the impact of between-sample technical variation. In addition, each plate contained a standard curve and a calibrator for each gene assayed. The standard curve comprised a set of four 1/10 serial dilutions (1/10, 1/100, 1/1000 and 1/10,000) of whole brain cDNA (human or mouse, as appropriate). The standard curve was used to quantify cDNA using the relative standard curve method (described in section *4.2.6.5*). Whole brain cDNA (1/200 dilution) was also used as the calibrator. Inclusion of a calibrator permitted correction for between-plate differences and thus the comparison and combination of results across plates.

qRT-PCR was carried out using TaqMan Fast Universal PCR Master Mix (2x), no AmpErase UNG. Reactions were carried out in a total volume of 10µl, comprising 5µl mastermix, 0.5µl of the appropriate probe and 4.5µl cDNA. Plates were sealed with a plastic sheet using a heat sealer and centrifuged for 30 seconds before being placed in the Applied Biosystems 7900HT Fast Real-Time PCR machine. Thermal cycling conditions differed for the TaqMan assays and the PerfectProbe assays, due to their different chemistries. Conditions for each probe type are described below:

TaqMan assays: 95°C for 10 minutes (enzyme activation), followed by 40 cycles of 95°C for 15 seconds (denaturation) and 60°C for 1 minute (annealing/extension).

PerfectProbes assays: 95°C for 10 minutes (enzyme activation), followed by 50 cycles of 95°C for 15 seconds (denaturation) and 50°C for 30 seconds (annealing), and 72°C for 15 seconds (extension).

4.3.5.6 Quality control and statistical analysis of qRT-PCR data

qRT-PCR produces a cycle threshold (Ct) value for each sample. The Ct value indicates how many PCR cycles were required before the quantity of a given sample reached the detection threshold.

Data were initially analysed using the Sequence Detection Systems (SDS) software version 2.3 (Applied Biosystems) for the 7900HT Fast Real-Time PCR System. First, the Ct values for each triplicate of technical replicates were checked for outliers: if any sample had a Ct value that differed by more than one from either of the other two samples in the triplicate, it was discarded. Then, the standard curves were checked to assess the reaction efficiency and the accuracy of each assay. If the PCR reaction works with 100% efficiency (i.e. the amount of product has doubled with every cycle) the slope of the standard curve will be -3.32. If the slope deviated from this value, Ct values were inspected. Sometimes it was clear that the assay did not work efficiently for the most dilute samples in the standard curve. In this case, these samples were excluded. The correlation of the curve reflects the accuracy of the reaction, with 100% accuracy denoted by $R^2 = 1$. Correlations of $R^2 \geq 0.98$ were considered acceptable.

Ct values for experimental samples were translated into quantity values using the relative standard curve method. This method requires a set of relative standards formed by serial dilution of a cDNA, RNA, or DNA sample known to express the appropriate target, and a calibrator sample, also expressing the appropriate target, to be run for each probe used on each plate. Preparation of the relative standards and calibrator sample used in this thesis has been described in section *4.2.6.4.I*. Quantity values for experimental samples were determined by interpolating from the standard curve. At this point, the data were exported to Microsoft Excel for further analyses.

In Excel, the following steps of analyses were implemented, as described in Hellemans et al. (2007).

1. Calculation of a normalisation factor (NF).

For each sample, *k*, a NF was calculated based on the geometric mean of the quantity values (QVs) of the reference genes, *p*:

$$NF_k = \sqrt[f]{\prod_{p=1}^{f} QV_{pk}}$$

For *Foxp2*, two or three reference genes were used, and for *DISC1*, the required number of reference genes was calculated using the geNorm algorithm described by Vandesompele et al. (2002).

2. Calculation of the normalised experimental sample QVs (NQVs).

NQVs were calculated by for each gene, *j*, measured in sample, *k*:

$$NQV_{jk} = \frac{QV_{jk}}{NF_k}$$

3. Calculation of the calibrator normalisation factor (CF).

For each gene, *j*, in each run, *l*, a CF was calculated:

$$CF_{jl} = \sqrt[c]{\prod_{m=1}^{c} NQV_{jlm}}$$

4. Calculation of calibrated NVQs (CNVQs).

For gene, *j*, measured in sample, *k*, a CNVQ was calculated:

$$CNQV_{jk} = \frac{NQV_{jkl}}{CF_{jl}}$$

Following CNVQ calculation, outlying samples within each condition were identified using the inter-quartile range method, whereby the CNVQ for a sample *x* must abide by the following restrictions, where Q_1 is the first quartile and Q_3 is the third quartile:

$$(Q_1 - 1.5 \times Q_1) \leq CNVQ_x \geq (Q_3 + 1.5 \times Q_3\ )$$

Samples with a CNVQ value violating these conditions were deemed to be outliers and were excluded from further analyses.

The mean, standard deviation, and standard error of the mean (SEM) were then calculated for each condition using the statistical functions of Microsoft Excel. Comparisons between genotypes were carried out using SPSS. Comparisons were performed separately for each developmental stage using two-tailed independent samples Student's *t*-tests, where data met the assumption of normality and homogeneity of variances. When analysing data with equal sample sizes, the *t*-test has been shown to be relatively robust to deviations from these assumptions (Boneau, 1960). However, as sample sizes in this analysis varied, compliance with the assumptions of normality and homogeneity of variances was formally assessed. Normality was assessed using the Shapiro-Wilk test and homogeneity of variances was assessed using Levene's test. If data were found to violate the assumption of normality, homogeneity of variances was assessed using a non-parametric Levene's test (Nordstokke and Zumbo, 2010). Deviations from these assumptions were deemed to have occurred when $p \leq 0.05$. None of the samples were found to violate the assumption of normality; however, deviations from the assumption of homogeneity of variance were corrected by carrying out a Welch's *t*-test in which the degrees of freedom are reduced according to the Welch-Satterthwaite method. Between-group differences were deemed to be statistically significant when $p \leq 0.05$.

4.3.6 Dual Luciferase Reporter (DLR) assay

4.3.6.1 Cell transfection and lysis

For the assessment of the effect of FOXP2 and the mutated forms of FOXP2, FOXP2.R553H and FOXP2.R328X, on *DISC1* promoter activity, HEK293 cells were co-transfected with three constructs: (i) 163.7µg of the appropriate reporter construct (promoterless pGL4.10 or pGL4.10-*DISC1* promoter construct (long, medium, or short)), (ii) 3.3µg of the transfection efficiency control pRL-TK, and (iii) 166.7µg of a FOXP2 expression construct (pcDNA4/HisMax-FOXP2/-FOXP2.R553H/-FOXP2.R328X), or the control plasmid, pcDNA3.1/HA (Invitrogen). Transfections were performed in triplicate and repeated in three experimental replicates. Twenty-four hours post-transfection, each well was supplemented with 100µl fresh DMEM with 10%FBS. Forty-eight hours post-transfection, cells were rinsed with phosphate buffered saline (PBS) and lysed with 20µl Passive Lysis Buffer (PLB; Promega).

4.3.6.2 Measurement of luciferase activity

The measurement of luciferase activity using the Dual-Luciferase Reporter Assay System (Promega) has been described previously in section 3.3.4.2.

4.3.6.3 Calculation of relative luciferase activity

The calculation of relative luciferase values was performed using a similar method to that described in section 3.3.4.3. The main differences were that (i) for the data presented in this chapter, only three technical replicates were performed for each of the three experimental replicates and (ii) the pGL4.13 (SV40/*luc2*) was not used as a positive control, as the SV40 promoter contains function FOXP2 binding sites and, therefore, the expression of this construct is regulated by FOXP2 (Vernes et al., 2006). Moreover, little between-plate variation was observed in the experiments described in chapter 3 and, therefore, an inter-plate calibrator was not considered necessary.

As described in section 3.3.4.3, for each technical replicate, firefly luciferase values were divided by *Renilla* luciferase values to obtain relative luciferase values. Background luciferase activity (the mean relative luciferase value obtained from the three technical

replicates transfected with the promoterless vector pGL4.10), was subtracted from each relative luciferase value.

Statistical analysis of the relative luciferase values was performed using SPSS version 17.0. Dual luciferase reporter assay data was assessed using either one-way or two-way analysis of variance (ANOVA), followed by Tukey's HSD post-hoc test. Significant interaction terms were investigated by splitting the data file by each independent variable and performing either two-tailed independent samples Student's *t*-tests or one-way ANOVAs followed by Tukey's HSD post-hoc test.

4.4 Results

4.4.1 Assessment of the effect of FOXP2 on DISC1 expression

4.4.1.1 Cell line selection and antibody characterisation

In order to assess the effects of the transient overexpression of FOXP2 on DISC1 protein expression and promoter activity in an appropriate cellular context, a cell line that endogenously expresses FOXP2 was required. Endogenous expression of FOXP2 has previously been demonstrated in HEK293 and HEK293T cells (Abcam datasheet for ab16046) (Vernes et al., 2006), but was not detected in SH-SY5Y cells (Vernes et al., 2006). The expression of endogenous FOXP2 in HEK293 cells was confirmed by western blotting (figure 4.4.A). Overexpressed and endogenous FOXP2 were detected by western blotting as a band of 80 kDa, consistent with the manufacturer's data sheet. SH-SY5Y cell lysates were probed for FOXP2 expression as a negative control; however, a band at 80 kDa was detected. This band was detected in two independent SH-SY5Y cell lysates (data not shown). In order to assess further whether this band represented FOXP2, FOXP2 mRNA expression was measured by RT-PCR. Expression was detected as a product of 152bp in SH-SY5Y cDNA (figure 4.4.B). HEK293 cDNA was assayed for FOXP2 expression at the same time as a positive control. It was, therefore, concluded that SH-SY5Y cells do endogenously express FOXP2.

Due to time constraints, the assessment of the effect of FOXP2 activity on *DISC1* promoter activity could only be carried out in one cell line. Thus, despite the evidence for the expression of FOXP2 in SH-SY5Y cells, it was decided that HEK293 cells would be used for future experiments, as this would permit comparison with previous studies in which the transcriptional effects of FOXP2 and the R553H and R328X mutant forms of FOXP2 have been investigated in HEK293 cells (Roll et al., 2010; Vernes et al., 2006).

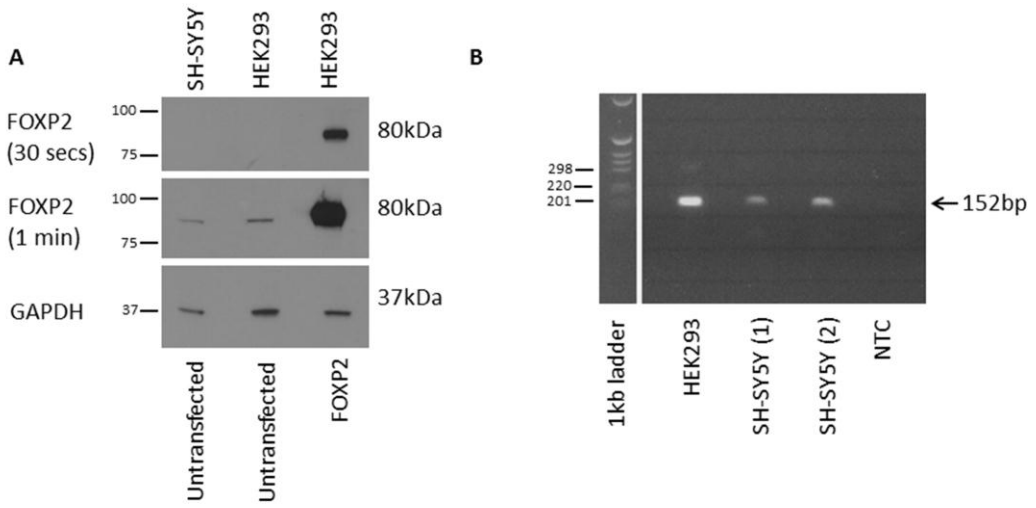


Figure 4.4. Detection of endogenous and over-expressed FOXP2. **A**. FOXP2 is detected as an 80kDa immunoreactive band by western blotting. Endogenous expression was detected in both SH-SY5Y and HEK293 cells, after a one minute exposure. Detection of over-expressed FOXP2 as an 80kDa band (30 second and 1 minute exposures) in HEK293 cells transfected with 2µg pcDNA4/HisMax-FOXP2 supported the identification of the 80kDa band as FOXP2. Blots were probed for GAPDH as an indicator of protein loading. The location and molecular weight (in kDa) of relevant bands from the Precision Plus Protein All Blue Standards molecular weight marker are indicated to the left of the image and expected product sizes are indicated to the right of the image. **B**. Gel electrophoresis of RT-PCR amplimers showing a 152bp product indicative of endogenous expression of FOXP2 in HEK293 and SH-SY5Y cells. Assessment of FOXP2 expression in SH-SY5Y cells was carried out in two independent replicates, due to the prior evidence suggesting that this cell line does not endogenously express FOXP2 (Vernes et al., 2006). No product was observed in the no template control (NTC) negative control reaction. The sizes of relevant bands from the 1 kb ladder (in bp) are indicated to the left of the image and the expected product sizes are indicated to the right of the image.

The DISC1 antibody used in this chapter has been previously shown to detect endogenous full-length 100 kDa DISC1 expression in SH-SY5Y cells (Murdoch et al., 2007; Ogawa et al., 2005), but there are no reports of its use in HEK293 cells. Therefore, its ability to detect DISC1 expression in HEK293 cells, which have previously been demonstrated to express DISC1 (James et al., 2004), was assessed by western blotting. Consistent with findings in SH-SY5Y cells, endogenous DISC1 was detected as a single 100 kDa immunoreactive band (figure 4.5.A). The expression of DISC1 by HEK293 cells was further confirmed by qRT-PCR (figure 4.4.B). qRT-PCR was carried out using the DISC1 TaqMan probe Hs009962131_m1 described in table 4.1. This probe detects 22 of the 23 RefSeq transcripts for DISC1, including the full-length transcript.

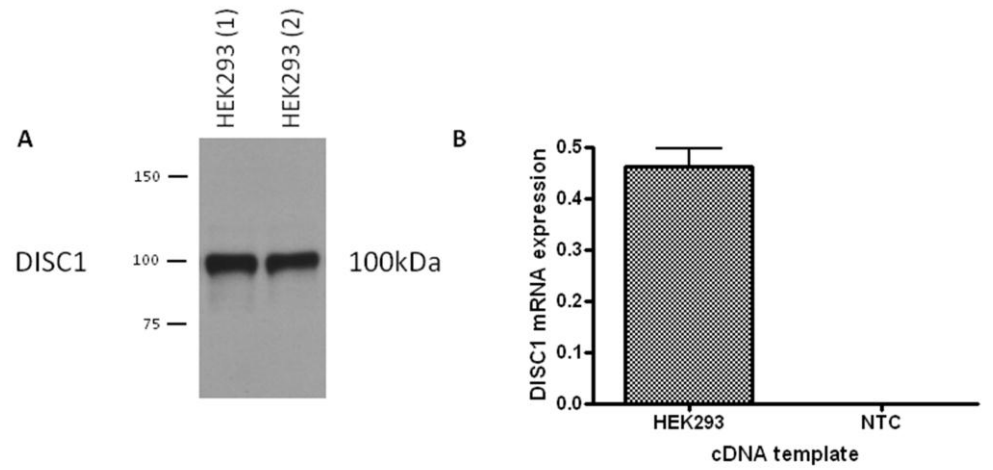


Figure 4.5. Detection of endogenous DISC1 expression in HEK293 cells. **A**. The ability of the C-terminal DISC1 antibody (Ogawa et al., 2005) to detect DISC1 endogenous expression in HEK293 cells was assessed by western blotting. DISC1 was detected as a 100 kDa immunoreactive band, consistent with previous reports in SH-SY5Y cells (Ogawa et al., 2005; Murdoch et al., 2007). The location and molecular weight (in kDa) of relevant bands from the Precision Plus Protein All Blue Standards molecular weight marker are indicated to the left of the image and the expected product size is indicated to the right of the image. **B**. The presence of endogenous DISC1 in HEK293 cells was confirmed by qRT-PCR. DISC1 mRNA expression was detected in HEK293 cDNA but not in a no template control (NTC) negative control reaction.

4.4.1.2 Selection of a control vector for the transfection of FOXP2

To assess the effect of FOXP2 on DISC1 RNA and protein expression and promoter activity, the empty vector, pcDNA3.1/HA (Invitrogen) was used as control for the effect of transfecting cells with foreign DNA. This vector was used in place of pcDNA4/HisMax, the vector into which the FOXP2 constructs were cloned, which would have been the ideal control vector, as this vector was not available in the laboratory at the time of carrying out these experiments. pcDNA3.1/HA was selected as an alternative as it is highly similar in size to pcDNA4/HisMax (pcDNA3.1/HA: 5.4kb; pcDNA4/HisMax: 5.3kb), and therefore provides a control for the molar quantity of foreign DNA transfected into the cell. pcDNA3.1/HA will henceforth be referred to as the "control plasmid".

4.4.1.3 Selection of a ratio of transfection reagent to DNA for transfections using X-tremeGENE HP

As described in section 4.3.2, a novel transfection reagent (X-tremeGENE HP) was used to transfect HEK293 cells for the qRT-PCR assays carried out in this chapter. As this transfection reagent operates in a similar manner to Fugene HD, the transfection reagent used for all other transfections presented in this chapter, it was used at the same ratio of transfection reagent to DNA as Fugene HD (4:1). In case this ratio resulted in excess toxicity, all transfections were additionally performed using a lower ratio of transfection reagent to DNA (3:1). Although cell death was not formally quantified, visual inspection did not reveal any obvious differences in toxicity between the two conditions and, therefore, gene expression was assayed in both conditions.

4.4.1.4 Assessment of gDNA contamination in HEK293 cDNA samples

Prior to using the HEK293 cDNA samples in the experiments presented in this chapter, the samples were assessed by Helen Torrance (HT) for contamination with gDNA. Each sample was PCR-amplified using intron-spanning primers targeting a region of the *WDR1* gene and the resulting amplimers fractionated on an agarose gel (figure 4.6). Control reactions were performed for both the reverse transcription reaction and the PCR. Samples were reverse transcribed in two batches, such that half the experimental replicates (replicates 1 and 2) were reverse transcribed in batch 1 and the other half (replicates 3 and 4) were reverse transcribed in batch 2. None of the HEK293 cDNA samples were found to be contaminated with gDNA.

For each batch of samples, two negative control reactions were carried out, one without any RNA (no RNA) and the other without the reverse transcriptase enzyme (-RT). Both the no RNA reactions were free from contamination; however, a band of 238 bp, indicating cDNA, was evident in the –RT reaction for Batch 1. Positive controls were carried out to assess whether the primers amplified products of the expected size in cDNA from the brain, where *WDR1* is known to be expressed, and gDNA. These reactions produced products of the expected sizes. A negative control PCR reaction (no template control) did not produce a product, indicative of a lack of DNA contamination in the PCR reagents.

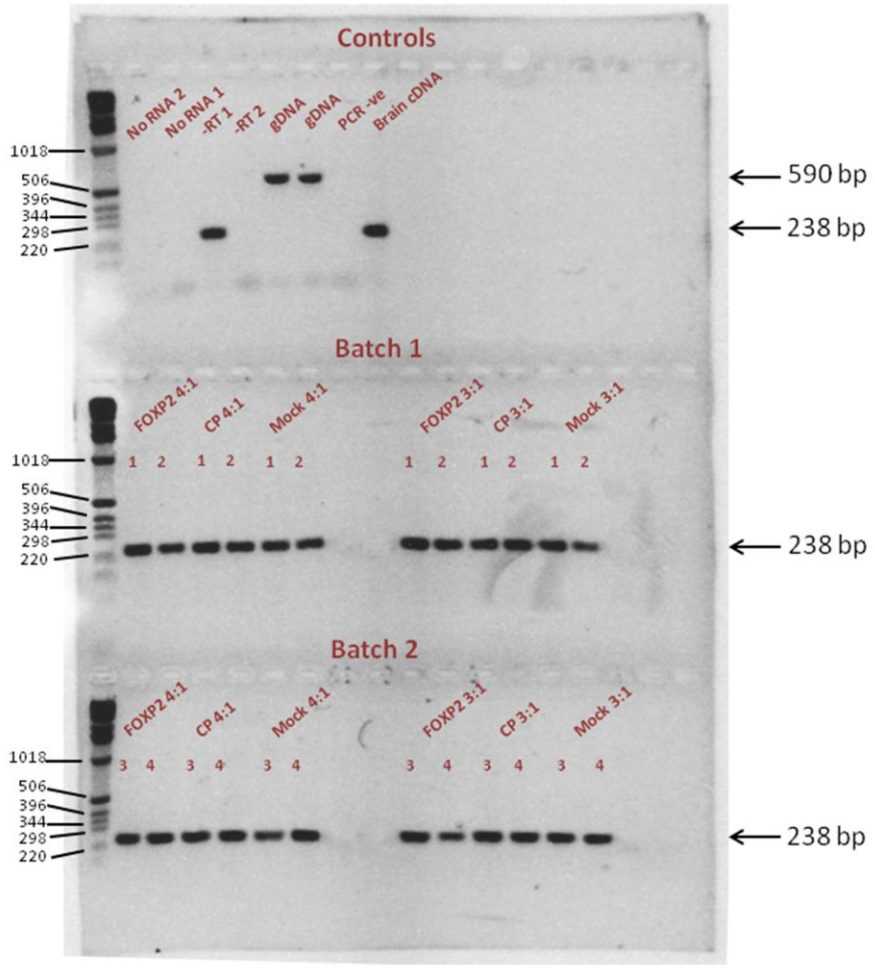


Figure 4.6. Assessment of gDNA contamination in HEK293 cDNA. HEK293 cDNAs were assessed for contamination with gDNA by performing PCR using a set of primers that amplify a region of *WDR1* to yield a product of 590bp from a gDNA template and 238bp from a cDNA template. cDNAs were assessed from cells that had been transfected with FOXP2, the pcDNA.31/HA control plasmid (CP), or mock transfected (Mock). Transfections were carried out using two different ratios of transfection reagent to DNA, 4:1 and 3:1. For each condition, cDNAs were made from cells transfected in four experimental replicates (1-4). The reverse transcription reactions were carried out in two batches: experimental replicates 1 and 2 were reverse transcribed in batch 1 and experimental replicates 3 and 4 were reverse transcribed in batch 2. For each batch, two negative control reactions for the reverse transcription reaction were carried out: no RNA and no reverse transcriptase (-RT; 1 = Batch 1, 2 = Batch 2). Positive control PCR reactions were carried out using (i) gDNA as the template (PCR performed in duplicate) and (ii) cDNA from the brain, where *WDR1* is known to be expressed, as the template. A PCR negative control (PCR -ve) was carried out in which no template was included. Samples were run with a 1 kb DNA ladder. The sizes of relevant bands from the 1 kb ladder (in bp) are indicated to the left of the image and the expected sizes of the products detected in cDNA and gDNA are indicated to the right of the image.

Regarding the 238 bp band observed in the –RT reaction for batch 1, the absence of contamination in the no RNA reactions indicates that the reverse transcription reagents were free from DNA or RNA contamination. The band in this sample could, therefore, indicate either (i) contamination of the RNA sample with cDNA contamination or (ii) contamination of the reverse transcription reagents with reverse transcriptase. The latter explanation seems unlikely as the same reagents were subsequently used for the second batch of reverse transcription reactions, for which the –RT reaction was free from DNA contamination. Another possibility is that the reverse transcriptase enzyme was accidently added to the –RT reaction. To differentiate between these possibilities, the sample used for the batch 1 –RT reaction was reverse transcribed again and another –RT reaction performed. As a control, the sample used for the –RT reaction for batch 2, which did not show evidence of contamination, was reverse transcribed as well, and another –RT reaction performed for this sample. PCR amplification of the products from these reactions using the same *WDR1* primers as used previously revealed the two reverse transcription reactions to contain a single band of 238 bp, indicative of cDNA free from gDNA contamination. A low level of cDNA contamination was evident in the –RT reactions for both batches 1 and 2; however, a similar level of cDNA contamination was present in the PCR negative control reaction (no template control). It, therefore, seems likely that the 238 bp bands indicative of cDNA in the –RT reactions are due to cDNA contamination of the PCR reagents. Importantly, the intensity of the band present in the –RT reaction for batch 1 is much lower than that observed initially (figure 4.7), suggesting, perhaps, that the reverse transcriptase enzyme was accidently added to the initial reaction. As such, these cDNA samples were deemed to be unlikely to be unduly affected by DNA contamination and were, therefore, used for qRT-PCR analysis.

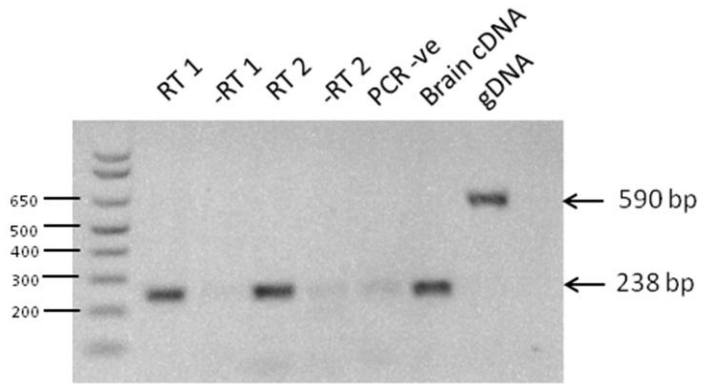


Figure 4.7. Assessment of the reverse transcriptase negative controls for DNA contamination. To clarify the source of a band indicative of cDNA contamination previously observed in the reverse transcriptase negative (-RT) control carried out for batch 1 of the reverse transcription reactions (-RT 1; figure 4.5), the reverse transcription (RT) and –RT reactions were repeated for this sample. As a control, RT and –RT reactions were repeated for the sample used in the –RT reaction for batch 2 of the reverse transcription reactions (-RT2), which previously showed no evidence of DNA contamination. Samples were PCR amplified using a set of primers that amplify a region of *WDR1* to yield a product of 590bp from a gDNA template and 238bp from a cDNA template. A PCR negative control (PCR –ve), which contained no cDNA template was performed together with two positive control PCR reactions in which (i) gDNA was used as the template and (ii) cDNA from the brain, where *WDR1* is known to be expressed, was used as the template. Samples were run with a 1 kb Plus DNA ladder. The size (in bp) of relevant ladder bands are shown to the left of the image and the expected sizes of the products detected in cDNA and gDNA are indicated to the right of the image.

4.4.1.5 Identification of suitable reference genes for qRT-PCR assays

Stably expressed reference genes were required for the normalisation of qRT-PCR assays carried out to compare DISC1 expression in HEK293 cells under the following conditions: (i) transfection with the control plasmid vs. mock transfection, and (ii) transfection with full-length FOXP2 (pcDNA4/HisMax-FOXP2) vs. transfection with the control plasmid. For both the above comparisons four reference genes were selected to be assessed for the stability of their expression. These genes were chosen as they fulfilled the criteria of (i) not having any ChIP-identified potential FOXP2 binding sites in their upstream region, according to the "ENCODE Integrated Regulation" track on the UCSC human genome browser (hg18 and hg19), and (ii) being targeted by an intron-spanning assay available from PrimerDesign in their hydrolysis probe (PerfectProbe) "geNorm kits". The selected genes

were: *glyceraldehyde 3-phosphate dehydrogenase* (*GAPDH*), s*uccinate dehydrogenase complex, subunit A* (*SDHA*), *splicing factor 3 subunit 1* (*SF3A1*), and *ubiquitin C* (*UBC*).

As discussed previously (section 4.1.3.1-4.1.3.3), there are several algorithms available for the assessment of the stability of candidate reference genes. Here, two algorithms, geNorm, and NormFinder were implemented. As described in sections 4.1.3.1 and 4.1.3.2, geNorm identifies the n most stable reference genes required for the variation in sequential normalisation factors to fall below a pre-determined threshold, while NormFinder identifies the most stably expressed pair of reference genes. The results of these two algorithms are summarised in table 4.4.

Ratio	Comparison	geNorm	NormFinder
3:1	CP vs. Mock	SDHA and UBC	SDHA and UBC
	FOXP2 vs. CP	SDHA and UBC	GAPDH and UBC
4:1	CP vs. Mock	SDHA and UBC	SDHA and UBC
	FOXP2 vs. CP	SDHA and GAPDH	SDHA and SF3A1

Table 4.4. Reference genes selected for the normalisation of DISC1 qRT-PCR data. The expression values of the gene of interest, DISC1, and four reference genes, GAPDH, SDHA, SF3A1, and UBC, were determined by qRT-PCR in HEK293 cells transfected with the control plasmid (CP), full-length FOXP2 or mock transfected using two different ratios of transfection reagent to DNA (3:1 and 4:1). Two programs, geNorm and NormFinder, were used to select the most stably expressed reference genes to permit the comparison of DISC1expression values under the following conditions: (i) CP vs. Mock and (ii) FOXP2 vs. CP.

4.4.1.6 Assessment of the effect of pcDNA3.1/HA on DISC1 expression

Before using the control plasmid, it was important to establish whether it had any effect on DISC1 expression. The effect of the control plasmid on DISC1 expression was assessed in HEK293 cells by western blotting and qRT-PCR using the DISC1 TaqMan probe Hs00962131_m1 (table 4.1). For the assessment at the RNA level, HT extracted the RNA, synthesised the cDNAs and performed the qRT-PCR; I performed the transfections and the data analysis.

4.4.1.6.1 Assessment of the effect of pcDNA3.1/HA on DISC1 protein expression

To assess the effect of pcDNA3.1/HA, the control plasmid, on DISC1 protein expression, HEK293 cells were transfected with either 2µg of the control plasmid or mock untransfected. Comparison of DISC1 protein expression in the control plasmid-transfected cells with expression in the mock transfected cells was carried out using a two-tailed independent samples *t*-test. This revealed no significant difference between the two conditions when assessed across three experimental replicates ($p = 0.272$; figure 4.8).

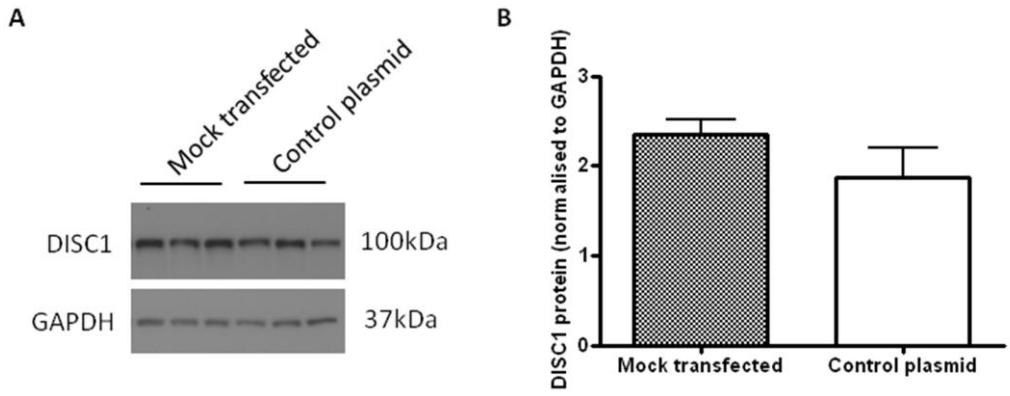


Figure 4.8. Assessment of the effect of the control plasmid, pcDNA3.1/HA on DISC1 protein expression. HEK293 cells were transiently transfected with either the control plasmid or subjected to mock transfection. **A**. DISC1 expression, together with the expression of the reference gene, GAPDH, was assayed 48 hours post-transfection by western blot. **B**. Immunoreactive bands were quantified by optical densitometry (OD) and the OD values for DISC1 normalised to the OD values for GAPDH. Statistical significance was assessed using a two-tailed independent samples Student's *t*-test; no statistically significant differences were detected. The experiment was performed in experimental triplicate.

4.4.1.6.2 Assessment of the effect of pcDNA3.1/HA on DISC1 mRNA expression

DISC1 expression was assessed by qRT-PCR in HEK293 cells transfected with either 2µg of the control plasmid, pcDNA3.1/HA, or mock transfected using two ratios of transfection reagent to DNA (3:1 and 4:1). For both ratios, the two reference gene selection programs, geNorm and NormFinder, identified the same two reference genes for calculation of the normalisation factor, *SDHA* and *UBC* (table 4.4). DISC1 expression was assayed using two TaqMan qRT-PCR assays, Hs00962131_m1 and Hs00962133_m1 (table 4.1). No significant

differences in normalised DISC1 expression values were identified between cells transfected with the control plasmid and mock transfected cells (all $p \geq 0.319$, two-tailed independent samples *t*-tests; figure 4.9).

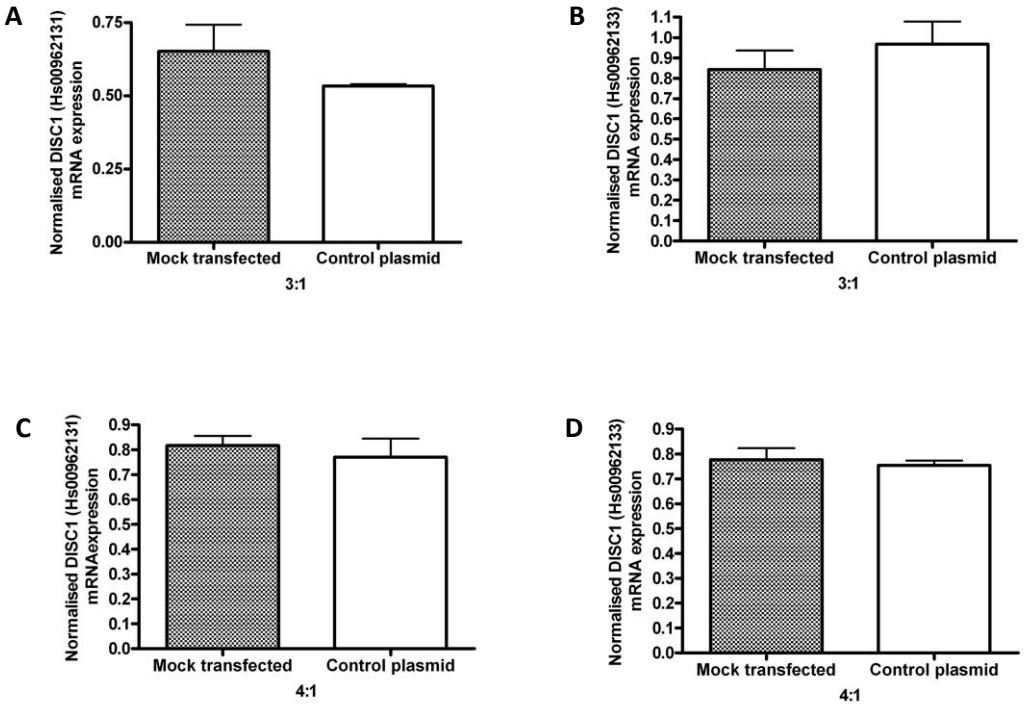


Figure 4.9. Assessment of the effect of the control plasmid, pcDNA3.1/HA on DISC1 mRNA expression. HEK293 cells were transfected with either 2µg of the control plasmid (pcDNA3.1/HA) or mock transfected using two different ratios of transfection reagent to DNA (3:1 and 4:1). Forty-eight hours post-transfection mRNA was extracted, cDNA synthesised and gene expression measured by qRT-PCR. Two different TaqMan gene expression assays were used to measure DISC1 expression, Hs00962131_m1 and Hs00962133_m1. Two algorithms, geNorm and NormFinder, were used to identify stably expressed reference genes. For both ratios of DNA to transfection reagent, the two algorithms identified *SDHA* and *UBC* as the most stable reference genes. DISC1 expression values were normalised to the geometric mean of the expression values of *SDHA* and *UBC*. Shown are normalised DISC1 mRNA expression values obtained under the following conditions (DISC1 gene expression assay, transfection ratio): **A**. Hs00962131, 3:1, **B**. Hs00962133, 3:1, **C**. Hs00962131, 4:1, **D**. Hs00862133, 4:1. Comparison of normalised DISC1 mRNA expression values was carried out using two-tailed independent samples Student's *t*-tests; no statistically significant differences were detected. Experiments were performed in experimental quadruplicate. Error bars represent standard error of the mean (S.E.M.).

To conclude, the available evidence suggests that the transfection of HEK293 cells with the control plasmid does not alter DISC1 expression at the protein or mRNA level.

4.4.1.7 Assessment of the effect of FOXP2 on DISC1 protein expression

The effect of FOXP2 on endogenous DISC1 protein expression was assessed by western blotting (figure 4.10A). For two of the five experimental replicates, Alice Newman (AN), a student under my supervision, performed the transfections, protein extraction and western blotting; I performed the data analysis.

HEK293 cells were transfected with 2µg of either the pcDNA4/HisMax vector containing full-length FOXP2 or the control plasmid. HEK293 cells transfected with the FOXP2-expressing construct showed a mean decrease in the expression of the 100 kDa full-length isoform of DISC1 of 35.3% (corresponding to a fold change of -1.54), which was statistically significant ($p = 0.00627$, two-tailed independent samples *t*-test; figure 4.10B).

