Genome Wide Gene Expression Analysis of Two ENU Mouse Models of Major Mental Illness

Sarah Mills Brown

PhD
The University of Edinburgh
2011
Declaration
I declare that this thesis has been composed by myself, 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 personal qualification.

Sarah Mills Brown
Acknowledgements

So many people have supported me, in one way or another, over the course of this PhD that I am afraid I cannot possibly do them all justice here. I am forever grateful and indebted to the friends, family and colleagues who have in many cases gone beyond the call of duty to help me get to where I am today. I could not have done it without you and I thank you all.

To my supervisors, Kathy and Pippa, for your support, supervision and patience. To David for your encouragement and support, and for pushing me to believe in myself. To Helen and Susan, for answering my unending questions and supporting and training me in the early stages. To Steve and Stewart for their computing expertise. I am sure you will be breathing a sigh of relief that I am finished! To John Roder and Steve Clapcote who welcomed me into their lab in Toronto while I was collecting samples and who taught me everything I know about managing mouse colonies! To Angie, Allison and Lee from the WTCRF for helping set up and run the microarray study. To William from the WTCRF for training and assistance for the real-time PCR, and ensuring there was always a space in the booking form for me to complete my experiments on time. To Tom Freeman for microarray statistical assistance, and Andrea Christoforou for all statistical assistance! To Sharon, Rosemary, Helen and Susan (again!) for taking care of absolutely everything else so I could concentrate on my work! To Becky, who was a constant sounding board and friend, and who on many occasions helped me see the good side when things weren’t going to plan.

To my wonderful friends who have provided both encouragement and support when I have sometimes been less than easy to be around, and to my partner Gordon, who must have the patience of a saint to put up with my stressful flapping! To Jake and Taff, for keeping me grounded and helping me to take time out to relax when things got a little hectic.

Finally, I would not have been able to do this without the support of my amazing family. You will possibly never know how much the seemingly pointless phone calls, emails and photos from home have meant to me. This thesis is dedicated to you.

Thank you all.
Abstract

Major mental illness is now recognised as one of the leading causes of adult morbidity. Of the adult onset psychiatric disorders, the functional psychoses (schizophrenia, bipolar disorder and recurrent major depression) are the most severe and most common in the general population. Evidence suggests that certain genetic factors influence an individual’s susceptibility to developing these disorders when combined with appropriate social and environmental conditions. Several good candidate genes have been identified. Of relevance to this study is Disrupted in Schizophrenia 1 (DISC1) which was identified in a large Scottish family that carried a balanced translocation (t1:11) and had a history of major mental illness. In 2008, two ENU mutant mouse models with missense mutations in exon 2 of Disc1 were characterised and found to have behavioural and neuroanatomical phenotypes consistent with schizophrenia and major depression. The primary aim of this thesis is to further analyse these mouse models by performing whole genome gene expression studies and secondary protein analysis to identify genes involved in the aetiology of schizophrenia and major depression.

My initial analysis used Illumina BeadChip microarray technology to identify 368 genes that were differentially expressed in ENU mutant animals under different biological conditions, compared to appropriate control animals. Nine biological groups were compared including one embryonic group at E13, and three groups treated with appropriate anti-psychotic or anti-depressant drugs. Of the 368 genes identified as differentially expressed, 46 were chosen for validation by qRT-PCR based on fold-change, p-value, functional significance, overenrichment of GO terms, pathway analysis and previous implications in major mental illness. NRXN1, NRXN3 and CDH11 were found to be significantly up-regulated in the schizophrenia mouse model with EGR4 significantly down-regulated compared to C57BL/6J wild-type controls. These findings were also replicated in an independent sample using wild-type littermates. The mental retardation gene PAK3 was up-regulated in the schizophrenia mouse model and expression levels were corrected to a level not significantly different to wild-type, when treated with the PDE4 inhibitor Rolipram. Semi-quantitative western blotting also confirmed the disregulation of EGR4 and PAK3 at the protein level in these animals. RNA expression profiles were also characterised for each of the genes above, and DISC1, through development.
In summary this thesis describes the striking disregulation of four prominent genetic candidates of major mental illness in an independent animal model. A first functional link between DISC1 and NRXN1 is described suggesting, for the first time, a DISC1-dependant mechanism for regulating neurexin gene expression.
Table of Contents

Declaration ............................................................................................................. ii
Acknowledgements ............................................................................................. iii
Abstract ................................................................................................................ iv
Table of Contents ............................................................................................... vi
List of Figures .................................................................................................... xiv
List of Tables ...................................................................................................... xvii
Abbreviations .................................................................................................... xix

1. Introduction

1.1 Major Mental Illness Overview ................................................................. 2

1.2 A genetic basis for major mental illness .................................................. 6

1.2.1 Linkage Analysis ................................................................................. 8

1.2.2 Association Studies ............................................................................. 8

1.2.3 Copy Number Variation ...................................................................... 11

1.2.4 Genome wide gene expression analysis ............................................. 12

1.2.5 Cytogenetics ....................................................................................... 14

1.3 Identification of Disrupted in Schizophrenia 1 ......................................... 14

1.4 The biology of DISC1 and its interactions ............................................. 18

1.5 Current Hypotheses of Major Mental Illness ........................................... 22

1.5.1 The neurotransmitter hypothesis ....................................................... 22

1.5.2 The developmental hypothesis .......................................................... 23

1.5.3 The cAMP hypothesis ....................................................................... 24

1.6 Genome-wide expression and pathway analysis ..................................... 25

1.7 The schizophrenia mouse ........................................................................ 26

1.8 Thesis Aims ............................................................................................... 35
2. Materials and Methods

2.1 Collection of Tissue Samples

2.1.1 Use of inbred strains

2.1.2 Extraction of Genomic DNA from adult tissue

2.1.3 Extraction of Genomic DNA from embryonic tissue

2.1.4 Genotyping by sequencing reaction

2.1.4.1 Polymerase Chain Reaction

2.1.4.2 Exosapit Clean-up

2.1.4.3 Sequencing reaction

2.1.4.4 Ethanol/EDTA Precipitation

2.1.5 Genotyping using the Transnetyx sequencing service

2.1.6 Genotyping Q31L het x hom pups

2.1.7 Dissection of embryos and pups

2.1.8 Dissection of adult brain samples

2.2 Sample Preparation

2.2.1 Extraction of total RNA from tissue samples

2.2.2 PARIS procedure

2.2.3 Post extraction DNase Treatment of RNA

2.2.4 On column DNase Treatment of RNA

2.2.5 Assessing RNA quality using the Agilent Bioanalyser

2.2.6 Vacuum Drying RNA samples

2.2.7 Synthesis of cDNA

2.2.8 Synthesis of cDNA from low quantity RNA
2.2.9 Determining cDNA concentration using the Quant-iT™ PicoGreen® dsDNA kit

2.2.10 Determining cDNA or RNA quantity on the NanoDrop

2.3 PCR

2.3.1 PCR Reagents

2.3.2 Design and Synthesis of Oligonucleotide Primer pairs

2.3.3 RT-PCR Cycling

2.3.4 Visualisation of PCR Products

2.4 Protein Preparations

2.4.1 Determining Protein Sample Concentrations

2.5 Pre-Pulse Inhibition

2.5.1 Calibration

2.5.2 Drug Administration

2.5.3 Behavioural Testing

2.5.4 Analysis of Behavioural Testing Results

2.6 Microarray Analysis of ENU mutant gene expression

2.6.1 cRNA amplification and Biotinylation

2.6.2 Quantification and Quality control of cRNA

2.6.3 Preparation of Pooled cRNA

2.6.4 Hybridisation and scanning of Illumina BeadChips at the Wellcome Trust Clinical Research Facility

2.6.5 Data Analysis

2.6.6 Microarray Quality Control Measures

2.6.6.1 Illumina direct hybridisation assay controls

2.6.6.2 Technical Replicates and Between Chip Controls
2.6.6.3 Cluster analysis for detection of experimental Artefacts……………………………………………………...59
2.6.6.4 Power Calculations……………………………………...59
2.6.6.5 Pre-processing……………………………………………59

2.7 Testing For Differential Expression………………………………………………...61
2.7.1 Differential expression analysis……………………………………...61
2.7.2 Functional Analysis (Including IPA Pathway analysis)…………61
2.7.3 GOTree Analysis of differentially expressed genes………………62
2.7.4 Testing for overlaps and gene expression corrections
from drug treatments……………………………………………………62
2.7.5 Comparing data with previously published work………………63

2.8 Follow up analysis of the microarray study……………………………………63
2.8.1 Identifying the genes to take forward for validation………………63
2.8.2 Preparation of Samples for two phase follow up…………………64
2.8.3 Identification of suitable probes for Taqman real time PCR…..64
2.8.4 Verification of genes by ABI Taqman analysis…………………65
2.8.5 Statistical Analysis of ABI Taqman output…………………66