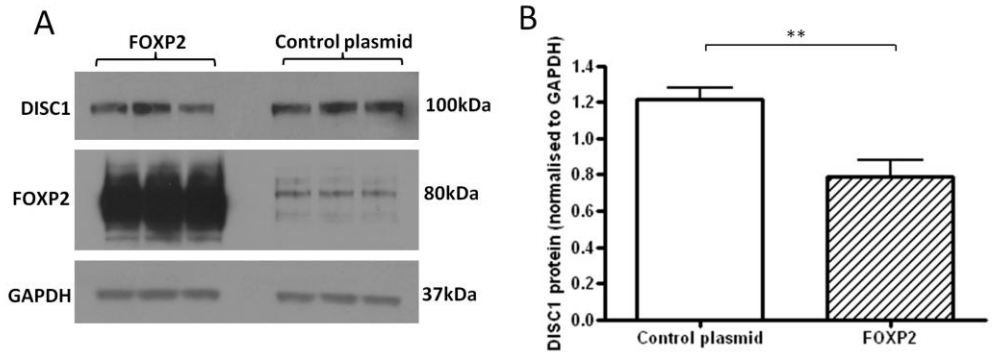


Figure 4.10. DISC1 protein expression is regulated by wild-type FOXP2. **A.** Representative image of western blot analysis of DISC1, FOXP2, and GAPDH protein expression in HEK293 cells transfected with either pcDNA4/HisMax-FOXP2 or the control plasmid (pcDNA3.1/HA). **B.** Quantification of DISC1 protein expression normalised to GAPDH. DISC1 expression was reduced in cells transfected with the FOXP2-expressing construct compared to the control plasmid, (n = 5). Statistical significance was assessed by a two-tailed independent samples Student's *t*-test, $^{**}p \leq 0.01$, error bars represent standard error of the mean (S.E.M.). Figure adapted from Walker et al. (2012).

4.4.1.8 Assessment of the effect of FOXP2 on DISC1 mRNA expression

The effect of FOXP2 on endogenous DISC1 mRNA expression was assessed using qRT-PCR. For this analysis, HT extracted the RNA, synthesised the cDNA and performed the qRT-PCR; I performed the transfections and the data analysis.

HEK293 cells were transfected with 2µg of either pcDNA4/HisMax-FOXP2 or the control plasmid, pcDNA3.1/HA, using two ratios of transfection reagent to DNA (3:1 and 4:1). As described in section 4.4.1.5, two algorithms, geNorm and NormFinder, were used to select the optimal reference genes for normalisation of DISC1 expression values. The genes selected by these algorithms are listed in table 4.4. DISC1 expression was measured using two DISC1 gene expression assays, Hs00962131_m1 and Hs00962133_m1 (table 4.1). Comparison of normalised DISC1 mRNA expression values between the FOXP2- and control plasmid-transfected cells revealed significant differences when DISC1 was measured using the Hs00962133_m1 assay in three of four combinations of reference gene selection algorithm and transfection reagent to DNA ratio assessed (table 4.5, figure 4.11). In all three cases, the direction of the change was such that DISC1 expression was decreased in the cells transfected with FOXP2. Consistent with these findings, a decrease in DISC1 expression was also observed in the fourth comparison for this probe; however the *p*-value fell just short of the threshold for statistical significance. When DISC1 was measured with the Hs00962131_m1 assay, no significant changes in DISC1 expression were observed.

	Ratio	DISC1 assay	Mean		S.E.M.		Fold change	Direction of change	T-test *p*-value
			CP	FOXP2	CP	FOXP2			
geNorm	3:1	Hs00962131	0.533	0.605	0.003	0.032	1.13	↑	0.117
		Hs00962133	0.967	0.693	0.113	0.050	-1.40	↓	0.068
	4:1	Hs00962131	0.318	0.264	0.025	0.008	-1.21	↓	0.081
		Hs00962133	0.314	0.233	0.019	0.026	-1.35	↓	**0.047**
NormFinder	3:1	Hs00962131	0.394	0.394	0.053	0.018	-1.00	↓	0.994
		Hs00962133	0.689	0.449	0.088	0.023	-1.23	↓	**0.023**
	4:1	Hs00962131	0.356	0.378	0.039	0.041	1.06	↑	0.728
		Hs00962133	0.348	0.307	0.007	0.002	-1.13	↓	**0.006**

Table 4.5. Assessment of the effect of FOXP2 on DISC1 mRNA expression. Summary of the data presented in figure 4.10. For each combination of reference gene selection algorithm (geNorm or NormFinder), ratio of transfection reagent to DNA (3:1 or 4:1), and DISC1 expression assay (Hs00962131, highlighted in yellow, or Hs00962133, highlighted in blue), mean DISC1 mRNA expression levels together with the standard error of the mean (S.E.M.) are shown for HEK293 cells transfected with either the control plasmid (CP; pcDNA3.1/HA) or full-length FOXP2 (pcDNA4/HisMax-FOXP2). For each comparison between control plasmid- and FOXP2-transfected cells the fold change in DISC1 expression, calculated relative to the control plasmid-transfected cells, is shown together with an arrow indicating the direction of change. Statistical significance was assessed using a two-tailed independent samples Student's *t*-test with the threshold for significance defined as $p \leq 0.05$. *P*-values meeting the significance threshold are indicated in bold.

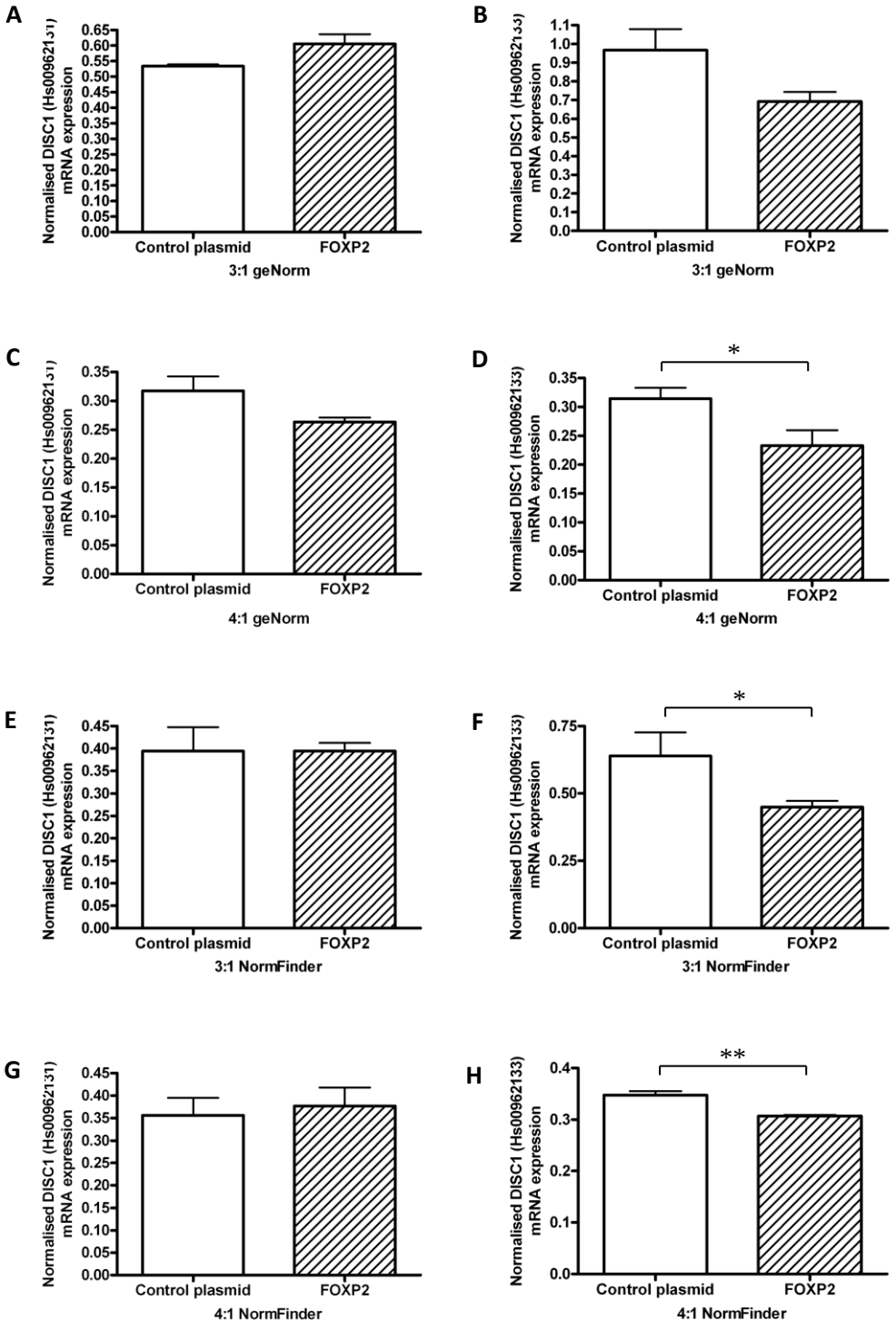


Figure 4.11. Assessment of the effect of FOXP2 on DISC1 mRNA expression. See next page for legend.

Figure 4.11. Assessment of the effect of FOXP2 on DISC1 mRNA expression. HEK293 cells were transfected with 2µg of either the control plasmid (CP; pcDNA3.1/HA) or full-length FOXP2 (pcDNA4/HisMax-FOXP2) using two different ratios of transfection reagent to DNA (3:1 and 4:1). Forty-eight hours post-transfection mRNA was extracted, cDNA synthesised and gene expression measured by qRT-PCR. Two different TaqMan gene expression assays were used to measure DISC1 expression, Hs00962131 and Hs00962133. Two algorithms, geNorm and NormFinder, were used to identify stably expressed reference genes for normalisation of the DISC1 mRNA expression values (table 4.4). Shown are normalised DISC1 mRNA expression values obtained under the following conditions (reference gene selection algorithm, DISC1 gene expression assay, transfection ratio): **A**. geNorm, Hs00962131, 3:1, **B**. geNorm, Hs00962133, 3:1, **C**. geNorm, Hs00962131, 4:1, **D**. geNorm, Hs00862133, 4:1., **E**. NormFinder, Hs00962131, 3:1, **F**. NormFinder, Hs00962133, 3:1, **G**. NormFinder, Hs00962131, 4:1, **H**. NormFinder, Hs00862133, 4:1. Experiments were performed in experimental quadruplicate. Error bars represent standard error of the mean (SEM). Comparison of normalised DISC1 mRNA expression values was carried out using two-tailed independent samples Student's *t*-tests. $*p \leq 0.05$ and $**p \leq 0.01$.

4.4.1.9 Assessment of the effect of FOXP2 on DISC1 *promoter activity*

4.4.1.9.1 Assessment of the effect of wildtype full-length FOXP2 on DISC1 *promoter activity*

To determine whether the decrease in DISC1 protein expression observed in cells overexpressing FOXP2 could be attributed to repression of transcriptional activity, dual luciferase reporter assays were carried out. The region of the *DISC1* promoter identified by ChIP-seq as being bound by FOXP2 (see section 4.1.1.4) overlaps with the short, medium and long *DISC1* promoter constructs; therefore, the effect of transient overexpression of FOXP2 on all three was assessed. HEK293 cells were co-transfected with: (i) a DISC1 promoter construct (pGL4.10-long, -medium, or –short, described in the previous chapter); (ii) either a FOXP2-containing vector (pcDNA4/HisMax-FOXP2) or the control plasmid; (iii) and the transfection efficiency control, pRL-TK (figure 4.12).

Two-way ANOVA with the independent variables "FOXP2 condition" (control plasmid or wildtype FOXP2) and "promoter construct" (long, medium, or short) revealed a significant reduction in *DISC1* promoter activity in cells transfected with the FOXP2-expressing construct compared to the control plasmid ($p \leq 0.001$). As reported in chapter three (figure 3.9), the effect of promoter construct was also significant ($p \leq 0.001$), and, in addition, the interaction between promoter construct and FOXP2 condition was also significant ($p = 0.00388$).

To investigate the interaction between promoter construct and FOXP2 condition, simple main effects were analysed. This was achieved by performing a series of one-way ANOVAs followed by Tukey's Honestly Significant Difference (HSD) post-hoc tests. Firstly, the effect of FOXP2 on promoter activity was assessed independently for each promoter construct. Tukey's HSD revealed a decrease in promoter activity for all constructs in FOXP2-transfected cells compared to control plasmid-transfected cells, with this decrease attaining statistical significance for both the long ($p = 0.0101$) and the medium ($p = 0.00923$) constructs, but not the short ($p = 0.114$) construct. Secondly, the effect of promoter construct was assessed independently for each FOXP2 condition. In keeping with our initial characterisation of the long, medium, and short promoter constructs, Tukey's HSD revealed that in both FOXP2-transfected and control plasmid-transfected cells, the activity of the medium construct was significantly greater than the activity of the short (control plasmid: $p \leq 0.001$; FOXP2: $p \leq 0.001$) and the long (control plasmid: $p = 0.00369$; FOXP2: $p = 0.0173$)

constructs. Additionally, in cells transfected with FOXP2, the activity of the long construct was significantly greater than the activity of the short construct (p = 0.0310), whilst in the control plasmid-transfected cells the same pattern of activity was observed but only approached statistical significance (p = 0.0577).

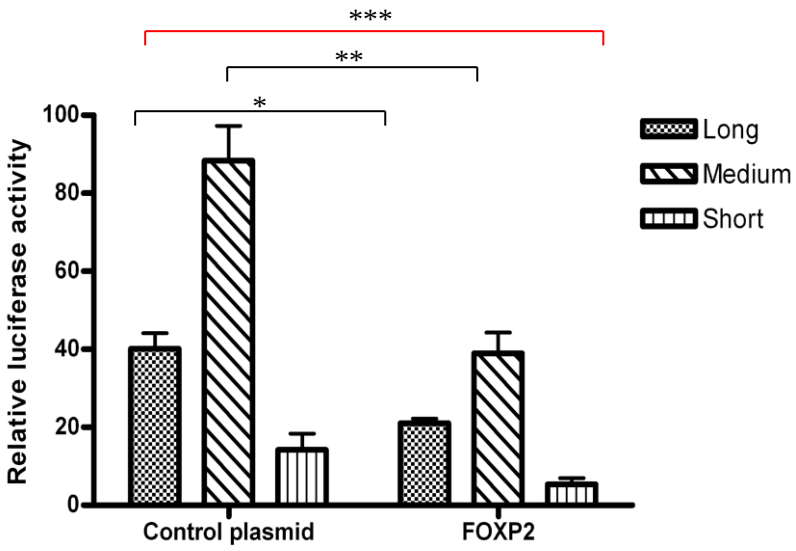


Figure 4.12. *DISC1* promoter activity is regulated by wild-type FOXP2. Dual luciferase reporter assays show down-regulation of *DISC1* promoter activity in cells transfected with the FOXP2-expressing construct compared to the control plasmid. The significant main effect of FOXP2 (***$p \leq 0.001$; indicated by a red bracket), determined by two-way ANOVA, and the significant effect of FOXP2 on the long (*$p \leq 0.05$) and the medium (**$p \leq 0.01$) constructs (indicated by black brackets), determined by two-tailed independent samples Student's *t*-tests are indicated on the graph (n = 3). Error bars represent standard error of the mean (S.E.M.). Figure adapted from

4.4.1.9.2 Assessment of the effect of FOXP2 mutations implicated in speech and language disorder on FOXP2-mediated transcriptional repression of DISC1

Two FOXP2-expressing constructs each containing one of the rare coding mutations, R553H or R328X, which are found only in individuals affected with developmental verbal dyspraxia (DVD) (Lai et al., 2001; MacDermot et al., 2005) were assessed for their ability to repress transcription of *DISC1* using the dual luciferase reporter assay. HEK293 cells were co-transfected with: (i) a *DISC1* promoter construct (pGL4.10-long, -medium, or -short); (ii) either a FOXP2-containing vector (pcDNA4/HisMax-FOXP2, -FOXP2.R553H, or-

FOXP2.R328X) or the control plasmid; (iii) and the transfection efficiency control, pRL-TK (figure 4.13).

Separate two-way ANOVAs were carried out for the two FOXP2 mutations with the independent variables "FOXP2 condition" (control plasmid, wildtype FOXP2, and FOXP2.R553H or FOXP2.R328X) and "promoter construct" (long, medium, and short). For both mutations ANOVA revealed significant main effects of FOXP2 condition and promoter construct ($p \leq 0.001$ for all comparisons). The interaction between FOXP2 condition and promoter construct was also significant for both FOXP2 mutations (R553H: $p = 0.00202$; R328X: $p = 0.0130$).

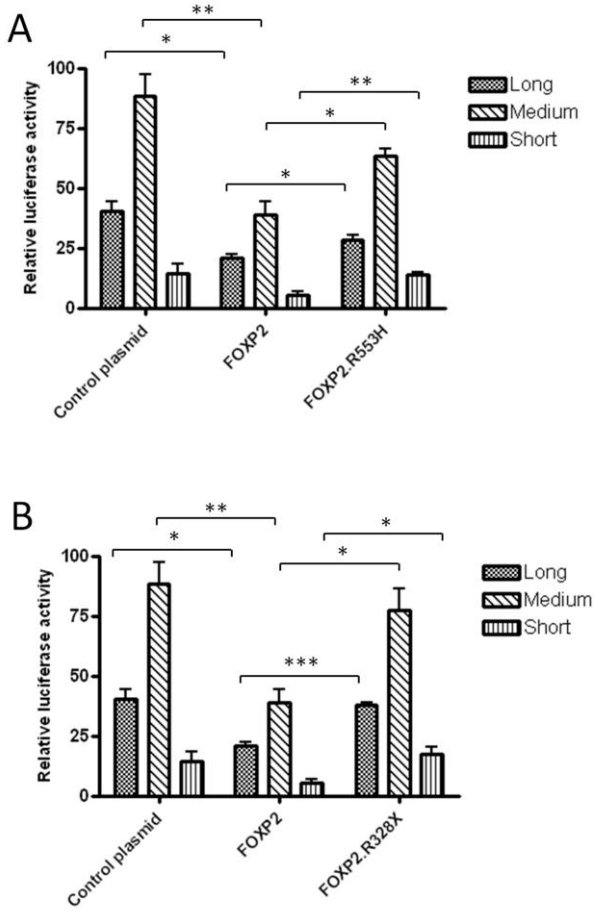


Figure 4.13. FOXP2-mediated inhibition of *DISC1* promoter activity is altered by the pathogenic point mutations R553H and R328X. **A** and **B**. Dual luciferase reporter assays reveal that both the R553H (pcDNA4/HisMax-FOXP2.R553H; A) and R328X (pcDNA4/HisMax-FOXP2.R328X; B) point mutations reduce the ability of FOXP2 to inhibit *DISC1* promoter activity. Both point mutations resulted in a significant reduction in the ability of FOXP2 to down-regulate the promoter activity of all three *DISC1* promoter constructs (n = 3). Statistical significance was assessed by two-tailed independent samples Student's *t*-tests. $*p \leq 0.05$, $**p \leq 0.01$, $***p \leq 0.001$. Error bars represent standard error of the mean (S.E.M.). This figure has been published previously in Walker et al. (2012).

In order to understand the significant interactions, the effect of the two FOXP2 point mutations, R553H and R328X, on the ability of FOXP2 to repress *DISC1* promoter activity was assessed for each promoter construct individually using two-tailed independent samples *t*-tests. A reduction in the ability of FOXP2 to repress *DISC1* promoter activity was observed for both mutations, with this effect attaining significance across all constructs ($p \leq 0.0251$ for all comparisons). The effect of promoter construct was then assessed individually for each FOXP2 condition. Consistent with previous observations, Tukey's HSD revealed that under both conditions the medium promoter construct displayed significantly greater promoter activity than either the short (R553H: $p \leq 0.001$; R328X: $p \leq 0.001$) or long (R553H: $p \leq 0.001$; R328X: $p = 0.00639$) constructs. As observed for cells transfected with FOXP2, the activity of the long construct was significantly greater than that of the short construct in cells transfected with FOXP2.R553H ($p = 0.006$). A trend in the same direction was observed in cells transfected with FOXP2.R328X; however this difference did not attain statistical significance ($p = 0.0932$).

4.4.2 Assessment of Foxp2 expression in mice carrying the L100P *Disc1* point mutation

In the introduction to this chapter (section 4.1.1.2), it was suggested that a bi-directional regulatory relationship might exist between FOXP2 and DISC1 expression. This hypothesis was based on the observation that pathways involved in transcriptional regulation (Wnt signalling, cAMP signalling, and EGR-mediated transcription), which have been linked to DISC1 function, are implicated in the regulation of FOXP2.

As a preliminary step in the investigation of this putative bi-directional regulatory relationship, Foxp2 expression was characterised in mice carrying an ENU-induced point mutation, resulting in an L100P amino acid substitution, in the *Disc1* gene (Clapcote et al., 2007). As described in the introduction to this chapter (section 4.1.2), the *Disc1* L100P mouse displays a "schizophrenia-like" phenotype, which may arise, in part, from the aberrant expression of several genes, including *Disc1*.

Foxp2 mRNA expression was assessed by qRT-PCR in a panel of samples representing different developmental stages collected by SB during her PhD, which has been described fully in Brown et al. (2011). Whole brain cDNA was available for embryonic (E) stages E13,

E15, and E18, and hippocampal cDNA was available for the postnatal (P) stages P1, P7, P20, and adult.

4.4.2.1 Assessment of Foxp2 mRNA expression in the Disc1 L100P mouse by qRT-PCR

4.4.2.1.1 Detection of gDNA contamination

cDNA samples synthesised by SB had previously been found to be free from gDNA contamination. As there was insufficient cDNA for some samples (P1 wildtypes, P7 *Disc1* L100P mice and wildtypes, and P20 wildtypes), RNA for these samples were reverse transcribed by HT and checked, by HT, for gDNA contamination by PCR amplification using intron-spanning primers. Evidence for gDNA contamination was observed in all of the samples from the wildtype mice at P1, P7 and P20. Unfortunately, no more RNA was available from these mice to repeat the cDNA synthesis, so it was decided to proceed with the qRT-PCR using only intron-spanning assays known not to amplify gDNA.

4.4.2.1.2 Selection of reference genes normalisation of qRT-PCR data

As there was only very limited cDNA available for the embryonic samples, and no more RNA available to synthesise additional cDNA, only three reference genes could be measured. To permit comparison with the gene expression profiles obtained from these mice by SB, it was decided that the two reference gene assays used by SB, *Hprt1* and *Gapdh*, should be used in addition to an assay measuring *Actb*.

After measuring the expression of these three reference genes by qRT-PCR, it was found that both the *Gapdh* and *Actb* assays were single-exon assays, which, therefore, would amplify gDNA. As evidence for gDNA contamination had been detected in the wildtype mice at P1, P7 and P20, these assays cannot be considered to provide valid reference gene expression data for normalisation. As such, the sections below report Foxp2 expression values normalised to Hprt1expression, which was measured using an intron-spanning assay, previously shown (by Applied Biosystems) not to amplify gDNA.
As an aside, it is worth noting that prior to realising that both the *Gapdh* and *Actb* assays would be affected by the gDNA contamination present in some samples, Foxp2 expression

values were normalised to the geometric mean of Hprt1, Gapdh and Actb. The results of this analysis were highly similar to those obtained when normalising to Hprt1 alone.

4.4.2.1.3 qRT-PCR analysis

Comparison of the developmental expression of Foxp2 in mice carrying the L100P point mutation with their wild-type counterparts revealed significant differences at E15, E18, P1, P7 and in the adult mice (figure 4.14 and table 4.6). As the uneven numbers of mice in each group may render the Student's *t*-test less robust to violations of the assumptions of normality and homogeneity of variances (Boneau, 1960), compliance with these assumptions was assessed. Normality was assessed using the Shapiro-Wilk test and homogeneity of variances using Levene's test. The data were found to comply with the assumption of normality (all $p \geq 0.170$); however, deviation from the homogeneity of variances assumption was detected at P1 ($p = 0.00678$) and P7 ($p = 0.048$). At these developmental stages significance was, therefore, assessed using Welch's *t*-test.

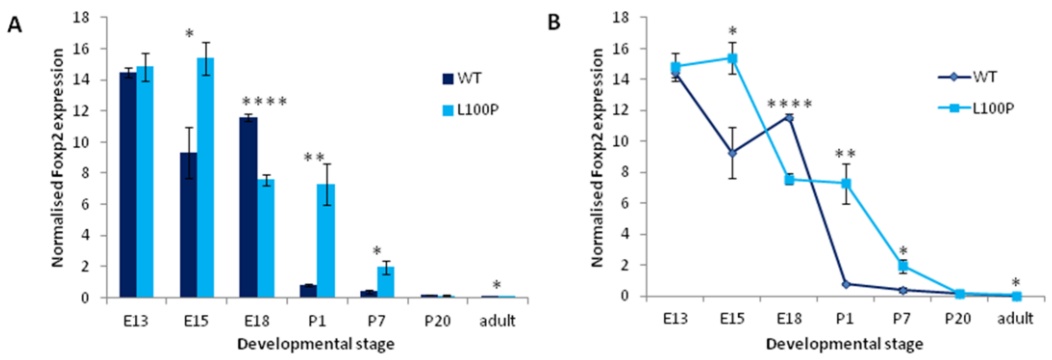


Figure 4.14. Developmental profile of Foxp2 mRNA expression in the L100P mouse obtained by the normalisation of Foxp2 expression values to the expression values of the reference gene Hprt1. Gene expression was measured in whole-brain cDNA at embryonic days (E) 13, 15, and 18, and hippocampal cDNA at postnatal days (P) 1, 7, 20, and in adulthood. Normalised Foxp2 expression values are represented in a bar chart (**A**) and a line graph (**B**). Each data point represents the mean of normalised expression values from a minimum of three mice. Error bars represent standard error of the mean (S.E.M.). Significance was assessed using the two-tailed independent samples Student's *t*-test for all developmental stages except from P1 and P7. As the homogeneity of variances assumption was violated at these stages, significance was determined using Welch's *t*-test. * $p \leq 0.05$, ** $p \leq 0.01$, **** $p \leq 0.0001$.

Dev. stage	N		Mean		S.E.M.		Fold change	Direction of change	*p*-value
	WT	L100P	WT	L100P	WT	L100P			
E13	5	6	14.5	14.8	0.34	0.89	1.02	↑	0.729
E15	6	5	9.31	15.4	1.65	1.03	1.65	↑	**0.0156**
E18	5	6	11.6	7.58	0.23	0.36	-1.53	↓	**8.84×10^{-6}**
P1	5	6	0.792	7.28	0.069	1.31	9.20	↑	**0.00427**†
P7	6	4	0.408	1.96	0.132	0.429	4.81	↑	**0.0307**†
P20	3	6	0.196	0.147	0.0208	0.0328	-1.34	↓	0.358
Adult	5	3	0.0245	0.0556	0.00643	0.00408	2.27	↑	**0.0141**

Table 4.6. Assessment of Foxp2 expression in *Disc1* L100P mice (L100P) and wildtype (WT) controls throughout development. Normalised *Foxp2* mRNA expression values were ascertained by qRT-PCR followed by normalisation to the reference gene *Hprt1*. Gene expression was measured in whole-brain cDNA at embryonic days (E) 13, 15, and 18, and hippocampal cDNA at postnatal days (P) 1, 7, 20, and in adulthood. Shown for each group are the number (N) of mice assessed, the mean normalised Foxp2 mRNA expression value and the standard error of the mean (S.E.M.). The fold change in Foxp2 expression was calculated relative to the wildtype control group. The direction of each change is indicated by an arrow (red upwards arrow = increase, green downwards arrow = decrease). *P*-values are displayed for the two-tailed independent samples Student's *t*-test. *P*-values meeting the significance threshold are indicated in bold. †As the homogeneity of variances assumption was violated for the P1 and P7 comparisons, significance was determined using Welch's *t*-test.

4.4.2.2 Comparison of Foxp2 and Disc1 developmental expression profiles

During her PhD, SB measured Disc1 expression in the same developmental panel of mRNA samples used here for the measurement of Foxp2. To permit comparison between the developmental pattern of Disc1 and Foxp2 mRNA expression, the Disc1 expression data was re-analysed, this time normalising only to Hprt1. This revealed significant differences in Disc1 expression between the *Disc1* L100P mice and the wildtype mice at E15, P1, P7 and P20 (figure 4.15; table 4.7). As there were unequal numbers of mice in several of the groups, the data were assessed for normality and homogeneity of variances. Normality was assessed using the Shapiro-Wilk test and homogeneity of variances using Levene's test. The distribution of the data for the wildtype mice at P20 was found to significantly differ from the normal distribution (p = 0.0292) and violation of the homogeneity of variances assumption was detected at P7 (p = 0.00338) and in the adult mice (p = 0.00467). As such,

the comparison at P20 was carried out using a Mann-Whitney U test and the comparisons at P7 and for the adult mice were carried out using Welch's *t*-test.

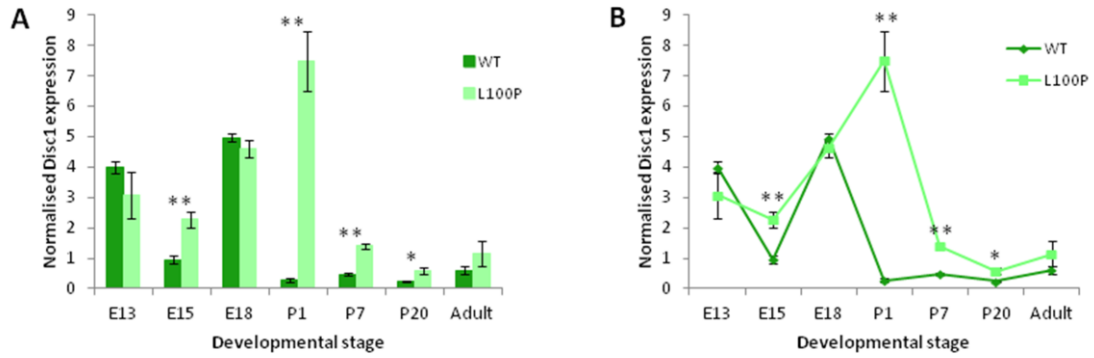


Figure 4.15. Developmental profile of Disc1 mRNA expression in the L100P mouse obtained by the normalisation of Disc1 expression values to the expression values of the reference gene Hprt1. Gene expression was measured in whole-brain cDNA at embryonic days (E) 13, 15, and 18, and hippocampal cDNA at postnatal days (P) 1, 7, 20, and in adulthood. Normalised Disc1 expression values are represented in a bar chart (**A**) and a line graph (**B**). Each data point represents the mean of normalised expression values from a minimum of three mice. Error bars represent standard error of the mean (S.E.M.). Significance was assessed using the two-tailed independent samples Student's *t*-test for all developmental stages except from P1, P7 and P20. At P20 the assumption of normality was violated, so significance was assessed using a Mann-Whitney U test. At P1 and P7 the homogeneity of variances assumption was violated, so significance was determined using Welch's *t*-test. * $p \leq 0.05$, ** $p \leq 0.01$.

Dev. stage	N		Mean		S.E.M.		Fold change	Direction of change	*p*-value
	WT	L100P	WT	L100P	WT	L100P			
E13	5	5	3.99	3.07	0.207	0.777	-1.30	↓	0.288
E15	6	3	0.953	2.27	0.136	0.272	2.38	↑	**0.00170**
E18	4	6	4.95	4.61	0.138	0.275	-1.08	↓	0.363
P1	5	5	0.281	7.49	0.0578	0.982	26.6	↑	**0.00255**†
P7	5	4	0.476	1.40	0.0401	0.106	2.93	↑	**0.00250**†
P20	3	5	0.243	0.568	0.0330	0.112	2.34	↑	**0.0360**‡
Adult	4	6	0.608	1.14	0.128	0.404	1.88	↑	0.254

Table 4.7. Assessment of Foxp2 expression in *Disc1* L100P mice (L100P) and wildtype (WT) controls throughout development. Normalised *Foxp2* mRNA expression values were ascertained by qRT-PCR followed by normalisation to the reference gene *Hprt1*. Gene expression was measured in whole-brain cDNA at embryonic days (E) 13, 15, and 18, and hippocampal cDNA at postnatal days (P) 1, 7, 20, and in adulthood. Shown for each group are the number (N) of mice assessed, the mean normalised Foxp2 mRNA expression value and the standard error of the mean (S.E.M.). The fold change in Foxp2 expression was calculated relative to the wildtype control group. The direction of each change is indicated by an arrow (red upwards arrow = increase, green downwards arrow = decrease). *P*-values are displayed for the two-tailed independent samples Student's *t*-test. *P*-values meeting the significance threshold are indicated in bold. †As the homogeneity of variances assumption was violated for the P1 and P7 comparisons, significance was determined using Welch's *t*-test. ‡As the assumption of normality was violated at P20, comparisons were carried out using a Man-Whitney U test.

Comparison of the Disc1 and Foxp2 developmental expression profiles reveals some intriguing similarities between the expression patterns of the two genes in wild-type mice (figure 4.15). Interestingly, the relationship between Disc1 and Foxp2 expression is altered in the L100P mice.

In the wildtype mice (figure 4.16 A and 4.16 B), both Disc1 and Foxp2 show a decrease in expression between E13 and E15, before increasing between E15 and E18. The expression of both in the postnatal mice is low, although there is a slight increase in the expression of Disc1, but not Foxp2, in the adult hippocampus between P20 and adulthood.

The expression of both genes and the relationship between their expression levels is altered in mice carrying the *Disc1* L100P mutation (figure 4.16 C and figure 4.15 D). In the embryonic stages, Disc1 and Foxp2 show almost inverse patterns of expression in the L100P mice, in contrast to the parallel patterns of expression found in the wildtype mice. In common with the wildtype mice, the expression of Disc1, but not Foxp2, increases from P20 to adulthood.

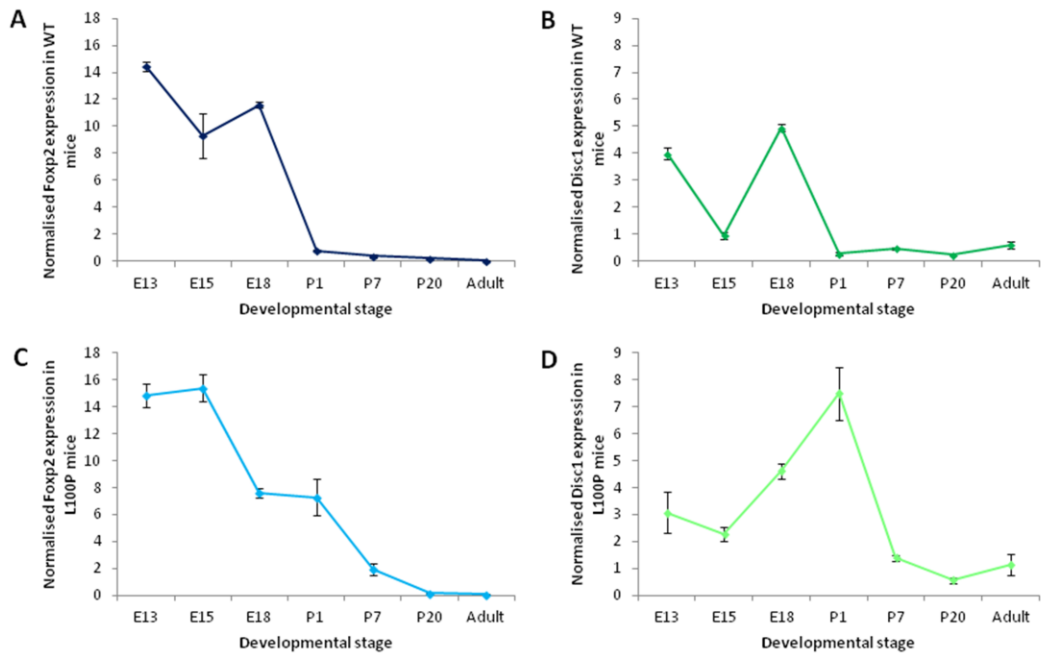


Figure 4.16. Comparison of Foxp2 and Disc1 mRNA developmental expression profiles in the *Disc1* L100P mice and wildtype (WT) controls. Gene expression was measured in whole-brain cDNA at embryonic days (E) 13, 15, and 18, and hippocampal cDNA at postnatal days (P) 1, 7, 20, and in adulthood**.** Disc1 and Foxp2expression values were normalised to expression values for the reference gene Hprt1. **A**. Foxp2 mRNA expression in the WT mice, **B**. Disc1 mRNA expression in the WT mice, **C**. Foxp2 mRNA expression in the *Disc1* L100P mice, and **D**. Disc1 mRNA expression in the *Disc1* L100P mice. Measurements were carried out in at least three biological replicates. Error bars represent standard error of the mean (S.E.M.).

4.4.2.3 Assessment of Foxp2 protein expression in the Disc1 *L100P mouse by western blotting*

In addition to assessing Foxp2 mRNA expression in the *Disc1* L100P mice, Foxp2 protein expression was measured. Expression was assessed in whole brain lysates from embryonic mice at embryonic days (E) 12.5, 15.5, and 17.5, and in region-specific lysates (cortex, hippocampus, striatum, and cerebellum) from postnatal mice at postnatal days (P) 1, 7, 21, and 90. Protein samples were extracted by EG, who also pooled the samples, such that the embryonic samples comprised the brains of approximately six littermates, and the postnatal samples comprised tissue from at least three mice. As protein samples were limited, it was only possible to assess Foxp2 expression in one experimental replicate. As such, these findings must be considered preliminary and in need of independent replication.

4.4.2.3.1 Embryonic stages

Foxp2 expression was measured in whole brain lysates from pooled litters of embryonic mice at E12.5, E15.5, and E17.5. The expression levels of three reference genes, α-tubulin, β-actin, and Gapdh were also measured to permit normalisation for between-lane differences in protein loading.

Foxp2 was identified as a single band of 80 kDa, consistent with the manufacturer's datasheet. Inspection of the immunoreactive bands (figure 4.17 A) revealed an apparent decrease in Foxp2 signal intensity in the *Disc1* L100P mice at each developmental stage. The signal intensities of the α-tubulin and β-actin immunoreactive bands was fairly even between the *Disc1* L100P and the wildtype mice at each developmental stage, with the exception of β-actin at E12.5: in the *Disc1* L100P mice, the β-actin E12.5 band was disrupted by an air bubble, thus preventing comparison. Developmental regulation of β-actin is suggested by an increase in signal intensity with age. In contrast, inspection of the Gapdh immunoreactive bands revealed an alternating pattern of signal intensity, most obvious at E15.5 and E17.5, such that the signal intensity of Gapdh was decreased in mice carrying the *Disc1* L100P mutation.

The signal intensities of each band were quantified by optical densitometry, and normalised Foxp2 expression values obtained by dividing the Foxp2 optical density (OD) values by the

corresponding OD values for each reference gene. In light of the lack of consistency between the signal intensities of the different reference genes, it was decided that Foxp2 should be normalised to each gene separately in order to permit an unbiased depiction of Foxp2 expression. Normalising to both α-tubulin and β-actin suggested a decrease in Foxp2 expression across all the embryonic stages assessed in the *Disc1* L100P mice compared to the wildtypes (figures 4.17 B and 4.17 C). In contrast, normalisation to Gapdh revealed the *Disc1* L100P mice to have a slight decrease in Foxp2 expression at E12.5, but increased Foxp2 expression at E15.5 and E17.5 (figure 4.17 D).

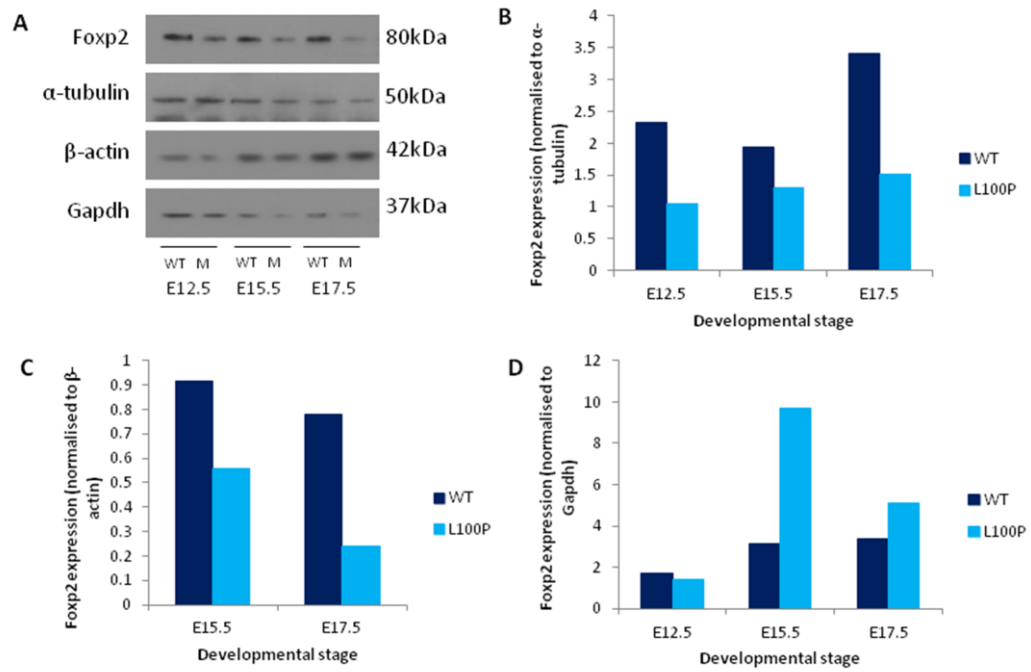


Figure 4.17. Foxp2 protein expression in embryonic mice carrying the *Disc1* L100P mutation at embryonic days (E) 12.5, 15.5, and 17.5. **A**. The expression of Foxp2, together with three reference genes, α-tubulin, β-actin, and Gapdh, was assessed in whole brain lysates pooled from a litter of either *Disc1* L100P (M) or wildtype (WT) embryos by western blotting. Densitometric analysis was performed and the acquired optical density values for the Foxp2 immunoreactive bands were normalised to the corresponding optical density values for each of α-tubulin (**B**), β-actin (**C**), and Gapdh (**D**).

The decrease in Gapdh signal intensity in the *Disc1* L100P compared to the wildtype mice was unexpected given (i) the lack of a comparable difference in the signal intensities of α-tubulin and β-actin, and (ii) the fact that equal loading of the samples was attempted. Unfortunately, it was not possible to obtain whole brain lysates from independent mice to

attempt independent replication of these results; however, to assess the contribution of technical variation to the present results, a second western blot of the same samples was probed as a technical replicate. EG performed the western blotting and probed the blot for β-actin and Gapdh, I probed the blot for Foxp2, and Susan Anderson (SA) probed the blot for α-tubulin.

Visual inspection of the technical replicate revealed some unexpected results (figure 4.17 A): whereas an apparent decrease in the signal intensity of the FOXP2 immunoreactive bands was again evident in the L100P mice, the pattern of the reference gene immunoreactive bands differed from the first technical replicate. At E12.5, for both the wildtype and the L100P mice, the expression of α-tubulin was undetectable (previously, this was highest at E12.5 and decreased with age). β-actin expression showed a similar pattern in the two replicates, while the expression of Gapdh appeared more equal between the wildtype and mutant mice at each developmental stage in the second technical replicate.

Normalisation of the Foxp2 OD values to the OD values of the reference genes (figure 4.18 B-D) revealed a trend for decreased Foxp2 expression in the L100P mice at every developmental stage. For normalisation to α-tubulin (figure 4.18 B) and β-actin (figure 4.18 C), the direction of the between-genotype expression changes were consistent between the two technical replicates, although, the relative expression levels of Foxp2 between the two genotypes, and the pattern of expression throughout development differed. Normalising to Gapdh produced results that contradicted the first Gapdh technical replicate: the L100P mice, instead of showing increased Foxp2 expression at E15.5 and E17.5, showed a decrease in Foxp2 expression at all embryonic stages.

Assessment of the fold changes in Foxp2 expression calculated using normalised Foxp2 values obtained from each of the three reference genes provides suggests a decrease in Foxp2 expression in embryonic *Disc1* L100P mice (table 4.8). When considering both technical replicates and all three reference genes, the normalised Foxp2 expression values determined by normalising to Gapdh in the first technical replicate appear anomalous. Nevertheless, further technical and experimental replicates, which were unfortunately not possible due to limited protein samples, are required before any firm conclusions can be drawn.

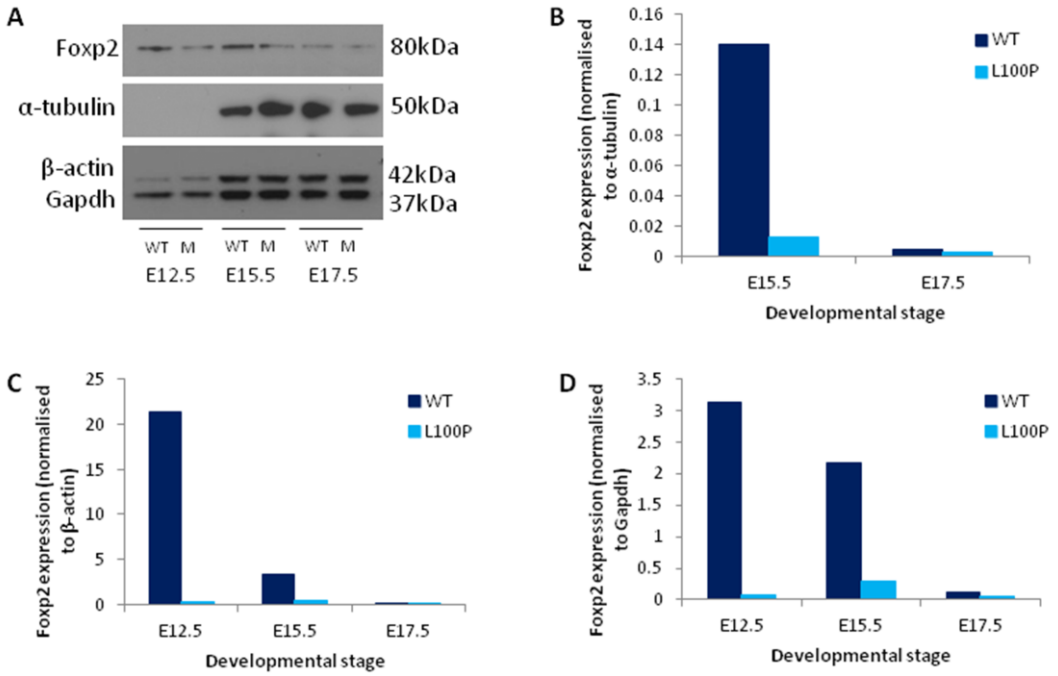


Figure 4.18. Technical replicate of the assessment of Foxp2 expression in embryonic mice carrying the *Disc1* L100P mutation at embryonic days (E) 12.5, 15.5, and 17.5. **A**. The expression of Foxp2, together with three reference genes, α-tubulin, β-actin, and Gapdh, was assessed in whole brain lysates pooled from a litter of either *Disc1* L100P (M) or wildtype (WT) embryos by western blotting. Densitometric analysis was performed and the acquired optical density values for the Foxp2 immunoreactive bands were normalised to the corresponding optical density values for each of α-tubulin (**B**), β-actin (**C**), and Gapdh (**D**).

Developmental stage	Fold change (replicate 1)			Fold change (replicate 2)		
	α-tubulin	β-actin	Gapdh	α-tubulin	β-actin	Gapdh
E12.5	-2.30 ↓	-2.01 ↓	1.30 ↑	N/A*	-77.0 ↓	-40.9 ↓
E15.5	-1.71 ↓	-1.16 ↓	12.7 ↑	-10.6 ↓	-6.96 ↓	-7.29 ↓
E17.5	-1.89 ↓	-3.58 ↓	6.04 ↑	-1.58 ↓	-1.88 ↓	-2.17 ↓

Table 4.8. Fold changes in Foxp2 protein expression in embryonic *Disc1* L100P mice. Fold changes for normalised Foxp2 protein expression values (determined by normalisation to either α-tubulin, β-actin, or Gapdh) in the *Disc1* L100P mice were calculated relative to the wildtype control mice: a positive fold change indicates increased expression in the *Disc1* L100P mice, while a negative fold change indicates decreased expression in the Disc1 L100P mice. Expression was assessed at embryonic (E) days 12.5, 15.5, and 17.5, in two technical replicates. Arrows indicating the direction of the fold change are included for clarity. *This value could not be calculated as the signal intensity of the α-tubulin immunoreactive bands was too low to quantify.

4.4.2.3.2 Postnatal stages

Foxp2 protein expression was measured in the striatum, cerebellum, cortex, and hippocampus at P1, P7, P21, and P90. The striatum, cerebellum, and cortex were assessed as these regions have previously been shown to express Foxp2 throughout development, with expression in the striatum being particularly strong (Ferland et al., 2003; Lai et al., 2003). In contrast, there is a lack of consensus regarding the expression of Foxp2 in the hippocampus: some studies have failed to detect hippocampal expression of Foxp2 at the mRNA (Ferland et al., 2003; Takahashi et al., 2003) or protein level (Ferland et al., 2003), whereas another reports weak Foxp2 immunoreactivity in the hippocampus (Fujita et al., 2008). In humans, FOXP2 expression has been detected in the hippocampus (Wilcke et al., 2012). As qRT-PCR assessment of Foxp2 mRNA expression in the Disc1 L100P mice revealed hippocampal expression (section 4.3.7.1.2), the hippocampus was also assessed for Foxp2 protein expression. As for the embryonic stages, the expression of α-tubulin, β-actin and Gaodh were measured to permit normalisation of Foxp2 expression values.

In all four brain regions an 80 kDa product indicative of Foxp2 expression was detected (figure 4.19). An additional, fainter, band, of a slightly lower molecular weight was also present in all four regions. Generally, this band was most evident at the earliest postnatal stages, disappearing by adulthood; it is, however, visible at P90 in the striatum (although only in the first technical replicate, STM (1)). As this band was not predicted by the manufacturer's datasheet, only the 80 kDa band was quantified.

As for the embryonic samples, normalisation of Foxp2 expression values was carried out for each reference gene individually rather than to the geometric mean of all three reference genes. This decision was made as (i) inconsistencies were evident between the relative expression levels of the three reference genes between the *Disc1* L100P and the wildtype conditions for a given comparison and (ii) sometimes the signal intensity of one or more of the reference genes for a given sample was too low to quantify, perhaps due to a technical artefact. An example of the first issue is the relationship between the expression levels of the three reference genes in the hippocampus (HPC (1)). Here, at P1 and P7 the signal intensity of Gapdh is lower in the *Disc1* L100P mice than the wildtype mice; a similar change is not, however, observed in the signal intensities of α-tubulin or β-actin. Unfortunately, Gapdh ran off the end of the gel during electrophoresis of the second technical replicate, so further assessment of the cause of this discrepancy was precluded. Regarding the second issue, there

are some incidences where it appears likely that a technical artefact, such as the presence of an air bubble during the transfer stage, resulted in a poorly formed band. One example of this is the band obtained for β-actin in the *Disc1* L100P hippocampal sample (HPC (2)) at P1. In other cases, the expression levels of one or two of the reference genes were too low to quantify, while the expression of the other reference genes was clearly evident. This occurred in the cerebellum at P21 and P90: here, the expression of α-tubulin and β-actin was unquantifiable in the wildtype mice, while Gapdh was clearly expressed. When the band for a reference gene was only partially formed, or unquantifiable, normalisation to that reference gene was not performed.

Comparison of the normalised Foxp2 values between the *Disc1* L100P mice and the wildtype controls (figure 4.19 and table 4.9), suggests that Foxp2 expression may be altered in the *Disc1* L100P mice at several developmental stages across multiple brain regions. Consistent with findings at the mRNA level, increased Foxp2 expression was observed in the *Disc1* L100P mice at P1 in the hippocampus. This increase was evident in both technical replicates for all reference genes. Slight increases in Foxp2 expression at P1 were also observed in the striatum (although only for the first technical replicate) and cortex. A consistent decrease in Foxp2 expression in the *Disc1* L100P mice at P7 was observed for all four brain regions when normalising to every reference gene. The findings at P20 and P90 were less clear-cut, in part because the lower level of Foxp2 immunoreactivity detected at these stages limited the number of comparisons that could be made. Nevertheless, the available evidence suggests decreased Foxp2 expression at both these developmental stages in the *Disc1* L100P mice, the exception being in the striatum, where normalisation to different reference genes yielded conflicting results.

Considering each brain region separately, a consistent downregulation in Foxp2 expression in the cerebellum was observed for the *Disc1* L100P mice at every developmental stage. In both the hippocampus and the cortex, the available evidence suggests an upregulation of Foxp2 expression in the *Disc1* L100P mice at P1 and a downregulation at subsequent developmental stages. The pattern of expression in the striatum is less clear due to inconsistencies between reference genes and technical replicates.

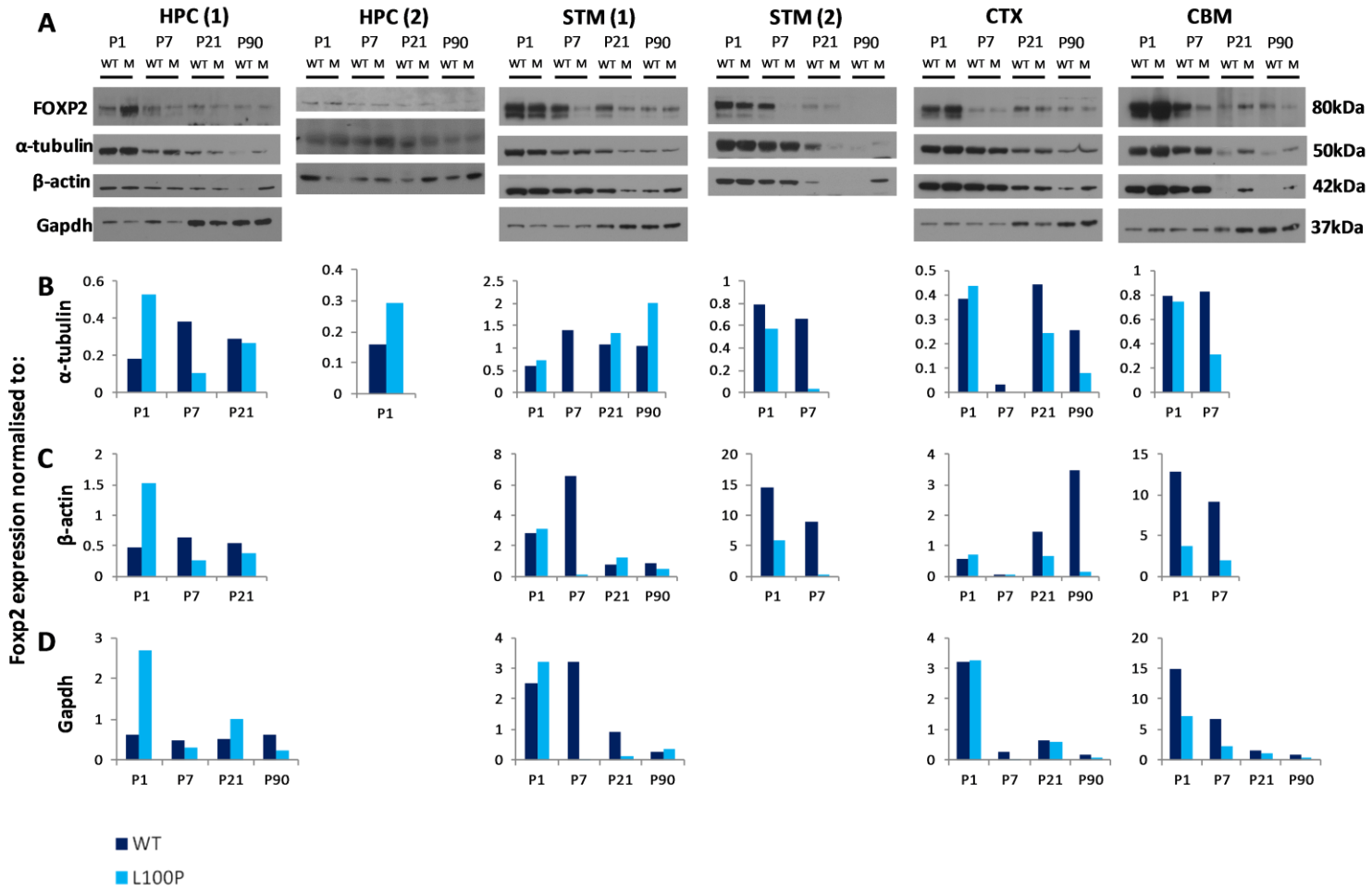


Figure 4.19. Foxp2 protein expression in postnatal mice carrying the *Disc1* L100P mutation at postnatal days (P) 1, 7, 20, and 90. See next page for full legend.

Figure 4.19. Foxp2 protein expression in postnatal mice carrying the *Disc1* L100P mutation at postnatal days (P) 1, 7, 20, and 90. **A**. The expression of Foxp2, together with three reference genes, α-tubulin, β-actin, and Gapdh, was assessed in the hippocampus (HPC), striatum (STM), cortex (CTX) and cerebellum (CBM) in *Disc1* L100P (M) mice and wildtype mice (WT) by western blotting. Densitometric analysis was performed and the acquired optical density values for the Foxp2 immunoreactive bands were normalised to the corresponding optical density values for each of α-tubulin (**B**), β-actin (**C**), and Gapdh (**D**). Western blotting of hippocampal and striatal lysates were performed in two technical replicates, indicated in parentheses; expression in the cortex and the cerebellum was assessed in one replicate. Assessment of Gapdh expression in the second technical replicates of the hippocampus and the striatum was precluded as proteins of this molecular weight had run off the end of the gel during electrophoresis.

	HPC						STM						CTX			CBM		
	Replicate 1			Replicate 2			Replicate 1			Replicate 2								
	AT	BT	G	AT	BT	G	AT	BT	G	AT	BT	G	AT	BT	G	AT	BT	G
P1	2.92 ↑	3.23 ↑	4.32 ↑	1.86 ↑	N/A	N/A	1.22 ↑	1.11 ↑	1.27 ↑	-1.36 ↓	-2.41 ↓	N/A	1.14 ↑	1.26 ↑	1.00 ↑	-1.05 ↓	-3.43 ↓	-2.06 ↓
P7	-3.73 ↓	-2.54 ↓	-1.45 ↓	N/A	N/A	N/A	-84.8 ↓	-124 ↓	-82.6 ↓	-18.2 ↓	-46.2 ↓	N/A	-5.92 ↓	-5.80 ↓	-6.34 ↓	-2.62 ↓	-4.49 ↓	-3.06 ↓
P20	-1.09 ↓	-1.40 ↓	-1.93 ↓	N/A	N/A	N/A	1.25 ↑	1.66 ↑	-6.36 ↓	N/A	N/A	N/A	-1.81 ↓	-2.28 ↓	-1.10 ↓	-1.60 ↓	N/A	-1.32 ↓
P90	N/A	-17.3 ↓	-2.65 ↓	N/A	N/A	N/A	2.00 ↑	-1.79 ↓	1.37 ↑	N/A	N/A	N/A	-3.17 ↓	-20.4 ↓	-2.45 ↓	-4.72 ↓	N/A	-1.94 ↓

Table 4.9. Fold changes in Foxp2 protein expression in postnatal *Disc1* L100P mice. Fold changes for normalised Foxp2 protein expression values (determined by normalisation to either α-tubulin (AT), β-actin (BA), or Gapdh (G)) in the *Disc1* L100P mice were calculated relative to the wildtype control mice: a positive fold change indicates increased expression in the *Disc1* L100P mice, while a negative fold change indicates decreased expression in the Disc1 L100P mice. Expression was assessed at postnatal (P) days 1, 7, 20, and 90 in the hippocampus (HPC), striatum (STM), cortex (CTX) and cerebellum (CBM). Assessment of protein expression in the hippocampus and the striatum was carried out in two technical replicates, while expression levels in the cortex and the cerebellum were both assessed in one replicate. Arrows indicating the direction of the fold change are included for clarity. N/A indicates values that could not be calculated due to the signal intensity of Foxp2 and/or the reference gene being too low to quantify.

4.5 Summary and discussion

The work described in this chapter has explored the FOXP2-DISC1 regulatory relationship. A summary and discussion of each part of this work is presented below.

4.5.1 The role of FOXP2 in the regulation of DISC1 expression and promoter activity

The role of FOXP2 in regulating the expression of DISC1 was assessed by western blotting and qRT-PCR and the effect of FOXP2 on DISC1 promoter activity was assessed using the dual luciferase reporter (DLR) assay. All experiments were carried out in HEK293 cells, which were shown to express both FOXP2 and DISC1 endogenously.

Western blotting revealed a significant decrease in the expression of endogenous DISC1 in HEK293 cells transfected with a vector encoding full-length FOXP2 compared to the control plasmid. This finding was partially supported by findings at the RNA level. Significant decreases in DISC1 mRNA expression were observed in cells transfected with FOXP2 when DISC1 expression was measured with one TaqMan gene expression assay but not with another. Dual luciferase reporter (DLR) assays demonstrated a significant reduction in the activity of the long and medium *DISC1* promoter constructs in cells transfected with FOXP2, with the activity of the short construct showing the same trend. Failure to detect a significant effect of FOXP2 on the short promoter construct appears to be due to the relatively high variance associated with the measurement of activity from this construct, as the percentage decrease in mean luciferase activity in the FOXP2 condition compared to the control plasmid condition was similar to that observed for the long and medium constructs. The observation of decreased activity from all three *DISC1* promoter constructs (albeit non-significant for the short construct) is compatible with the findings of chromatin immunoprecipitation (ChIP) screens in this region (Rosenbloom et al., 2010; Spiteri et al., 2007): both ChIP studies identified FOXP2-bound regions that span all three promoter constructs.

Taken together with the ChIP data (Rosenbloom et al., 2010; Spiteri et al., 2007), the most parsimonious interpretation of the data presented in this chapter is that FOXP2 is a transcriptional repressor of *DISC1*, which, when overexpressed, results in reduced DISC1 mRNA and protein expression.

The main caveat to this interpretation is the discrepancy between the results obtained using the two gene expression assays. The assay, Hs00962133, which detected reduced DISC1 expression following FOXP2 transfection detects six RefSeq transcripts formed from the intergenic splicing of *DISC1* and the upstream gene *TSNAX* (Millar et al., 2000a; Nakata et al., 2009), in addition to fifteen RefSeq DISC1 transcripts. The assay Hs00962131, which did not detect any difference in expression, detects 22 DISC1 transcripts, including the 15 detected by Hs00962133. Of the seven transcripts detected by Hs00962131 but not Hs00962133, six encode short DISC1 isoforms. The FOXP2-mediated decrease in DISC1 protein expression was detected using a C-terminal antibody, which only detects transcripts containing exons 10-13. This antibody would not, therefore, detect protein products produced from six of the seven DISC1 transcripts detected uniquely by Hs00962131. Before discussing possible explanations for the discrepant qRT-PCR results, it is worth noting that it would be useful to repeat this experiment using another pan-DISC1gene expression assay to determine whether the discrepancy might result from a technical issue affecting one of the assays (although, both assays had been validated by Applied Biosystems).

If both assays are assumed to have functioned correctly, one explanation for their contradictory findings is that the FOXP2-mediated decrease in expression detected by the Hs00962133 assay was attributable to altered expression of the TSNAX-DISC1 transcripts. This possibility could be tested by performing gel electrophoresis of the qRT-PCR reaction products, followed by sequencing to confirm their identity. While this possibility should be formally assessed, an observation that implicates a FOXP2-mediated effect on DISC1 expression rather than TSNAX-DISC1 expression is that FOXP2 has not been identified as binding in the region proximal (1 kb upstream or downstream) to the TSNAX transcription start site according to either the ENCODE Transcription Factor ChIP track on the UCSC human genome browser or genome-wide FOXP2/Foxp2 ChIP-chip studies (Spiteri et al., 2007; Vernes et al., 2011; Vernes et al., 2007). Furthermore, altered TSNAX-DISC1 expression cannot account for the change in expression observed at the protein level. Only one protein product of 100 kDa, known to correspond to full-length DISC1, was detected by western blotting. Moreover, according to RefSeq, all DISC1-TSNAX transcripts are predicted to be non-coding and to undergo nonsense mediated decay. As such, if a FOXP2 mediated decrease in the expression of the TSNAX-DISC1 transcripts accounts for the discrepant qRT-PCR findings, then separate mechanisms would be required to explain both FOXP2's effect on DISC1 protein expression and its effect on *DISC1* promoter activity.

Another possible explanation for the disagreement between the two qRT-PCR assays is that additional *DISC1* transcription start sites (TSSs) exist that remain to be characterised. As such, it would be possible for some DISC1 transcripts but not others to be regulated by FOXP2. Thus, DISC1 transcripts derived from different TSSs could have contributed differentially to the signal measured by each assay. This hypothesis could be assessed by gel electrophoresis followed by sequencing of the qRT-PCR products. If there is evidence suggestive of the use of alternative TSSs, then 5' rapid amplification of cDNA ends (RACE) could be carried out to assess their existence.

The existence of two additional *DISC1* TSSs was suggested by the previous release of the UCSC human genome browser (hg18). Here, two DISC1 transcripts, one utilising a TSS in a putative exon between *DISC1* exons 1 and 2 (exon 1b) and the other derived from a TSS in exon 2 were identified as non-RefSeq transcripts. However, following a series of 5' RACE experiments carried out by Nakata et al. (2009), which identified only one *DISC1* TSS (the exon 1 TSS used by all RefSeq transcripts), these two transcripts were excluded from the most recent release of the UCSC human genome browser (hg19). As Nakata et al. (2009) only assessed human foetal brain, it is possible that alternative *DISC1* TSSs are used at other developmental stages, or that transcripts from alternative TSSs are expressed in a cell-type-specific or region-specific manner not detectable when analysing whole-brain cDNA.

4.5.1.1 FOXP2 as a transcriptional repressor of DISC1 expression

The data presented in this chapter identify FOXP2 as a transcriptional repressor of *DISC1*. This adds *DISC1* to a list of experimentally validated transcriptional targets of FOXP2 that includes *CNTNAP2* (Vernes et al., 2008), *SRPX2*/*uPAR* (Roll et al., 2010) and *MET* (Mukamel et al., 2011). All three of these transcriptional targets have been implicated in neurodevelopmental conditions. Genetic variation in *CNTNAP2* has been associated with several conditions including specific language impairment (Vernes et al., 2008), autism spectrum disorder (ASD) (Penagarikano and Geschwind, 2012), severe intellectual disability (Gregor et al., 2011), schizophrenia (Friedman et al., 2008; ISC, 2008; Wang et al., 2010), and bipolar disorder (Wang et al., 2010). Mutations in the *SPRX2* gene have been found to cause speech dysfunction resulting from epilepsy of the rolandic speech areas, developmental verbal dyspraxia and bilateral perisylvian polymicrogyria (Roll et al., 2006). Bilateral perisylvian polymicrogyria is a clinical syndrome that manifests as mild mental retardation, epilepsy and pseudobulbar palsy, a lesion of the upper motor neuron that causes

slurred speech as well as difficulties in chewing and swallowing (Verrotti et al., 2010). Genetic variation in *MET* has been associated with ASD in several independent studies (Judson et al., 2011). The nature of the impairments resulting from genetic variation in these genes suggests FOXP2 to play an important role in neurodevelopment. This assertion is supported by the findings of genome-wide ChIP studies: FOXP2's ChIP-identified targets are enriched for genes involved in synaptic plasticity and in neurite outgrowth, processes essential for normal brain development (Spiteri et al., 2007; Vernes et al., 2011; Vernes et al., 2007).

Several lines of evidence implicate FOXP2 as being particularly important for the development and function of the corticostriatal and olivocerebellar circuits, which are involved in motor control. During development, FOXP2 is expressed in several brain regions, including the basal ganglia, thalamus, cerebellum, cortical plate and inferior olives (Lai et al., 2003). In the developing mouse brain, Foxp2 expression is highest in the basal ganglia (Ferland et al., 2003). These sites of expression coincide with regions shown to function abnormally in affected members of the KE family who suffer from a severe speech and language disorder resulting from an amino acid substitution in FOXP2 (R553H). MRI and PET analyses of carriers of the R553H mutation have revealed structural and functional abnormalities in the caudate nucleus of the basal ganglia (Vargha-Khadem et al., 1998; Watkins et al., 2002b). Mice carrying a "humanised" version of Foxp2, in which the two amino acid changes that have occurred in human FOXP2 but not in chimpanzee FOXP2 have been knocked-in, show an increase in dendritic length in the striatum, cerebral cortex, and thalamus and an increase in long term depression in the medium spiny neurons of the striatum (Enard et al., 2009; Reimers-Kipping et al., 2011). Moreover, mice carrying an R552H Foxp2 amino acid substitution, which is equivalent to the human R553H mutation, show deficits in motor learning and abnormal synaptic plasticity in striatal and cerebellar circuits (Groszer et al., 2008). These findings have led to the idea that FOXP2 and its regulatory targets play a key role in sensorimotor learning, and that the speech and language deficits evident in individuals carrying point mutations in the gene arise from a core deficit in learning, planning and executing complex motor sequences (Fisher and Scharff, 2009; Vargha-Khadem et al., 1998; Watkins et al., 2002a).

In light of this evidence, a key question to address, having identified FOXP2 as a transcriptional repressor of *DISC1*, is can any of the phenotypes associated with disturbed DISC1 function be explained by abnormal function of the basal ganglia circuitry? A review

of the literature revealed few studies explicitly investigating the expression or the role of DISC1 in the basal ganglia. In the mouse, Disc1 is expressed pre- and post-synaptically in striatal synapses (Ramsey et al., 2011) and mice carrying the *Disc1* L100P amino acid substitution show an increase in striatal dopamine (D_2) receptors (Lipina et al., 2010), although whether this is a direct effect of altered striatal Disc1 function is not known. An association between genetic variation in *DISC1* and striatal volume has been suggested by one study; however, the association did not survive a stringent correction for multiple testing (Chakravarty et al., 2012). The observation of disturbed prepulse inhibition and/or latent inhibition in some *Disc1* mouse models (Clapcote et al., 2007; Hikida et al., 2007) suggests that certain alterations to Disc1 function can disrupt sensorimotor gating, a process that depends, in part, on striatal function (Baldan Ramsey et al., 2011; Swerdlow and Geyer, 1998). The evidence, therefore, suggests that DISC1 may play a role in basal ganglia function; however, it is clear that further studies are required before any conclusions can be drawn. Given the repressive regulatory relationship that exists between FOXP2 and *DISC1* and the high level of FOXP2 expression in the basal ganglia, it may be that the maintenance of DISC1 expression at a low level in the basal ganglia is important for normal function. This possibility should be formally assessed by characterising the developmental relationship between FOXP2 and DISC1 expression in the basal ganglia. If the basal ganglia appear to be characterised by a low level of DISC1 expression, then the effects of basal ganglia-specific overexpression of Disc1 could be assessed in a mouse model.

An inverse relationship between the expression of FOXP2 and the expression of its regulatory targets that it represses might be expected. Such a pattern of expression is observed between FOXP2 and one of its transcriptional targets, MET (Mukamel et al., 2011); however, it is likely that, when considering expression in the context of the entire brain, a neat inverse pattern of expression will not exist between FOXP2 and all of its regulatory targets. To some extent, there is evidence of an inverse pattern of expression between DISC1 and FOXP2: in the cortex, the expression of Foxp2 is restricted to layer VI (Ferland et al., 2003), whereas Disc1 expression is highest in layers II/III and IV/V (Schurov et al., 2004). Disc1 is expressed prominently in the hippocampus (Schurov et al., 2004), whereas, as mentioned previously (section 4.4.8.3.2), the evidence for Foxp2 expression in the hippocampus is mixed (Ferland et al., 2003; Fujita et al., 2008) but, when taken in the context of the results presented in this chapter, suggestive of a low level of expression in this region. The basal ganglia are characterised by a particularly high level of Foxp2 expression throughout development (Ferland et al., 2003). As mentioned above, little is known about

Disc1 expression in the basal ganglia relative to other brain regions; however, Disc1 expression has been detected in synaptic fractions from striatal preparations (Ramsey et al., 2011). In contrast, in the cerebellum, the expression of both Foxp2 and Disc1 is predominantly localised to Purkinje neurons (Ferland et al., 2003; Schurov et al., 2004). Further investigation of the cellular and subcellular expression patterns of Foxp2 and Disc1 throughout development and in adulthood would be useful in furthering understanding of the relationship between the two genes.

4.5.2 Assessment of the effect of two FOXP2 coding mutations on the ability of FOXP2 to regulate *DISC1* expression

The R553H and R328X *FOXP2* point mutations, which have previously been implicated in developmental verbal dyspraxia (DVD) (Lai et al., 2001; MacDermot et al., 2005), were both found to reduce the ability of FOXP2 to exert transcriptional repression on *DISC1* promoter activity. Both mutations significantly reduced FOXP2-mediated repression of all three *DISC1* promoter constructs. These findings identify a regulatory link between genes implicated in disorders that have traditionally been considered as diagnostically distinct, thus furthering support for the view that clinically separate neuropsychiatric disorders might share overlapping aetiologies (Carroll and Owen, 2009; Mitchell and Porteous, 2011).

The possibility of shared aetiologies was initially suggested by the partially overlapping behavioural characteristics and cognitive deficits observed in schizophrenia, bipolar disorder, and autism (Carroll and Owen, 2009). Further support came from the finding of overlap in the genetic variants that predispose to these disorders, suggesting that variation in certain genes alters neurological processes whose abnormal functioning results in phenotypes common to multiple conditions (Mitchell and Porteous, 2011).

That there is overlap between the genetic aetiologies of clinically distinct neurodevelopmental disorders is not surprising: these conditions are phenotypically complex, essentially comprising a collection of co-occurring symptoms. It is plausible that a genetic variant predisposing to a particular symptom in one condition might predispose to the same symptom in another condition. Although, it should be noted that a strict relationship between genetic variation and observable phenotypes should not be expected as the relationship between the two will be modified by other genetic, epigenetic, and environmental factors.