2.9 Protein analysis and Antibody Staining……………………………………67
2.9.1 Identification of suitable antibodies……………………………67
2.9.2 Western Blotting……………………………………………….68
2.9.3 Quantification of Westerns using Image J………………….69
2.9.4 Culture of Primary Neurons……………………………………69
2.9.5 ICC of Primary Neurons………………………………………70
3. Sample collection and processing of a genome wide microarray analysis of the Disc1 ENU mouse mutants

3.1 Introduction..............................................................................75
3.1.1 Background and motivation.................................................75

3.2 Partial Replication of Clapcote et al’s Behavioural Analysis
of the Disc1 ENU mouse mutants...............................................80
3.2.1 Effect of Gender on PPI in non-drug treated animals..........80
3.2.2 Effect of Genotype on PrePulse Inhibition and
Startle Response in Mice.............................................................80
3.2.3 Effect of Drug treatment on PrePulse Inhibition and
Acoustic Startle Response in Mice..............................................82

3.3 Whole Genome Gene Expression Study Experimental Design........83
3.3.1 Sample size and power.........................................................83
3.3.2 Sample Layout and Technical Replication............................86

3.4 Sample Preparation and Quality.............................................87
3.4.1 Integrity of total RNA..........................................................88
3.4.2 Integrity of cRNA.................................................................91

3.5 Data Generation using the Illumina BeadArray Platform..........92

3.6 Quality Control.......................................................................95
3.6.1 Illumina BeadChip Built in Controls.................................95
3.6.2 Technical Replicates .............................................. 101
3.6.3 Quality Control using Cluster Analysis .................... 103

3.7 Data Pre-processing .................................................. 106
3.7.1 Variance Stabilisation .......................................... 106
3.7.2 Normalisation ...................................................... 107
3.7.3 Data Filtering ...................................................... 109

3.8 Conclusions .......................................................... 112

4. Differential Expression Analysis of the Disc1 ENU Mutant Mouse Microarray
   4.1 Introduction .......................................................... 117
   4.2 Differentially Expressed Genes .................................. 117
      4.2.2 Differentially expressed genes: primary
             annotation and gene ontology ............................ 119
      4.2.3 Differential Expression between the L100P
             adult ENU mutant and the C57BL/6J adult controls .... 121
      4.2.4 Differential Expression between the Q31L
             adult ENU mutant and the C57BL/6J adult controls ... 136
      4.2.5 Differential Expression between the Embryonic
             ENU mutant and the C57BL/6J adult controls .......... 136
      4.2.6 The effect of Drug Treatment on gene expression
             in adult ENU mutant mice ................................. 138
   4.3 Selection of Genes of Interest for Validation by qRT PCR .... 145
   4.4 Identification of L100P outlier .................................. 147
      4.4.1 Reanalysis of array dataset ................................ 148
      4.4.2 New Gene Ontology, Pathway analysis and
             Comparisons with previous study ......................... 157
   4.5 New list of genes for qRT PCR .................................. 160
   4.6 Conclusions .......................................................... 162
5. Quantitative real-time PCR analysis of candidate differentially expressed genes in the L100P and Q31L mutant mice

5.1 Introduction........................................................................................................169
5.2 Validating the Disc1 microarray result................................................................170
  5.2.1 Initial Validations and Outlier removal of the L100P drug naïve adult microarray........170
  5.2.2 Validation of the Q31L drug naïve adult Microarray.........................................173
5.2.3 Validation of the L100P and Q31L embryonic Microarray..................................174
  5.2.4 Validation of the L100P and Q31L drug treatment Microarray..............................174
5.3 Conclusions.......................................................................................................177

6. Establishing a developmental profile for the genes of interest in L100P mutant mice and wild type controls..................................................................................................................................183
6.1 The Disc1 developmental profile.............................................................................183
6.2 Establishing developmental profiles for the genes of interest...............................186
  6.2.1 The Disc1 developmental profile on pooled samples...........................................186
  6.2.2 Developmental Profile for EGR4.................................................................189
  6.2.3 Developmental Profile for NRXN1 and NRXN3..............................................190
  6.2.4 Developmental Profile of PAK3.................................................................191
  6.2.5 Developmental Profile of Cdh11....................................................................193
  6.2.6 Developmental Profile of Sort1........................................................................194
6.3 Conclusions from the developmental profiles.....................................................195

7. Preliminary protein analysis of validated genes in L100P drug naïve (saline) adult mouse samples........................................................................................................................................199
7.1 Protein analysis of Egr4.......................................................................................199
7.2 Protein analysis of Pak3..........................................................................................201
7.3 Protein analysis of NRXN1....................................................................................202
7.4 Protein analysis of Nrxn3.....................................................................................204
7.5 Protein analysis of Sort1.......................................................................................206
7.6 Conclusions........................................................................................................207
8. Discussion

8.1 Summary of findings .....................................................211

8.1.1 Chapter 3 .................................................................211

8.1.2 Chapter 4 .................................................................212

8.1.3 Chapter 5 .................................................................216

8.1.4 Chapter 6 .................................................................219

8.1.5 Chapter 7 .................................................................221

8.1.6 Conclusions ..............................................................223

8.2 Caveats .........................................................................223

8.3 Relevance of this study to the field .....................................227

8.4 Differential networking and other future work .........................229

8.5 Final Remarks ..............................................................231

Appendix ............................................................................232

References ........................................................................233
List of Figures

Figure 1.1: The Disc1 protein........................................................................................................32

Figure 2.1: Examples of Agilent 2100 Bioanalyser output Electropherographs........46
Figure 2.2: Picogreen standard curve for assessing cDNA concentration......................49
Figure 2.3: Image of a PrePulse inhibition chamber..........................................................53
Figure 3.1: Image depicting the effective reduction in an animals startle
response when a strong stimulus is preceded by a weaker prestimulus..............78
Figure 3.2: Total PrePulse Inhibition and Acoustic Startle Response
in non-drug treated adult mice from each line.............................................................81
Figure 3.3: Power Curves.....................................................................................................85
Figure 3.4: Layout of the seven Illumina mouse ref-6 version 2 BeadChips.............87
Figure 3.5: Examples of Agilent 2100 Bioanalyser output Electropherographs.........89
Figure 3.6: Agilent Bioanalyser output of cRNA Integrity.............................................92
Figure 3.7: Signal Intensities for Illumina built-in Quality control measures..........96
Figure 3.8: Signal Intensities for Illumina built-in Quality control
measures by BeadChip.......................................................................................................99
Figure 3.9: Signal Intensities for Illumina built-in Quality control
measures by BeadChip post chip5 repeat.................................................................100
Figure 3.10: Pairwise Comparisons of raw data for C57BL6 F1 pool samples.....102
Figure 3.11: Pairwise comparisons of raw data for C57BL6 F2 pool samples.....103
Figure 3.12: Hierarchical cluster dendrogram of all samples based on the
raw dataset.....................................................................................................................105
Figure 3.13: Hierarchical cluster dendrogram of all samples based on the
stabilized and normalized dataset..............................................................................108
Figure 3.14: Hierarchical cluster dendrogram of all samples based on
the filtered, transformed and stabilized dataset……………………………………..111
Figure 3.15: Venn diagrams of probes surviving the 1.3 fold-change
filter in six comparisons.................................................................112
Figure 4.1 Venn diagram of significant probes in the genotype comparisons……119
Figure 4.2: Bar graph of GOTree categories for L100P adult vs C57BL6 adult………132
Figure 4.3: Ingenuity Pathway Analysis networks for L100P adult deregulated
Genes………………………………………………………………………………134
Figure 4.5: Ingenuity Pathway Analysis of Embryonic deregulated gene lists………138
Figure 4.6: Ingenuity Pathway Analysis of Drug treated groups…………………144
Figure 4.7: PrePulse Inhibition data for L100P saline and C57BL/6J mice………148
Figure 4.8: Venn diagram of significant probes in the genotype comparisons……156
Figure 4.9: Ingenuity Pathway Analysis..................................................158
Figure 4.10: Canonical pathways for IPA analysis on new deregulated gene list…159
Figure 6.1: Developmental Profile of Disc1 in the mouse brain………………..184
Figure 6.2: Comparing Disc1 expression in pooled and unpooled
developmental samples…………………………………………………………187
Figure 6.3: Developmental profile of Disc1 in the L100P mutant mouse………..188
Figure 6.4: Developmental profile of Egr4 in cDNA pools from the
L100P mutant mouse……………………………………………………………189
Figure 6.5: Developmental profile of Nrxn1 and Nrxn3 in cDNA
pools from the L100P mutant mouse…………………………………………190
Figure 6.6: Developmental profile of PAK3 in cDNA pools from the
L100P mutant mouse…………………………………………………………..192
Figure 6.7: Developmental profile of Cdh11 in cDNA pools from the L100P mutant mouse……………………………………………………………………………………………………193
Figure 6.8: Developmental profile of Sort1 in cDNA pools from the L100P mutant mouse…………………………………………………………………………………………194
Figure 6.9 Composite of the developmental profiles compared to Disc1……………196
Figure 7.1: Western blotting and ICC for EGR4 in L100P ENU mutant mice………200
Figure 7.2: Western blotting and ICC for Pak3 in L100P ENU mutant mice………201
Figure 7.3: Western blotting and ICC for Nrxn1 in L100P ENU mutant mice………203
Figure 7.4: Western blotting and ICC for NRXN3 in L100P ENU mutant mice…..205
Figure 7.5: Western blotting and ICC for Sort1 in L100P ENU mutant mice………206
**List of Tables**