With this caveat in mind, it is interesting to consider the areas of phenotypic overlap between the condition resulting from the R328X and R553H *FOXP2* point mutations (Lai et al., 2001; MacDermot et al., 2005) and those conditions, such as schizophrenia, bipolar disorder, and autism, associated with variation in the *DISC1* (Bradshaw and Porteous, 2012; Chubb et al., 2008). The phenotype arising from the FOXP2 point mutations has been described in section 4.1.1.3. Briefly, the core phenotype associated with point mutations in the *FOXP2* gene is speech and language dysfunction. Additionally, members of the KE family affected by the R553H mutation show lowering of non-verbal IQ, although the effects on verbal cognition appear more severe and wide-ranging. To varying extents, cognitive dysfunction and abnormalities in some aspects of language function are observed in autism, schizophrenia, and bipolar disorder (Barch and Ceaser, 2012; Barrett et al., 2009; Carroll and Owen, 2009; Mefford et al., 2012; Owen et al., 2011). Further studies are required to assess whether overlap in the observable phenotypes of these conditions is attributable to the shared disruption of components of the FOXP2 regulatory network. Assessment of DISC1 expression in carriers of the R553H or the R328X point mutations would be useful in determining whether altered DISC1 expression is likely to contribute to the phenotype displayed by individuals carrying these mutations.

Some of the overlap in phenotypes arising from genetic variation in *FOXP2* and *DISC1* might be attributable to the fact that both genes play an important role at the synapse (Brandon and Sawa, 2011; Fisher and Scharff, 2009). Cognition relies on normal synaptic function (Grant, 2003), and synaptic dysfunction has been identified in schizophrenia (Harrison and Weinberger, 2005; Owen et al., 2005), autism (van Spronsen and Hoogenraad, 2010) and bipolar disorder (Martinowich et al., 2009). Both DISC1 and FOXP2 have been implicated in regulating members of the neurexin (NRXN) family (Brown et al., 2011; Vernes et al., 2008). NRXNs are cell adhesion molecules that connect pre- and post-synaptic terminals and regulate synaptic transmission (Sudhof, 2008). FOXP2 has been shown to repress transcription of the NRXN family member *CNTNAP2*, which encodes CASPR2 (Vernes et al., 2008). As discussed in section 4.5.1.1, polymorphisms in *CNTNAP2* have been associated with specific language impairment, autistic spectrum disorder, schizophrenia and bipolar disorder. Recently, our group identified a role for DISC1 in the regulation of NRXN expression: disturbed developmental expression of Nrxn1 and Nrxn3 was observed in the *Disc1* L100P mouse model of schizophrenia (Brown et al., 2011). Genetic variation in *NRXN1* (Feng et al., 2006; Kim et al., 2008; Marshall et al., 2008; Szatmari et al., 2007;

Wisniowiecka-Kowalnik et al., 2011; Yan et al., 2008), and *NRXN2* (Gauthier et al., 2011), has been implicated in autism, and variation in *NRXN1* has been linked to schizophrenia (Gauthier et al., 2011; Kirov et al., 2008; Rujescu et al., 2009; Yue et al., 2011). Moreover, a ChIP-chip study has identified *NRXN3* as a putative transcriptional target of FOXP2 (Vernes et al., 2007).

Taken together, the findings presented in this chapter and the published literature suggest that FOXP2, DISC1, and the NRXN family are linked in a molecular network that, when altered confers risk for neurodevelopmental conditions in which various aspects of linguistic and cognitive function are disturbed, possibly due to aberrant synaptic function. As discussed in section 4.5.1.1, the evidence implicating the FOXP2 regulatory network in basal ganglia function suggests this region to be worthy of investigation in the context of these conditions. Future studies should (i) further characterise the relationship between these genes, for example by determining the regional and temporal specificity of regulatory relationships, and (ii) aim to refine understanding of the phenotypes arising from the altered function of these genes.

4.5.3 Assessment of the expression of Foxp2 in mice carrying the *Disc1* L100P point mutation

The expression of Foxp2 in mice carrying the *Disc1* L100P point mutation was assessed as an initial step in testing the hypothesis that Disc1 is involved in the regulation of Foxp2 expression. This hypothesis was based on the observations that the expression of FOXP2 may be regulated by pathways linked to DISC1 function, including Wnt signalling, and cAMP- and EGR-dependent transcription (section 4.1.1.2). Although the experiments carried out here were preliminary due to limitations in the availability of mRNA and protein samples, several significant differences were detected between the L100P mice and their wildtype counterparts at the mRNA level and investigation of Foxp2 protein expression suggested that differences in expression might occur at this level too.

At the mRNA level, significant differences in expression were observed at E15, E18, P1, P7, and in the adult mice. At all stages except E18, Foxp2 expression was increased in the *Disc1* L100P mice. The decrease in mRNA expression at E18 and the increase in mRNA expression at P1 were supported by consistent changes in protein expression at these stages. At other developmental stages, the mRNA and protein data were less consistent; however, it

is important to note that as the qRT-PCR analysis was carried out using only one reference gene and protein expression assessed in only one technical replicate, the results presented in this chapter must be replicated before any firm conclusions are drawn. Ferland et al. (2003) have previously measured Foxp2 mRNA and protein expression in the developing mouse brain and shown changes in protein expression to sometimes lag behind changes in mRNA expression, as might be expected. It is not immediately obvious that this explanation can explain the discrepancies in mRNA and protein expression observed in this chapter; however, it is a possibility that should be considered if further replicates are obtained.

Consistency between mRNA and protein expression should not necessarily be expected: in a comparison of the mRNA and protein expression profiles of 1066 genes measured in 23 human cell lines, Gry et al. (2009) found the mean correlation between mRNA and protein expression to be 0.25 or 0.20 depending on whether mRNA expression was measured using an cDNA- or oligo-based microarray. Moreover, the correlation coefficients between mRNAs and their corresponding protein products varied widely.

Notwithstanding the need for replication, it seems reasonable to suggest that when consistent changes are observed in Foxp2 protein expression, either in terms of developmental stage or brain region, confidence in the validity of the results is increased. Foxp2 expression at P7 was reduced in all four brain regions assessed (hippocampus, striatum, cortex, and cerebellum), and in the cerebellum expression was reduced at every postnatal stage assessed (P1, P7, P20, and P90). The functional consequences of these changes in expression require further investigation; however, by considering the processes occurring at a given developmental stage and/or in a given region, some insight can be gained into their likely effects. At E18, transcriptomic analysis of the mouse brain has revealed an up-regulation of genes involved in synapse development, neurogenesis, and cell survival and growth (Han et al., 2009; Matsuki et al., 2005). P1 is also a time of synaptogenesis (Matsuki et al., 2005; Mody et al., 2001). Comparison of the transcriptome at E18 and P7 has revealed a decrease in the expression of genes related to neurogenesis and an increase in the expression of genes involved in synaptogenesis (Han et al., 2009). Genes encoding cell signalling molecules, such as adenylate cyclase 1, were also up-regulated at P7 compared to E18. The pattern of changes in the expression of genes encoding neurotransmitter receptors is more complex, although there was a general increase in expression at P7 compared to E18. The *Disc1* L100P mice have been reported to have a decrease in neural proliferation and in the density of dendritic spines, the postsynaptic target of synaptic transmission (Lee et al., 2011). It can

be speculated that altered Foxp2 expression at these developmental stages might contribute to these abnormalities; however, this theory clearly requires experimental validation and refinement. It would also be useful to carry out electrophysiological assessment of the *Disc1* L100P mice to assess synaptic function.

As Foxp2 generally confers a repressive effect on its regulatory targets (Shu et al., 2001), the decrease in Foxp2 expression in the cerebellum detected at all postnatal stages might result in a more permissive transcriptional environment. Given the evidence for the involvement of Foxp2 in regulating neurite outgrowth and synaptogenesis, decreased Foxp2 expression might result in over-connectivity or aberrant connectivity in this region; however, this hypothesis clearly requires experimental assessment. There are currently no reports of cerebellar morphology or function in the *Disc1* L100P mice; however, in light of the current results, such studies would be warranted. As the cerebellum is involved in both motor and cognitive functions (Middleton and Strick, 2000), the abnormal function of this region could have wide-spread consequences.

Comparison of the Foxp2 and Disc1 mRNA expression profiles between the *Disc1* L100P and wildtype mice revealed an interesting similarity in the shape of the expression profiles of the two genes in wildtype mice but not in *Disc1* L100P mice. Having demonstrated that Foxp2 negatively regulates the transcription and protein expression of Disc1, this result might appear surprising; however, it is, of course, the case that several factors will be involved in the regulation of Disc1 expression, and perhaps, therefore, a direct inverse correlation with Foxp2 expression should not be expected. It is also the case that, as it is Foxp2 protein that regulates the transcription of the *Disc1* gene, evidence for a negative correlation between Foxp2 and Disc1 expression might be more obvious when considering Foxp2 protein expression and Disc1 mRNA expression. Intriguingly, in the *Disc1* L100P mice, the developmental expression patterns of Foxp2 and Disc1 show almost inverse patterns of expression between developmental stages E13 and P1. Future studies should aim to establish whether it is the *Disc1* L100P mutation *per se* or whether it is the dysregulation of Disc1 expression that the L100P mutation entails (Brown et al., unpublished) that results in the aberrant expression of Foxp2 in these mice. These possibilities could be distinguished *in vitro* by overexpressing wildtype and L100P *Disc1* in a cell culture system and assessing Foxp2 expression.

4.5.3.1 Caveats

A significant caveat of the experiments discussed in this section is that as limited amounts of mRNA and protein were available from the mice, mRNA expression could only be normalised to one reference gene and western blotting was only carried out in one experimental replicate. As such, the results presented here require replication.

Comparison of the Foxp2 mRNA developmental expression profile observed in the wildtype animals in this study with a previous study reveals only partial agreement. Here, the highest level of Foxp2 expression was observed at E13. Expression then decreased between E13 and E15, before increasing again between E15 and E18 and decreasing throughout postnatal development. In contrast, Lai et al. (2003) found an increase in the expression of Foxp2 in the mouse brain at E16.5 compared to E11.5, E15.5 and newborn animals. Differences in these findings might be attributable to differences in the method used (qRT-PCR vs. in situ hybridisation) or strain differences in the mice studied. It is also possible that as only one reference gene, which had not been assessed for the stability of its expression, was used to normalise Foxp2 expression values that the developmental profile of Foxp2 expression was influenced by a poor choice of reference gene. Future studies should measure the expression of a large number of reference genes in the *Disc1* L100P mice and wildtype controls to determine a stably expressed subset (as recommended by Vandesompele et al., 2002) that can be used for future qRT-PCR experiments.

The second immunoreactive band that was sometimes evident on western blots probed for Foxp2 expression was not expected given the information available on the manufacturer's datasheet. A review of the literature identified only a few studies in which Foxp2 had been detected in mice by western blotting. Of these studies, one contains an image of a western blot (of cerebellar lysates) in which a second band similar to that observed here is visible (Fatemi et al., 2008a). Interestingly, this study used the same C-terminal Foxp2 antibody as used here. Foxp2 has also previously been detected as a double band in the striatum, nidopallium and area X of the zebra finch brain (Miller et al., 2008). The manufacturer's datasheet for the Foxp2 antibody used here (http://www.abcam.com/FOXP2-antibody-ab16046.htmlP) states that additional bands have been identified at 56 kDa and 70 kDa but that the identity of these bands is not known. One possibility is that the second band represents an alternatively spliced Foxp2 isoform. Using a different C-terminal Foxp2 antibody to that used here, Tanabe et al. (2012) identified two Foxp2 isoforms of 80 kDa and

68 kDa in western blots of mouse cerebellum at P0 and P10. Similar to observations in this chapter, the lower molecular weight band decreased in signal intensity with development and was no longer evident at P15. Both bands disappeared when the Foxp2 antibody was pre-absorbed with a specific peptide. A Foxp2 band of approximately 68 kDa corresponding to Foxp2 isoform III has been observed previously (Vernes et al., 2006). This isoform lacks the N-terminal domain, which is consistent with the fact that Tanabe et al. (2012) did not detect the 68 kDa band using an N-terminal antibody. As members of the Foxp family share a high level of homology at the amino acid level (Lu et al., 2002; Teufel et al., 2003), another possibility is that the second band might result from cross-reactivity with Foxp1, which is detected as a protein of 75 kDa, or Foxp4, which is detected as a protein of 73 kDa. Further experiments, are required to confirm the identity of the second band. It would be useful to pre-absorb the antibody with a blocking peptide to determine whether the lower molecular weight band corresponds to Foxp2. Additionally, it would be useful to determine whether this antibody detects products in samples where Foxp2 has been knocked down.

The western blotting performed in this chapter was clearly affected by certain technical issues, as indicated by the partially contradictory results produced by technical replicates of the hippocampal and striatal blots. Disagreement between technical replicates could be caused by several factors. For example, it is possible that the compositions of the samples loaded onto the two gels differed. The samples were mixed prior to loading on the gel but it is possible that more mixing would have reduced the variation between replicates. Other factors that may have contributed include the temperature at which the antibody was used, differences between individual gels and between polyvinylidene fluoride (PVDF) membranes, differences in electrophoresis conditions, transfer and/or developing conditions, and the interval between performing the western blot and assessing protein expression. Attempts were made to standardise all of these variables; however, evidently, there were between-replicate differences. This variability highlights the need for further replicates to be performed.

Another issue affecting the western blots carried out here was the inconsistency between the expression levels of the three reference genes measured. Initially, it was intended that Foxp2 optical density values would be normalised to the geometric mean of the optical density values for the reference genes, a technique recommended for the normalisation of qRT-PCR results (Vandesompele et al., 2002). However, the clear inconsistencies between the expression levels of the reference genes, which in some incidences were so extreme that

certain reference genes were unquantifiable, meant that normalising to each reference gene separately was considered a more informative approach. This issue highlights the importance of (i) selecting reference genes unlikely to be affected by the experimental condition and (ii) confirming western blot data using multiple reference genes (or, complimentary approaches).

In the first technical replicates of the western blots of the embryonic samples and the postnatal hippocampal samples, relative to the expression of α-tubulin and β-actin, the expression of Gapdh appeared to be reduced in the *Disc1* L100P mice. As this pattern was not observed in the second technical replicate of the embryonic samples, and in the absence of experimental replicates, it is difficult to know to what extent this effect was caused by technical factors. Clearly, further experiments, perhaps using qRT-PCR or ELISA, are required to assess Gapdh expression in the *Disc1* L100P mice.

It is important to note that western blotting is a semi-quantitative technique. Even in a well-controlled experiment, many variables can affect the signal intensities of the observed immunoreactive bands. As such, where accurate protein quantification is required, a more quantitative technique, such as ELISA or mass spectrometry, should be employed.

A final caveat, affecting both the qRT-PCR and western blotting performed in this chapter, is the cellular heterogeneity of the samples assessed. Measurement of gene expression in embryonic animals was carried out using RNA or protein extracted from the whole brain, while gene expression in postnatal animals was carried out in the hippocampus. Thus, at each developmental stage, recorded gene expression levels represent the combined effects of multiple different cell types. It is, therefore, impossible to determine whether the observed changes in gene expression (i) are representative of the entire population of cells sampled or (ii) represent an averaging of changes of gene expression of different magnitudes and, potentially, directions occurring in different cell types. Moreover, implicit in the comparison of gene expression changes across the whole brain or hippocampus between genotypes is the assumption that the cellular composition of these gross anatomical structures is identical between genotypes at each developmental stage. This assumption may not be justified: *Disc1* L100P mice have been found to demonstrate reduced cortical neuronal proliferation at E12 and E15 and a reduction in neocortical neuron number in adulthood (Lee et al., 2011). In order to obtain a more informative depiction of the expression of genes of interest in the *Disc1* L100P mice, methods that enable the study of discrete populations of cells, such as laser-capture microdissection, should be employed. Measurements of gene expression in the

adult hippocampus are affected by the additional potential problem of variability in the accuracy of hippocampal dissection. Particularly at the early postnatal stages, it is possible that inaccuracies in hippocampal dissection, and, thus, the cellular composition of the samples, might result in spurious differences in gene expression. Again, the study of isolated cellular populations would aid in mitigating this confound.

4.5.4 Conclusions

To conclude, DISC1 protein expression and transcriptional activity appear to be negatively regulated by FOXP2. The findings for the assessment of the effect of FOXP2 on DISC1 mRNA expression were less clear cut; however, they suggest that the overexpression of FOXP2 results in a reduction in the expression of at least some DISC1 transcripts. Further experiments are required to clarify the effect of FOXP2 on DISC1 mRNA expression. The discovery that two FOXP2 point mutations known to cause developmental verbal dyspraxia reduce FOXP2-mediated repression of DISC1 promoter activity expands the list of neurodevelopmental disorders in which dysregulated DISC1 expression may play a part. Profiling of DISC1 expression in individuals carrying these mutations would now be of great interest. Assessment of Foxp2 expression in the *Disc1* L100P mice provided preliminary evidence for dysregulated expression at the mRNA and protein levels, although further experiments are required both to confirm these findings and to investigate whether any changes are the result of altered Disc1 expression or of the L100P mutation. Taken together, the results presented in this chapter link DISC1 and FOXP2, two genes known to play important roles in neurodevelopment, in a regulatory network, which when disturbed confers risk for a range of disorders that share varying degrees of phenotypic overlap. Future studies should aim to (i) elucidate the molecular and cellular consequences of the altered expression of these genes and (ii) characterise their expression in well-phenotyped samples, in order to investigate the relationship between gene expression and phenotypes relevant to neurodevelopmental disorders.

Chapter 5

Discussion

Chapter 5: Discussion

The aim of the work presented in this thesis was to investigate putative genetic risk loci for psychiatric illness, thought to harbour variants potentially exerting a pathogenic effect via altered regulation of gene expression. This work was predicated on the growing evidence for the role of abnormal gene expression in the aetiology of psychiatric illness. Altered gene expression has received support as a candidate mechanism from several lines of evidence, including the frequent implication of variants in non-coding regions of the genome by association studies, evidence for altered gene expression in affected individuals, and the observed changes in the function of regulatory mechanisms, such as epigenetic modification and miRNA-mediated regulation, in patients. The focus of this thesis has been on *NRG1* and *DISC1*, two of the leading candidate genes for schizophrenia, which have also been implicated in bipolar disorder. Here, the findings presented in the previous chapters will be summarised, their potential limitations considered and suggestions made for future experiments. The relevance of the work presented in this thesis to the wider field will then be considered.

5.1.1 Chapter 2: Association analysis of *NRG1*

5.1.2 Summary of findings

In Chapter 2 association analysis was carried out to assess the involvement of two haplotypes in the *NRG1* gene in schizophrenia and bipolar disorder. These two haplotypes had previously been found to be associated with schizophrenia and/or bipolar disorder in a sample from the Scottish population (Thomson et al., 2007). One haplotype, the region A haplotype, located at the 5' end of the gene, comprised three intronic SNPs and the associated region, defined by linkage disequilibrium (LD), overlapped with the putative promoter region and extended into the first intron. The second haplotype, the region B haplotype also comprised three intronic SNPs and spanned a largely intronic region at the 3' end of the gene, although it did also encompass the coding region of the SMDF NRG1 isoform and 3' exons of all other isoforms.

Here, a two-SNP haplotype in region B showed significant association with both schizophrenia and a combined schizophrenia and bipolar disorder case group in a sample formed from combining Thomson et al.'s (2007) Scottish sample (Scottish 1) with a second Scottish sample (Scottish 2). This combined sample comprised 765 control subjects, 682

schizophrenic patients and 601 bipolar disorder patients. Assessment of association to the three-SNP region B haplotype was precluded by the failure of one of the SNPs in this region to genotype in the Scottish 2 sample; however, assessment of LD in the Scottish 1 sample suggested that little information was lost by the absence of the third SNP.

Association analysis of the Scottish 2 sample (n = 307 control subjects, 303 schizophrenic patients, and 239 bipolar disorder patients) independent of the Scottish 1 sample did not yield any significant associations. Likewise, association analysis of the three region B SNPs in a German sample (n = 397 control subjects, 396 schizophrenic patients, and 400 bipolar disorder patients), in which association was assessed with schizophrenia, bipolar disorder, and the combined schizophrenia and bipolar disorder case group, did not identify any significant associations following correction for multiple testing.

Finally, no significant associations were detected in a combined sample that comprised the Scottish 1 sample, the Scottish 2 sample and the German sample.

5.1.3 Caveats

The main caveats of the association analyses carried out in this chapter are the lack of certainty over the underlying genetic architecture of schizophrenia and bipolar disorder and the issue of phenotypic heterogeneity. These issues make the suitability of association analysis as a method for detecting the variants involved in psychiatric disorders unclear.

Until the relative contributions of common vs. rare variants to the aetiology of psychiatric illness are known, the possibility that association analysis might not be the most suitable method for the detection of causal variants must be considered when interpreting the findings of association studies. Thus, failure to detect association must not be treated as evidence for the absence of causative variants. Likewise, the meaning of significant associations is difficult to ascertain as several possible explanations for their occurrence exist: such findings might reflect the existence of a variant(s) that confers increased risk for the disorder of interest at the population level; they might indicate the presence of a variant(s) that confers increased risk but only in a few individuals, which was, therefore, only detected in a particular sample due to random sampling bias; or they might have arisen entirely due to chance. Replication of association findings can assist in determining whether random sampling error is likely to have affected an initial finding; it cannot, however, rule out the possibility that there are causal variants in a given region.

Integrating association data with functional information can aid in identifying variants likely to play a causal role. Such variants can then be prioritised for further assessment. Recent years have seen significant advances in the functional annotation of the genome, particularly by the ENCODE Project (Bernstein et al., 2012), this information will permit variants implicated by association studies, including variants in LD with the associated variants, to be assessed for their functional potential. Genome-wide annotation of features such as chromatin states, transcription factor bound regions, and areas of DNase hypersensitivity will greatly aid in the identification of potential aetiological variants in non-coding regions. The potential for the study of DNase1 hypersensitivity sites (DHS) to yield insights into the functional effects of associated variants, and those implicated by LD, was recently demonstrated by Maurano et al. (2012). Here, Maurano et al. (2012) showed that variants located in DHSs may exert their phenotype via an effect on distant regulatory targets and are frequently predicted to alter transcription factor binding and chromatin state.

Several tools are available online for the assessment of the functional potential of variants. One such tool, HaploReg, permits the user to search a set of haplotype-tagging variants for various features, including their chromatin state in nine cell lines, conservation across mammals and predicted effect on regulatory motifs (Ward and Kellis, 2012). The assignment of predicted functions to associated variants can, to some extent, mitigate some of the issues affecting association studies. For example: should multiple variants in a given gene showing inconsistent association between studies be predicted to affect the same function (e.g. splicing or transcription factor binding), then, despite the lack of direct between-study replication, insight can be gained into likely pathogenic mechanisms.

If phenotypic heterogeneity reflects heterogeneity in aetiology, it is likely to represent a major limiting factor for analyses of schizophrenia and bipolar disorder carried out using diagnostic category as the sole method of case selection. As such, the lack of significant associations in the Scottish 2 sample and the German sample might have reflected differences in the composition of the case groups. The selection of cases using endophenotypes is a popular approach to limiting the extent of phenotypic heterogeneity. However, while the use of endophenotypes in the field of psychiatric genetics has received increasing attention in recent years, there is a lack of consensus regarding how best to define an endophenotype (Kendler and Neale, 2010). One definition states that endophenotypes should be heritable, co-segregate with the illness, be present in the absence of the disease and

be observed in family members at a higher rate than in the population (Gottesman and Gould, 2003). Various other criteria have been suggested, including the requirement that the endophenotype should be a causal factor for the disease (Lavori et al., 2002) or should be involved in a biologically plausible mechanism (Castellanos and Tannock, 2002; Tsuang et al., 1993). In a review of the subject, Kendler and Neale (2010) suggest that interest in the study of endophenotypes arises from two factors: (i) endophenotypes appear aetiologically simpler than diagnostic categories and (ii) endophenotypes appear to be "closer" to changes at the DNA level. Both of these factors suggest that endophenotypes will be more genetically tractable than their related diagnostic categories. In practice, however, association studies of psychiatric endophenotypes have rarely yielded effect sizes any larger than when assessing association to the diagnostic category itself (Flint and Munafo, 2007).

GWASs have identified association to variants associated with several endophenotypes relevant to psychiatric illness, including resting state electroencephalogram activity (Hodgkinson et al., 2010), hippocampal volume (Bis et al., 2012; Stein et al., 2012), and neurocognitive function (LeBlanc et al., 2012); although an earlier study failed to identify any variants showing genome-wide significant association to neurocognitive function (Need et al., 2009). Stein et al.'s (2012) meta-analysis of GWASs of hippocampal volume identified one variant, significant at the genome-wide level, which, when analysed under several different models, explained no more than 0.458% of the variance in hippocampal volume. A meta-analysis of three GWASs assessing association to speed of information processing failed to identify any variants attaining genome-wide significance, although several nominally significant variants were detected (Luciano et al., 2011).

These findings of these studies, while not representing an exhaustive discussion of the literature, are consistent with those of Flint and Munafò, (2007) and suggest that the study of endophenotypes will not necessarily circumvent the problems associated with the analysis of diagnostic categories. This might be, in part, because many of the endophenotypes studied are, in fact, no simpler in terms of their genetic architecture than their related diagnostic categories. Flint and Munafò (2007) suggest that the use of simpler endophenotypes, such as gene expression levels, may yield the desired advantages in genetic tractability. The potential of this approach to uncover genetic loci harbouring variants involved in the aetiology of schizophrenia has been demonstrated by Bray et al. (2008). By performing genome-wide linkage analysis on the expression level of DTNBP1, a schizophrenia susceptibility gene previously found to show dysregulated expression in patients, Bray et al. (2008) identified

evidence for loci involved in *cis*- and *trans*-regulation. Identifying the variants underlying the observed linkage peaks will lead to the identification of schizophrenia candidate genes for further assessment. Moreover, Richards et al.'s (2012) demonstration that schizophrenia susceptibility alleles are enriched for variants known to affect gene expression in the adult brain supports the use of gene expression as a schizophrenia endophenotype.

5.1.4 Suggestions for future work

The lack of clarity surrounding the validity, and therefore the interpretation, of association studies for psychiatric illness suggests the sequencing of regulatory and coding regions of the *NRG1* locus might be a worthwhile pursuit in determining the existence and the nature of risk variants at this locus. The recent functional annotation of the genome by the ENCODE project (Bernstein et al., 2012) and Maurano et al.'s (2012) characterisation of DHSs are likely to facilitate understanding of associated variants in non-coding regions of the genome.

5.2 Chapters 3: Characterisation of the *DISC1* promoter region

5.2.1 Summary of findings

Chapter 3 described work carried out to identify and characterise the previously unstudied *DISC1* promoter region. Here, the *DISC1* candidate promoter region was initially defined as the region extending 1 kb upstream of the *DISC1* transcription start site (TSS) to the translation start site. Bioinformatic analysis of this region revealed it to be devoid of canonical promoter motifs, but to contain a CpG island and a complex dinucleotide repeat region. Assessment of the potential for the region to form Z-DNA, a more open DNA conformation associated with transcriptional activity, revealed the *DISC1* candidate promoter region to have high Z-DNA forming potential.

An initial series of three nested *DISC1* promoter fragments (short, medium and long) was designed for assessment for regulatory potential using dual luciferase reporter assay. These fragments were selected to taking into account sequence features of the DISC1 candidate promoter region (i.e. the CpG island and the dinucleotide repeat region), conserved regions of high regulatory potential identified by the ESPERR 7 x regulatory potential track of the UCSC human genome browser, and epigenetic modifications identified by ChIP studies. The findings of a study carried out as part of the ENCODE project in which the promoter regions of a large number of genes were characterised were also considered. This study found that

the region located -300 to -50 bp relative to the TSS usually contributes positively to promoter activity, while the region located -1000 bp to -500 bp relative to the TSS often confers a repressive effect (Cooper et al., 2006).

Assessment of the promoter activity of the three *DISC1* promoter constructs revealed the medium construct to have the highest level promoter activity in all cell lines assessed. This region corresponds with the region identified by the ENCODE project as usually contributing positively to promoter activity (Cooper et al., 2006). The observation that the long promoter construct, which extends approximately 1kb upstream of the *DISC1* TSS, resulted in a reduced level of reporter gene expression compared to the medium construct was, again, consistent with the findings of ENCODE (Cooper et al., 2006), and suggested the existence of repressive elements in this region. The assessment of a fourth promoter construct designed to assess the contribution of the complex dinucleotide repeat region to the activity of the medium promoter construct suggested that this region did not underlie the high activity level of this construct; although, the caveat that the method employed here would not have identified any effects dependent on the genomic context should be noted.

Assessment of publicly available ChIP-seq (Rosenbloom et al., 2010) data to identify transcription factors identified as potentially binding sites within the DISC1 candidate promoter region identified several interesting candidates for experimental follow-up. These included FOXP2, TCF4 (TCF7L2), NF-KB and PU.1. The potential role of FOXP2 in the regulation of DISC1 expression was considered particularly interesting in light of FOXP2's involvement in speech and language function (Fisher and Scharff, 2009), suggested role in neurodevelopment (Vernes et al., 2011) and link, via its regulatory targets, with autism (Mukamel et al., 2011; Vernes et al., 2008). As such, the region was searched for known FOXP2 binding sites, leading to the identification of 11 putative binding sites.

5.2.2 Caveats

A key limitation work presented in this chapter is the artificial nature of performing dual luciferase reporter assays to assess the activity of transiently transfected reporter constructs. By studying a regulatory region outside of its genomic context, some regulatory effects are likely to remain unidentified, and those that are observed might be an artefact of the method. Furthermore, transient transfection of the reporter construct results in hundreds or even thousands of copies of the construct being present in a cell. As such, a shortage of transcription factors relative to constructs might result in some constructs being regulated by only a subset of the transcription factors that should regulate them (Carey and Smale, 2000).

Regarding the experiments performed in this chapter, the observation of replicable results between technical and experimental replicates, and in different cell lines, suggests that random effects resulting from the scarcity of transcription factors relative to reporter constructs are unlikely to have exerted a large effect on the observed results. Nevertheless, confirmation of results obtained using dual luciferase reporter assays using other, complimentary, methods should be sought.

An alternative approach to the assessment of reporter constructs that overcomes some problems of transient transfection has been described by Karimi et al. (2007). This approach uses the FLP-FRT recombinase system to generate cell lines in which the reporter construct has been integrated at the same genomic locus, in the same orientation in all of the cells in a given culture. As each cell only expresses one copy of the reporter construct, there is no need to normalise for transfection efficiency, thus avoiding one potential source of random variation. A further advantage of this method is that the reporter construct is assessed in the context of chromatin. Chromatin-dependent events, which would have been missed using transient transfection, are, therefore, potentially observable. A caveat of this approach is that the observed effects may be dependent on the specific chromatin environment at the site of integration and, therefore, it is recommended that assessments are carried out using multiple cell lines with FRT sites in different genomic locations (Karimi et al., 2007).

The used of immortalised cell lines represents an additional limitation. These cell lines often have an abnormal karyotype: HEK293 cells, for example carry three copies of the X chromosome and a modal number of 64 chromosomes per cell (European Collection of Cell Cultures datasheet: http://www.hpacultures.org.uk/products/celllines/generalcell/detail.jsp?refId=85120602&collection=ecacc_gc), while SH-SY5Y cells are affected by several chromosomal rearrangements including a complex duplication of the long arm of one chromosome 1 (Spengler et al., 2002). Clearly, karyotypic abnormalities detract from the physiological validity of a cell-based assay; however, while not representing an ideal solution, replication of findings using multiple cell lines with different karyotype aberrations is likely to lend credence to the results. As such, it would appear fair to conclude that, having observed the same relative pattern of luciferase activity from the short, medium and long *DISC1* promoter constructs in three different cell lines (albeit with non-significant results in the LAN-5 cells), it is likely that the observed pattern of results was not unduly influenced by any particular karyotype, or, more generally, cellular environment.

5.2.3 Suggestions for future work

The work presented in this chapter represents only a preliminary characterisation of the mechanisms involved in regulating the expression of *DISC1* and there are many further experiments that could be carried out to further understanding of this process. In light of the fact that the repeat length of the dinucleotide repeat region is known to be polymorphic (Hayesmoore et al., 2008), it would be usefully to further characterise its role. This could be achieved by generating a series of reporter constructs in which the number of dinucleotide repeats is varied. To negate the potential confounds associated with transient transfection, these reporter constructs could be assessed using the FLP-FRT site-specific integration system described above (Karimi et al., 2007). Assessment of the association between dinucleotide repeat length and DISC1 expression level in humans would provide a further assessment of the role of this region. It would also be of interest to assess the effect of the duplication of this region identified by Hayesmoore et al. (2008) on the transcriptional activity of the *DISC1* promoter region in reporter assays. Additionally, as a growing number of whole-genome sequencing projects are performed, the assessment of the effect of *DISC1* promoter sequence variants, including those implicated in psychiatric disorders, on the transcriptional activity of the promoter constructs would be informative.

Analysis of the findings of ChIP studies suggested several interesting factors for experimental follow-up. Of particular interest is the transcription factor TCF4 (TCF7L2), a member of the Wnt-responsive family of transcription factors, which has been implicated in schizophrenia (Alkelai et al., 2012; Hansen et al., 2011). Thus, it would be of great interest to (i) experimentally validate the role of TCF4 in regulating *DISC1* transcription and to assess the nature of any regulatory relationship, (ii) characterise the effects of Wnt-signalling through TCF4 on DISC1 expression and (iii) assess the effects of drugs, such as lithium, known to affect components of the Wnt-signalling pathway.

In more general terms, it would be useful to characterise other mechanisms involved in determining the level at which the DISC1 protein is ultimately expressed. For example, the growing evidence for the role of miRNAs in schizophrenia and bipolar disorder (Beveridge and Cairns, 2012; Forero et al., 2010) highlights the identification miRNAs involved in the regulation of DISC1 expression as a worthwhile line of enquiry.

5.3 Chapter 4: Assessment of the FOXP2-DISC1 Regulatory relationship

5.3.1 Summary of findings

In Chapter 4 the regulatory relationship between DISC1 and FOXP2 was explored: the potential role of FOXP2 in regulating transcription from the *DISC1* promoter was investigated and preliminary assessment of a putative role for DISC1 in the regulation of FOXP2 was carried out.

Transient overexpression of FOXP2 in HEK293 cells resulted in reduced DISC1 protein expression. A decrease in DISC1 mRNA expression was also detected when using one TaqMan gene expression assay but not another. Assessment of the promoter activity of the short, medium, and long DISC1 promoter constructs in response to the transient overexpression of FOXP2 revealed a decrease in the activity of all three constructs (although this decrease failed to attain statistical significance for the short construct). FOXP2 constructs carrying non-synonymous point mutations identified in individuals affected by developmental verbal dyspraxia showed a reduced ability to repress transcription from the *DISC1* promoter compared to wildtype FOXP2.

The *Disc1* L100P mouse, a mouse model of schizophrenia (Clapcote et al., 2007), was used to carry out a preliminary assessment of the putative bi-directional relationship between FOXP2 and DISC1 expression. Assessment of Disc1 mRNA and protein expression revealed evidence suggestive of altered expression. The strongest evidence for consistent changes in Disc1 mRNA and protein expression was observed at P7, P1, and at E18/E17.5 (mRNA expression was measured at E18, protein expression was measured at E17). At P1 and P7, expression was increased in the *Disc1* L100P mouse, while at E18/E17.5, the evidence suggested a decrease in expression. Significant changes in mRNA expression were also observed at E15 and in the adult mouse, when expression was increased in the *Disc1* L100P mouse; however, the nature of these changes at the protein level was unclear, either due to inconsistency between the results obtained when normalising to different reference genes, inconsistency between technical replicates, or poor quality western blots precluding assessment.

5.3.2 Caveats

The caveats of the work performed in Chapter 3 (section 5.2.2) also apply to the dual luciferase reporter assays performed in this chapter. Similarly, the limitations associated with the use of transient transfection discussed previously with respect to the dual luciferase reporter assays also apply to the assessments of FOXP2 overexpression on DISC1 mRNA and protein expression performed in this chapter. The transient overexpression of FOXP2 would have resulted in FOXP2 being expressed at a level far exceeding that which occurs under normal physiological circumstances. As such, it is the fact that FOXP2 was shown to down-regulate DISC1 expression and promoter activity, rather than the extent of this down-regulation, which is of interest. While each of the methods employed in assessing the effect of FOXP2 on DISC1 expression suffers from some limitations, the fact that the different approaches yielded compatible results (with the exception of DISC1 expression measured using one TaqMan assay) suggests that a true effect was observed.