Table 1.1: Phenotypes of *DISC1/Disc1* mouse models………………………………30

Table 1.2: Common behavioural traits and the effects of the ENU missense mutation on phenotype when compared to the C57B6/J parent line .................................33

Table 1.3: The effects of drug treatment on behavioural rescue in Q31L and L100P missense mutant mouse lines………………………………………………………..34

Table 2.1: BeadChip layout with key.................................................................56

Table 3.1: Behavioural Phenotypes of the L100P and Q31L Disc1 mutant mice…..79

Table 3.2: Mean results on PPI at all 3 levels tested, plus total.........................82

Table 3.3: Mean acoustic startle response and average PPI over all ranges tested….82

Table 3.4 Summary of Adult Mouse Groups used in the genome-wide expression analysis......................................................................................................................83

Table 3.5 Summary of developmental mouse groups used in the genome wide expression analysis..........................................................84

Table 3.6: Mouse samples used in the microarray analysis...............................90

Table 3.7: Summary of raw expression data..................................................94

Table 4.1 Summary of probes differentially expressed by genotype of drug treatment..........................................................................................118

Table 4.2: Overlap with previous studies.....................................................131

Table 4.3: Complete list of disregulated genes in the L100P adult group...........122

Table 4.4: Disregulated genes in the Q31L adult group compared to C57BL/6J controls..........................................................136

Table 4.5: Disregulated genes in the Embryonic sample groups......................137

Table 4.6.1: Disregulated genes in Clozapine treated L100P adults.................139
Table 4.6.2: Disregulated genes in Rolipram treated L100P adults………………..140
Table 4.6.3: Disregulated genes in Bupropion treated Q31L adults………………141
Table 4.7: Genes for validation by qRT PCR................................................146
Table 4.8 Summary of probes differentially expressed by genotype of drug treatment.................................................................149
Table 4.9: New list of disregulated genes post outlier removal…………………..150
Table 4.10: Table of biological processes over-enriched in the L100PSaline group.................................................................157
Table 4.11: Overlap with previous studies.............................................................160
Table 4.12: List of genes to be analysed by qRT PCR......................................161
Table 5.1: Table of genes differentially expressed in the L100P drug naïve adult that were tested by qRT-PCR........................................172
Table 5.2: Table of genes differentially expressed in the Q31L drug naïve adult that were tested by qRT-PCR. ........................................173
Table 5.3: Table of genes differentially expressed in the L100P and Q31L embryo that were tested by qRT-PCR.................................174
Table 5.4: Table of genes differentially expressed in the L100P and Q31L drug treated adults that were tested by qRT-PCR..................175
Table 5.5: Table of genes differentially expressed in the C57BL/6J drug treated adults that were tested by qRT-PCR..........................176
<table>
<thead>
<tr>
<th>Abbreviations</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td>Analysis of Variance</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complimentary DNA</td>
</tr>
<tr>
<td>CNS</td>
<td>Central nervous system</td>
</tr>
<tr>
<td>CNV</td>
<td>Copy Number Variant</td>
</tr>
<tr>
<td>ENU</td>
<td>N-ethyl-N-nitrosourea</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic reticulum</td>
</tr>
<tr>
<td>FDA</td>
<td>Food and drugs agency</td>
</tr>
<tr>
<td>FDR</td>
<td>False discovery rate</td>
</tr>
<tr>
<td>fMRI</td>
<td>functional magnetic resonance imaging</td>
</tr>
<tr>
<td>GO</td>
<td>Gene ontology</td>
</tr>
<tr>
<td>HapMap</td>
<td>Haplotype map</td>
</tr>
<tr>
<td>ICC</td>
<td>Immunocytochemistry</td>
</tr>
<tr>
<td>IQR</td>
<td>Interquartile range</td>
</tr>
<tr>
<td>LOD</td>
<td>Logarithm of the odds</td>
</tr>
<tr>
<td>MAANOVA</td>
<td>Microarray analysis of variance</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>PPI</td>
<td>Prepulse inhibition</td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>Quantitative real time polymerase chain reaction</td>
</tr>
<tr>
<td>RIN</td>
<td>RNA integrity number</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>Vst</td>
<td>Variance stabilising transformation</td>
</tr>
</tbody>
</table>
Chapter 1
Introduction
1.1 Major Mental Illness Overview

Major mental illness is now recognised as one of the leading causes of adult morbidity, affecting populations throughout the world [1] with little demographic variability. Of the adult onset psychiatric disorders, the functional psychoses (schizophrenia, bipolar disorder and recurrent major depression) are the most severe and the most common in the general population. Functional psychoses are serious disorders with, as yet, no known organic cause [2] often rendering the sufferer unable to maintain normal day to day activities. They encompass disorders of thought, perception, often leading to delusion and hallucination, mood and behaviour. Evidence suggests that genetic factors influence an individual’s susceptibility to developing these disorders [3] when combined with appropriate social or environmental conditions. Over the last century the functional psychoses have classically been divided into two diagnostic groups; schizophrenia and affective disorders (recurrent major depression and bipolar disorder). More recently, it has been suggested that the disorders may not be as distinct as was once thought but form a continuum with recurrent major depression at one end and schizophrenia at the other [4]. This view is supported by epidemiological evidence that shows genetic sharing of risk between schizophrenia and bipolar disorder [5], increased risk of both schizophrenia and bipolar disorder in first degree probands with schizophrenia or bipolar disorder [6], recent genome wide association studies [7] and indeed earlier work from our own laboratory [8].

Schizophrenia is the most common form of psychotic illness with a lifetime morbid risk of around 1% (more if the many schizophrenia spectrum disorders are also considered) and roughly equal proportions of males and females affected. The illness is characterised by behavioural and cognitive symptoms, which can be categorised as either positive or negative (referring to an increase or absence in normal functioning). Positive symptoms include heightened sensory perception, often leading to delusion and hallucination, disorganisation, rapidity of speech, uncoordinated gait and altered cognitive functioning. Negative symptoms include social withdrawal, lack of care over personal appearance or wellbeing and the inability to act to achieve simple goals [9]. Diagnosis of schizophrenia is based on patient interview with presentation of at least two symptoms over a period of
2-6 months affecting ability to work, socialise or provide basic self care. Many schizophrenia patients are permanently or temporarily hospitalised as they may pose a danger to themselves, and current statistics show 10% of schizophrenia patients commit suicide and many more attempt to [10]. The need for long term social and medical care of schizophrenia patients costs the health service around £1 billion per year in England and Wales alone [11]. Onset of psychosis is normally in adolescence to mid-twenties after the end of puberty, with male patients being admitted on average 4-5 years earlier than female patients [12]. Some studies have concluded that the symptomology of schizophrenia differs between the sexes, with females often displaying less extreme symptoms than males [13] (also reviewed in [14] [15]). As well as presenting around 4 years after males, females have a second peak of onset around the age of 50. Animal studies suggest this is due to the protective effects of oestrogen until menopause [16] and treatment with combined oestrogen and antipsychotics has been shown to accelerate symptom remission, when compared to antipsychotic treatment alone, in both males and females.

Brain morphology of affected individuals is often abnormal with increased ventricle size and reduction in anterior hippocampal volume commonly shown by MRI scanning [17]. Patients on continual medication had less reductions in hippocampal volume over time, and those with a longer duration of psychosis prior to first treatment had greater reductions in temporal lobe volume, suggesting that either the reduced volume is due to developmental effects or breakdown during the initial onset of psychosis. It is proposed that the increases in ventricle size are not due to confounding treatment effects [18] as patients with better disease outcome and drug response had less changes in ventricular size. Recent meta-analysis studies of brain morphology in schizophrenia have found reductions in grey matter volume in frontal, temporal, insular and cingulate cortex and thalamus [19]. Overall cerebral volume is reduced in schizophrenia with an increase in lateral ventricular volume and bilateral reduction in amygdala and hippocamal volume relative to the overall decrease in brain size [20]. Further meta-analysis has also shown bilateral reductions in thalamic volume in both first-episode and chronic schizophrenia patients compared to control subjects [21]. In addition, reductions in frontal white matter
have been observed through voxel based morphometry and have also been confirmed by meta-analysis [22]. MRI studies of high risk children and non-psychotic relatives of individuals with a schizophrenia diagnosis reveal volumetric abnormalities of the prefrontal and temporal regions to a lesser extent than affected individuals, but still significantly different to controls. Further high risk studies indicate reductions in amygdala and hippocampal volumes and in superior temporal gyrus suggesting these volumetric changes may be indicative of disease susceptibility (reviewed in [23]).

The main treatment for patients with schizophrenia is anti-psychotic drugs. Classic treatments involve the use of Thorazine plus chlorpromazine and Haloperidol which are effective in treating the positive symptoms while newer ‘atypical’ (second generation) treatments such as Risperidone and Clozapine are also partially effective in reducing the negative symptoms [24]. Clozapine acts by blocking receptors for some key neurotransmitters in the brain, including dopamine type 4 receptors, serotonin type 2 receptors, norepinephrine receptors, acetylcholine receptors, and histamine receptors [25]. It is also a partial blocker of the dopamine type 2 receptor. Typical (first generation) antipsychotics such as Haloperidol act primarily on the dopamine type 2 receptor and can have sedative effects. Approximately one third of patients do not respond to treatment with first generation antipsychotics. Clozapine was the first second generation antipsychotic to be licenced by the FDA for treatment of treatment resistant schizophrenia [26] and is currently the only effective treatment for refractory schizophrenia. The main side effect of Clozapine use is the development of agranulocytosis, a reduction in the white blood cell count, which can be fatal due to the high risk of infection resulting from a suppressed immune system [27]. Risk can be dramatically reduced through monitoring of white blood cell counts. Side effects of treatment with first generation antipsychotics include dysphoria, dystonia, akathisia, dyskinesia, and Parkinsonian motor symptoms. There may also be an increase in the depressive/demoralization aspects of illness course, impaired learning, and slow information processing, and increased hostility, aggression, and suicide [28-30]. While the majority of side effects seen in the first generation antipsychotics are not present in
the second generation antipsychotics, they do carry their own risks. These include metabolic syndrome, hyperlipidemia and reduced insulin sensitivity with an increased incidence of diabetes. A reduced lifespan due to exposure to adverse drug effects is also observed [31-33]. Current work is concentrating on the discovery of drug targets outwith the dopamine system, that may reduce the negative and cognitive symptoms not helped by the current anti-psychotics used [34] and may have more favourable side effects than current treatments. The use of multiple antipsychotics and/or augmentation strategies (such as combined treatment with estrogen in males [35]) are also currently being studied [36]. Gender differences in presentation of schizophrenia have been observed for a long time, and it is proposed that estrogen and other female sex hormones may have a neuroprotective effect [37], delaying presentation of symptoms in females til later in life when hormonal changes occur due to menopause [38]. Psychotherapy is occasionally used along-side a programme of drug treatment in schizophrenia patients but, however effective these treatments may be in controlling symptoms, there is currently no cure for schizophrenia and treatment is lifelong.

Like schizophrenia, bipolar disorder has a lifetime prevalence of approximately 1% and affects an equal proportion of males and females [39]. Bipolar disorder is a severe mood disorder characterised by mood disturbances ranging from severe depression to extreme elation and is often accompanied by psychotic features and cognitive changes. Pathological mood swings may occur spontaneously or be cyclic in nature while some individuals will have predominant manic or depressive episodes with few mood swings. In some cases states can be mixed, with characteristics of mania intruding upon the depressive episode or vice versa. When in the depressive phase sufferers will, for example, display social withdrawal, anhedonia, difficulty with concentration and decision making, fatigue and loss of self esteem. The prevalence of suicide in bipolar individuals is high when in the depressive phase and many require constant monitoring and care to prevent self injury [40]. As an almost a direct result of this, 35% of the £200 million a year spent on bipolar disorder by the NHS is the result of hospital admissions alone [41]. In contrast, while in the manic phase sufferers will display elevated mood, delusions, hallucinations, extreme restlessness and inflated self esteem. Males and females do not
display sex differences in age of onset (mean age ~ 25) but show differences in type of episode at first presentation. Males are more likely to present with mania prior to initial diagnosis while females more commonly present with major depression [42]. Bipolar patients normally receive treatment in the form of lithium and other mood stabilisers. Lithium is the most favoured of the mood stabilisers as it does not have a sedative effect while still being effective, and is effective for both manic and depressive phases [43]. The exact mechanism by which lithium works is still unknown but studies of animal models suggests a neuroprotective role against glutamate induced excitotoxicity [44]. Psychotherapy is also widely used as treatment for bipolar disorder.

Recurrent major depression carries a lifetime risk of 16.2% [45] and often associates with schizophrenia or bipolar disorder in high risk families [5, 8]. Recurrent major depression is unipolar and characterised by phases of severe depression without mania. Symptoms are similar to those of bipolar individuals in the depressed phase. Treatment of recurrent major depression with anti-depressant drugs is somewhat effective (around 53% improvement rate) [46], however, residual symptoms including insomnia, fatigue and anhedonia persist in many patients [47], increasing the risk of full relapse [48]. Overall one third of patients will make a full recovery, one third will have persistent reduced symptoms and one third do not respond to treatment.

1.2 A Genetic Basis for Major Mental Illness
The definitive causes of schizophrenia and severe mood disorders are as yet unknown, but it has been shown that genetic control is involved [49]. Familial studies have shown that an individual’s risk of developing schizophrenia increases relative to how closely related they are to a sufferer [50]. Adoption studies have shown that children with one schizophrenic parent have around a 13% chance of developing the illness later in life (compared to 1% in the general population) even when raised by mentally healthy adoptive parents [51]. It has been claimed that the heritability estimates are incompatible
with transmission through a single major locus in the majority of cases, but become plausible when modelled with multiple loci [52].

Bipolar disorder is also heritable; concordance rates in monozygotic twins are ~43% while in dizygotic twins this drops to ~6% [53]. Around half of patients with bipolar disorder have a parent with a major mood disorder [54]. Family studies have shown that families with probands of schizophrenia have a higher prevalence of unipolar depression, and aggregation of psychotic affective disorders occurs in families of probands with either schizophrenia or bipolar disorder [55].

In 2009, Lichtenstein et al [5] carried out a large scale study of the Swedish population to assess environmental and genetic factors involved in the predisposition to schizophrenia and bipolar disorder. Using multivariate generalised linear mixed model analysis of information from the multi-generation register and the hospital discharge register, they could investigate these risk factors for schizophrenia and bipolar disorder and the comorbidity of the disorders. They found an increased risk of schizophrenia or bipolar disorder in individuals with a first degree relative who suffered from the same disorder. The risk was higher in full siblings than half siblings with heritability estimates of 64% for schizophrenia and 59% for bipolar disorder. There was also an increased risk of schizophrenia in relatives of probands with bipolar disorder, including in children adopted and raised by healthy parents. Common additive genetic effects contributed highly to the comorbidity between the disorders (63%). Patterns of inheritance of major mental illness are not simple and it is likely many interacting factors are involved. Large families that display quasi-Mendelian segregation do exist [56], which highlight potential genetic risk factors of major effect conferring susceptibility to major mental illness in some cases. Both genetic and allelic heterogeneity occur in major mental illness. Genetic heterogeneity accounts for the large number of genes that give rise to the same classification of mental illness. Allelic heterogeneity occurs when multiple different single mutations affecting the same gene give rise to the same disorder, such as the case of DISC1, described in section 1.3.
1.2.1 Linkage Analysis
Linkage analysis involves the study of closely related, high risk families (ie families with multiply affected members where genetic factor(s) show high penetrance), to identify a region of the chromosome that co-segregates with the disease. This information can then be used to study genetic mutations within the linkage regions to identify potential candidate genes. Linkage analysis tends to identify large regions of the genome due to the sharing of large sections of the chromosome between subjects, hence the subsequent identification of susceptibility genes can be difficult.

None the less, several linkage regions identified for schizophrenia and bipolar disorder have been replicated in multiple studies. 1p13-q23, 1q42, 2p12-q22, 5q21-q33, 6p24-p22, 6q16-q25, 8p22-p11, 8p22-p21, 10p15-p11, 13q32-q34 and 22q11-q12 [57, 58] are now well established linkage regions for schizophrenia, while 5p15, 6q21, 8q24, 10q26, 11p15, 12q24, 17q25, 18p11, 18q22, 20q13, 22q12 and Xq26 [59] have been established as strong linkage regions for bipolar disorder. What is most evident from linkage studies to date, is the large number of putative susceptibility regions that have been identified, scattered throughout the entire genome, which supports the theory of multiple genes of small effect [60].