The work carried out in the *Disc1* L100P mice to assess the putative role of Disc1 in the regulation of Foxp2 expression was affected by several limitations that mean the work must be considered preliminary in nature and in need of replication. Perhaps the most significant problem affecting these experiments was the limited nature of the samples available from the *Disc1* L100P mice. Unfortunately, as I did not have access to these mice, the only RNA and cDNA samples available were those that had been collected by a previous PhD student (Sarah Brown). Protein samples were donated by Ellen Grünewald but, as she required protein for her own experiments, she could only spare limited amounts. The scarcity of the RNA and cDNA samples meant that only one reference gene could be used to normalise the qRT-PCR results obtained from these mice. Valid normalisation of qRT-PCR data is best performed by normalising to the geometric mean of multiple reference genes, which have, ideally been selected on the basis of their stability under the experimental conditions from a larger pool of reference genes (Vandesompele et al., 2002). As such, before pursuing the qRT-PCR results obtained in this chapter, the experiment should be repeated using additional reference genes. As assessment at the protein level was only carried out in one experimental replicate, further replicates should be performed. Additionally, a more quantitative technique for protein measure should be sought, particularly in light of the inconsistencies observed between the intensities of the immunoreactive bands for different reference genes. The interpretation of the findings arising from these experiments is also limited by the fact that altered gene expression in the *Disc1* L100P mice could either result from the L100P mutation, or the change in Disc1 expression that it confers (Brown et al., unpublished). As

such, it is imperative that, should the findings presented here be replicated, future experiments identify the mechanism underlying the altered expression of Foxp2. Finally, as Foxp2 expression profiling was carried out using RNA or protein derived from either the whole brain or the hippocampus, it is impossible to determine whether the change in expression is localised to any particular cell type(s). Moreover, it is possible that between-genotype differences in the cellular composition of samples might have confounded the results observed in this chapter. These limitations highlight the need for cell-type specific assessment of Foxp2 expression in the *Disc1* L100P mice, using a technique such as laser-capture microdissection.

5.3.3 Suggestions for future work

The observation that the transient overexpression of FOXP2 resulted in the downregulation of DISC1 mRNA expression when measured with one TaqMan gene expression assay but not another warrants further investigation. Examining the qRT-PCR products by gel electrophoresis and sequencing would be an informative next step in determining the reason behind these apparently conflicting results. If these analyses yield results suggestive of the use of alternative transcription start sites, it would be useful to perform 5' RACE in an attempt to identify any additional transcription start sites. As HEK293 cells, the cells used for the qRT-PCR analysis, have an unusual karyotype, repeating the experiment in other cell lines, including more neuronal-like cell lines, would aid in determining whether the apparently conflicting results are cell-line dependent. When assessing the existence of alternative transcription start sites, it would be useful to assess tissue from multiple brain regions from several developmental stages, as Nakata et al. (2009) have shown the expression of DISC1 isoforms to be developmentally regulated. An additional approach to assessing the role of FOXP2 in regulating DISC1 expression would be to measure DISC1 expression in cells following siRNA- or shRNA-mediated knockdown of FOXP2. This would offer a complimentary approach to the overexpression experiments performed here.

Assessment of Disc1 expression in genetic mouse models in which Foxp2 expression is either decreased or increased would represent a useful addition to the experiments performed for this thesis. This approach would negate the issues associated with the artificial nature of *in vitro* assays discussed above; however, it would be significantly more time-consuming and costly. A further advantage to this approach is the opportunity to carry out assessments at the behavioural level; however, particularly when studying phenotypically complex conditions, such as psychiatric disorders, extrapolation from mouse behaviour to human

phenotypes must, clearly, be carried out with caution. More generally, any results obtained in mice come with the caveat that the extent to which they can be generalised to humans is uncertain. This caveat may be of particular relevance to psychiatric genetics: there is evidence that many candidate genes for schizophrenia have been subject to recent selective sweeps (Crespi et al., 2007). Thus, the potential limitations associated with the use of mouse models highlight the need to carry out research into psychiatric disorders using multiple, complimentary methods, including bioinformatic, *in vitro*, and *in vivo* approaches, as well as studies of the post-mortem human brain.

The measurement of DISC1 expression in individuals known to carry either of the two *FOXP2* point mutations, R553H and R328X, shown to reduce FOXP2's ability to regulate DISC1 expression would provide an indication of whether these mutations exert an effect on DISC1 expression in a physiological context. This would provide an indication of whether altered DISC1 expression is likely to contribute to the phenotype observed in individuals carrying either of these mutations. Moreover, the developmental nature of the phenotypes resulting from these two mutations suggests that, ideally, DISC1 expression should be measured in these individuals at various stages throughout development.

As mentioned previously, the experiments carried out to assess the role of DISC1 in the regulation of FOXP2 were very preliminary. In addition to the replication of the experiments performed in this thesis recommended previously, it would be useful to attempt to delineate the effects of the L100P mutation from the change in Disc1 expression it confers (Brown et al., unpublished). To address the question of whether DISC1 is involved in the regulation of FOXP2 expression in a more direct fashion, overexpression and knockdown of DISC1 could be performed *in vitro* and the effects on FOXP2 mRNA and protein expression assessed. Additionally, to obtain a better understanding of the relationship between Foxp2 and Disc1 expression in a physiological context, Foxp2 expression could be assessed in mouse models in which the expression of wildtype Disc1 has been manipulated. Inducible and region- or cell-type-specific manipulation of Disc1 expression would be particularly valuable in assessing the effects of developmental stage, developmental history, and brain region on any observed regulatory relationship.

More generally, in light of the strong evidence for the regulatory link between FOXP2 and DISC1, future studies should focus on: (i) identifying and experimentally validating other members of this regulatory network, (ii) characterising the expression of the genes involved

in the network in individuals with psychiatric disorders, (iii) investigating the role of DISC1 in physiological functions suggested by its identification as a transcriptional target of FOXP2, and (iv) determining how drugs commonly used in the treatment of these conditions affect the absolute and relative expression levels of network members.

5.4 Relevance of this thesis to the wider field

The work presented in this thesis sought to aid understanding of the role of aberrant gene regulation in the pathogenesis of psychiatric illness. This aim was inspired by the growing evidence for the role of altered gene expression in both schizophrenia and bipolar disorder. Transcriptional profiling of the brains of schizophrenic patients has revealed consistent changes in the expression of genes involved in several functions including myelination and oligodendrocyte function, GABA and glutamate transmission, synaptic plasticity, mitochondrial function, and immune- and stress-response (Sequeira et al., 2012). Using a polygenic score approach first described by Purcell et al. (2009), Richards et al. (2012) demonstrated that variants conferring a risk of schizophrenia are enriched amongst those affecting gene expression. Altered gene expression has also been reported in patients with bipolar disorder, with affected functions including myelination, mitochondrial function and immune function (Konradi et al., 2012). In keeping with the observed overlaps between risk variants for the two disorders, common changes in gene expression have been identified (Shao and Vawter, 2008).

The likely involvement of altered gene expression in the pathogenesis of schizophrenia and bipolar disorder highlights the investigation of mechanisms involved in the regulation of candidate genes as a useful endeavour in understanding the genetic bases of these conditions. The work carried out in this thesis highlights the important insights that can be gained from studying mechanisms involved in gene regulation. Consideration of candidate genes in the context of their regulatory networks is more informative than considering a gene in isolation. Knowing how a gene is regulated and what it regulates allows (i) predictions to be made about pathogenic mechanisms, (ii) the identification of plausible candidate genes and interesting gene x gene interactions for further assessment, and (iii) the identification of potential drug targets that can either prevent or rectify the pathogenic consequences of a given mutation. For example, one hypothesis that could be derived from the work presented here is that some phenotypes resulting from *DISC1* variants and/or changes in DISC1 expression levels will be due to altered basal ganglia function. This hypothesis is based on the fact that the FOXP2 network plays a key role in synaptic plasticity at cortico-basal

ganglia synapses, a process necessary for sensory-guided motor learning (Scharff and Petri, 2011). Taken together with the observations of altered sensorimotor gating in the *Disc1* L100P mouse (Clapcote et al., 2007; Lipina et al., 2010), the investigation of the role of DISC1 in basal ganglia function appears to warrant further investigation. Moreover, greater understanding of a gene's regulation will facilitate the prediction of the consequences of pathogenic variants, and thus aid the selection of variants identified by re-sequencing projects for experimental follow-up.

The linking together of FOXP2 and DISC1 in a regulatory network supports the notion that diagnostically distinct psychiatric conditions are likely to share some genetic risk factors (Carroll and Owen, 2009; Owen et al., 2007). This notion is supported by the recent findings of Maurano et al. (2012). Here, it was shown that non-coding SNPs that have attained genome-wide significant association with at least one neuropsychiatric condition (schizophrenia, bipolar disorder, attention deficit hyperactivity disorder (ADHD), panic disorder, conduct disorder, or depression) are repeatedly located in predicted transcription factor binding sites for a network of interacting transcription factors. Interestingly, this network included four members of the FOX family, binding sites for which were predicted to be affected by variants associated with schizophrenia, bipolar disorder, and ADHD.

The likely aetiological overlap between neuropsychiatric conditions prompts the question of whether reliance on diagnostic categories defined by the DSM-IV actually limits the power of a study to detect genotype-phenotype relationships. The proposed use of psychopathological dimensions in the DSM-V (American Psychiatric Association DSM-V Development, http://www.dsm5.org/Pages/Default.aspx) may result in diagnoses being better grounded in the underlying genetic aetiology; however, the benefits of this approach will depend on the validity of the selected psychopathological dimensions. Given the inconclusive evidence regarding the usefulness of endophenotypes in genetic research (Flint and Munafo, 2007), it would be naive to assume that the study of individuals selected on psychopathological dimensions will necessarily result in greater genetic tractability. Moreover, the success of any method of determining diagnosis will ultimately be tested by its usefulness in a clinical setting.

5.5 Recent developments

Recent developments in the field are likely to greatly enhance future studies exploring the role of gene expression in psychiatric disorders. As alluded to previously, of particular note

is the publication of the ENCODE project's "Integrated encyclopaedia of DNA elements in the human genome" (Bernstein et al., 2012), a genome-wide annotation of functional elements. According to ENCODE, a functional element is defined as "a discrete genome segment that encodes a defined product (for example, protein or non-coding RNA) or displays a reproducible biochemical signature (for example, protein binding, or a specific chromatin structure)". The ENCODE project have characterised a comprehensive range of functional elements, including transcription start sites, protein-bound DNA regions, regions of histone modification, and long-range chromosome interactions in over 100 different cell types. This information will facilitate the generation of hypotheses both about the regulation of individual genes of interest but also about commonalities that may unite sets of genes thought to be implicated in a given disorder or biological process.

Another development which is likely to be particularly beneficial to the study of neurological conditions is the possibility of studying human induced pluripotent stem cells (hIPSCs). The difficulties associated with acquiring live neurons from the human brain represents a key limitation in the study of psychiatric disorders; however, recent studies have demonstrated the possibility of reprogramming fibroblasts derived from schizophrenic patients into human induced pluripotent stem cells (hIPSCs) and then differentiating these hIPSCs into neurons (Brennand et al., 2011; Chiang et al., 2011). Although the application of hIPSCs to the study of psychiatric disorders is in its infancy, early findings appear to support many observations made previously using other approaches, including post-mortem studies, animal models, genetic association studies and gene expression profiling (Brennand et al., 2011). Compared to hIPSCs derived from control subjects, Brennand et al. (2011) found hIPSCs from schizophrenic patients to show reduced neuronal connectivity, reduced neurite outgrowth and wide-spread alterations in gene expression, including reduced expression of key synaptic proteins, such as PSD-95 and certain glutamate receptors. Furthermore, genes showing altered expression included members of the Wnt, cAMP, and glutamate signalling pathways, all of which, as discussed in Chapter 1, are believed to be relevant to pathogenesis of schizophrenia and bipolar disorder. Thus the study of hIPSCs is likely to yield important insights into the role of altered gene expression in psychiatric illness. Future studies should focus on identifying the genetic determinants of altered gene expression in hIPSCs derived from patients with schizophrenia and bipolar disorder, as well as the impact of altered expression at a network level.

5.6 Concluding remarks

The findings presented in this thesis (i) suggest that further investigation of the contribution of variation in the *NRG1* gene, particularly at the 3' end of the gene, to the aetiology of schizophrenia and bipolar disorder by re-sequencing is warranted and (ii) link DISC1 and FOXP2 in a regulatory network implicated in a range of psychiatric disorders that share varying degrees of phenotypic overlap. This latter finding suggests that the consideration of candidate genes in the context of their regulatory networks can provide valuable information about the genetic bases of the shared and distinct aspects of psychiatric conditions.

References

Abazyan B, Nomura J, Kannan G, Ishizuka K, Tamashiro KL, Nucifora F, et al. Prenatal interaction of mutant DISC1 and immune activation produces adult psychopathology. Biol Psychiatry 2011; 68: 1172-81.

Abdolmaleky HM, Cheng KH, Russo A, Smith CL, Faraone SV, Wilcox M, et al. Hypermethylation of the reelin (RELN) promoter in the brain of schizophrenic patients: a preliminary report. Am J Med Genet B Neuropsychiatr Genet 2005; 134B: 60-6.

Abe Y, Namba H, Kato T, Iwakura Y, Nawa H. Neuregulin-1 signals from the periphery regulate AMPA receptor sensitivity and expression in GABAergic interneurons in developing neocortex. J Neurosci 2011; 31: 5699-709.

Abi-Dargham A, Mawlawi O, Lombardo I, Gil R, Martinez D, Huang Y, et al. Prefrontal dopamine D1 receptors and working memory in schizophrenia. J Neurosci 2002; 22: 3708-19.

Adler CM, Malhotra AK, Elman I, Goldberg T, Egan M, Pickar D, et al. Comparison of ketamine-induced thought disorder in healthy volunteers and thought disorder in schizophrenia. Am J Psychiatry 1999; 156: 1646-9.

AhnAllen CG. The role of the alpha7 nicotinic receptor in cognitive processing of persons with schizophrenia. Curr Opin Psychiatry 2012; 25: 103-8.

Akbarian S, Kim JJ, Potkin SG, Hagman JO, Tafazzoli A, Bunney WE, Jr., et al. Gene expression for glutamic acid decarboxylase is reduced without loss of neurons in prefrontal cortex of schizophrenics. Arch Gen Psychiatry 1995; 52: 258-66.

Akey J, Jin L, Xiong M. Haplotypes vs single marker linkage disequilibrium tests: what do we gain? Eur J Hum Genet 2001; 9: 291-300.

Aleman A, Kahn RS, Selten JP. Sex differences in the risk of schizophrenia: evidence from meta-analysis. Arch Gen Psychiatry 2003; 60: 565-71.

Alkelai A, Greenbaum L, Lupoli S, Kohn Y, Sarner-Kanyas K, Ben-Asher E, et al. Association of the type 2 diabetes mellitus susceptibility gene, TCF7L2, with schizophrenia in an Arab-Israeli family sample. PLoS One 2012; 7: e29228.

Alloy LB, Abramson LY, Urosevic S, Walshaw PD, Nusslock R, Neeren AM. The psychosocial context of bipolar disorder: environmental, cognitive, and developmental risk factors. Clin Psychol Rev 2005; 25: 1043-75.

Alsabban S, Rivera M, McGuffin P. Genome-wide searches for bipolar disorder genes. Curr Psychiatry Rep 2011; 13: 522-7.

Altar CA, Jurata LW, Charles V, Lemire A, Liu P, Bukhman Y, et al. Deficient hippocampal neuron expression of proteasome, ubiquitin, and mitochondrial genes in multiple schizophrenia cohorts. Biol Psychiatry 2005; 58: 85-96.

Andersen CL, Jensen JL, Orntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res 2004; 64: 5245-50.

Angermeyer MC, Kuhn L. Gender differences in age at onset of schizophrenia. An overview. Eur Arch Psychiatry Neurol Sci 1988; 237: 351-64.

Anton ES, Marchionni MA, Lee KF, Rakic P. Role of GGF/neuregulin signaling in interactions between migrating neurons and radial glia in the developing cerebral cortex. Development 1997; 124: 3501-10.

APA. Diagnostic and Statistical Manual of Mental Disorder, 4th Edition. Washington DC: American Psychiatric Press, 1994.

Arai M, Obata N, Kockelkorn TT, Yamada K, Toyota T, Haga S, et al. Lack of association between polymorphisms in the 5' upstream region of the DISC1 gene and mood disorders. Psychiatr Genet 2007; 17: 357.

Arnold SE, Hyman BT, Van Hoesen GW, Damasio AR. Some cytoarchitectural abnormalities of the entorhinal cortex in schizophrenia. Arch Gen Psychiatry 1991; 48: 625-32.

Arnone D, Cavanagh J, Gerber D, Lawrie SM, Ebmeier KP, McIntosh AM. Magnetic resonance imaging studies in bipolar disorder and schizophrenia: meta-analysis. Br J Psychiatry 2009; 195: 194-201.

Arnone D, McIntosh AM, Chandra P, Ebmeier KP. Meta-analysis of magnetic resonance imaging studies of the corpus callosum in bipolar disorder. Acta Psychiatr Scand 2008; 118: 357-62.

Athanasiu L, Mattingsdal M, Kahler AK, Brown A, Gustafsson O, Agartz I, et al. Gene variants associated with schizophrenia in a Norwegian genome-wide study are replicated in a large European cohort. J Psychiatr Res 2010; 44: 748-53.

Aubin HJ, Rollema H, Svensson TH, Winterer G. Smoking, quitting, and psychiatric disease: a review. Neurosci Biobehav Rev 2012; 36: 271-84.

Austin CP, Ky B, Ma L, Morris JA, Shughrue PJ. Expression of Disrupted-In-Schizophrenia-1, a schizophrenia-associated gene, is prominent in the mouse hippocampus throughout brain development. Neuroscience 2004; 124: 3-10.

Awadalla P, Gauthier J, Myers RA, Casals F, Hamdan FF, Griffing AR, et al. Direct measure of the de novo mutation rate in autism and schizophrenia cohorts. Am J Hum Genet 2010; 87: 316-24.

Babulas V, Factor-Litvak P, Goetz R, Schaefer CA, Brown AS. Prenatal exposure to maternal genital and reproductive infections and adult schizophrenia. Am J Psychiatry 2006; 163: 927-9.

Baiano M, David A, Versace A, Churchill R, Balestrieri M, Brambilla P. Anterior cingulate volumes in schizophrenia: a systematic review and a meta-analysis of MRI studies. Schizophr Res 2007; 93: 1-12.

Baiano M, Perlini C, Rambaldelli G, Cerini R, Dusi N, Bellani M, et al. Decreased entorhinal cortex volumes in schizophrenia. Schizophr Res 2008; 102: 171-80.

Baldan Ramsey LC, Xu M, Wood N, Pittenger C. Lesions of the dorsomedial striatum disrupt prepulse inhibition. Neuroscience 2011; 180: 222-8.

Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell Res 2011; 21: 381-95.

Bao J, Wolpowitz D, Role LW, Talmage DA. Back signaling by the Nrg-1 intracellular domain. J Cell Biol 2003; 161: 1133-41.

Barch DM. The cognitive neuroscience of schizophrenia. Annu Rev Clin Psychol 2005; 1: 321-53.

Barch DM, Ceaser A. Cognition in schizophrenia: core psychological and neural mechanisms. Trends Cogn Sci 2012; 16: 27-34.

Barnett JH, Salmond CH, Jones PB, Sahakian BJ. Cognitive reserve in neuropsychiatry. Psychol Med 2006; 36: 1053-64.

Barrett SL, Mulholland CC, Cooper SJ, Rushe TM. Patterns of neurocognitive impairment in first-episode bipolar disorder and schizophrenia. Br J Psychiatry 2009; 195: 67-72.

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell 2007; 129: 823-37.

Bartos M, Vida I, Jonas P. Synaptic mechanisms of synchronized gamma oscillations in inhibitory interneuron networks. Nat Rev Neurosci 2007; 8: 45-56.

Baum AE, Akula N, Cabanero M, Cardona I, Corona W, Klemens B, et al. A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. Mol Psychiatry 2008; 13: 197-207.

Beekman L, Tohver T, Dardari R, Leguillette R. Evaluation of suitable reference genes for gene expression studies in bronchoalveolar lavage cells from horses with inflammatory airway disease. BMC Mol Biol 2011; 12: 5.

Benayoun BA, Caburet S, Veitia RA. Forkhead transcription factors: key players in health and disease. Trends Genet 2011; 27: 224-32.

Benes FM, Berretta S. GABAergic interneurons: implications for understanding schizophrenia and bipolar disorder. Neuropsychopharmacology 2001; 25: 1-27.

Benes FM, Lim B, Matzilevich D, Subburaju S, Walsh JP. Circuitry-based gene expression profiles in GABA cells of the trisynaptic pathway in schizophrenics versus bipolars. Proc Natl Acad Sci U S A 2008; 105: 20935-40.

Beneyto M, Kristiansen LV, Oni-Orisan A, McCullumsmith RE, Meador-Woodruff JH. Abnormal glutamate receptor expression in the medial temporal lobe in schizophrenia and mood disorders. Neuropsychopharmacology 2007; 32: 1888-902.

Benito E, Barco A. CREB's control of intrinsic and synaptic plasticity: implications for CREB-dependent memory models. Trends Neurosci 2010; 33: 230-40.

Berger SL. The complex language of chromatin regulation during transcription. Nature 2007; 447: 407-12.

Berk M, Dodd S, Kauer-Sant'anna M, Malhi GS, Bourin M, Kapczinski F, et al. Dopamine dysregulation syndrome: implications for a dopamine hypothesis of bipolar disorder. Acta Psychiatr Scand Suppl 2007: 41-9.

Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. Nature 2012; 489: 57-74.

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 2006; 125: 315-26.

Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 2001; 409: 363-6.

Berrettini W. Bipolar disorder and schizophrenia: not so distant relatives? World Psychiatry 2003; 2: 68-72.

Betz BC, Jordan-Williams KL, Wang C, Kang SG, Liao J, Logan MR, et al. Batf coordinates multiple aspects of B and T cell function required for normal antibody responses. J Exp Med 2010; 207: 933-42.

Beveridge NJ, Cairns MJ. MicroRNA dysregulation in schizophrenia. Neurobiol Dis 2012; 46: 263-71.

Beveridge NJ, Gardiner E, Carroll AP, Tooney PA, Cairns MJ. Schizophrenia is associated with an increase in cortical microRNA biogenesis. Mol Psychiatry 2010; 15: 1176-89.

Bilder RM, Volavka J, Czobor P, Malhotra AK, Kennedy JL, Ni X, et al. Neurocognitive correlates of the COMT Val(158)Met polymorphism in chronic schizophrenia. Biol Psychiatry 2002; 52: 701-7.

Bird A. DNA methylation patterns and epigenetic memory. Genes Dev 2002; 16: 6-21.

Birdsill AC, Walker DG, Lue L, Sue LI, Beach TG. Postmortem interval effect on RNA and gene expression in human brain tissue. Cell Tissue Bank 2011; 12: 311-8.

Bis JC, DeCarli C, Smith AV, van der Lijn F, Crivello F, Fornage M, et al. Common variants at 12q14 and 12q24 are associated with hippocampal volume. Nat Genet 2012; 44: 545-51.

Bjarnadottir M, Misner DL, Haverfield-Gross S, Bruun S, Helgason VG, Stefansson H, et al. Neuregulin1 (NRG1) signaling through Fyn modulates NMDA receptor phosphorylation: differential synaptic function in NRG1+/- knock-outs compared with wild-type mice. J Neurosci 2007; 27: 4519-29.

Blackwood DH, Fordyce A, Walker MT, St Clair DM, Porteous DJ, Muir WJ. Schizophrenia and affective disorders--cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family. Am J Hum Genet 2001; 69: 428-33.

Blake MC, Jambou RC, Swick AG, Kahn JW, Azizkhan JC. Transcriptional initiation is controlled by upstream GC-box interactions in a TATAA-less promoter. Mol Cell Biol 1990; 10: 6632-41.

Boneau CA. The effects of violations of assumptions underlying the test. Psychol Bull 1960; 57: 49-64.

Bonkowsky JL, Wang X, Fujimoto E, Lee JE, Chien CB, Dorsky RI. Domain-specific regulation of foxP2 CNS expression by lef1. BMC Dev Biol 2008; 8: 103.

Bord L, Wheeler J, Paek M, Saleh M, Lyons-Warren A, Ross CA, et al. Primate disrupted-in-schizophrenia-1 (DISC1): high divergence of a gene for major mental illnesses in recent evolutionary history. Neurosci Res 2006; 56: 286-93.

Borrmann L, Seebeck B, Rogalla P, Bullerdiek J. Human HMGA2 promoter is coregulated by a polymorphic dinucleotide (TC)-repeat. Oncogene 2003; 22: 756-60.

Bowden NA, Scott RJ, Tooney PA. Altered expression of regulator of G-protein signalling 4 (RGS4) mRNA in the superior temporal gyrus in schizophrenia. Schizophr Res 2007; 89: 165-8.

Bradshaw NJ, Porteous DJ. DISC1-binding proteins in neural development, signalling and schizophrenia. Neuropharmacology 2012; 62: 1230-41.

Braff DL, Geyer MA. Sensorimotor gating and schizophrenia. Human and animal model studies. Arch Gen Psychiatry 1990; 47: 181-8.

Bramon E, Dempster E, Frangou S, Shaikh M, Walshe M, Filbey FM, et al. Neuregulin-1 and the P300 waveform--a preliminary association study using a psychosis endophenotype. Schizophr Res 2008; 103: 178-85.

Bramon E, McDonald C, Croft RJ, Landau S, Filbey F, Gruzelier JH, et al. Is the P300 wave an endophenotype for schizophrenia? A meta-analysis and a family study. Neuroimage 2005; 27: 960-8.

Bramon E, Rabe-Hesketh S, Sham P, Murray RM, Frangou S. Meta-analysis of the P300 and P50 waveforms in schizophrenia. Schizophr Res 2004; 70: 315-29.

Brandon NJ, Sawa A. Linking neurodevelopmental and synaptic theories of mental illness through DISC1. Nat Rev Neurosci 2011; 12: 707-22.

Bray NJ. Gene expression in the etiology of schizophrenia. Schizophr Bull 2008; 34: 412-8.

Bray NJ, Holmans PA, van den Bree MB, Jones L, Elliston LA, Hughes G, et al. Cis- and trans- loci influence expression of the schizophrenia susceptibility gene DTNBP1. Hum Mol Genet 2008; 17: 1169-74.

Breathnach R, Chambon P. Organization and expression of eucaryotic split genes coding for proteins. Annu Rev Biochem 1981; 50: 349-83.

Brennand KJ, Simone A, Jou J, Gelboin-Burkhart C, Tran N, Sangar S, et al. Modelling schizophrenia using human induced pluripotent stem cells. Nature 2011; 473: 221-5.

Brinkmann BG, Agarwal A, Sereda MW, Garratt AN, Muller T, Wende H, et al. Neuregulin-1/ErbB signaling serves distinct functions in myelination of the peripheral and central nervous system. Neuron 2008; 59: 581-95.

Broadbelt K, Ramprasaud A, Jones LB. Evidence of altered neurogranin immunoreactivity in areas 9 and 32 of schizophrenic prefrontal cortex. Schizophr Res 2006; 87: 6-14.

Brown AS, Hooton J, Schaefer CA, Zhang H, Petkova E, Babulas V, et al. Elevated maternal interleukin-8 levels and risk of schizophrenia in adult offspring. Am J Psychiatry 2004; 161: 889-95.

Brown AS, Schaefer CA, Quesenberry CP, Jr., Liu L, Babulas VP, Susser ES. Maternal exposure to toxoplasmosis and risk of schizophrenia in adult offspring. Am J Psychiatry 2005; 162: 767-73.

Brown R, Colter N, Corsellis JA, Crow TJ, Frith CD, Jagoe R, et al. Postmortem evidence of structural brain changes in schizophrenia. Differences in brain weight, temporal horn area, and parahippocampal gyrus compared with affective disorder. Arch Gen Psychiatry 1986; 43: 36-42.

Brown SM, Clapcote SJ, Millar JK, Torrance HS, Anderson SM, Walker R, et al. Synaptic modulators Nrxn1 and Nrxn3 are disregulated in a Disc1 mouse model of schizophrenia. Mol Psychiatry 2011; 16: 585-7.

Bruce HA, Margolis RL. FOXP2: novel exons, splice variants, and CAG repeat length stability. Hum Genet 2002; 111: 136-44.

Bruton CJ, Crow TJ, Frith CD, Johnstone EC, Owens DG, Roberts GW. Schizophrenia and the brain: a prospective clinico-neuropathological study. Psychol Med 1990; 20: 285-304.

Buizer-Voskamp JE, Muntjewerff JW, Strengman E, Sabatti C, Stefansson H, Vorstman JA, et al. Genome-wide analysis shows increased frequency of copy number variation deletions in Dutch schizophrenia patients. Biol Psychiatry 2011; 70: 655-62.

Bull HC, Venables PH. Speech perception in schizophrenia. Br J Psychiatry 1974; 125: 350-4.

Bulyk ML. Computational prediction of transcription-factor binding site locations. Genome Biol 2003; 5: 201.

Burdick KE, Hodgkinson CA, Szeszko PR, Lencz T, Ekholm JM, Kane JM, et al. DISC1 and neurocognitive function in schizophrenia. Neuroreport 2005; 16: 1399-402.

Burke LJ, Baniahmad A. Co-repressors 2000. Faseb J 2000; 14: 1876-88.

Burke TW, Kadonaga JT. Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes Dev 1996; 10: 711-24.

Bustin SA, Nolan T. Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. J Biomol Tech 2004; 15: 155-66.

Cai C, Langfelder P, Fuller TF, Oldham MC, Luo R, van den Berg LH, et al. Is human blood a good surrogate for brain tissue in transcriptional studies? BMC Genomics 2010; 11: 589.

Callicott JH, Egan MF, Mattay VS, Bertolino A, Bone AD, Verchinksi B, et al. Abnormal fMRI response of the dorsolateral prefrontal cortex in cognitively intact siblings of patients with schizophrenia. Am J Psychiatry 2003; 160: 709-19.

Camargo LM, Collura V, Rain JC, Mizuguchi K, Hermjakob H, Kerrien S, et al. Disrupted in Schizophrenia 1 Interactome: evidence for the close connectivity of risk genes and a potential synaptic basis for schizophrenia. Mol Psychiatry 2007; 12: 74-86.

Campbell H, Manolio T. Commentary: rare alleles, modest genetic effects and the need for collaboration. Int J Epidemiol 2007; 36: 445-8.

Cannon TD, Hennah W, van Erp TG, Thompson PM, Lonnqvist J, Huttunen M, et al. Association of DISC1/TRAX haplotypes with schizophrenia, reduced prefrontal gray matter, and impaired short- and long-term memory. Arch Gen Psychiatry 2005; 62: 1205-13.

Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, et al. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science 2002; 298: 1039-43.

Cardno AG, Gottesman, II. Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. Am J Med Genet 2000; 97: 12-7.

Cardno AG, Rijsdijk FV, Sham PC, Murray RM, McGuffin P. A twin study of genetic relationships between psychotic symptoms. Am J Psychiatry 2002; 159: 539-45.

Carey M, Smale ST. Transcriptional Regulation in Eukaryotes. NY: CSHL Press, 2000.

Carless MA, Glahn DC, Johnson MP, Curran JE, Bozaoglu K, Dyer TD, et al. Impact of DISC1 variation on neuroanatomical and neurocognitive phenotypes. Mol Psychiatry 2011; 16: 1096-104, 1063.

Carlyle BC, Mackie S, Christie S, Millar JK, Porteous DJ. Co-ordinated action of DISC1, PDE4B and GSK3beta in modulation of cAMP signalling. Mol Psychiatry 2011; 16: 693-4.

Carotta S, Wu L, Nutt SL. Surprising new roles for PU.1 in the adaptive immune response. Immunol Rev 2010; 238: 63-75.

Carroll LS, Owen MJ. Genetic overlap between autism, schizophrenia and bipolar disorder. Genome Med 2009; 1: 102.

Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, et al. MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics 2005; 21: 2933-42.

Castellanos FX, Tannock R. Neuroscience of attention-deficit/hyperactivity disorder: the search for endophenotypes. Nat Rev Neurosci 2002; 3: 617-28.

Castensson A, Emilsson L, Sundberg R, Jazin E. Decrease of serotonin receptor 2C in schizophrenia brains identified by high-resolution mRNA expression analysis. Biol Psychiatry 2003; 54: 1212-21.

Chakirova G, Whalley HC, Thomson PA, Hennah W, Moorhead TW, Welch KA, et al. The effects of DISC1 risk variants on brain activation in controls, patients with bipolar disorder and patients with schizophrenia. Psychiatry Res 2011; 192: 20-8.

Chakravarty MM, Felsky D, Tampakeras M, Lerch JP, Mulsant BH, Kennedy JL, et al. DISC1 and Striatal Volume: A Potential Risk Phenotype For mental Illness. Front Psychiatry 2012; 3: 57.

Cherng JY, Schuurmans-Nieuwenbroek NM, Jiskoot W, Talsma H, Zuidam NJ, Hennink WE, et al. Effect of DNA topology on the transfection efficiency of poly((2-dimethylamino)ethyl methacrylate)-plasmid complexes. J Control Release 1999; 60: 343-53.

Chiang CH, Su Y, Wen Z, Yoritomo N, Ross CA, Margolis RL, et al. Integration-free induced pluripotent stem cells derived from schizophrenia patients with a DISC1 mutation. Mol Psychiatry 2011; 16: 358-60.

Chuang DM, Chen RW, Chalecka-Franaszek E, Ren M, Hashimoto R, Senatorov V, et al. Neuroprotective effects of lithium in cultured cells and animal models of diseases. Bipolar Disord 2002; 4: 129-36.

Chubb JE, Bradshaw NJ, Soares DC, Porteous DJ, Millar JK. The DISC locus in psychiatric illness. Mol Psychiatry 2008; 13: 36-64.

Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. Genetics 1994; 138: 963-71.

Cichon S, Craddock N, Daly M, Faraone SV, Gejman PV, Kelsoe J, et al. Genomewide association studies: history, rationale, and prospects for psychiatric disorders. Am J Psychiatry 2009; 166: 540-56.

Cichon S, Muhleisen TW, Degenhardt FA, Mattheisen M, Miro X, Strohmaier J, et al. Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. Am J Hum Genet 2011; 88: 372-81.

Clapcote SJ, Lipina TV, Millar JK, Mackie S, Christie S, Ogawa F, et al. Behavioral phenotypes of Disc1 missense mutations in mice. Neuron 2007; 54: 387-402.

Condray R, Steinhauer SR, van Kammen DP, Kasparek A. The language system in schizophrenia: effects of capacity and linguistic structure. Schizophr Bull 2002; 28: 475-90.

Connor CM, Crawford BC, Akbarian S. White matter neuron alterations in schizophrenia and related disorders. Int J Dev Neurosci 2011; 29: 325-34.

Connor CM, Guo Y, Akbarian S. Cingulate white matter neurons in schizophrenia and bipolar disorder. Biol Psychiatry 2009; 66: 486-93.

Cooper SJ, Trinklein ND, Anton ED, Nguyen L, Myers RM. Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. Genome Res 2006; 16: 1-10.

Cordell HJ, Clayton DG. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. Am J Hum Genet 2002; 70: 124-41.

Corvin AP, Morris DW, McGhee K, Schwaiger S, Scully P, Quinn J, et al. Confirmation and refinement of an 'at-risk' haplotype for schizophrenia suggests the EST cluster, Hs.97362, as a potential susceptibility gene at the Neuregulin-1 locus. Mol Psychiatry 2004; 9: 208-13.

Craddock N, Owen MJ. The beginning of the end for the Kraepelinian dichotomy. Br J Psychiatry 2005; 186: 364-6.

Craddock N, Owen MJ. Rethinking psychosis: the disadvantages of a dichotomous classification now outweigh the advantages. World Psychiatry 2007; 6: 84-91.

Crepel A, Breckpot J, Fryns JP, De la Marche W, Steyaert J, Devriendt K, et al. DISC1 duplication in two brothers with autism and mild mental retardation. Clin Genet 2010; 77: 389-94.

Crespi B, Summers K, Dorus S. Adaptive evolution of genes underlying schizophrenia. Proc Biol Sci 2007; 274: 2801-10.

Cress WD, Seto E. Histone deacetylases, transcriptional control, and cancer. J Cell Physiol 2000; 184: 1-16.

Crow TJ. Craddock & Owen vs Kraepelin: 85 years late, mesmerised by "polygenes". Schizophr Res 2008; 103: 156-60.

Cruceanu C, Alda M, Nagy C, Freemantle E, Rouleau GA, Turecki G. H3K4 tri-methylation in synapsin genes leads to different expression patterns in bipolar disorder and major depression. Int J Neuropsychopharmacol 2012: 1-11.

Cuesta MJ, Ugarte MD, Goicoa T, Eraso S, Peralta V. A taxometric analysis of schizophrenia symptoms. Psychiatry Res 2007; 150: 245-53.

D'Sa C, Duman RS. Antidepressants and neuroplasticity. Bipolar Disord 2002; 4: 183-94.

Davis KL, Kahn RS, Ko G, Davidson M. Dopamine in schizophrenia: a review and reconceptualization. Am J Psychiatry 1991; 148: 1474-86.

De Rienzo G, Bishop JA, Mao Y, Pan L, Ma TP, Moens CB, et al. Disc1 regulates both beta-catenin-mediated and noncanonical Wnt signaling during vertebrate embryogenesis. Faseb J 2011; 25: 4184-97.

Dean B, Keriakous D, Scarr E, Thomas EA. Gene expression profiling in Brodmann's area 46 from subjects with schizophrenia. Aust N Z J Psychiatry 2007; 41: 308-20.