1.2.2 Association Studies
Classical association analysis differs from linkage analysis as it involves the study of unrelated individuals in a population for the frequency of alleles (particularly SNPs – single nucleotide polymorphisms). Up until recently, individual functional candidate genes or in positional candidate genes within identified linkage peaks were studied. The completion of the human genome sequence has greatly facilitated the identification and study of genes for complex human diseases. In addition, the International HapMap project (www.hapmap.org), which characterises SNP frequency and correlation in samples from four geographically diverse populations, has allowed the identification of patterns of linkage disequilibrium within and between populations (The International HapMap Consortium 2003). The basis of the HapMap project is that within populations,
the human genome is divided into linkage disequilibrium regions (LD) between hotspots of recombination [61] and that the extent of LD will determine the tendency of alleles at two or more loci to be inherited together at a higher degree than that expected by chance [62]. In numerical terms, this is the difference between observed and expected allelic frequencies. LD is influenced by rate of mutation, rate of recombination, genetic drift, population structure and genetic linkage. The use of LD to define tagging SNPs can, therefore, cover more of the variation within the region of interest than the genotyping of randomly spread SNPs [63]. Association studies for schizophrenia have identified *neuregulin 1 (NRG1)* and *disrupted in schizophrenia 1 (DISC1)* as strong candidate genes ([64-66] and others). Originally identified by linkage studies they have been supported by findings from classic association studies and copy number variants, suggesting particular polymorphisms confer risk population wide and are not confined to distinct family groups. mRNA analysis confirms the expression of these genes in distinct brain regions involved in the pathology of schizophrenia, including the prefrontal cortex and hippocampus, and functional data links many of these genes to processes which are thought to be involved in major mental illness[67].

The combination of commercial high throughput genotyping and knowledge of linkage disequilibrium means that association studies can now be carried out in a genome wide fashion, and are no longer restricted to distinct genes or chromosomal regions. GWAS (Genome Wide Association Study) combines classical association analysis with the positional cloning flexibility of genome-wide linkage scans to provide a powerful tool for the identification of candidate genes, with genome wide significance that requires no prior hypothesis about the role or function of the candidate genes in disease pathology [68]. While some possible susceptibility genes for bipolar disorder have been identified by classical association studies, the GWAS method has generally been more successful for schizophrenia. To date, the largest schizophrenia GWAS tested approximately one million SNPs in 3322 schizophrenia and 3587 controls [7]. They found association with *MYO18B* and 450 SNPs in the major histocompatibility complex in schizophrenia cases. The major histocompatibility complex had also been implicated by Steffanson et al (2009) [69] using combined SNP data from several large scale genome wide scans. Kirov
et al (2008) [70] carried out a GWAS study using parent-offspring trios, to minimise the risk of population stratification, with 574 schizophrenia patients, their parents and 605 unaffected individuals and identified \textit{RBP1} and \textit{CCDC60} as susceptibility genes for schizophrenia. \textit{RBP1} inhibits PI3K/Akt signalling which has previously been implicated in schizophrenia pathology [71]. In addition, \textit{IL3RA}, \textit{CSF2RA} and \textit{RELN} have also been associated with schizophrenia [72-74]. In 2010, one GWAS study tested 572888 markers for association with schizophrenia and found no genome wide significance in the sample set from a Norwegian population [75] but found significance when expanded to a large European cohort. What is apparent from these studies is that there are multiple genes involved and little overlap of genes identified between sample sets. This apparent lack of reproducibility could be due to a number of factors, including the use of different SNP sets, genotyping platforms, sample size and genetic or phenotypic heterogeneity between samples. This second factor was addressed in part by the Wellcome Trust Case Control Consortium in 2007. They examined seven common diseases, including bipolar disorder, and demonstrated the use of a carefully selected and consistent control group was crucial in the reproducibility of results [76].

The Psychiatric GWAS Consortium (PGC) was set up in 2007 with the aim of consolidating the data from GWAS studies by statistically robust meta-analysis. To date there are five disease working groups (schizophrenia, bipolar disorder, major depressive disorder, autism and ADHD), a statistical analysis group and a cross disease working group. Early studies proved that small sample sizes (<1000 cases and 1000 controls) were often not large enough to reach genome wide significance and that larger samples (around 3000 cases and 3000 controls) were required for greater power. The formation of the PGC has allowed researchers access to far more sample sets and enabled them to increase the power of their analysis. The four main aims were to harmonize data, carry out within disorder meta-analysis, increase availability of shared data and to identify convincing associations common to two or more of the five disease groups [77]. A number of genes have been associated with schizophrenia and bipolar disorder through GWAS meta-analysis. Zinc-finger protein 804A (\textit{ZNF804A}) was originally associated with schizophrenia and bipolar disorder through genome wide association. Further meta-
analysis has replicated this association with high statistical significance ($p=2.5 \times 10^{-11}$ schizophrenia, $p=4.1 \times 10^{-13}$ schizophrenia and bipolar disorder combined) supporting the theory of genetic overlap between disorders [78]. Additional meta-analysis (n=5142 cases/6561 controls) have identified a SNP mapping 85kDa from fibroblast growth factor receptor 2 ($FGFR2$) with high association with schizophrenia ($p=0.0009$) [79]. The accumulation of data from multiple studies has led to the emergence of some interesting and consistent patterns. It is proposed that certain risk haplotypes in these genes may have a detrimental effect on brain function which is modulated by poorly understood environmental variables and social factors [80], and the spectrum of clinical features are suggestive of risk through overlapping sets of genes [81].

1.2.3 Copy Number Variation

Much recent work has investigated the role of copy number variants (CNVs) in neurodevelopmental disorders (reviewed in Kirov 2010) [82]. As the aetiology of schizophrenia and other major mental illness is so varied, with multiple genetic components identified, the general consensus is that multiple small components are likely to be responsible, and as CNVs account for a substantial proportion of human genetic variation, they are likely to play an important role.

A study of 418 individuals, including 150 individuals affected with schizophrenia or schizoaffective disorder and 268 healthy controls, identified 24 previously unreported copy number variants in schizophrenia cases [83]. These included deletions in $NRXN1$, $ERBB4$, $GRM7$ and $SLC1A$, all known synaptic genes. Additionally, Xu et al (2008) [84] identified de novo and inherited copy number variations in schizophrenia cases associated with neural development, small GTPase activity and RNA binding/processing. In a study of 724 patients with psychiatric disorders and 314 healthy controls, Saus et al (2010) [85] studied the dosage effects of 68 known candidate genes for psychiatric conditions overlapping with CNVs. Contrary to previous reports, no statistically significant difference was found in the overall burden of gains or losses in psychiatric patients compared to control individuals. However, 47% of rare structural variants identified were found only
in psychiatric disorder patients and not in control individuals. These included variants in \textit{GRM7}, and \textit{COMT}, both previously implicated in schizophrenia. Need \textit{et al} (2009) were also unable to confirm the ‘load burden’ previously reported but did identify a number of large deletions in cases but not controls. Using samples from 1013 cases and 1084 controls from schizophrenia cohorts in Aberdeen, Munich and America, Need \textit{et al} found that deletions $>$2Mb were not present in control subjects but were in many schizophrenia cases. They propose that large deletions confer risk to psychiatric illness. In the Aberdeen cohort, deletions affecting \textit{NDE1, MPV17L, ABCC1, KIAA0430, KIAA0866, MRP6} and \textit{SPRY2} were found to be significantly associated with schizophrenia, and in the Munich cohort deletions affecting \textit{TUSC3, PCM1, NAT1, NAT2} and \textit{ASAH1} were significantly associated [86]. Many of these genes have been implicated in schizophrenia previously.

1.2.4 Genome wide gene expression analysis

From current research it is clear that schizophrenia and related synaptopathies are complex trait disorders, attributable to multiple genes and epigenetic factors [87]. Microarray technology allows relatively rapid, large scale screening of genes with the potential to identify candidate genes and pathways involved in complex trait disorders.

Recent microarray studies of human schizophrenia patients have identified multiple presynaptic and myelin related genes dysregulated in the prefrontal cortex [88, 89]. Mirnics \textit{et al} (2000) identified a group of genes involved in presynaptic secretory function whose RNA expression was decreased in schizophrenia patients. Genes involved in GABA and glutamate neurotransmission were also reduced. These findings were consistent with previous targeted gene expression studies in hippocampus, which found GABA and glutamate receptor function was decreased in schizophrenia patients [90, 91]. Vawter \textit{et al} (2002) [89] carried out a targeted microarray for 1127 brain related genes in the dorsolateral prefrontal cortex of post mortem schizophrenia brains and matched controls and reported an overlap in 5 of the presynaptic secretory genes, GABA and glutamate receptor genes identified by Mirnics \textit{et al}. In a similar study, Sugai \textit{et al} (2004) identified oligodendrocyte and astrocyte related genes, and growth/neurotrophic factors.
and their receptors, as being altered in the prefrontal cortex of schizophrenia patients [92]. In the superior temporal gyrus, which has connections to the thalamus, the limbic system and the prefrontal cortex, altered expression of genes involved in neurotransmission, neurodevelopment and presynaptic function were identified in post mortem brain tissue of schizophrenia patients [93]. Two studies have also identified genes involved in mitochondrial function and oxidative stress as being differentially expressed in schizophrenia patients [94, 95].

While it is clear that microarray studies are somewhat useful in identifying candidate genes for the complex psychiatric disorders, there are a number of drawbacks to the method. Bunney et al (2003) [96] highlight the possible pitfalls of using microarray technology for studying psychiatric disorders. One major limiting factor is the availability of post mortem brain tissue from psychiatric patients. Post mortem interval and heterogeneity of human samples will effect gene expression and are difficult to control for. Brain tissue from subjects who had previously been medicated will display altered expression in some genes relative to unmedicated, otherwise matched, individuals. It is therefore important to note the medical history of the subject prior to making conclusions about gene expression patterns [96]. An alternative approach using samples from live patients, such as lymphoblastoid cell lines derived from blood samples, can increase the availability of subjects and allow ongoing assessment. Lymphoblastoid cell lines overcome the problems associated with drug treatments and post mortem interval, however, many genes that have been identified by studies using brain tissue are not expressed in blood [97] and studies comparing gene expression in brain and that in blood suggest the different tissue types will yield very different results. It is therefore suggested that while brain tissue can provide candidate genes, blood samples will provide possible biomarkers for disease.

It is of most interest that the same genes appear to recur in genetic studies of schizophrenia and mood disorders suggesting a causative link between the two illnesses [98]. The genetic overlaps between schizophrenia, recurrent major depression and bipolar disorder are becoming increasingly more convincing through the results of high
throughput microarray studies [99] and some researchers have suggested they may not be separate disorders but variations on a spectrum of a single disorder[98].

1.2.5 Cytogenetics

Cytogenetics is the study of chromosomal abnormalities that cause disease. This can be the addition of an entire chromosome (as in Down’s syndrome), the absence of a chromosome (Turner’s syndrome) or the translocation of one part of a chromosome to another. Diseases resulting from chromosomal abnormalities tend to be Mendelian in nature, however a number of genes for major mental illness have been discovered through the use of cytogenetics [100] suggesting that single genes of large effect do exist for these disorders, and that the polygenic theory of psychiatric illness may not be the whole story. This project focuses on the DISC1 gene, which was first identified in a balanced translocation in a large Scottish family with multiple individuals affected by schizophrenia.

1.3 Identification of Disrupted in Schizophrenia-1

A Scottish family was first reported by Jacobs (1970) [101] who noted a balanced translocation – (t1:11) (q42;q14.3) – in an individual with an adolescent conduct disorder and subsequently in members of four generations of the extended family. A 20 year follow up study [102] observed increased occurrence of major psychiatric disorders including schizophrenia, bipolar disorder and major depression in translocation carriers, but not in non-carrier relatives (LOD=7.1)[8]. Of interest was one individual diagnosed with bipolar disorder whose child went on to develop schizophrenia, adding to the evidence that the functional psychoses have overlapping genetic predispositions. Further studies determined that the translocation breakpoint directly disrupts two genes on chromosome 1 (DISC1 and DISC2) [56] making them ideal candidates for further research. This translocation breakpoint occurs in intron 8 of DISC1 leading to the transfer of the coding sequence for 257 C-terminal amino acids (exons 9-13) from chromosome 1.
to chromosome 11 in translocation carriers[103]. There is, however, no evidence of a C-terminal truncated protein in translocation carriers [104]. Instead the expression of DISC1 in the brains of affected individuals is reduced by half compared to karyotypically normal controls, suggesting haploinsufficiency is the most likely mechanism of susceptibility to schizophrenia and other major mental illness in (t1;11) (q42;q14.3) translocation carriers. Zhou et al [105] later identified a novel gene that was disrupted by the translocation on chromosome 11. This gene (termed Boymaw) was suggested to form fusion proteins with DISC1 in translocation carriers. While these fusion proteins have been generated and their function studied in cell cultures [105], work to determine whether fusion proteins are transcribed in translocation carriers is still ongoing.

Translocation carriers (as well as unrelated patients with schizophrenia without the translocation) show prolonged latency and reduced amplitude of P300 event related potential (a measure of attention dependant information processing which has been shown to be impaired in schizophrenia patients) compared with controls, and karyotypically normal relatives[8, 103, 106]. This finding led to the proposal that the balanced (t1;11)(q42.1;q14.3) translocation conferred predisposition to slower and more inefficient processing of stimuli in short term memory. Thus DISC1 is likely to participate in pathways important for cognitive function. DISC2 is antisense to DISC1, and it is also disrupted by the translocation. Sequence analysis of DISC2 suggests it is transcribed in the opposite direction to DISC1 and has no protein coding potential. It appears to be an anti-sense RNA gene, which may regulate activity of DISC1 [103]. So far DISC2 has only been detected in humans, suggesting relatively recent evolutionary origins.

While the translocation is confined to one family, many association and linkage studies of other populations have discovered other variants within the DISC1 locus that also associate with schizophrenia, bipolar disorder and recurrent major depression [8, 107-110]. Three independent studies have, however, failed to find association between DISC1 locus variations and schizophrenia [111-113] although it is proposed that this can be
explained by small sample size, too few SNPs being studied and ethnic differences in susceptibility [113].

Replication of linkage in a Finish sample has shown under-transmission of three common \textit{DISC1} haplotypes and major mental illness (HEP2, HEP3 and HEP4), and the over-transmission of one haplotype (HEP1). The under-transmitted haplotypes represent regions spanning intron 1 to exon 2, exon 4 and exon 13 respectively. HEP3 occurred in the control group at a frequency of 1% and a frequency of 8.8% in the schizophrenia patients group [114]. HEP3 is also associated with poor visual working memory and attention [115] and HEP1 has been associated with impaired long-term verbal memory and total hippocampal volume reduction [116]. In addition, the HEP3 haplotype is strongly associated with schizophrenia in a Scottish case-control group [117] and also shows overlap with a three-SNP haplotype associated with Aspersgers syndrome in cohort of families with infantile autism and Aspergers syndrome [118]. This study also found association with an intragenic single nucleotide polymorphism (SNP) of \textit{DISC1} (rs1322784) and Aspersgers syndrome. They also established association between autism and a \textit{DISC1} intragenic microsatellite (D1S2709) suggesting a strong role for \textit{DISC1} in early onset neuropsychiatric conditions and neurodevelopment. In a genome wide linkage scan of schizoaffective disorder, Hamshere \textit{et al} (2005) [119] reported the highest linkage peak (LOD 3.54) occurring at 1q42, close to the \textit{DISC1} locus. Furthermore, a missense allele in exon 9 of \textit{DISC1} was reported to be overrepresented in patients with schizoaffective disorder [120] suggesting \textit{DISC1} may confer risk to this disorder.

\textit{DISC1} has been associated with neurocognitive functioning [121] in measures of verbal working memory and rapid visual searching. The association of variations in \textit{DISC1} with cognitive function was investigated in a large birth cohort tested using a general mental ability test at ages 11 and 79 [122]. Cognitive ability was compared with genotype within age groups, and between ages 11 and 79 to test for effects on cognitive ageing. After adjustment for cognitive ability at age 11, it was found that females homozygous for the \textit{DISC1} Cys (Ser704Cys SNP) allele had significantly lower cognitive ability score than males, suggesting a sex specific role of \textit{DISC1} genotypes on cognitive ageing [122].
Callicott et al (2005) also report that a nonsynonymous SNP in exon 11 (Ser704Cys) of *DISC1* is associated with schizophrenia. Using fMRI they observed abnormal engagement of the hippocampus during cognitive tasks, and a reduction in hippocampal grey matter volume [123]. Further to this, Di Giorgio et al (2008) [124] investigated the effects of the polymorphism on grey matter volume and formation, hippocampal formation and functional coupling during memory recognition tasks. Using fMRI they showed that individuals with two Ser alleles had greater engagement of the hippocampus and greater functional coupling to the frontal cortex during memory recognition tasks, but in contrast with Callicott et al, observed higher grey matter volume than heterozygous individuals. While these results are in part contradictory, they both support a role for *DISC1* in hippocampal formation and function. The authors suggest this inconsistency is due to the ‘flip-flop’ effect, where the direction of the allele association may be dependent on interaction with another allele at a different locus, thus the same allele could exert different effects in different populations depending on their genetic pressures. The Cys allele has also been shown to reduce activation of the left medial and superior frontal gyrus during a verbal fluency task, relative to Ser homozygotes [125]. No brain regions were found to be significantly more activated in Cys carriers. These results support a strong physiological effect of the *DISC1* polymorphism in both the hippocampus and the frontal cortex. At the molecular level, sRNAi knockdown of *DISC1* results in suppressed phosphorylation of ERK and Akt. The Ser704 allele phosphorylated ERK to a greater extent than the Cys704 allele, implicating reduced biological activation of ERK due to polymorphisms in *DISC1* in major mental illness pathology [126].

Current estimates suggest that approximately 2% of individuals with schizophrenia carry a *DISC1* missense mutation [127]. As discussed in the next section, DISC1 acts as a hub protein, with multiple interactors, and as such this estimate may be under representative. Interplay between SNP variants [128], common cis-variants [128], and interacting proteins [129] has been shown to result in modest reductions of *DISC1* gene expression, which may exert subtle effects on neurodevelopment, neurophysiology and neural circuitry. The DISC1 protein has also been shown to interact with transcriptional modulators of cAMP signaling, cytoskeletal, synaptogenic, neurodevelopmental and
sensory perception proteins [128]. Current targets of psychiatric drug development are highly enriched in this group [128]. With recent evidence suggesting there may be as many as 50 DISC1 mRNA isoforms [130], showing dramatic differences in their brain expression profiles, it is proposed the combination of common variants of low penetrance, and rare variants of high penetrance within the DISC1 pathway may contribute a much larger fraction of the genetic variance in schizophrenia and related disorders than previously thought [127].