Demjaha A, Murray RM, McGuire PK, Kapur S, Howes OD. Dopamine Synthesis Capacity in Patients With Treatment-Resistant Schizophrenia. Am J Psychiatry 2012; 169: 1203-1210.

Dempster EL, Pidsley R, Schalkwyk LC, Owens S, Georgiades A, Kane F, et al. Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. Hum Mol Genet 2011; 20: 4786-96.

Deng C, Huang XF. Decreased density of muscarinic receptors in the superior temporal gyrusin schizophrenia. J Neurosci Res 2005; 81: 883-90.

Devon RS, Anderson S, Teague PW, Burgess P, Kipari TM, Semple CA, et al. Identification of polymorphisms within Disrupted in Schizophrenia 1 and Disrupted in Schizophrenia 2, and an investigation of their association with schizophrenia and bipolar affective disorder. Psychiatr Genet 2001; 11: 71-8.

Dheda K, Huggett JF, Chang JS, Kim LU, Bustin SA, Johnson MA, et al. The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. Anal Biochem 2005; 344: 141-3.

Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. PLoS Biol 2010; 8: e1000294.

DiLeone RJ, Marcus GA, Johnson MD, Kingsley DM. Efficient studies of long-distance Bmp5 gene regulation using bacterial artificial chromosomes. Proc Natl Acad Sci U S A 2000; 97: 1612-7.

Dong E, Nelson M, Grayson DR, Costa E, Guidotti A. Clozapine and sulpiride but not haloperidol or olanzapine activate brain DNA demethylation. Proc Natl Acad Sci U S A 2008; 105: 13614-9.

Duan J, Sanders AR, Gejman PV. Genome-wide approaches to schizophrenia. Brain Res Bull 2010; 83: 93-102.

Duan X, Chang JH, Ge S, Faulkner RL, Kim JY, Kitabatake Y, et al. Disrupted-In-Schizophrenia 1 regulates integration of newly generated neurons in the adult brain. Cell 2007; 130: 1146-58.

Dudbridge F. Pedigree disequilibrium tests for multilocus haplotypes. Genet Epidemiol 2003; 25: 115-21.

Eastwood SL, Burnet PW, Gittins R, Baker K, Harrison PJ. Expression of serotonin 5-HT(2A) receptors in the human cerebellum and alterations in schizophrenia. Synapse 2001; 42: 104-14.

Eastwood SL, Burnet PW, Harrison PJ. Decreased hippocampal expression of the susceptibility gene PPP3CC and other calcineurin subunits in schizophrenia. Biol Psychiatry 2005; 57: 702-10.

Eastwood SL, Harrison PJ. Markers of glutamate synaptic transmission and plasticity are increased in the anterior cingulate cortex in bipolar disorder. Biol Psychiatry 2010; 67: 1010-6.

Egan MF, Goldberg TE, Kolachana BS, Callicott JH, Mazzanti CM, Straub RE, et al. Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. Proc Natl Acad Sci U S A 2001; 98: 6917-22.

Einat H, Yuan P, Gould TD, Li J, Du J, Zhang L, et al. The role of the extracellular signal-regulated kinase signaling pathway in mood modulation. J Neurosci 2003; 23: 7311-6.

Enard W, Gehre S, Hammerschmidt K, Holter SM, Blass T, Somel M, et al. A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. Cell 2009; 137: 961-71.

Enomoto A, Asai N, Namba T, Wang Y, Kato T, Tanaka M, et al. Roles of disrupted-in-schizophrenia 1-interacting protein girdin in postnatal development of the dentate gyrus. Neuron 2009; 63: 774-87.

Eranti SV, Maccabe JH, Bundy H, Murray RM. Gender difference in age at onset of schizophrenia: a meta-analysis. Psychol Med 2012: 1-13.

Erdely HA, Tamminga CA, Roberts RC, Vogel MW. Regional alterations in RGS4 protein in schizophrenia. Synapse 2006; 59: 472-9.

Etain B, Henry C, Bellivier F, Mathieu F, Leboyer M. Beyond genetics: childhood affective trauma in bipolar disorder. Bipolar Disord 2008; 10: 867-76.

Eykelenboom JE, Briggs GJ, Bradshaw NJ, Soares DC, Ogawa F, Christie S, et al. 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. Hum Mol Genet 2012; 21: 3374-86.

Falconer DS. Introduction to quantitive genetics. Harlow: Longman Scientific and Technical, 1989.

Falkai P, Schneider-Axmann T, Honer WG. Entorhinal cortex pre-alpha cell clusters in schizophrenia: quantitative evidence of a developmental abnormality. Biol Psychiatry 2000; 47: 937-43.

Falls DL. Neuregulins: functions, forms, and signaling strategies. Exp Cell Res 2003; 284: 14-30.

Fatemi SH, Folsom TD. The neurodevelopmental hypothesis of schizophrenia, revisited. Schizophr Bull 2009; 35: 528-48.

Fatemi SH, Reutiman TJ, Folsom TD, Huang H, Oishi K, Mori S, et al. Maternal infection leads to abnormal gene regulation and brain atrophy in mouse offspring: implications for genesis of neurodevelopmental disorders. Schizophr Res 2008a; 99: 56-70.

Fatemi SH, Reutiman TJ, Folsom TD, Lee S. Phosphodiesterase-4A expression is reduced in cerebella of patients with bipolar disorder. Psychiatr Genet 2008b; 18: 282-8.

Fatemi SH, Stary JM, Earle JA, Araghi-Niknam M, Eagan E. GABAergic dysfunction in schizophrenia and mood disorders as reflected by decreased levels of glutamic acid decarboxylase 65 and 67 kDa and Reelin proteins in cerebellum. Schizophr Res 2005; 72: 109-22.

Feng J, Schroer R, Yan J, Song W, Yang C, Bockholt A, et al. High frequency of neurexin 1beta signal peptide structural variants in patients with autism. Neurosci Lett 2006; 409: 10-3.

Ferguson BS, Nam H, Hopkins RG, Morrison RF. Impact of reference gene selection for target gene normalization on experimental outcome using real-time qRT-PCR in adipocytes. PLoS One 2010; 5: e15208.

Ferland RJ, Cherry TJ, Preware PO, Morrisey EE, Walsh CA. Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. J Comp Neurol 2003; 460: 266-79.

Fernandez PA, Tang DG, Cheng L, Prochiantz A, Mudge AW, Raff MC. Evidence that axon-derived neuregulin promotes oligodendrocyte survival in the developing rat optic nerve. Neuron 2000; 28: 81-90.

Ferreira MA, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L, et al. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. Nat Genet 2008; 40: 1056-8.

Feuk L, Kalervo A, Lipsanen-Nyman M, Skaug J, Nakabayashi K, Finucane B, et al. Absence of a paternally inherited FOXP2 gene in developmental verbal dyspraxia. Am J Hum Genet 2006; 79: 965-72.

Fisher SE, Scharff C. FOXP2 as a molecular window into speech and language. Trends Genet 2009; 25: 166-77.

Fisher SE, Vargha-Khadem F, Watkins KE, Monaco AP, Pembrey ME. Localisation of a gene implicated in a severe speech and language disorder. Nat Genet 1998; 18: 168-70.

Flint J, Munafo MR. The endophenotype concept in psychiatric genetics. Psychol Med 2007; 37: 163-80.

Forero DA, van der Ven K, Callaerts P, Del-Favero J. miRNA genes and the brain: implications for psychiatric disorders. Hum Mutat 2010; 31: 1195-204.

Frans EM, Sandin S, Reichenberg A, Lichtenstein P, Langstrom N, Hultman CM. Advancing paternal age and bipolar disorder. Arch Gen Psychiatry 2008; 65: 1034-40.

Frantseva MV, Fitzgerald PB, Chen R, Moller B, Daigle M, Daskalakis ZJ. Evidence for impaired long-term potentiation in schizophrenia and its relationship to motor skill learning. Cereb Cortex 2008; 18: 990-6.

Friedman JI, Vrijenhoek T, Markx S, Janssen IM, van der Vliet WA, Faas BH, et al. CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. Mol Psychiatry 2008; 13: 261-6.

Fujita E, Tanabe Y, Shiota A, Ueda M, Suwa K, Momoi MY, et al. Ultrasonic vocalization impairment of Foxp2 (R552H) knockin mice related to speech-language disorder and abnormality of Purkinje cells. Proc Natl Acad Sci U S A 2008; 105: 3117-22.

Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes. J Mol Biol 1987; 196: 261-82.

Gauthier J, Siddiqui TJ, Huashan P, Yokomaku D, Hamdan FF, Champagne N, et al. Truncating mutations in NRXN2 and NRXN1 in autism spectrum disorders and schizophrenia. Hum Genet 2011; 130: 563-73.

Gavin DP, Akbarian S. Epigenetic and post-transcriptional dysregulation of gene expression in schizophrenia and related disease. Neurobiol Dis 2012; 46: 255-62.

Gavin DP, Kartan S, Chase K, Grayson DR, Sharma RP. Reduced baseline acetylated histone 3 levels, and a blunted response to HDAC inhibition in lymphocyte cultures from schizophrenia subjects. Schizophr Res 2008; 103: 330-2.

Gavin DP, Rosen C, Chase K, Grayson DR, Tun N, Sharma RP. Dimethylated lysine 9 of histone 3 is elevated in schizophrenia and exhibits a divergent response to histone deacetylase inhibitors in lymphocyte cultures. J Psychiatry Neurosci 2009; 34: 232-7.

Geiger T, Cox J, Mann M. Proteomic changes resulting from gene copy number variations in cancer cells. PLoS Genet 2010; 6.

Georgieva L, Dimitrova A, Ivanov D, Nikolov I, Williams NM, Grozeva D, et al. Support for neuregulin 1 as a susceptibility gene for bipolar disorder and schizophrenia. Biol Psychiatry 2008; 64: 419-27.

Gerber DJ, Hall D, Miyakawa T, Demars S, Gogos JA, Karayiorgou M, et al. Evidence for association of schizophrenia with genetic variation in the 8p21.3 gene, PPP3CC, encoding the calcineurin gamma subunit. Proc Natl Acad Sci U S A 2003; 100: 8993-8.

Gershenzon NI, Ioshikhes IP. Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. Bioinformatics 2005; 21: 1295-300.

Glaser RL, Jabs EW. Dear old dad. Sci Aging Knowledge Environ 2004; 2004: re1.

Glatt SJ, Everall IP, Kremen WS, Corbeil J, Sasik R, Khanlou N, et al. Comparative gene expression analysis of blood and brain provides concurrent validation of SELENBP1 up-regulation in schizophrenia. Proc Natl Acad Sci U S A 2005; 102: 15533-8.

Glessner JT, Reilly MP, Kim CE, Takahashi N, Albano A, Hou C, et al. Strong synaptic transmission impact by copy number variations in schizophrenia. Proc Natl Acad Sci U S A 2010; 107: 10584-9.

Goes FS, Willour VL, Zandi PP, Belmonte PL, MacKinnon DF, Mondimore FM, et al. Family-based association study of Neuregulin 1 with psychotic bipolar disorder. Am J Med Genet B Neuropsychiatr Genet 2009; 150B: 693-702.

Gold JH. Gender differences in psychiatric illness and treatments: a critical review. J Nerv Ment Dis 1998; 186: 769-75.

Goldstein DB. Common genetic variation and human traits. N Engl J Med 2009; 360: 1696-8.

Gorisch SM, Wachsmuth M, Toth KF, Lichter P, Rippe K. Histone acetylation increases chromatin accessibility. J Cell Sci 2005; 118: 5825-34.

Gottesman, II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. Am J Psychiatry 2003; 160: 636-45.

Gottesman, II, Shields J. A polygenic theory of schizophrenia. Proc Natl Acad Sci U S A 1967; 58: 199-205.

Grant SG. Synapse signalling complexes and networks: machines underlying cognition. Bioessays 2003; 25: 1229-35.

Grayson DR, Jia X, Chen Y, Sharma RP, Mitchell CP, Guidotti A, et al. Reelin promoter hypermethylation in schizophrenia. Proc Natl Acad Sci U S A 2005; 102: 9341-6.

Grayson DR, Kundakovic M, Sharma RP. Is there a future for histone deacetylase inhibitors in the pharmacotherapy of psychiatric disorders? Mol Pharmacol 2010; 77: 126-35.

Green EK, Raybould R, Macgregor S, Gordon-Smith K, Heron J, Hyde S, et al. Operation of the schizophrenia susceptibility gene, neuregulin 1, across traditional diagnostic boundaries to increase risk for bipolar disorder. Arch Gen Psychiatry 2005; 62: 642-8.

Green MF, Kern RS, Braff DL, Mintz J. Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the "right stuff"? Schizophr Bull 2000; 26: 119-36.

Gregor A, Albrecht B, Bader I, Bijlsma EK, Ekici AB, Engels H, et al. Expanding the clinical spectrum associated with defects in CNTNAP2 and NRXN1. BMC Med Genet 2011; 12: 106.

Grice EA, Rochelle ES, Green ED, Chakravarti A, McCallion AS. Evaluation of the RET regulatory landscape reveals the biological relevance of a HSCR-implicated enhancer. Hum Mol Genet 2005; 14: 3837-45.

Grillo RW, Ottoni GL, Leke R, Souza DO, Portela LV, Lara DR. Reduced serum BDNF levels in schizophrenic patients on clozapine or typical antipsychotics. J Psychiatr Res 2007; 41: 31-5.

Groen WB, Zwiers MP, van der Gaag RJ, Buitelaar JK. The phenotype and neural correlates of language in autism: an integrative review. Neurosci Biobehav Rev 2008; 32: 1416-25.

Groszer M, Keays DA, Deacon RM, de Bono JP, Prasad-Mulcare S, Gaub S, et al. Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. Curr Biol 2008; 18: 354-62.

Grozeva D, Kirov G, Ivanov D, Jones IR, Jones L, Green EK, et al. Rare copy number variants: a point of rarity in genetic risk for bipolar disorder and schizophrenia. Arch Gen Psychiatry 2010; 67: 318-27.

Gry M, Rimini R, Stromberg S, Asplund A, Ponten F, Uhlen M, et al. Correlations between RNA and protein expression profiles in 23 human cell lines. BMC Genomics 2009; 10: 365.

Guan F, Zhang C, Wei S, Zhang H, Gong X, Feng J, et al. Association of PDE4B polymorphisms and schizophrenia in Northwestern Han Chinese. Hum Genet 2012; 131: 1047-56.

Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. Cell 2007; 130: 77-88.

Guidotti A, Auta J, Davis JM, Di-Giorgi-Gerevini V, Dwivedi Y, Grayson DR, et al. Decrease in reelin and glutamic acid decarboxylase67 (GAD67) expression in schizophrenia and bipolar disorder: a postmortem brain study. Arch Gen Psychiatry 2000; 57: 1061-9.

Guidotti A, Ruzicka W, Grayson DR, Veldic M, Pinna G, Davis JM, et al. S-adenosyl methionine and DNA methyltransferase-1 mRNA overexpression in psychosis. Neuroreport 2007; 18: 57-60.

Guo L, Lu Z. The fate of miRNA* strand through evolutionary analysis: implication for degradation as merely carrier strand or potential regulatory molecule? PLoS One 2010; 5: e11387.

Guryev V, Saar K, Adamovic T, Verheul M, van Heesch SA, Cook S, et al. Distribution and functional impact of DNA copy number variation in the rat. Nat Genet 2008; 40: 538-45.

Gustavsson A, Svensson M, Jacobi F, Allgulander C, Alonso J, Beghi E, et al. Cost of disorders of the brain in Europe 2010. Eur Neuropsychopharmacol 2011; 21: 718-79.

Gustincich S, Sandelin A, Plessy C, Katayama S, Simone R, Lazarevic D, et al. The complexity of the mammalian transcriptome. J Physiol 2006; 575: 321-32.

Haenschel C, Bittner RA, Waltz J, Haertling F, Wibral M, Singer W, et al. Cortical oscillatory activity is critical for working memory as revealed by deficits in early-onset schizophrenia. J Neurosci 2009; 29: 9481-9.

Hakak Y, Walker JR, Li C, Wong WH, Davis KL, Buxbaum JD, et al. Genome-wide expression analysis reveals dysregulation of myelination-related genes in chronic schizophrenia. Proc Natl Acad Sci U S A 2001; 98: 4746-51.

Hall D, Gogos JA, Karayiorgou M. The contribution of three strong candidate schizophrenia susceptibility genes in demographically distinct populations. Genes Brain Behav 2004; 3: 240-8.

Hall J, Whalley HC, Job DE, Baig BJ, McIntosh AM, Evans KL, et al. A neuregulin 1 variant associated with abnormal cortical function and psychotic symptoms. Nat Neurosci 2006; 9: 1477-8.

Hallikas O, Palin K, Sinjushina N, Rautiainen R, Partanen J, Ukkonen E, et al. Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. Cell 2006; 124: 47-59.

Han X, Wu X, Chung WY, Li T, Nekrutenko A, Altman NS, et al. Transcriptome of embryonic and neonatal mouse cortex by high-throughput RNA sequencing. Proc Natl Acad Sci U S A 2009; 106: 12741-6.

Hancock ML, Canetta SE, Role LW, Talmage DA. Presynaptic type III neuregulin1-ErbB signaling targets {alpha}7 nicotinic acetylcholine receptors to axons. J Cell Biol 2008; 181: 511-21.

Hancock ML, Nowakowski DW, Role LW, Talmage DA, Flanagan JG. Type III neuregulin 1 regulates pathfinding of sensory axons in the developing spinal cord and periphery. Development 2011; 138: 4887-98.

Hansen T, Ingason A, Djurovic S, Melle I, Fenger M, Gustafsson O, et al. At-risk variant in TCF7L2 for type II diabetes increases risk of schizophrenia. Biol Psychiatry 2011; 70: 59-63.

Harrison PJ, Heath PR, Eastwood SL, Burnet PW, McDonald B, Pearson RC. The relative importance of premortem acidosis and postmortem interval for human brain gene expression studies: selective mRNA vulnerability and comparison with their encoded proteins. Neurosci Lett 1995; 200: 151-4.

Harrison PJ, Kleinman JE. Methodological Issues. In: PJ Harrison GR, editor. The Neuropathology of Schizophrenia. New York: Oxford University Press, 2000.

Harrison PJ, Law AJ. Neuregulin 1 and schizophrenia: genetics, gene expression, and neurobiology. Biol Psychiatry 2006; 60: 132-40.

Harrison PJ, Weinberger DR. Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. Mol Psychiatry 2005; 10: 40-68; image 5.

Hasan A, Nitsche MA, Rein B, Schneider-Axmann T, Guse B, Gruber O, et al. Dysfunctional long-term potentiation-like plasticity in schizophrenia revealed by transcranial direct current stimulation. Behav Brain Res 2011; 224: 15-22.

Hashimoto R, Straub RE, Weickert CS, Hyde TM, Kleinman JE, Weinberger DR. Expression analysis of neuregulin-1 in the dorsolateral prefrontal cortex in schizophrenia. Mol Psychiatry 2004; 9: 299-307.

Hashimoto T, Bazmi HH, Mirnics K, Wu Q, Sampson AR, Lewis DA. Conserved regional patterns of GABA-related transcript expression in the neocortex of subjects with schizophrenia. Am J Psychiatry 2008; 165: 479-89.

Hashimoto T, Volk DW, Eggan SM, Mirnics K, Pierri JN, Sun Z, et al. Gene expression deficits in a subclass of GABA neurons in the prefrontal cortex of subjects with schizophrenia. J Neurosci 2003; 23: 6315-26.

Havas K, Flaus A, Phelan M, Kingston R, Wade PA, Lilley DM, et al. Generation of superhelical torsion by ATP-dependent chromatin remodeling activities. Cell 2000; 103: 1133-42.

Hayashi-Takagi A, Takaki M, Graziane N, Seshadri S, Murdoch H, Dunlop AJ, et al. Disrupted-in-Schizophrenia 1 (DISC1) regulates spines of the glutamate synapse via Rac1. Nat Neurosci 2010; 13: 327-32.

Hayden MS, West AP, Ghosh S. NF-kappaB and the immune response. Oncogene 2006; 25: 6758-80.

Hayesmoore JB, Bray NJ, Owen MJ, O'Donovan MC. DISC1 mRNA expression is not influenced by common Cis-acting regulatory polymorphisms or imprinting. Am J Med Genet B Neuropsychiatr Genet 2008; 147B: 1065-9.

Heffner JL, Strawn JR, DelBello MP, Strakowski SM, Anthenelli RM. The co-occurrence of cigarette smoking and bipolar disorder: phenomenology and treatment considerations. Bipolar Disord 2011; 13: 439-53.

Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol 2007; 8: R19.

Heng S, Song AW, Sim K. White matter abnormalities in bipolar disorder: insights from diffusion tensor imaging studies. J Neural Transm 2010; 117: 639-54.

Henikoff S, Henikoff JG, Sakai A, Loeb GB, Ahmad K. Genome-wide profiling of salt fractions maps physical properties of chromatin. Genome Res 2009; 19: 460-9.

Hennah W, Porteous D. The DISC1 pathway modulates expression of neurodevelopmental, synaptogenic and sensory perception genes. PLoS One 2009; 4: e4906.

Hennah W, Thomson P, McQuillin A, Bass N, Loukola A, Anjorin A, et al. DISC1 association, heterogeneity and interplay in schizophrenia and bipolar disorder. Mol Psychiatry 2009; 14: 865-73.

Hennah W, Varilo T, Kestila M, Paunio T, Arajarvi R, Haukka J, et al. Haplotype transmission analysis provides evidence of association for DISC1 to schizophrenia and suggests sex-dependent effects. Hum Mol Genet 2003; 12: 3151-9.

Henrichsen CN, Vinckenbosch N, Zollner S, Chaignat E, Pradervand S, Schutz F, et al. Segmental copy number variation shapes tissue transcriptomes. Nat Genet 2009; 41: 424-9.

Hess J, Angel P, Schorpp-Kistner M. AP-1 subunits: quarrel and harmony among siblings. J Cell Sci 2004; 117: 5965-73.

Hikida T, Jaaro-Peled H, Seshadri S, Oishi K, Hookway C, Kong S, et al. Dominant-negative DISC1 transgenic mice display schizophrenia-associated phenotypes detected by measures translatable to humans. Proc Natl Acad Sci U S A 2007; 104: 14501-6.

Ho JW, Bishop E, Karchenko PV, Negre N, White KP, Park PJ. ChIP-chip versus ChIP-seq: lessons for experimental design and data analysis. BMC Genomics 2011; 12: 134.

Ho PS, Ellison MJ, Quigley GJ, Rich A. A computer aided thermodynamic approach for predicting the formation of Z-DNA in naturally occurring sequences. Embo J 1986; 5: 2737-44.

Hobara T, Uchida S, Otsuki K, Matsubara T, Funato H, Matsuo K, et al. Altered gene expression of histone deacetylases in mood disorder patients. J Psychiatr Res 2010; 44: 263-70.

Hodgkinson CA, Enoch MA, Srivastava V, Cummins-Oman JS, Ferrier C, Iarikova P, et al. Genome-wide association identifies candidate genes that influence the human electroencephalogram. Proc Natl Acad Sci U S A 2010; 107: 8695-700.

Hong L, Schroth GP, Matthews HR, Yau P, Bradbury EM. Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA. J Biol Chem 1993; 268: 305-14.

Horiuchi Y, Ishiguro H, Koga M, Inada T, Iwata N, Ozaki N, et al. Support for association of the PPP3CC gene with schizophrenia. Mol Psychiatry 2007; 12: 891-3.

Hosking L, Lumsden S, Lewis K, Yeo A, McCarthy L, Bansal A, et al. Detection of genotyping errors by Hardy-Weinberg equilibrium testing. Eur J Hum Genet 2004; 12: 395-9.

Hotta Y, Ohnuma T, Hanzawa R, Shibata N, Maeshima H, Baba H, et al. Association study between Disrupted-in-Schizophrenia-1 (DISC1) and Japanese patients with treatment-resistant schizophrenia (TRS). Prog Neuropsychopharmacol Biol Psychiatry 2011; 35: 636-9.

Howes OD, Kapur S. The dopamine hypothesis of schizophrenia: version III--the final common pathway. Schizophr Bull 2009; 35: 549-62.

Hu J, Matsui M, Gagnon KT, Schwartz JC, Gabillet S, Arar K, et al. Allele-specific silencing of mutant huntingtin and ataxin-3 genes by targeting expanded CAG repeats in mRNAs. Nat Biotechnol 2009; 27: 478-84.

Huntzinger E, Izaurralde E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat Rev Genet 2011; 12: 99-110.

Hur EM, Zhou FQ. GSK3 signalling in neural development. Nat Rev Neurosci 2010; 11: 539-51.

Hurst JA, Baraitser M, Auger E, Graham F, Norell S. An extended family with a dominantly inherited speech disorder. Dev Med Child Neurol 1990; 32: 352-5.

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. Nat Genet 2004; 36: 949-51.

Ibi D, Nagai T, Koike H, Kitahara Y, Mizoguchi H, Niwa M, et al. Combined effect of neonatal immune activation and mutant DISC1 on phenotypic changes in adulthood. Behav Brain Res 2010; 206: 32-7.

Ingason A, Soeby K, Timm S, Wang AG, Jakobsen KD, Fink-Jensen A, et al. No significant association of the 5' end of neuregulin 1 and schizophrenia in a large Danish sample. Schizophr Res 2006; 83: 1-5.

Ingraham LJ, Kety SS. Adoption studies of schizophrenia. Am J Med Genet 2000; 97: 18-22.

ISC. Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature 2008; 455: 237-41.

Ishizuka K, Kamiya A, Oh EC, Kanki H, Seshadri S, Robinson JF, et al. DISC1-dependent switch from progenitor proliferation to migration in the developing cortex. Nature 2011; 473: 92-6.

Jakob H, Beckmann H. Circumscribed malformation and nerve cell alterations in the entorhinal cortex of schizophrenics. Pathogenetic and clinical aspects. J Neural Transm Gen Sect 1994; 98: 83-106.

James R, Adams RR, Christie S, Buchanan SR, Porteous DJ, Millar JK. Disrupted in Schizophrenia 1 (DISC1) is a multicompartmentalized protein that predominantly localizes to mitochondria. Mol Cell Neurosci 2004; 26: 112-22.

Javitt DC, Zukin SR. Recent advances in the phencyclidine model of schizophrenia. Am J Psychiatry 1991; 148: 1301-8.

Jia P, Wang L, Meltzer HY, Zhao Z. Common variants conferring risk of schizophrenia: a pathway analysis of GWAS data. Schizophr Res 2010; 122: 38-42.

Jin C, Felsenfeld G. Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes Dev 2007; 21: 1519-29.

Jindal RD, Pillai AK, Mahadik SP, Eklund K, Montrose DM, Keshavan MS. Decreased BDNF in patients with antipsychotic naive first episode schizophrenia. Schizophr Res 2010; 119: 47-51.

John JP, Halahalli HN, Vasudev MK, Jayakumar PN, Jain S. Regional brain activation/deactivation during word generation in schizophrenia: fMRI study. Br J Psychiatry 2011; 198: 213-22.

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, et al. Haplotype tagging for the identification of common disease genes. Nat Genet 2001; 29: 233-7.

Johnson GD, Lalancette C, Linnemann AK, Leduc F, Boissonneault G, Krawetz SA. The sperm nucleus: chromatin, RNA, and the nuclear matrix. Reproduction 2011; 141: 21-36.

Jonsson EG, Saetre P, Vares M, Andreou D, Larsson K, Timm S, et al. DTNBP1, NRG1, DAOA, DAO and GRM3 polymorphisms and schizophrenia: an association study. Neuropsychobiology 2009; 59: 142-50.

Joseph RM, Tager-Flusberg H, Lord C. Cognitive profiles and social-communicative functioning in children with autism spectrum disorder. J Child Psychol Psychiatry 2002; 43: 807-21.

Judson MC, Eagleson KL, Levitt P. A new synaptic player leading to autism risk: Met receptor tyrosine kinase. J Neurodev Disord 2011; 3: 282-92.

Jung WH, Jang JH, Byun MS, An SK, Kwon JS. Structural brain alterations in individuals at ultra-high risk for psychosis: a review of magnetic resonance imaging studies and future directions. J Korean Med Sci 2010; 25: 1700-9.

Kadonaga JT. Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. Cell 1998; 92: 307-13.

Kadonaga JT. The DPE, a core promoter element for transcription by RNA polymerase II. Exp Mol Med 2002; 34: 259-64.

Kaern M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. Nat Rev Genet 2005; 6: 451-64.

Kamiya A, Kubo K, Tomoda T, Takaki M, Youn R, Ozeki Y, et al. A schizophrenia-associated mutation of DISC1 perturbs cerebral cortex development. Nat Cell Biol 2005; 7: 1167-78.

Kamnasaran D, Muir WJ, Ferguson-Smith MA, Cox DW. Disruption of the neuronal PAS3 gene in a family affected with schizophrenia. J Med Genet 2003; 40: 325-32.

Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, et al. Spatio-temporal transcriptome of the human brain. Nature 2011; 478: 483-9.

Kantrowitz J, Javitt DC. Glutamatergic transmission in schizophrenia: from basic research to clinical practice. Curr Opin Psychiatry 2012; 25: 96-102.

Kapur S, Remington G. Serotonin-dopamine interaction and its relevance to schizophrenia. Am J Psychiatry 1996; 153: 466-76.

Karimi M, Goldie LC, Cruickshank MN, Moses EK, Abraham LJ. A critical assessment of the factors affecting reporter gene assays for promoter SNP function: a reassessment of -308 TNF polymorphism function using a novel integrated reporter system. Eur J Hum Genet 2009; 17: 1454-62.

Karimi M, Goldie LC, Ulgiati D, Abraham LJ. Integration site-specific transcriptional reporter gene analysis using Flp recombinase targeted cell lines. Biotechniques 2007; 42: 217-24.

Katsel P, Tan W, Abazyan B, Davis KL, Ross C, Pletnikov MV, et al. Expression of mutant human DISC1 in mice supports abnormalities in differentiation of oligodendrocytes. Schizophr Res 2011; 130: 238-49.

Keefe RS, Fenton WS. How should DSM-V criteria for schizophrenia include cognitive impairment? Schizophr Bull 2007; 33: 912-20.

Kendler KS, Neale MC. Endophenotype: a conceptual analysis. Mol Psychiatry 2010; 15: 789-97.

Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, et al. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. Arch Gen Psychiatry 1994; 51: 8-19.

Kieseppa T, Partonen T, Haukka J, Kaprio J, Lonnqvist J. High concordance of bipolar I disorder in a nationwide sample of twins. Am J Psychiatry 2004; 161: 1814-21.

Kilpinen H, Ylisaukko-Oja T, Hennah W, Palo OM, Varilo T, Vanhala R, et al. Association of DISC1 with autism and Asperger syndrome. Mol Psychiatry 2008; 13: 187-96.

Kim HG, Kishikawa S, Higgins AW, Seong IS, Donovan DJ, Shen Y, et al. Disruption of neurexin 1 associated with autism spectrum disorder. Am J Hum Genet 2008; 82: 199-207.

Kim JS, Kornhuber HH, Schmid-Burgk W, Holzmuller B. Low cerebrospinal fluid glutamate in schizophrenic patients and a new hypothesis on schizophrenia. Neurosci Lett 1980; 20: 379-82.

Kim JW, Lee YS, Cho EY, Jang YL, Park DY, Choi KS, et al. Linkage and association of schizophrenia with genetic variations in the locus of neuregulin 1 in Korean population. Am J Med Genet B Neuropsychiatr Genet 2006; 141B: 281-6.

Kim JY, Duan X, Liu CY, Jang MH, Guo JU, Pow-anpongkul N, et al. DISC1 regulates new neuron development in the adult brain via modulation of AKT-mTOR signaling through KIAA1212. Neuron 2009; 63: 761-73.

King DC, Taylor J, Elnitski L, Chiaromonte F, Miller W, Hardison RC. Evaluation of regulatory potential and conservation scores for detecting cis-regulatory modules in aligned mammalian genome sequences. Genome Res 2005; 15: 1051-60.

Kirkpatrick B, Xu L, Cascella N, Ozeki Y, Sawa A, Roberts RC. DISC1 immunoreactivity at the light and ultrastructural level in the human neocortex. J Comp Neurol 2006; 497: 436-50.

Kirmizis A, Bartley SM, Kuzmichev A, Margueron R, Reinberg D, Green R, et al. Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. Genes Dev 2004; 18: 1592-605.

Kirov G, Grozeva D, Norton N, Ivanov D, Mantripragada KK, Holmans P, et al. Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. Hum Mol Genet 2009; 18: 1497-503.

Kirov G, Gumus D, Chen W, Norton N, Georgieva L, Sari M, et al. Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. Hum Mol Genet 2008; 17: 458-65.

Kirov G, Pocklington AJ, Holmans P, Ivanov D, Ikeda M, Ruderfer D, et al. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol Psychiatry 2012; 17: 142-53.

Klein PS, Melton DA. A molecular mechanism for the effect of lithium on development. Proc Natl Acad Sci U S A 1996; 93: 8455-9.

Klose RJ, Bird AP. Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci 2006; 31: 89-97.

Knight HM, Pickard BS, Maclean A, Malloy MP, Soares DC, McRae AF, et al. A cytogenetic abnormality and rare coding variants identify ABCA13 as a candidate gene in schizophrenia, bipolar disorder, and depression. Am J Hum Genet 2009; 85: 833-46.

Kockelkorn TT, Arai M, Matsumoto H, Fukuda N, Yamada K, Minabe Y, et al. Association study of polymorphisms in the 5' upstream region of human DISC1 gene with schizophrenia. Neurosci Lett 2004; 368: 41-5.

Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, et al. Rate of de novo mutations and the importance of father's age to disease risk. Nature 2012; 488: 471-5.

Konick LC, Friedman L. Meta-analysis of thalamic size in schizophrenia. Biol Psychiatry 2001; 49: 28-38.

Konopka G, Bomar JM, Winden K, Coppola G, Jonsson ZO, Gao F, et al. Human-specific transcriptional regulation of CNS development genes by FOXP2. Nature 2009; 462: 213-7.

Konradi C, Heckers S. Antipsychotic drugs and neuroplasticity: insights into the treatment and neurobiology of schizophrenia. Biol Psychiatry 2001; 50: 729-42.

Konradi C, Sillivan SE, Clay HB. Mitochondria, oligodendrocytes and inflammation in bipolar disorder: evidence from transcriptome studies points to intriguing parallels with multiple sclerosis. Neurobiol Dis 2012; 45: 37-47.

Korostishevsky M, Kaganovich M, Cholostoy A, Ashkenazi M, Ratner Y, Dahary D, et al. Is the G72/G30 locus associated with schizophrenia? single nucleotide polymorphisms, haplotypes, and gene expression analysis. Biol Psychiatry 2004; 56: 169-76.

Kouzarides T. Chromatin modifications and their function. Cell 2007; 128: 693-705.

Kovalenko S, Bergmann A, Schneider-Axmann T, Ovary I, Majtenyi K, Havas L, et al. Regio entorhinalis in schizophrenia: more evidence for migrational disturbances and suggestions for a new biological hypothesis. Pharmacopsychiatry 2003; 36 Suppl 3: S158-61.

Kraepelin E. Die Erscheinungsformen des Irreseins (Patterns of mental disorder). In: Shepherd SRHM, editor. Themes and Variations in European Psychiatry. Bristol: Wright, 1920: 7-30.

Kringlen E, Cramer G. Offspring of monozygotic twins discordant for schizophrenia. Arch Gen Psychiatry 1989; 46: 873-7.

Krystal JH, Karper LP, Seibyl JP, Freeman GK, Delaney R, Bremner JD, et al. Subanesthetic effects of the noncompetitive NMDA antagonist, ketamine, in humans. Psychotomimetic, perceptual, cognitive, and neuroendocrine responses. Arch Gen Psychiatry 1994; 51: 199-214.

Kuschner ES, Bennetto L, Yost K. Patterns of nonverbal cognitive functioning in young children with autism spectrum disorders. J Autism Dev Disord 2007; 37: 795-807.

Kutach AK, Kadonaga JT. The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. Mol Cell Biol 2000; 20: 4754-64.

Kwon E, Wang W, Tsai LH. Validation of schizophrenia-associated genes CSMD1, C10orf26, CACNA1C and TCF4 as miR-137 targets. Mol Psychiatry 2011.

Kwon OB, Paredes D, Gonzalez CM, Neddens J, Hernandez L, Vullhorst D, et al. Neuregulin-1 regulates LTP at CA1 hippocampal synapses through activation of dopamine D4 receptors. Proc Natl Acad Sci U S A 2008; 105: 15587-92.

Kyogoku C, Yanagi M, Nishimura K, Sugiyama D, Morinobu A, Fukutake M, et al. Association of calcineurin A gamma subunit (PPP3CC) and early growth response 3 (EGR3) gene polymorphisms with susceptibility to schizophrenia in a Japanese population. Psychiatry Res 2011; 185: 16-9.

Lachman HM, Papolos DF, Saito T, Yu YM, Szumlanski CL, Weinshilboum RM. Human catechol-O-methyltransferase pharmacogenetics: description of a functional polymorphism and its potential application to neuropsychiatric disorders. Pharmacogenetics 1996; 6: 243-50.