1.4 The Biology of DISC1 and its interactions

DISC1 acts as a molecular scaffold and has been shown to interact with multiple proteins including NUDEL and PDE4B, suggesting roles in neurodevelopmental and signalling pathways [131, 132]. The known isoforms of DISC1 are produced by alternate splicing of 13 major exons, of which exon 2 is the longest (955bp) and present in all isoforms. Exon 2 encodes most of the protein head domain, and haplotypes spanning this region have been shown to associate with schizophrenia and working memory [107, 108].

Recently an interaction network of 127 proteins and 158 interactions has suggested DISC1 may also be involved in synapse function and development through these complex interactions [132]. The majority of these interactors can be loosely classified as cell cycle, signal transduction, cytoskeleton, intracellular transport and central nervous system development genes, clearly demonstrating the importance of DISC1 in multiple diverse and critical brain functions. A direct role for DISC1 in early brain development in vivo was first demonstrated by Kamiya et al (2005) [133] who reported that in utero application of short hairpin loop RNA oligonucleotides (shRNA) targeting mouse Disc1 can repress expression of the gene, resulting in reduced migration of neurons out of the sub-ventricular zone to the cortical plate. This is accompanied by altered cell polarity and reduced dendritic arborisation. This correlates well with evidence from mouse studies where Disc1 is upregulated at crucial developmental time points both embryonically (E13), when neurogenesis and cell migration commence, and at the onset of puberty.
(P35) when synaptogenesis is again occurring [134]. From puberty onwards the expression of *Disc1* in the mouse remains steady. Interestingly, in adult hippocampal neurogenesis, downregulation of *Disc1* results in increased acceleration of neuronal integration, resulting in mispositioning of dentate granule cells [135]. Newborn *Disc1* knockdown dentate granule cells display increased excitability, dendritic development and synapse formation. Frequency of GABAergic and glutamatergic spontaneous synaptic currents in the *Disc1* knockdown cells was consistently higher, indicative of increased synaptic excitability[135]. This would suggest that *Disc1* has distinct functions during development that differ pre and post-natally, and that it is likely to play a role in sustained and simultaneous firing of neurons.

While DISC1 has been shown to have multiple interactors, only around 25% of these have been confirmed by further study. Of those which have been confirmed however, many have been implicated in schizophrenia and related disorders in their own right, including NDE1, NDEL1, PDE4B [65] PCM1 [129] and FEZ1 [136], and it has been suggested that in some cases where DISC1 is not directly implicated, variants acting on its binding partners may act instead to confer risk of major mental illness [65]. In the following paragraphs I will discuss the major confirmed interactors of DISC1 with links to psychiatric illness; NDE1, NDEL1, LIS1, GRB2, PCM1, GSK3β, FEZ1, ATF4 and ATF5.

NDE1, NDEL1 and LIS1 form part of a conserved nuclear distribution pathway involved in neurogenesis and neuronal migration [137-140]. Mutations in *Lis1* lead to deficits in neuronal migration, neuroblast proliferation and cortical layering [141, 142]. Similarly, mice with mutations in *Nde1* and *Ndel1* show defective neurogenesis and neuronal migration [139, 140]. DISC1 and NDEL1 have been shown to be selectively upregulated during neurite outgrowth in PC12 cells differentiated with neuronal growth factor (NGF) (Kamiya 2006) [143]. Inhibition or disturbance of the DISC1-NDEL interaction has been shown to inhibit neurite outgrowth in cell culture. This would suggest the resulting complex is crucial for successful neurite outgrowth [143].
These genes have also been associated with schizophrenia in their own right. mRNA studies of post mortem brain implicate both LIS1 and NDEL1 [144] in schizophrenia, while NDE1 and NDEL1 have also been associated with schizophrenia in a Finnish cohort [145, 146].

GRB2 competes with NDEL1 for binding to DISC1 [147] giving two functionally distinct interactions. The DISC1/GRB2 interaction is required for neurotrophin induced axonal elongation [147]. GRB2 has been shown to bind several brain proteins including PDE4D [148], huntingtin [149], amyloid precursor protein [150] and presenilin 1 [151] as well as dopamine receptors 3 and 4 [152]. As such, its involvement in Alzheimers disease and molecular pathways involved in other psychiatric syndromes has been suggested [153].

PCM1 is a centrosomal protein required for the targeting of multiple proteins to the centrosome for the regulation of microtubular dynamics [154]. Co-immunoprecipitation in HEK-293 cells suggests PCM1 and DISC1 may interact to regulate neuronal migration. In addition, knockdown of DISC1 expression reduced accumulation of PCM1 at the centrosome. Further RNAi studies of both DISC1 and PCM1 show delayed radial neuronal migration if either protein is knocked down [133, 155]. Gurling et al (2006) [156] reported an association between haplotypes of PCM1 and orbitofrontal grey matter defects in schizophrenia family and case-control samples although this has as yet failed to replicate in other samples.

GSK3ß is a presynaptic protein kinase that is proposed to be involved in the negative regulation of synaptic vesicle fusion events [157]. GS3Kß has been shown to suppress long-term potentiation and presynaptic release of excitatory glutamate in neurons. Mao et al (2009) [158] observed that suppression of Disc1 by shRNA in embryonic mouse brains resulted in reduced neural progenitor proliferation, that could be rescued using the GSK3ß specific inhibitor SB216763. GSK3ß is inhibited by direct DISC1 interaction suggesting that a loss of DISC1 results in increased GSK3ß and leads to reduced neural progenitor proliferation, leading to premature cell cycle exit and differentiation.
Therefore it is suggested that the DISC1-GSK3β interaction is required for successful neural proliferation.

Many current typical and atypical antipsychotics increase phosphorylation of GSK3β and enhance AKT signalling though GSK3, or by activating AKT directly. AKT1 itself has been tentatively associated with schizophrenia [159] although further studies are required. The actin binding protein Girdin (KIA1212) was initially identified as a potential DISC1 interactor by Camargo et al 2007 [132]. Since then, its interaction with DISC1 has been confirmed both in vivo and in vitro. Interaction of DISC1 with Girdin in HEK293 cells suppresses AKT signaling [160]. When Kim et al [160] knocked down Disc1 in mouse hippocampal neurons, Akt activation increased, and increased dendritic growth and abnormal migration were observed. Similar observations were made when Girdin was selectively overexpressed in these neurons [160]. Interestingly, the effects of Disc1 suppression could be rescued by rapamycin, which inhibits an AKT activated effector pathway. Enomoto et al (2009) [161] showed that Girdin suppression also mimicked the effects of Disc1 supression in cultured hippocampal neurons. Girdin appears to be anchored at the growth cone by DISC1, and regulates the migration and positioning of dentate gyrus cells from this position. Girdin has also been shown to interact with NDEL, another known interactor if DISC1.

Fasciculation and Elongation factor Zeta 1 (FEZ1) is involved in axon outgrowth and has been shown to be expressed in a pattern similar to that of DISC1 in the rat brain [162]. In addition, protein overexpression leads to robust binding of DISC1 to FEZ1 and co-localisation to growth cones in rat cultured hippocampal neurons [163]. Expression of FEZ1 in post-mortem brain samples is reduced in subjects with schizophrenia [144] however genetic/association evidence for FEZ1 as a susceptibility gene for schizophrenia is weak [164-166].

ATF4 and ATF5 have been associated with schizophrenia through the use of yeast two hybrid screens of DISC1 [167]. ATF4 has been shown to interact directly with nuclear DISC1 and a corepressor, N-Cor, to modulate CRE-mediated gene transcription [168].
There is currently little evidence to suggest ATF4 and ATF5 are involved in psychiatric illness in their own right. As they are both cAMP-responsive transcription factors, however, and because ATF4 is partially regulated by \textit{NRG1}, another well characterised schizophrenia susceptibility gene [169] they are of interest for further study. Furthermore, it has been shown that DISC1 and PDE4B interact dynamically to regulate cAMP signalling [131]. Disruption of this interaction, as predicted by some mouse models (section 1.7) would, therefore, be predicted to alter modulation of cAMP signalling and may result in abnormal gene transcription. PDE4s are orthologous to the Drosophila \textit{Dunce}, which is involved in learning and memory and known to affect synaptic plasticity, which in turn requires alteration of gene expression profiles [170, 171]. DISC1 also possibly binds to chromatin remodelling factors, such as SMARCE1 [132]. These interactions with transcription factors, together with the fact that DISC1 localises to the nucleus [172, 173] is consistent with a role for DISC1 in transcriptional regulation. These are discussed further in the next section.