Lachman HM, Pedrosa E, Nolan KA, Glass M, Ye K, Saito T. Analysis of polymorphisms in AT-rich domains of neuregulin 1 gene in schizophrenia. Am J Med Genet B Neuropsychiatr Genet 2006; 141B: 102-9.

Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 2001; 413: 519-23.

Lai CS, Gerrelli D, Monaco AP, Fisher SE, Copp AJ. FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. Brain 2003; 126: 2455-62.

Lai CY, Yu SL, Hsieh MH, Chen CH, Chen HY, Wen CC, et al. MicroRNA expression aberration as potential peripheral blood biomarkers for schizophrenia. PLoS One 2011; 6: e21635.

Laruelle M, Abi-Dargham A, van Dyck CH, Gil R, D'Souza CD, Erdos J, et al. Single photon emission computerized tomography imaging of amphetamine-induced dopamine release in drug-free schizophrenic subjects. Proc Natl Acad Sci U S A 1996; 93: 9235-40.

Lavori PW, Krause-Steinrauf H, Brophy M, Buxbaum J, Cockroft J, Cox DR, et al. Principles, organization, and operation of a DNA bank for clinical trials: a Department of Veterans Affairs cooperative study. Control Clin Trials 2002; 23: 222-39.

Law AJ, Lipska BK, Weickert CS, Hyde TM, Straub RE, Hashimoto R, et al. Neuregulin 1 transcripts are differentially expressed in schizophrenia and regulated by 5' SNPs associated with the disease. Proc Natl Acad Sci U S A 2006; 103: 6747-52.

Lawrie SM, Abukmeil SS. Brain abnormality in schizophrenia. A systematic and quantitative review of volumetric magnetic resonance imaging studies. Br J Psychiatry 1998; 172: 110-20.

Leckman JF, Sholomskas D, Thompson WD, Belanger A, Weissman MM. Best estimate of lifetime psychiatric diagnosis: a methodological study. Arch Gen Psychiatry 1982; 39: 879-83.

Lee FH, Fadel MP, Preston-Maher K, Cordes SP, Clapcote SJ, Price DJ, et al. Disc1 point mutations in mice affect development of the cerebral cortex. J Neurosci 2011; 31: 3197-206.

Lee SH, DeCandia TR, Ripke S, Yang J, Sullivan PF, Goddard ME, et al. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nat Genet 2012; 44: 247-50.

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, et al. The nuclear RNase III Drosha initiates microRNA processing. Nature 2003; 425: 415-9.

Lennon PA, Cooper ML, Peiffer DA, Gunderson KL, Patel A, Peters S, et al. Deletion of 7q31.1 supports involvement of FOXP2 in language impairment: clinical report and review. Am J Med Genet A 2007; 143A: 791-8.

Leonard S, Gault J, Hopkins J, Logel J, Vianzon R, Short M, et al. Association of promoter variants in the alpha7 nicotinic acetylcholine receptor subunit gene with an inhibitory deficit found in schizophrenia. Arch Gen Psychiatry 2002; 59: 1085-96.

Lewis DA. The human brain revisited: opportunities and challenges in postmortem studies of psychiatric disorders. Neuropsychopharmacology 2002; 26: 143-54.

Li T, Li Z, Chen P, Zhao Q, Wang T, Huang K, et al. Common variants in major histocompatibility complex region and TCF4 gene are significantly associated with schizophrenia in Han Chinese. Biol Psychiatry 2010; 68: 671-3.

Li T, Zeng Z, Zhao Q, Wang T, Huang K, Li J, et al. FoxP2 is significantly associated with schizophrenia and major depression in the Chinese Han Population. World J Biol Psychiatry 2012.

Liang Y, Ridzon D, Wong L, Chen C. Characterization of microRNA expression profiles in normal human tissues. BMC Genomics 2007; 8: 166.

Lichtenstein P, Yip BH, Bjork C, Pawitan Y, Cannon TD, Sullivan PF, et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. Lancet 2009; 373: 234-9.

Lieberman JA, Kane JM, Alvir J. Provocative tests with psychostimulant drugs in schizophrenia. Psychopharmacology (Berl) 1987; 91: 415-33.

Lim CY, Santoso B, Boulay T, Dong E, Ohler U, Kadonaga JT. The MTE, a new core promoter element for transcription by RNA polymerase II. Genes Dev 2004; 18: 1606-17.

Lindstrom LH, Gefvert O, Hagberg G, Lundberg T, Bergstrom M, Hartvig P, et al. Increased dopamine synthesis rate in medial prefrontal cortex and striatum in schizophrenia indicated by L-(beta-11C) DOPA and PET. Biol Psychiatry 1999; 46: 681-8.

Lipina TV, Kaidanovich-Beilin O, Patel S, Wang M, Clapcote SJ, Liu F, et al. Genetic and pharmacological evidence for schizophrenia-related Disc1 interaction with GSK-3. Synapse 2011; 65: 234-48.

Lipina TV, Niwa M, Jaaro-Peled H, Fletcher PJ, Seeman P, Sawa A, et al. Enhanced dopamine function in DISC1-L100P mutant mice: implications for schizophrenia. Genes Brain Behav 2010; 9: 777-89.

Lipina TV, Wang M, Liu F, Roder JC. Synergistic interactions between PDE4B and GSK-3: DISC1 mutant mice. Neuropharmacology 2012; 62: 1252-62.

Lipska BK, Peters T, Hyde TM, Halim N, Horowitz C, Mitkus S, et al. Expression of DISC1 binding partners is reduced in schizophrenia and associated with DISC1 SNPs. Hum Mol Genet 2006; 15: 1245-58.

Lisman JE, Coyle JT, Green RW, Javitt DC, Benes FM, Heckers S, et al. Circuit-based framework for understanding neurotransmitter and risk gene interactions in schizophrenia. Trends Neurosci 2008; 31: 234-42.

Liu H, Mulholland N, Fu H, Zhao K. Cooperative activity of BRG1 and Z-DNA formation in chromatin remodeling. Mol Cell Biol 2006; 26: 2550-9.

Liu LF, Wang JC. Supercoiling of the DNA template during transcription. Proc Natl Acad Sci U S A 1987; 84: 7024-7.

Liu R, Liu H, Chen X, Kirby M, Brown PO, Zhao K. Regulation of CSF1 promoter by the SWI/SNF-like BAF complex. Cell 2001a; 106: 309-18.

Liu TY, Hsieh JC, Chen YS, Tu PC, Su TP, Chen LF. Different patterns of abnormal gamma oscillatory activity in unipolar and bipolar disorder patients during an implicit emotion task. Neuropsychologia 2012; 50: 1514-20.

Liu Y, Ford B, Mann MA, Fischbach GD. Neuregulins increase alpha7 nicotinic acetylcholine receptors and enhance excitatory synaptic transmission in GABAergic interneurons of the hippocampus. J Neurosci 2001b; 21: 5660-9.

Liu YL, Fann CS, Liu CM, Chang CC, Yang WC, Hung SI, et al. More evidence supports the association of PPP3CC with schizophrenia. Mol Psychiatry 2007; 12: 966-74.

Lopez-Bendito G, Cautinat A, Sanchez JA, Bielle F, Flames N, Garratt AN, et al. Tangential neuronal migration controls axon guidance: a role for neuregulin-1 in thalamocortical axon navigation. Cell 2006; 125: 127-42.

Lopez-Figueroa AL, Norton CS, Lopez-Figueroa MO, Armellini-Dodel D, Burke S, Akil H, et al. Serotonin 5-HT1A, 5-HT1B, and 5-HT2A receptor mRNA expression in subjects with major depression, bipolar disorder, and schizophrenia. Biol Psychiatry 2004; 55: 225-33.

Lopez de Lara C, Jaitovich-Groisman I, Cruceanu C, Mamdani F, Lebel V, Yerko V, et al. Implication of synapse-related genes in bipolar disorder by linkage and gene expression analyses. Int J Neuropsychopharmacol 2010; 13: 1397-410.

Lubow RE, Gewirtz JC. Latent inhibition in humans: data, theory, and implications for schizophrenia. Psychol Bull 1995; 117: 87-103.

Lubow RE, Moore AU. Latent inhibition: the effect of nonreinforced pre-exposure to the conditional stimulus. J Comp Physiol Psychol 1959; 52: 415-9.

Luciano M, Hansell NK, Lahti J, Davies G, Medland SE, Raikkonen K, et al. Whole genome association scan for genetic polymorphisms influencing information processing speed. Biol Psychol 2011; 86: 193-202.

Ma L, Liu Y, Ky B, Shughrue PJ, Austin CP, Morris JA. Cloning and characterization of Disc1, the mouse ortholog of DISC1 (Disrupted-in-Schizophrenia 1). Genomics 2002; 80: 662-72.

MacDermot KD, Bonora E, Sykes N, Coupe AM, Lai CS, Vernes SC, et al. Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. Am J Hum Genet 2005; 76: 1074-80.

Maeda K, Nwulia E, Chang J, Balkissoon R, Ishizuka K, Chen H, et al. Differential expression of disrupted-in-schizophrenia (DISC1) in bipolar disorder. Biol Psychiatry 2006; 60: 929-35.

Mahon K, Burdick KE, Szeszko PR. A role for white matter abnormalities in the pathophysiology of bipolar disorder. Neurosci Biobehav Rev 2010; 34: 533-54.

Malhotra AK, Kestler LJ, Mazzanti C, Bates JA, Goldberg T, Goldman D. A functional polymorphism in the COMT gene and performance on a test of prefrontal cognition. Am J Psychiatry 2002; 159: 652-4.

Malhotra D, McCarthy S, Michaelson JJ, Vacic V, Burdick KE, Yoon S, et al. High frequencies of de novo CNVs in bipolar disorder and schizophrenia. Neuron 2011; 72: 951-63.

Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell 2012; 148: 1223-41.

Mansur RB, Zugman A, Asevedo Ede M, da Cunha GR, Bressan RA, Brietzke E. Cytokines in schizophrenia: possible role of anti-inflammatory medications in clinical and preclinical stages. Psychiatry Clin Neurosci 2012; 66: 247-60.

Mao Y, Ge X, Frank CL, Madison JM, Koehler AN, Doud MK, et al. Disrupted in schizophrenia 1 regulates neuronal progenitor proliferation via modulation of GSK3beta/beta-catenin signaling. Cell 2009; 136: 1017-31.

Mar JC, Matigian NA, Mackay-Sim A, Mellick GD, Sue CM, Silburn PA, et al. Variance of gene expression identifies altered network constraints in neurological disease. PLoS Genet 2011; 7: e1002207.

Marballi K, Cruz D, Thompson P, Walss-Bass C. Differential neuregulin 1 cleavage in the prefrontal cortex and hippocampus in schizophrenia and bipolar disorder: preliminary findings. PLoS One 2012; 7: e36431.

Marin O. Interneuron dysfunction in psychiatric disorders. Nat Rev Neurosci 2012; 13: 107-20.

Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, et al. Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet 2008; 82: 477-88.

Martin N, Patel S, Segre JA. Long-range comparison of human and mouse Sprr loci to identify conserved noncoding sequences involved in coordinate regulation. Genome Res 2004; 14: 2430-8.

Martinowich K, Schloesser RJ, Manji HK. Bipolar disorder: from genes to behavior pathways. J Clin Invest 2009; 119: 726-36.

Matheson SL, Shepherd AM, Laurens KR, Carr VJ. A systematic meta-review grading the evidence for non-genetic risk factors and putative antecedents of schizophrenia. Schizophr Res 2011; 133: 133-42.

Matsuki T, Hori G, Furuichi T. Gene expression profiling during the embryonic development of mouse brain using an oligonucleotide-based microarray system. Brain Res Mol Brain Res 2005; 136: 231-54.

Matthews PR, Eastwood SL, Harrison PJ. Reduced myelin basic protein and actin-related gene expression in visual cortex in schizophrenia. PLoS One 2012; 7: e38211.

Matthysse S. Antipsychotic drug actions: a clue to the neuropathology of schizophrenia? Fed Proc 1973; 32: 200-5.

Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science 2012; 337: 1190-5.

Maycox PR, Kelly F, Taylor A, Bates S, Reid J, Logendra R, et al. Analysis of gene expression in two large schizophrenia cohorts identifies multiple changes associated with nerve terminal function. Mol Psychiatry 2009; 14: 1083-94.

McClellan JM, Susser E, King MC. Schizophrenia: a common disease caused by multiple rare alleles. Br J Psychiatry 2007; 190: 194-9.

McGrath JJ. Myths and plain truths about schizophrenia epidemiology--the NAPE lecture 2004. Acta Psychiatr Scand 2005; 111: 4-11.

McGuffin P. Nature and nurture interplay: schizophrenia. Psychiatr Prax 2004; 31 Suppl 2: S189-93.

McGuffin P, Rijsdijk F, Andrew M, Sham P, Katz R, Cardno A. The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. Arch Gen Psychiatry 2003; 60: 497-502.

McIntosh AM, Moorhead TW, Job D, Lymer GK, Munoz Maniega S, McKirdy J, et al. The effects of a neuregulin 1 variant on white matter density and integrity. Mol Psychiatry 2008; 13: 1054-9.

McQueen MB, Devlin B, Faraone SV, Nimgaonkar VL, Sklar P, Smoller JW, et al. Combined analysis from eleven linkage studies of bipolar disorder provides strong evidence of susceptibility loci on chromosomes 6q and 8q. Am J Hum Genet 2005; 77: 582-95.

McQuillin A, Bass N, Anjorin A, Lawrence J, Kandaswamy R, Lydall G, et al. Analysis of genetic deletions and duplications in the University College London bipolar disorder case control sample. Eur J Hum Genet 2011; 19: 588-92.

Meador-Woodruff JH, Healy DJ. Glutamate receptor expression in schizophrenic brain. Brain Res Brain Res Rev 2000; 31: 288-94.

Mefford HC, Batshaw ML, Hoffman EP. Genomics, intellectual disability, and autism. N Engl J Med 2012; 366: 733-43.

Mei L, Xiong WC. Neuregulin 1 in neural development, synaptic plasticity and schizophrenia. Nat Rev Neurosci 2008; 9: 437-52.

Mellios N, Huang HS, Grigorenko A, Rogaev E, Akbarian S. A set of differentially expressed miRNAs, including miR-30a-5p, act as post-transcriptional inhibitors of BDNF in prefrontal cortex. Hum Mol Genet 2008; 17: 3030-42.

Mendlewicz J, Rainer JD. Adoption study supporting genetic transmission in manic--depressive illness. Nature 1977; 268: 327-9.

Merikangas KR, Chakravarti A, Moldin SO, Araj H, Blangero JC, Burmeister M, et al. Future of genetics of mood disorders research. Biol Psychiatry 2002; 52: 457-77.

Merikangas KR, Jin R, He JP, Kessler RC, Lee S, Sampson NA, et al. Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. Arch Gen Psychiatry 2011; 68: 241-51.

Mexal S, Berger R, Adams CE, Ross RG, Freedman R, Leonard S. Brain pH has a significant impact on human postmortem hippocampal gene expression profiles. Brain Res 2006; 1106: 1-11.

Mexal S, Berger R, Logel J, Ross RG, Freedman R, Leonard S. Differential regulation of alpha7 nicotinic receptor gene (CHRNA7) expression in schizophrenic smokers. J Mol Neurosci 2010; 40: 185-95.

Meyer D, Birchmeier C. Multiple essential functions of neuregulin in development. Nature 1995; 378: 386-90.

Meyer KD, Morris JA. Immunohistochemical analysis of Disc1 expression in the developing and adult hippocampus. Gene Expr Patterns 2008; 8: 494-501.

Meyer U, Feldon J. Neural basis of psychosis-related behaviour in the infection model of schizophrenia. Behav Brain Res 2009; 204: 322-34.

Middleton FA, Strick PL. Basal ganglia and cerebellar loops: motor and cognitive circuits. Brain Res Brain Res Rev 2000; 31: 236-50.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 2007; 448: 553-60.

Mill J, Tang T, Kaminsky Z, Khare T, Yazdanpanah S, Bouchard L, et al. Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. Am J Hum Genet 2008; 82: 696-711.

Millar JK, Christie S, Semple CA, Porteous DJ. Chromosomal location and genomic structure of the human translin-associated factor X gene (TRAX; TSNAX) revealed by intergenic splicing to DISC1, a gene disrupted by a translocation segregating with schizophrenia. Genomics 2000a; 67: 69-77.

Millar JK, Pickard BS, Mackie S, James R, Christie S, Buchanan SR, et al. DISC1 and PDE4B are interacting genetic factors in schizophrenia that regulate cAMP signaling. Science 2005; 310: 1187-91.

Millar JK, Wilson-Annan JC, Anderson S, Christie S, Taylor MS, Semple CA, et al. Disruption of two novel genes by a translocation co-segregating with schizophrenia. Hum Mol Genet 2000b; 9: 1415-23.

Miller BH, Zeier Z, Xi L, Lanz TA, Deng S, Strathmann J, et al. MicroRNA-132 dysregulation in schizophrenia has implications for both neurodevelopment and adult brain function. Proc Natl Acad Sci U S A 2012; 109: 3125-30.

Miller JE, Spiteri E, Condro MC, Dosumu-Johnson RT, Geschwind DH, White SA. Birdsong decreases protein levels of FoxP2, a molecule required for human speech. J Neurophysiol 2008; 100: 2015-25.

Mirnics K, Middleton FA, Marquez A, Lewis DA, Levitt P. Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex. Neuron 2000; 28: 53-67.

Mirnics K, Middleton FA, Stanwood GD, Lewis DA, Levitt P. Disease-specific changes in regulator of G-protein signaling 4 (RGS4) expression in schizophrenia. Mol Psychiatry 2001; 6: 293-301.

Mitchell KJ. The genetics of brain wiring: from molecule to mind. PLoS Biol 2007; 5: e113.

Mitchell KJ. What is complex about complex disorders? Genome Biol 2012; 13: 237.

Mitchell KJ, Porteous DJ. Rethinking the genetic architecture of schizophrenia. Psychol Med 2011; 41: 19-32.

Mody M, Cao Y, Cui Z, Tay KY, Shyong A, Shimizu E, et al. Genome-wide gene expression profiles of the developing mouse hippocampus. Proc Natl Acad Sci U S A 2001; 98: 8862-7.

Money TT, Scarr E, Udawela M, Gibbons AS, Jeon WJ, Seo MS, et al. Treating schizophrenia: novel targets for the cholinergic system. CNS Neurol Disord Drug Targets 2010; 9: 241-56.

Moon E, Rollins B, Mesen A, Sequeira A, Myers RM, Akil H, et al. Lack of association to a NRG1 missense polymorphism in schizophrenia or bipolar disorder in a Costa Rican population. Schizophr Res 2011; 131: 52-7.

Morris JA, Kandpal G, Ma L, Austin CP. 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. Hum Mol Genet 2003; 12: 1591-608.

Mortensen PB, Pedersen CB, McGrath JJ, Hougaard DM, Norgaard-Petersen B, Mors O, et al. Neonatal antibodies to infectious agents and risk of bipolar disorder: a population-based case-control study. Bipolar Disord 2011; 13: 624-9.

Mortensen PB, Pedersen CB, Melbye M, Mors O, Ewald H. Individual and familial risk factors for bipolar affective disorders in Denmark. Arch Gen Psychiatry 2003; 60: 1209-15.

Mueller JC, Lohmussaar E, Magi R, Remm M, Bettecken T, Lichtner P, et al. Linkage disequilibrium patterns and tagSNP transferability among European populations. Am J Hum Genet 2005; 76: 387-98.

Mukamel Z, Konopka G, Wexler E, Osborn GE, Dong H, Bergman MY, et al. Regulation of MET by FOXP2, genes implicated in higher cognitive dysfunction and autism risk. J Neurosci 2011; 31: 11437-42.

Murdoch H, Mackie S, Collins DM, Hill EV, Bolger GB, Klussmann E, et al. Isoform-selective susceptibility of DISC1/phosphodiesterase-4 complexes to dissociation by elevated intracellular cAMP levels. J Neurosci 2007; 27: 9513-24.

Murphy DL, Brodie HK, Goodwin FK, Bunney WE, Jr. Regular induction of hypomania by L-dopa in "bipolar" manic-depressive patients. Nature 1971; 229: 135-6.

Muzina DJ, Elhaj O, Gajwani P, Gao K, Calabrese JR. Lamotrigine and antiepileptic drugs as mood stabilizers in bipolar disorder. Acta Psychiatr Scand Suppl 2005: 21-8.

Nadeau JH. Modifier genes in mice and humans. Nat Rev Genet 2001; 2: 165-74.

Nakata K, Lipska BK, Hyde TM, Ye T, Newburn EN, Morita Y, et al. DISC1 splice variants are upregulated in schizophrenia and associated with risk polymorphisms. Proc Natl Acad Sci U S A 2009; 106: 15873-8.

Need AC, Ge D, Weale ME, Maia J, Feng S, Heinzen EL, et al. A genome-wide investigation of SNPs and CNVs in schizophrenia. PLoS Genet 2009; 5: e1000373.

Nelson MD, Saykin AJ, Flashman LA, Riordan HJ. Hippocampal volume reduction in schizophrenia as assessed by magnetic resonance imaging: a meta-analytic study. Arch Gen Psychiatry 1998; 55: 433-40.

Ng MY, Levinson DF, Faraone SV, Suarez BK, DeLisi LE, Arinami T, et al. Meta-analysis of 32 genome-wide linkage studies of schizophrenia. Mol Psychiatry 2009; 14: 774-85.

Nicodemus KK, Law AJ, Luna A, Vakkalanka R, Straub RE, Kleinman JE, et al. A 5' promoter region SNP in NRG1 is associated with schizophrenia risk and type III isoform expression. Mol Psychiatry 2009; 14: 741-3.

Nishi A, Kuroiwa M, Miller DB, O'Callaghan JP, Bateup HS, Shuto T, et al. Distinct roles of PDE4 and PDE10A in the regulation of cAMP/PKA signaling in the striatum. J Neurosci 2008; 28: 10460-71.

Niwa M, Kamiya A, Murai R, Kubo K, Gruber AJ, Tomita K, et al. Knockdown of DISC1 by in utero gene transfer disturbs postnatal dopaminergic maturation in the frontal cortex and leads to adult behavioral deficits. Neuron 2010; 65: 480-9.

Nobrega MA, Ovcharenko I, Afzal V, Rubin EM. Scanning human gene deserts for long-range enhancers. Science 2003; 302: 413.

Nordstokke DW, Zumbo BD. A New Nonparametric Levene Test for Equal Variances. Psicologica 2010; 32: 401-430.

Nudmamud-Thanoi S, Reynolds GP. The NR1 subunit of the glutamate/NMDA receptor in the superior temporal cortex in schizophrenia and affective disorders. Neurosci Lett 2004; 372: 173-7.

Numata S, Ueno S, Iga J, Song H, Nakataki M, Tayoshi S, et al. Positive association of the PDE4B (phosphodiesterase 4B) gene with schizophrenia in the Japanese population. J Psychiatr Res 2008; 43: 7-12.

Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. Am J Hum Genet 2004; 74: 765-9.

O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V, et al. Identification of loci associated with schizophrenia by genome-wide association and follow-up. Nat Genet 2008; 40: 1053-5.

Ogawa F, Kasai M, Akiyama T. A functional link between Disrupted-In-Schizophrenia 1 and the eukaryotic translation initiation factor 3. Biochem Biophys Res Commun 2005; 338: 771-6.

Oh G, Petronis A. Environmental studies of schizophrenia through the prism of epigenetics. Schizophr Bull 2008; 34: 1122-9.

Okada M, Corfas G. Neuregulin1 downregulates postsynaptic GABAA receptors at the hippocampal inhibitory synapse. Hippocampus 2004; 14: 337-44.

Okerlund ND, Cheyette BN. Synaptic Wnt signaling-a contributor to major psychiatric disorders? J Neurodev Disord 2011; 3: 162-74.

Okochi T, Ikeda M, Kishi T, Kawashima K, Kinoshita Y, Kitajima T, et al. Meta-analysis of association between genetic variants in COMT and schizophrenia: an update. Schizophr Res 2009; 110: 140-8.

Olincy A, House R, Gao B, Recksiek P, Phang TL, Sullivan B, et al. Elevated DISC1 transcript levels in PBMCs during acute psychosis in patients with schizophrenia. Transl Biomed 2012; 2.

Olney JW, Farber NB. Glutamate receptor dysfunction and schizophrenia. Arch Gen Psychiatry 1995; 52: 998-1007.

Ortega MC, Bribian A, Peregrin S, Gil MT, Marin O, de Castro F. Neuregulin-1/ErbB4 signaling controls the migration of oligodendrocyte precursor cells during development. Exp Neurol 2012; 235: 610-20.

Overdier DG, Porcella A, Costa RH. The DNA-binding specificity of the hepatocyte nuclear factor 3/forkhead domain is influenced by amino-acid residues adjacent to the recognition helix. Mol Cell Biol 1994; 14: 2755-66.

Owen MJ, Craddock N, Jablensky A. The genetic deconstruction of psychosis. Schizophr Bull 2007; 33: 905-11.

Owen MJ, O'Donovan MC, Harrison PJ. Schizophrenia: a genetic disorder of the synapse? Bmj 2005; 330: 158-9.

Owen MJ, O'Donovan MC, Thapar A, Craddock N. Neurodevelopmental hypothesis of schizophrenia. Br J Psychiatry 2011; 198: 173-5.

Ozaki M, Sasner M, Yano R, Lu HS, Buonanno A. Neuregulin-beta induces expression of an NMDA-receptor subunit. Nature 1997; 390: 691-4.

Ozsolak F, Song JS, Liu XS, Fisher DE. High-throughput mapping of the chromatin structure of human promoters. Nat Biotechnol 2007; 25: 244-8.

Pakkenberg B. Post-mortem study of chronic schizophrenic brains. Br J Psychiatry 1987; 151: 744-52.

Palo OM, Antila M, Silander K, Hennah W, Kilpinen H, Soronen P, et al. Association of distinct allelic haplotypes of DISC1 with psychotic and bipolar spectrum disorders and with underlying cognitive impairments. Hum Mol Genet 2007; 16: 2517-28.

Pan B, Huang XF, Deng C. Antipsychotic treatment and neuregulin 1-ErbB4 signalling in schizophrenia. Prog Neuropsychopharmacol Biol Psychiatry 2011; 35: 924-30.

Parlapani E, Schmitt A, Wirths O, Bauer M, Sommer C, Rueb U, et al. Gene expression of neuregulin-1 isoforms in different brain regions of elderly schizophrenia patients. World J Biol Psychiatry 2010; 11: 243-50.

Pearlson GD, Wong DF, Tune LE, Ross CA, Chase GA, Links JM, et al. In vivo D2 dopamine receptor density in psychotic and nonpsychotic patients with bipolar disorder. Arch Gen Psychiatry 1995; 52: 471-7.

Pedrosa E, Nolan KA, Stefanescu R, Herskovits P, Novak T, Zukov I, et al. Analysis of a promoter polymorphism in the SMDF neuregulin 1 isoform in Schizophrenia. Neuropsychobiology 2009; 59: 205-12.

Pedrosa E, Shah A, Tenore C, Capogna M, Villa C, Guo X, et al. beta-catenin promoter ChIP-chip reveals potential schizophrenia and bipolar disorder gene network. J Neurogenet 2010; 24: 182-93.

Penagarikano O, Geschwind DH. What does CNTNAP2 reveal about autism spectrum disorder? Trends Mol Med 2012; 18: 156-63.

Pennacchio LA, Rubin EM. Genomic strategies to identify mammalian regulatory sequences. Nat Rev Genet 2001; 2: 100-9.

Perez R, Tupac-Yupanqui I, Dunner S. Evaluation of suitable reference genes for gene expression studies in bovine muscular tissue. BMC Mol Biol 2008; 9: 79.

Perl O, Ilani T, Strous RD, Lapidus R, Fuchs S. The alpha7 nicotinic acetylcholine receptor in schizophrenia: decreased mRNA levels in peripheral blood lymphocytes. Faseb J 2003; 17: 1948-50.

Perry W, Braff DL. Information-processing deficits and thought disorder in schizophrenia. Am J Psychiatry 1994; 151: 363-7.

Petroni M, Veschi V, Prodosmo A, Rinaldo C, Massimi I, Carbonari M, et al. MYCN sensitizes human neuroblastoma to apoptosis by HIPK2 activation through a DNA damage response. Mol Cancer Res 2011; 9: 67-77.

Petryshen TL, Middleton FA, Kirby A, Aldinger KA, Purcell S, Tahl AR, et al. Support for involvement of neuregulin 1 in schizophrenia pathophysiology. Mol Psychiatry 2005; 10: 366-74, 328.

Phiel CJ, Zhang F, Huang EY, Guenther MG, Lazar MA, Klein PS. Histone deacetylase is a direct target of valproic acid, a potent anticonvulsant, mood stabilizer, and teratogen. J Biol Chem 2001; 276: 36734-41.

Pickard BS, Malloy MP, Christoforou A, Thomson PA, Evans KL, Morris SW, et al. Cytogenetic and genetic evidence supports a role for the kainate-type glutamate receptor gene, GRIK4, in schizophrenia and bipolar disorder. Mol Psychiatry 2006; 11: 847-57.

Pickard BS, Malloy MP, Porteous DJ, Blackwood DH, Muir WJ. Disruption of a brain transcription factor, NPAS3, is associated with schizophrenia and learning disability. Am J Med Genet B Neuropsychiatr Genet 2005; 136B: 26-32.

Pittenger C, Duman RS. Stress, depression, and neuroplasticity: a convergence of mechanisms. Neuropsychopharmacology 2008; 33: 88-109.

Prabakaran S, Swatton JE, Ryan MM, Huffaker SJ, Huang JT, Griffin JL, et al. Mitochondrial dysfunction in schizophrenia: evidence for compromised brain metabolism and oxidative stress. Mol Psychiatry 2004; 9: 684-97, 643.

Prabhakar S, Poulin F, Shoukry M, Afzal V, Rubin EM, Couronne O, et al. Close sequence comparisons are sufficient to identify human cis-regulatory elements. Genome Res 2006; 16: 855-63.

Prata DP, Breen G, Osborne S, Munro J, St Clair D, Collier DA. An association study of the neuregulin 1 gene, bipolar affective disorder and psychosis. Psychiatr Genet 2009; 19: 113-6.

Priebe L, Degenhardt FA, Herms S, Haenisch B, Mattheisen M, Nieratschker V, et al. Genome-wide survey implicates the influence of copy number variants (CNVs) in the development of early-onset bipolar disorder. Mol Psychiatry 2011; 17: 421-32.

Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 2009; 460: 748-52.

Quednow BB, Brinkmeyer J, Mobascher A, Nothnagel M, Musso F, Grunder G, et al. Schizophrenia risk polymorphisms in the TCF4 gene interact with smoking in the modulation of auditory sensory gating. Proc Natl Acad Sci U S A 2012; 109: 6271-6.

Rajkowska G, Selemon LD, Goldman-Rakic PS. Neuronal and glial somal size in the prefrontal cortex: a postmortem morphometric study of schizophrenia and Huntington disease. Arch Gen Psychiatry 1998; 55: 215-24.

Ramirez-Carrozzi VR, Braas D, Bhatt DM, Cheng CS, Hong C, Doty KR, et al. A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. Cell 2009; 138: 114-28.

Ramsey AJ, Milenkovic M, Oliveira AF, Escobedo-Lozoya Y, Seshadri S, Salahpour A, et al. Impaired NMDA receptor transmission alters striatal synapses and DISC1 protein in an age-dependent manner. Proc Natl Acad Sci U S A 2011; 108: 5795-800.

Rastogi A, Zai C, Likhodi O, Kennedy JL, Wong AH. Genetic association and post-mortem brain mRNA analysis of DISC1 and related genes in schizophrenia. Schizophr Res 2009; 114: 39-49.

Ray BK, Dhar S, Shakya A, Ray A. Z-DNA-forming silencer in the first exon regulates human ADAM-12 gene expression. Proc Natl Acad Sci U S A 2010; 108: 103-8.

Reimers-Kipping S, Hevers W, Paabo S, Enard W. Humanized Foxp2 specifically affects cortico-basal ganglia circuits. Neuroscience 2011; 175: 75-84.

Rich A, Nordheim A, Azorin F. Stabilization and detection of natural left-handed Z-DNA. J Biomol Struct Dyn 1983; 1: 1-19.

Richards AL, Jones L, Moskvina V, Kirov G, Gejman PV, Levinson DF, et al. Schizophrenia susceptibility alleles are enriched for alleles that affect gene expression in adult human brain. Mol Psychiatry 2012; 17: 193-201.

Rieff HI, Raetzman LT, Sapp DW, Yeh HH, Siegel RE, Corfas G. Neuregulin induces GABA(A) receptor subunit expression and neurite outgrowth in cerebellar granule cells. J Neurosci 1999; 19: 10757-66.

Riley B, Thiselton D, Maher BS, Bigdeli T, Wormley B, McMichael GO, et al. Replication of association between schizophrenia and ZNF804A in the Irish Case-Control Study of Schizophrenia sample. Mol Psychiatry 2010; 15: 29-37.

Rio C, Rieff HI, Qi P, Khurana TS, Corfas G. Neuregulin and erbB receptors play a critical role in neuronal migration. Neuron 1997; 19: 39-50.

Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA, et al. Genome-wide association study identifies five new schizophrenia loci. Nat Genet 2011; 43: 969-76.

Risch N. Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 1990; 46: 222-8.

Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, Shannon PT, et al. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science 2010; 328: 636-9.

Robb JC, Young LT, Cooke RG, Joffe RT. Gender differences in patients with bipolar disorder influence outcome in the medical outcomes survey (SF-20) subscale scores. J Affect Disord 1998; 49: 189-93.

Roll P, Rudolf G, Pereira S, Royer B, Scheffer IE, Massacrier A, et al. SRPX2 mutations in disorders of language cortex and cognition. Hum Mol Genet 2006; 15: 1195-207.

Roll P, Vernes SC, Bruneau N, Cillario J, Ponsole-Lenfant M, Massacrier A, et al. Molecular networks implicated in speech-related disorders: FOXP2 regulates the SRPX2/uPAR complex. Hum Mol Genet 2010; 19: 4848-60.

Rosenbloom KR, Dreszer TR, Pheasant M, Barber GP, Meyer LR, Pohl A, et al. ENCODE whole-genome data in the UCSC Genome Browser. Nucleic Acids Res 2010; 38: D620-5.

Ross CA, Margolis RL, Reading SA, Pletnikov M, Coyle JT. Neurobiology of schizophrenia. Neuron 2006; 52: 139-53.

Ross KA. Evidence for somatic gene conversion and deletion in bipolar disorder, Crohn's disease, coronary artery disease, hypertension, rheumatoid arthritis, type-1 diabetes, and type-2 diabetes. BMC Med 2011; 9: 12.

Ruano D, Aulchenko YS, Macedo A, Soares MJ, Valente J, Azevedo MH, et al. Association of the gene encoding neurogranin with schizophrenia in males. J Psychiatr Res 2008; 42: 125-33.

Ruderfer DM, Kirov G, Chambert K, Moran JL, Owen MJ, O'Donovan MC, et al. A family-based study of common polygenic variation and risk of schizophrenia. Mol Psychiatry 2011; 16: 887-8.

Rujescu D, Ingason A, Cichon S, Pietilainen OP, Barnes MR, Toulopoulou T, et al. Disruption of the neurexin 1 gene is associated with schizophrenia. Hum Mol Genet 2009; 18: 988-96.

Russo-Neustadt A, Ha T, Ramirez R, Kesslak JP. Physical activity-antidepressant treatment combination: impact on brain-derived neurotrophic factor and behavior in an animal model. Behav Brain Res 2001; 120: 87-95.

Ryan MM, Lockstone HE, Huffaker SJ, Wayland MT, Webster MJ, Bahn S. Gene expression analysis of bipolar disorder reveals downregulation of the ubiquitin cycle and alterations in synaptic genes. Mol Psychiatry 2006; 11: 965-78.

Saetre P, Agartz I, De Franciscis A, Lundmark P, Djurovic S, Kahler A, et al. Association between a disrupted-in-schizophrenia 1 (DISC1) single nucleotide polymorphism and schizophrenia in a combined Scandinavian case-control sample. Schizophr Res 2008; 106: 237-41.

Saetre P, Emilsson L, Axelsson E, Kreuger J, Lindholm E, Jazin E. Inflammation-related genes up-regulated in schizophrenia brains. BMC Psychiatry 2007; 7: 46.

Saha S, Chant D, Welham J, McGrath J. A systematic review of the prevalence of schizophrenia. PLoS Med 2005; 2: e141.

Sanches M, Keshavan MS, Brambilla P, Soares JC. Neurodevelopmental basis of bipolar disorder: a critical appraisal. Prog Neuropsychopharmacol Biol Psychiatry 2008; 32: 1617-27.

Sandelin A, Carninci P, Lenhard B, Ponjavic J, Hayashizaki Y, Hume DA. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nat Rev Genet 2007; 8: 424-36.

Sanders AR, Duan J, Levinson DF, Shi J, He D, Hou C, et al. No significant association of 14 candidate genes with schizophrenia in a large European ancestry sample: implications for psychiatric genetics. Am J Psychiatry 2008; 165: 497-506.

Sanjuan J, Tolosa A, Gonzalez JC, Aguilar EJ, Molto MD, Najera C, et al. FOXP2 polymorphisms in patients with schizophrenia. Schizophr Res 2005; 73: 253-6.

Sanjuan J, Tolosa A, Gonzalez JC, Aguilar EJ, Perez-Tur J, Najera C, et al. Association between FOXP2 polymorphisms and schizophrenia with auditory hallucinations. Psychiatr Genet 2006; 16: 67-72.

Sawamura N, Sawamura-Yamamoto T, Ozeki Y, Ross CA, Sawa A. A form of DISC1 enriched in nucleus: altered subcellular distribution in orbitofrontal cortex in psychosis and substance/alcohol abuse. Proc Natl Acad Sci U S A 2005; 102: 1187-92.

Saykin AJ, Shtasel DL, Gur RE, Kester DB, Mozley LH, Stafiniak P, et al. Neuropsychological deficits in neuroleptic naive patients with first-episode schizophrenia. Arch Gen Psychiatry 1994; 51: 124-31.

Scarr E, Cowie TF, Kanellakis S, Sundram S, Pantelis C, Dean B. Decreased cortical muscarinic receptors define a subgroup of subjects with schizophrenia. Mol Psychiatry 2009; 14: 1017-23.

Scarr E, Pavey G, Sundram S, MacKinnon A, Dean B. Decreased hippocampal NMDA, but not kainate or AMPA receptors in bipolar disorder. Bipolar Disord 2003; 5: 257-64.

Scharff C, Petri J. Evo-devo, deep homology and FoxP2: implications for the evolution of speech and language. Philos Trans R Soc Lond B Biol Sci 2011; 366: 2124-40.

Schloesser RJ, Huang J, Klein PS, Manji HK. Cellular plasticity cascades in the pathophysiology and treatment of bipolar disorder. Neuropsychopharmacology 2008; 33: 110-33.

Schosser A, Gaysina D, Cohen-Woods S, Chow PC, Martucci L, Craddock N, et al. Association of DISC1 and TSNAX genes and affective disorders in the depression case-control (DeCC) and bipolar affective case-control (BACCS) studies. Mol Psychiatry 2010; 15: 844-9.

Schroeder DI, Myers RM. Multiple transcription start sites for FOXP2 with varying cellular specificities. Gene 2008; 413: 42-8.

Schug J, Schuller WP, Kappen C, Salbaum JM, Bucan M, Stoeckert CJ, Jr. Promoter features related to tissue specificity as measured by Shannon entropy. Genome Biol 2005; 6: R33.

Schumacher J, Laje G, Abou Jamra R, Becker T, Muhleisen TW, Vasilescu C, et al. The DISC locus and schizophrenia: evidence from an association study in a central European sample and from a meta-analysis across different European populations. Hum Mol Genet 2009; 18: 2719-27.

Schurov IL, Handford EJ, Brandon NJ, Whiting PJ. Expression of disrupted in schizophrenia 1 (DISC1) protein in the adult and developing mouse brain indicates its role in neurodevelopment. Mol Psychiatry 2004; 9: 1100-10.

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, et al. Large-scale copy number polymorphism in the human genome. Science 2004; 305: 525-8.

Seeman P. Dopamine receptor sequences. Therapeutic levels of neuroleptics occupy D2 receptors, clozapine occupies D4. Neuropsychopharmacology 1992; 7: 261-84.

Seeman P, Lee T. Antipsychotic drugs: direct correlation between clinical potency and presynaptic action on dopamine neurons. Science 1975; 188: 1217-9.

Segurado R, Detera-Wadleigh SD, Levinson DF, Lewis CM, Gill M, Nurnberger JI, Jr., et al. Genome scan meta-analysis of schizophrenia and bipolar disorder, part III: Bipolar disorder. Am J Hum Genet 2003; 73: 49-62.

Sei Y, Ren-Patterson R, Li Z, Tunbridge EM, Egan MF, Kolachana BS, et al. Neuregulin1-induced cell migration is impaired in schizophrenia: association with neuregulin1 and catechol-o-methyltransferase gene polymorphisms. Mol Psychiatry 2007; 12: 946-57.

Seifuddin F, Mahon PB, Judy J, Pirooznia M, Jancic D, Taylor J, et al. Meta-analysis of genetic association studies on bipolar disorder. Am J Med Genet B Neuropsychiatr Genet 2012; 159: 508-18.

Selemon LD, Goldman-Rakic PS. The reduced neuropil hypothesis: a circuit based model of schizophrenia. Biol Psychiatry 1999; 45: 17-25.

Sequeira PA, Martin MV, Vawter MP. The first decade and beyond of transcriptional profiling in schizophrenia. Neurobiol Dis 2012; 45: 23-36.

Serretti A, Mandelli L. The genetics of bipolar disorder: genome 'hot regions,' genes, new potential candidates and future directions. Mol Psychiatry 2008; 13: 742-71.

Seshadri S, Kamiya A, Yokota Y, Prikulis I, Kano S, Hayashi-Takagi A, et al. Disrupted-in-Schizophrenia-1 expression is regulated by beta-site amyloid precursor protein

cleaving enzyme-1-neuregulin cascade. Proc Natl Acad Sci U S A 2010; 107: 5622-7.

Shaffer AL, Emre NC, Romesser PB, Staudt LM. IRF4: Immunity. Malignancy! Therapy? Clin Cancer Res 2009; 15: 2954-61.

Shaikh M, Hall MH, Schulze K, Dutt A, Li K, Williams I, et al. Effect of DISC1 on the P300 Waveform in Psychosis. Schizophr Bull 2011.

Shaldubina A, Agam G, Belmaker RH. The mechanism of lithium action: state of the art, ten years later. Prog Neuropsychopharmacol Biol Psychiatry 2001; 25: 855-66.

Sham P. Statistics in Human Genetics. New York: Oxford University Press, 1998.

Shao H, Burrage LC, Sinasac DS, Hill AE, Ernest SR, O'Brien W, et al. Genetic architecture of complex traits: large phenotypic effects and pervasive epistasis. Proc Natl Acad Sci U S A 2008; 105: 19910-4.

Shao L, Vawter MP. Shared gene expression alterations in schizophrenia and bipolar disorder. Biol Psychiatry 2008; 64: 89-97.

Sharma RP, Grayson DR, Gavin DP. Histone deactylase 1 expression is increased in the prefrontal cortex of schizophrenia subjects: analysis of the National Brain Databank microarray collection. Schizophr Res 2008; 98: 111-7.

Shastry BS. Bipolar disorder: an update. Neurochem Int 2005; 46: 273-9.

Shen YC, Tsai HM, Cheng MC, Hsu SH, Chen SF, Chen CH. Genetic and functional analysis of the gene encoding neurogranin in schizophrenia. Schizophr Res 2012; 137: 7-13.

Shoji H, Toyama K, Takamiya Y, Wakana S, Gondo Y, Miyakawa T. Comprehensive behavioral analysis of ENU-induced Disc1-Q31L and -L100P mutant mice. BMC Res Notes 2012; 5: 108.

Shriberg LD, Ballard KJ, Tomblin JB, Duffy JR, Odell KH, Williams CA. Speech, prosody, and voice characteristics of a mother and daughter with a 7;13 translocation affecting FOXP2. J Speech Lang Hear Res 2006; 49: 500-25.

Shu W, Yang H, Zhang L, Lu MM, Morrisey EE. Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors. J Biol Chem 2001; 276: 27488-97.

Siegel DJ, Minshew NJ, Goldstein G. Wechsler IQ profiles in diagnosis of high-functioning autism. J Autism Dev Disord 1996; 26: 389-406.

Silver H, Feldman P, Bilker W, Gur RC. Working memory deficit as a core neuropsychological dysfunction in schizophrenia. Am J Psychiatry 2003; 160: 1809-16.

Singh KK, De Rienzo G, Drane L, Mao Y, Flood Z, Madison J, et al. Common DISC1 polymorphisms disrupt Wnt/GSK3beta signaling and brain development. Neuron 2011; 72: 545-58.

Sklar P, Ripke S, Scott LJ, Andreassen OA, Cichon S, Craddock N, et al. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat Genet 2011; 43: 977-83.

Sklar P, Smoller JW, Fan J, Ferreira MA, Perlis RH, Chambert K, et al. Whole-genome association study of bipolar disorder. Mol Psychiatry 2008; 13: 558-69.

Smale ST, Kadonaga JT. The RNA polymerase II core promoter. Annu Rev Biochem 2003; 72: 449-79.

Smith CL, Hager GL. Transcriptional regulation of mammalian genes in vivo. A tale of two templates. J Biol Chem 1997; 272: 27493-6.

Snitz BE, Macdonald AW, 3rd, Carter CS. Cognitive deficits in unaffected first-degree relatives of schizophrenia patients: a meta-analytic review of putative endophenotypes. Schizophr Bull 2006; 32: 179-94.

Snyder SH. Action of anticholinergic drugs on striatal dopamine. Pharmacol Ther B 1976; 2: 65-70.

Soares DC, Carlyle BC, Bradshaw NJ, Porteous DJ. DISC1: Structure, Function, and Therapeutic Potential for Major Mental Illness. ACS Chem Neurosci 2011; 2: 609-632.

Solomon DL, Amati B, Land H. Distinct DNA binding preferences for the c-Myc/Max and Max/Max dimers. Nucleic Acids Res 1993; 21: 5372-6.

Song XQ, Lv LX, Li WQ, Hao YH, Zhao JP. The interaction of nuclear factor-kappa B and cytokines is associated with schizophrenia. Biol Psychiatry 2009; 65: 481-8.

Spaniel F, Horacek J, Tintera J, Ibrahim I, Novak T, Cermak J, et al. Genetic variation in FOXP2 alters grey matter concentrations in schizophrenia patients. Neurosci Lett 2011; 493: 131-5.

Spengler BA, Biedler JL, Ross RA. A corrected karyotype for the SH-SY5Y human neuroblastoma cell line. Cancer Genet Cytogenet 2002; 138: 177-8.

Spiegelman BM, Heinrich R. Biological control through regulated transcriptional coactivators. Cell 2004; 119: 157-67.

Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA. Interchromosomal associations between alternatively expressed loci. Nature 2005; 435: 637-45.

Spiteri E, Konopka G, Coppola G, Bomar J, Oldham M, Ou J, et al. Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. Am J Hum Genet 2007; 81: 1144-57.

Spitzer RL, Williams JB, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R (SCID). I: History, rationale, and description. Arch Gen Psychiatry 1992; 49: 624-9.

Sprooten E, Lymer GK, Munoz Maniega S, McKirdy J, Clayden JD, Bastin ME, et al. The relationship of anterior thalamic radiation integrity to psychosis risk associated neuregulin-1 variants. Mol Psychiatry 2009; 14: 237-8, 233.

Sprooten E, Sussmann JE, Clugston A, Peel A, McKirdy J, Moorhead TW, et al. White matter integrity in individuals at high genetic risk of bipolar disorder. Biol Psychiatry 2011a; 70: 350-6.

Sprooten E, Sussmann JE, Moorhead TW, Whalley HC, Ffrench-Constant C, Blumberg HP, et al. Association of white matter integrity with genetic variation in an exonic DISC1 SNP. Mol Psychiatry 2011b; 16: 685, 688-9.

St Clair D, Blackwood D, Muir W, Carothers A, Walker M, Spowart G, et al. Association within a family of a balanced autosomal translocation with major mental illness. Lancet 1990; 336: 13-6.

Stefansson H, Sarginson J, Kong A, Yates P, Steinthorsdottir V, Gudfinnsson E, et al. Association of neuregulin 1 with schizophrenia confirmed in a Scottish population. Am J Hum Genet 2003; 72: 83-7.

Stefansson H, Sigurdsson E, Steinthorsdottir V, Bjornsdottir S, Sigmundsson T, Ghosh S, et al. Neuregulin 1 and susceptibility to schizophrenia. Am J Hum Genet 2002; 71: 877-92.

Stein JL, Medland SE, Vasquez AA, Hibar DP, Senstad RE, Winkler AM, et al. Identification of common variants associated with human hippocampal and intracranial volumes. Nat Genet 2012; 44: 552-61.

Steinberg S, Mors O, Borglum AD, Gustafsson O, Werge T, Mortensen PB, et al. Expanding the range of ZNF804A variants conferring risk of psychosis. Mol Psychiatry 2011; 16: 59-66.

Steinecke A, Gampe C, Valkova C, Kaether C, Bolz J. Disrupted-in-Schizophrenia 1 (DISC1) is necessary for the correct migration of cortical interneurons. J Neurosci 2012; 32: 738-45.

Steinthorsdottir V, Stefansson H, Ghosh S, Birgisdottir B, Bjornsdottir S, Fasquel AC, et al. Multiple novel transcription initiation sites for NRG1. Gene 2004; 342: 97-105.

Stober G, Kocher I, Franzek E, Beckmann H. First-trimester maternal gestational infection and cycloid psychosis. Acta Psychiatr Scand 1997; 96: 319-24.

Stormo GD. DNA binding sites: representation and discovery. Bioinformatics 2000; 16: 16-23.

Stothard P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 2000; 28: 1102, 1104.

Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 2007; 315: 848-53.

Stroud JC, Wu Y, Bates DL, Han A, Nowick K, Paabo S, et al. Structure of the forkhead domain of FOXP2 bound to DNA. Structure 2006; 14: 159-66.

Sudhof TC. Neuroligins and neurexins link synaptic function to cognitive disease. Nature 2008; 455: 903-11.

Sugai T, Kawamura M, Iritani S, Araki K, Makifuchi T, Imai C, et al. Prefrontal abnormality of schizophrenia revealed by DNA microarray: impact on glial and neurotrophic gene expression. Ann N Y Acad Sci 2004; 1025: 84-91.

Sullivan PF. The genetics of schizophrenia. PLoS Med 2005; 2: e212.

Sullivan PF, Fan C, Perou CM. Evaluating the comparability of gene expression in blood and brain. Am J Med Genet B Neuropsychiatr Genet 2006; 141B: 261-8.

Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch Gen Psychiatry 2003; 60: 1187-92.

Sullivan PF, Lin D, Tzeng JY, van den Oord E, Perkins D, Stroup TS, et al. Genomewide association for schizophrenia in the CATIE study: results of stage 1. Mol Psychiatry 2008; 13: 570-84.

Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. Am J Psychiatry 2000; 157: 1552-62.

Sun J, Kuo PH, Riley BP, Kendler KS, Zhao Z. Candidate genes for schizophrenia: a survey of association studies and gene ranking. Am J Med Genet B Neuropsychiatr Genet 2008; 147B: 1173-81.

Sun Y, Wang Y, Hu Y, Chen G, Ma H. Comparative analysis of neural transcriptomes and functional implication of unannotated intronic expression. BMC Genomics 2011; 12: 494.

Sussman CR, Vartanian T, Miller RH. The ErbB4 neuregulin receptor mediates suppression of oligodendrocyte maturation. J Neurosci 2005; 25: 5757-62.

Swerdlow NR, Geyer MA. Using an animal model of deficient sensorimotor gating to study the pathophysiology and new treatments of schizophrenia. Schizophr Bull 1998; 24: 285-301.

Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ, et al. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat Genet 2007; 39: 319-28.

Takahashi H, Takahashi K, Liu FC. FOXP genes, neural development, speech and language disorders. Adv Exp Med Biol 2009; 665: 117-29.

Takahashi K, Liu FC, Hirokawa K, Takahashi H. Expression of Foxp2, a gene involved in speech and language, in the developing and adult striatum. J Neurosci Res 2003; 73: 61-72.

Takahata R, Moghaddam B. Target-specific glutamatergic regulation of dopamine neurons in the ventral tegmental area. J Neurochem 2000; 75: 1775-8.

Talbot K, Eidem WL, Tinsley CL, Benson MA, Thompson EW, Smith RJ, et al. Dysbindin-1 is reduced in intrinsic, glutamatergic terminals of the hippocampal formation in schizophrenia. J Clin Invest 2004; 113: 1353-63.

Tan W, Wang Y, Gold B, Chen J, Dean M, Harrison PJ, et al. Molecular cloning of a brain-specific, developmentally regulated neuregulin 1 (NRG1) isoform and identification of a functional promoter variant associated with schizophrenia. J Biol Chem 2007; 282: 24343-51.

Tanabe Y, Fujiwara Y, Matsuzaki A, Fujita E, Kasahara T, Yuasa S, et al. Temporal expression and mitochondrial localization of a Foxp2 isoform lacking the forkhead domain in developing Purkinje cells. J Neurochem 2012; 122: 72-80.

Tang J, LeGros RP, Louneva N, Yeh L, Cohen JW, Hahn CG, et al. Dysbindin-1 in dorsolateral prefrontal cortex of schizophrenia cases is reduced in an isoform-specific manner unrelated to dysbindin-1 mRNA expression. Hum Mol Genet 2009; 18: 3851-63.

Tang JX, Chen WY, He G, Zhou J, Gu NF, Feng GY, et al. Polymorphisms within 5' end of the Neuregulin 1 gene are genetically associated with schizophrenia in the Chinese population. Mol Psychiatry 2004; 9: 11-2.

Taveggia C, Zanazzi G, Petrylak A, Yano H, Rosenbluth J, Einheber S, et al. Neuregulin-1 type III determines the ensheathment fate of axons. Neuron 2005; 47: 681-94.

Taylor J, Tyekucheva S, King DC, Hardison RC, Miller W, Chiaromonte F. ESPERR: learning strong and weak signals in genomic sequence alignments to identify functional elements. Genome Res 2006; 16: 1596-604.

Taylor MS, Devon RS, Millar JK, Porteous DJ. Evolutionary constraints on the Disrupted in Schizophrenia locus. Genomics 2003; 81: 67-77.

Thellin O, Zorzi W, Lakaye B, De Borman B, Coumans B, Hennen G, et al. Housekeeping genes as internal standards: use and limits. J Biotechnol 1999; 75: 291-5.

Thomason ME, Thompson PM. Diffusion imaging, white matter, and psychopathology. Annu Rev Clin Psychol 2011; 7: 63-85.

Thompson Ray M, Weickert CS, Wyatt E, Webster MJ. Decreased BDNF, trkB-TK+ and GAD67 mRNA expression in the hippocampus of individuals with schizophrenia and mood disorders. J Psychiatry Neurosci 2011; 36: 195-203.

Thomsen MS, Weyn A, Mikkelsen JD. Hippocampal alpha7 nicotinic acetylcholine receptor levels in patients with schizophrenia, bipolar disorder, or major depressive disorder. Bipolar Disord 2011; 13: 701-7.

Thomson PA, Christoforou A, Morris SW, Adie E, Pickard BS, Porteous DJ, et al. Association of Neuregulin 1 with schizophrenia and bipolar disorder in a second cohort from the Scottish population. Mol Psychiatry 2007; 12: 94-104.

Thomson PA, Harris SE, Starr JM, Whalley LJ, Porteous DJ, Deary IJ. Association between genotype at an exonic SNP in DISC1 and normal cognitive aging. Neurosci Lett 2005; 389: 41-5.

Tijssen MR, Cvejic A, Joshi A, Hannah RL, Ferreira R, Forrai A, et al. Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. Dev Cell 2011; 20: 597-609.

Tkachev D, Mimmack ML, Ryan MM, Wayland M, Freeman T, Jones PB, et al. Oligodendrocyte dysfunction in schizophrenia and bipolar disorder. Lancet 2003; 362: 798-805.

Tochigi M, Iwamoto K, Bundo M, Komori A, Sasaki T, Kato N, et al. Methylation status of the reelin promoter region in the brain of schizophrenic patients. Biol Psychiatry 2008; 63: 530-3.

Tomblin JB, O'Brien M, Shriberg LD, Williams C, Murray J, Patil S, et al. Language features in a mother and daughter of a chromosome 7;13 translocation involving FOXP2. J Speech Lang Hear Res 2009; 52: 1157-74.

Tomita H, Vawter MP, Walsh DM, Evans SJ, Choudary PV, Li J, et al. Effect of agonal and postmortem factors on gene expression profile: quality control in microarray analyses of postmortem human brain. Biol Psychiatry 2004; 55: 346-52.

Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG, Ioannidis JP. Establishment of genetic associations for complex diseases is independent of early study findings. Eur J Hum Genet 2004; 12: 762-9.

Tsuang MT, Faraone SV, Lyons MJ. Identification of the phenotype in psychiatric genetics. Eur Arch Psychiatry Clin Neurosci 1993; 243: 131-42.

Uddin MJ, Cinar MU, Tesfaye D, Looft C, Tholen E, Schellander K. Age-related changes in relative expression stability of commonly used housekeeping genes in selected porcine tissues. BMC Res Notes 2011; 4: 441.

Uhlhaas PJ, Singer W. Abnormal neural oscillations and synchrony in schizophrenia. Nat Rev Neurosci 2010; 11: 100-13.

Uranova NA, Vostrikov VM, Orlovskaya DD, Rachmanova VI. Oligodendroglial density in the prefrontal cortex in schizophrenia and mood disorders: a study from the Stanley Neuropathology Consortium. Schizophr Res 2004; 67: 269-75.

van Os J, Kapur S. Schizophrenia. Lancet 2009; 374: 635-45.

van Spronsen M, Hoogenraad CC. Synapse pathology in psychiatric and neurologic disease. Curr Neurol Neurosci Rep 2010; 10: 207-14.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002; 3: RESEARCH0034.

Vargha-Khadem F, Watkins K, Alcock K, Fletcher P, Passingham R. Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. Proc Natl Acad Sci U S A 1995; 92: 930-3.

Vargha-Khadem F, Watkins KE, Price CJ, Ashburner J, Alcock KJ, Connelly A, et al. Neural basis of an inherited speech and language disorder. Proc Natl Acad Sci U S A 1998; 95: 12695-700.

Vartanian T, Fischbach G, Miller R. Failure of spinal cord oligodendrocyte development in mice lacking neuregulin. Proc Natl Acad Sci U S A 1999; 96: 731-5.

Vasudevan S, Tong Y, Steitz JA. Switching from repression to activation: microRNAs can up-regulate translation. Science 2007; 318: 1931-4.

Vawter MP, Tomita H, Meng F, Bolstad B, Li J, Evans S, et al. Mitochondrial-related gene expression changes are sensitive to agonal-pH state: implications for brain disorders. Mol Psychiatry 2006; 11: 615, 663-79.

Veldic M, Caruncho HJ, Liu WS, Davis J, Satta R, Grayson DR, et al. DNA-methyltransferase 1 mRNA is selectively overexpressed in telencephalic GABAergic interneurons of schizophrenia brains. Proc Natl Acad Sci U S A 2004; 101: 348-53.

Veldic M, Guidotti A, Maloku E, Davis JM, Costa E. In psychosis, cortical interneurons overexpress DNA-methyltransferase 1. Proc Natl Acad Sci U S A 2005; 102: 2152-7.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science 2001; 291: 1304-51.

Vernes SC, Newbury DF, Abrahams BS, Winchester L, Nicod J, Groszer M, et al. A functional genetic link between distinct developmental language disorders. N Engl J Med 2008; 359: 2337-45.

Vernes SC, Nicod J, Elahi FM, Coventry JA, Kenny N, Coupe AM, et al. Functional genetic analysis of mutations implicated in a human speech and language disorder. Hum Mol Genet 2006; 15: 3154-67.

Vernes SC, Oliver PL, Spiteri E, Lockstone HE, Puliyadi R, Taylor JM, et al. Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. PLoS Genet 2011; 7: e1002145.

Vernes SC, Spiteri E, Nicod J, Groszer M, Taylor JM, Davies KE, et al. High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. Am J Hum Genet 2007; 81: 1232-50.

Verrotti A, Spalice A, Ursitti F, Papetti L, Mariani R, Castronovo A, et al. New trends in neuronal migration disorders. Eur J Paediatr Neurol 2010; 14: 1-12.

Vita A, De Peri L, Sacchetti E. Gray matter, white matter, brain, and intracranial volumes in first-episode bipolar disorder: a meta-analysis of magnetic resonance imaging studies. Bipolar Disord 2009; 11: 807-14.

Vita A, De Peri L, Silenzi C, Dieci M. Brain morphology in first-episode schizophrenia: a meta-analysis of quantitative magnetic resonance imaging studies. Schizophr Res 2006; 82: 75-88.

Vostrikov VM, Uranova NA, Orlovskaya DD. Deficit of perineuronal oligodendrocytes in the prefrontal cortex in schizophrenia and mood disorders. Schizophr Res 2007; 94: 273-80.

Walker RM, Christoforou A, Thomson PA, McGhee KA, Maclean A, Muhleisen TW, et al. Association analysis of Neuregulin 1 candidate regions in schizophrenia and bipolar disorder. Neurosci Lett 2010; 478: 9-13.

Walker RM, Hill AE, Newman AC, Hamilton G, Torrance HS, Anderson SM, et al. The DISC1 promoter: characterization and regulation by FOXP2. Hum Mol Genet 2012; 21: 2862-72.

Walsh J, Desbonnet L, Clarke N, Waddington JL, O'Tuathaigh CM. Disruption of exploratory and habituation behavior in mice with mutation of DISC1: an ethologically based analysis. J Neurosci Res 2012; 90: 1445-53.

Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 2008; 320: 539-43.

Wang AY, Lohmann KM, Yang CK, Zimmerman EI, Pantazopoulos H, Herring N, et al. Bipolar disorder type 1 and schizophrenia are accompanied by decreased density of parvalbumin- and somatostatin-positive interneurons in the parahippocampal region. Acta Neuropathol 2011a; 122: 615-26.

Wang B, Lin D, Li C, Tucker P. Multiple domains define the expression and regulatory properties of Foxp1 forkhead transcriptional repressors. J Biol Chem 2003; 278: 24259-68.

Wang B, Ren J, Ooi LL, Chong SS, Lee CG. Dinucleotide repeats negatively modulate the promoter activity of Cyr61 and is unstable in hepatocellular carcinoma patients. Oncogene 2005; 24: 3999-4008.

Wang JY, Miller SJ, Falls DL. The N-terminal region of neuregulin isoforms determines the accumulation of cell surface and released neuregulin ectodomain. J Biol Chem 2001; 276: 2841-51.

Wang KS, Liu XF, Aragam N. A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. Schizophr Res 2010; 124: 192-9.

Wang Q, Charych EI, Pulito VL, Lee JB, Graziane NM, Crozier RA, et al. The psychiatric disease risk factors DISC1 and TNIK interact to regulate synapse composition and function. Mol Psychiatry 2011b; 16: 1006-23.

Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 2012; 40: D930-4.

Watkins KE, Dronkers NF, Vargha-Khadem F. Behavioural analysis of an inherited speech and language disorder: comparison with acquired aphasia. Brain 2002a; 125: 452-64.

Watkins KE, Vargha-Khadem F, Ashburner J, Passingham RE, Connelly A, Friston KJ, et al. MRI analysis of an inherited speech and language disorder: structural brain abnormalities. Brain 2002b; 125: 465-78.

Weickert CS, Hyde TM, Lipska BK, Herman MM, Weinberger DR, Kleinman JE. Reduced brain-derived neurotrophic factor in prefrontal cortex of patients with schizophrenia. Mol Psychiatry 2003; 8: 592-610.

Weickert CS, Rothmond DA, Hyde TM, Kleinman JE, Straub RE. Reduced DTNBP1 (dysbindin-1) mRNA in the hippocampal formation of schizophrenia patients. Schizophr Res 2008; 98: 105-10.

Weickert CS, Straub RE, McClintock BW, Matsumoto M, Hashimoto R, Hyde TM, et al. Human dysbindin (DTNBP1) gene expression in normal brain and in schizophrenic prefrontal cortex and midbrain. Arch Gen Psychiatry 2004; 61: 544-55.

Weickert CS, Tiwari Y, Schofield PR, Mowry BJ, Fullerton JM. Schizophrenia-associated HapICE haplotype is associated with increased NRG1 type III expression and high nucleotide diversity. Transl Psychiatry 2012; 2: e104.

Weigelt K, Carvalho LA, Drexhage RC, Wijkhuijs A, de Wit H, van Beveren NJ, et al. TREM-1 and DAP12 expression in monocytes of patients with severe psychiatric disorders. EGR3, ATF3 and PU.1 as important transcription factors. Brain Behav Immun 2011; 25: 1162-9.

Weinberger DR. Implications of normal brain development for the pathogenesis of schizophrenia. Arch Gen Psychiatry 1987; 44: 660-9.

Weinberger DR, Egan MF, Bertolino A, Callicott JH, Mattay VS, Lipska BK, et al. Prefrontal neurons and the genetics of schizophrenia. Biol Psychiatry 2001; 50: 825-44.

Wender PH, Kety SS, Rosenthal D, Schulsinger F, Ortmann J, Lunde I. Psychiatric disorders in the biological and adoptive families of adopted individuals with affective disorders. Arch Gen Psychiatry 1986; 43: 923-9.

Westerink BH, Enrico P, Feimann J, De Vries JB. The pharmacology of mesocortical dopamine neurons: a dual-probe microdialysis study in the ventral tegmental area and prefrontal cortex of the rat brain. J Pharmacol Exp Ther 1998; 285: 143-54.

Whalley HC, Sussmann JE, Johnstone M, Romaniuk L, Redpath H, Chakirova G, et al. Effects of a mis-sense DISC1 variant on brain activation in two cohorts at high risk of bipolar disorder or schizophrenia. Am J Med Genet B Neuropsychiatr Genet 2012; 159B: 343-53.

Whitington T, Perkins AC, Bailey TL. High-throughput chromatin information enables accurate tissue-specific prediction of transcription factor binding sites. Nucleic Acids Res 2009; 37: 14-25.

WHO. The World Health Report 2001-Mental Health: New Understanding, New Hope. . Geneva: World Health Organisation, 2001.

Wilcke A, Ligges C, Burkhardt J, Alexander M, Wolf C, Quente E, et al. Imaging genetics of FOXP2 in dyslexia. Eur J Hum Genet 2012; 20: 224-9.

Williams HJ, Norton N, Dwyer S, Moskvina V, Nikolov I, Carroll L, et al. Fine mapping of ZNF804A and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder. Mol Psychiatry 2011; 16: 429-41.

Williams JM, Beck TF, Pearson DM, Proud MB, Cheung SW, Scott DA. A 1q42 deletion involving DISC1, DISC2, and TSNAX in an autism spectrum disorder. Am J Med Genet A 2009; 149A: 1758-62.

Williams NM, Preece A, Spurlock G, Norton N, Williams HJ, Zammit S, et al. Support for genetic variation in neuregulin 1 and susceptibility to schizophrenia. Mol Psychiatry 2003; 8: 485-7.

Wirgenes KV, Sonderby IE, Haukvik UK, Mattingsdal M, Tesli M, Athanasiu L, et al. TCF4 sequence variants and mRNA levels are associated with neurodevelopmental characteristics in psychotic disorders. Transl Psychiatry 2012; 2: e112.

Wisniowiecka-Kowalnik B, Nesteruk M, Peters SU, Xia Z, Cooper ML, Savage S, et al. Intragenic rearrangements in NRXN1 in three families with autism spectrum disorder, developmental delay, and speech delay. Am J Med Genet B Neuropsychiatr Genet 2011; 153B: 983-93.

Wittchen HU, Jacobi F, Rehm J, Gustavsson A, Svensson M, Jonsson B, et al. The size and burden of mental disorders and other disorders of the brain in Europe 2010. Eur Neuropsychopharmacol 2011; 21: 655-79.

Wittig B, Wolfl S, Dorbic T, Vahrson W, Rich A. Transcription of human c-myc in permeabilized nuclei is associated with formation of Z-DNA in three discrete regions of the gene. Embo J 1992; 11: 4653-63.

Woo TU, Shrestha K, Amstrong C, Minns MM, Walsh JP, Benes FM. Differential alterations of kainate receptor subunits in inhibitory interneurons in the anterior

cingulate cortex in schizophrenia and bipolar disorder. Schizophr Res 2007; 96: 46-61.

Woo TU, Walsh JP, Benes FM. Density of glutamic acid decarboxylase 67 messenger RNA-containing neurons that express the N-methyl-D-aspartate receptor subunit NR2A in the anterior cingulate cortex in schizophrenia and bipolar disorder. Arch Gen Psychiatry 2004; 61: 649-57.

Wood JD, Bonath F, Kumar S, Ross CA, Cunliffe VT. Disrupted-in-schizophrenia 1 and neuregulin 1 are required for the specification of oligodendrocytes and neurones in the zebrafish brain. Hum Mol Genet 2009; 18: 391-404.

Woolf CM. Does the genotype for schizophrenia often remain unexpressed because of canalization and stochastic events during development? Psychol Med 1997; 27: 659-68.

Wright IC, Rabe-Hesketh S, Woodruff PW, David AS, Murray RM, Bullmore ET. Meta-analysis of regional brain volumes in schizophrenia. Am J Psychiatry 2000; 157: 16-25.

Wu Y, Borde M, Heissmeyer V, Feuerer M, Lapan AD, Stroud JC, et al. FOXP3 controls regulatory T cell function through cooperation with NFAT. Cell 2006; 126: 375-87.

Wyatt RJ, Alexander RC, Egan MF, Kirch DG. Schizophrenia, just the facts. What do we know, how well do we know it? Schizophr Res 1988; 1: 3-18.

Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of de novo copy number mutations with sporadic schizophrenia. Nat Genet 2008; 40: 880-5.

Yamada K, Gerber DJ, Iwayama Y, Ohnishi T, Ohba H, Toyota T, et al. Genetic analysis of the calcineurin pathway identifies members of the EGR gene family, specifically EGR3, as potential susceptibility candidates in schizophrenia. Proc Natl Acad Sci U S A 2007; 104: 2815-20.

Yamashita N, Miyashiro M, Baba J, Sawa A. Rolipram, a selective inhibitor of phosphodiesterase type 4, pronouncedly enhanced the forskolin-induced promotion of dopamine biosynthesis in primary cultured rat mesencephalic neurons. Jpn J Pharmacol 1997; 75: 91-5.

Yan J, Noltner K, Feng J, Li W, Schroer R, Skinner C, et al. Neurexin 1alpha structural variants associated with autism. Neurosci Lett 2008; 438: 368-70.

Yang JZ, Si TM, Ruan Y, Ling YS, Han YH, Wang XL, et al. Association study of neuregulin 1 gene with schizophrenia. Mol Psychiatry 2003; 8: 706-9.

Yarden Y, Sliwkowski MX. Untangling the ErbB signalling network. Nat Rev Mol Cell Biol 2001; 2: 127-37.

Yin DM, Chen YJ, Sathyamurthy A, Xiong WC, Mei L. Synaptic dysfunction in schizophrenia. Adv Exp Med Biol 2011; 970: 493-516.

Young-Pearse TL, Suth S, Luth ES, Sawa A, Selkoe DJ. Biochemical and functional interaction of disrupted-in-schizophrenia 1 and amyloid precursor protein regulates neuronal migration during mammalian cortical development. J Neurosci 2010; 30: 10431-40.

Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, et al. Genome-scale identification of nucleosome positions in S. cerevisiae. Science 2005; 309: 626-30.

Yue W, Yang Y, Zhang Y, Lu T, Hu X, Wang L, et al. A case-control association study of NRXN1 polymorphisms with schizophrenia in Chinese Han population. Behav Brain Funct 2011; 7: 7.

Zaykin DV, Westfall PH, Young SS, Karnoub MA, Wagner MJ, Ehm MG. Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. Hum Hered 2002; 53: 79-91.

Zeesman S, Nowaczyk MJ, Teshima I, Roberts W, Cardy JO, Brian J, et al. Speech and language impairment and oromotor dyspraxia due to deletion of 7q31 that involves FOXP2. Am J Med Genet A 2006; 140: 509-14.

Zhang D, Cheng L, Qian Y, Alliey-Rodriguez N, Kelsoe JR, Greenwood T, et al. Singleton deletions throughout the genome increase risk of bipolar disorder. Mol Psychiatry 2009; 14: 376-80.

Zhang X, Tochigi M, Ohashi J, Maeda K, Kato T, Okazaki Y, et al. Association study of the DISC1/TRAX locus with schizophrenia in a Japanese population. Schizophr Res 2005; 79: 175-80.

Zhao X, Shi Y, Tang J, Tang R, Yu L, Gu N, et al. A case control and family based association study of the neuregulin1 gene and schizophrenia. J Med Genet 2004; 41: 31-4.

Zheng F, Wang L, Jia M, Yue W, Ruan Y, Lu T, et al. Evidence for association between Disrupted-in-Schizophrenia 1 (DISC1) gene polymorphisms and autism in Chinese Han population: a family-based association study. Behav Brain Funct 2011; 7: 14.

Zhong C, Du C, Hancock M, Mertz M, Talmage DA, Role LW. Presynaptic type III neuregulin 1 is required for sustained enhancement of hippocampal transmission by nicotine and for axonal targeting of alpha7 nicotinic acetylcholine receptors. J Neurosci 2008; 28: 9111-6.

Zhou X, Geyer MA, Kelsoe JR. Does disrupted-in-schizophrenia (DISC1) generate fusion transcripts? Mol Psychiatry 2008; 13: 361-3.

Zipfel PF, Decker EL, Holst C, Skerka C. The human zinc finger protein EGR-4 acts as autoregulatory transcriptional repressor. Biochim Biophys Acta 1997; 1354: 134-44.

Publications