\textbf{1.5 Current hypotheses of Major Mental Illness}

\textbf{1.5.1 The Neurotransmitter Hypotheses}

Selective targeting of individual neurotransmitters has not yet yielded many therapeutic results bringing us to the obvious conclusion that no single neurotransmitter hypothesis can account for the wide range of symptoms present in schizophrenia. However, the combined effect of neurotransmitter dysfunction is likely to account for a large proportion of disease pathology. The serotonergic system has been implicated in both schizophrenia and major depression. While no abnormalities in receptor density have been found in schizophrenia cases [174], the affinity of successful atypical antipsychotics for the serotonin receptor 5-HT2A support a role for the receptor in schizophrenia pathology [175]. The formation of the glutamate hypothesis of schizophrenia came from the observation that PCP, an NMDA receptor antagonist, could induce schizophrenia-like symptoms in healthy subjects [176]. This is also true of other NMDA antagonists [177-179]. An increase in NMDA receptor binding has been observed in the cortex of post-
mortem brain samples from schizophrenia patients [180] and a decreased release of glutamate has been reported in synaptosomes prepared from frozen brain samples of schizophrenia patients [181]. Genes involved in NMDA receptor function have been associated with schizophrenia adding further weight to the glutamate hypothesis. One of the most established associations, NRG1, regulates the expression of NMDA receptors through ErbB4. GRIN2B, the gene coding for the NR2B subunit of the NMDA receptor also shows a small but significant association with schizophrenia [182].

Dysfunction of dopamine neurotransmission has also been implicated in the pathology of schizophrenia based on the fact that dopamine-mimetic drugs elicit hallucinations, and that neuroleptics caused rigidity [183]. The original hypothesis stated that dopamine pathways may be overactive in schizophrenia. Recent studies have suggested an increase in dopamine 2 receptors in the striata of schizophrenia patients, which, while not statistically significant, indicated a substantial positive trend in dopamine activity [184] [185]. In addition, a transgenic mouse with selectively inducible over-expression of dopamine 2 receptors displays selective cognitive impairment in working memory tasks, but does not exhibit a global cognitive deficit [186]. This working memory deficit is not reversed when the transgene is switched off indicating that it is the expression level during development and not continued expression that is of key importance. Deficits in spacial working memory in individuals with schizophrenia have been well characterized ([187, 188] and others)

1.5.2 The Developmental Hypothesis
The neurodevelopmental hypothesis states that small neurodevelopmental deficits in key circuitry during brain development could lie dormant until puberty, when normal molecular changes could facilitate the onset of disease in affected individuals [189]. Studies of children from high risk family groups suggest that delays in motor and neurological development, deficits in attention and verbal short-term memory and poor social skills are evident in children who later go on to develop schizophrenia [190]. Individuals with high genetic liability have been shown to have reductions in grey matter
density in the prefrontal and temporal cortex, amygdala, hippocampus and superior temporal gyrus (reviewed in [191]). High risk individuals with subsequent follow up scans revealed a significantly greater reduction in temporal lobe size in those who developed psychotic symptoms, suggestive that the reduction in grey matter volume may be an early indicator of disease. It has also been reported that early onset human schizophrenia patients display a delay in brain development with respect to overall brain volume [192]. Evidence from animal models of schizophrenia identifies clear neurodevelopmental deficits in migration and proliferation of neurons in the hippocampus and cortex [133, 193], two regions implicated in schizophrenia pathology. Some very compelling evidence from transient knock-down studies of Disc1 suggests that a loss of Disc1, particularly in pyrimidal cells of the prefrontal cortex, results in behavioural abnormalities after puberty [194]. There was a reduction in parvalbumin-positive cells (a marker of fast-spiking GABA interneurons) and reduced dopamine expression after puberty but not before. Knock down mice also displayed a marked deficit in prepulse inhibition, which was not observed at earlier ages. The authors suggest that a lack of Disc1 during development results in abnormalities in the postnatal maturation of dopaminergic neurons which causes dendritic abnormalities and an overall disturbance in neural circuitry evident after puberty [194].

1.5.3 The cAMP hypothesis

PDE4B and PDE4D belong to the family of phosphodiesterase encoding genes whose protein products are involved in the inactivation of adenosine 3′,5′-monophosphate (cAMP) in the cell. cAMP is part of a second messenger system thought to be critically involved in learning and memory, and mood. Induction of cAMP carried out by Millar et al[104] resulted in decreased binding of PDE4B to DISC1 suggesting that this interaction is cAMP dependent. Immunoprecipitation of PDE4B showed that binding to DISC1 was direct and that it required an intact DISC1 N-terminal “head domain”. Therefore it is possible that functional or structural variation in DISC1 will modulate the interaction with PDE4B, affecting cAMP catabolism and thus altering expression of other downstream genes. In addition, a chromosomal rearrangement disrupting PDE4B was
identified in a family with schizophrenia and related psychoses [131]. PDE4 activity has been linked to memory formation in the fly (cognate *dunce* mutation) and to mood disorders in the mouse [195]. It is also the target for rolipram, a potent prototypic antipsychotic and antidepressant [196]. Furthermore, Clapcote et al [197] describe two mouse models of major mental illness that both have mutations in a Pde4b binding site of Disc1.

1.6 Genome-wide expression and pathway analysis

Analysis of gene expression datasets using pathway and network analysis is now providing insight into the regulatory mechanisms of disease. Porteous and Hennah (2009) [128] employed pathway analysis to determine correlations between variants in *DISC1* pathway genes and levels of gene expression in public domain datasets. Data mining of the four available HapMap population cohorts (GSE6536 in NCBI GEO) [198] identified six cis-acting variants that showed association in at least three of the four cohorts. In all cases the minor allele was associated with reduced *DISC1* expression. Further analysis incorporating the six cis-acting variants, three known missense mutations, and three known genetic interplay SNPs (conferring risk, neutral or protective effects) identified 100 genes that could be connected in a pathway of interacting molecules. Gene ontology analysis of this pathway identified over-enrichment of genes involved in cytoskeletal function, transcription factor function, synaptogenesis and sensory perception. While the data from Porteous and Hennah’s study pertains to normal genetic variance, its relevance to unraveling the function of *DISC1* is undeniable, and the identification of regulatory networks gives insight into the mechanisms of psychiatric disease. The use of pathway analysis to further test the validity of candidate genes and their likely contribution to psychiatric disorders is without doubt a valuable tool for gene expression analysis.

Recently, Torkamani et al (2010) have used co expression network analysis to determine gene clusters of functional significance in schizophrenia [199]. Genes were organised into functional modules, which were co-regulated and as such likely to be involved in
similar cellular processes and pathways. Using standard analysis of differential gene expression they found that genes with altered expression in schizophrenia clustered into five distinct gene modules, four of which were preferentially associated with neuronal function. They also found that changes in gene expression between patients with schizophrenia and controls differ with age. Harris et al [200] previously reported genes related to CNS development, neuron guidance and neurotransmitter secretion were down-regulated in healthy individuals during post natal development in a sample set from individuals from birth to early twenties. Torkamani et al propose that the downregulation observed in their control subjects is a continuation of this process in normal individuals. These changes were not observed in subjects with schizophrenia however, suggesting a progressive neurodevelopmental deficit. It is suggested that the lack of normal down-regulation of these CNS developmental genes may act as a trigger for the onset of disease [199]. As schizophrenia is not normally diagnosed until the late teens, gene expression measurements of individuals with schizophrenia during their early post-natal development is not possible. The use of animal models may be useful in shedding light on normal age-related changes in neuronal gene expression levels.

The use of animal models is becoming increasingly popular in psychiatric research and allows further investigation of candidate genes in a system less limited by sample availability and tissue type. The effect of alterations in specific candidate genes can also be assessed taking investigation into the mechanisms of psychiatric disorders forward more rapidly.

1.7 The Schizophrenia Mouse
One question I have been repeated asked is ‘so how do you know if a mouse is schizophrenic?’. This is of course a loaded question, mice are not schizophrenic. We can however genetically or pharmacologically manipulate mice to display certain behavioural phenotypes which model the human condition in an effort to further understand the underlying mechanisms, but major psychiatric syndromes are uniquely human conditions
and without a defining pathology, and as such, modeling is limited to examination of a specific dimension of the disorder [201]. Endophenotypes are heritable, intermediate phenotypes between genes and expression of a disorder and are often used as ‘biomarkers’ for disease. It is generally considered that each complex disorder phenotype is made up of a number of endophenotypes, which are under different genetic control, so by studying single endophenotypes the problem of finding genes involved becomes less complex [202]. For example, genes influencing liability to mental disorders such as schizophrenia are likely to act on the multiple neural systems that have already been associated with this disorder. These could include serotonergic, glutamatergic and dopaminergic systems mediating processes such as learning and memory, social cognition, emotion and sensorimotor gating. Neurophysiological, neuropsychological and biochemical measurements from these systems give quantifiable measures separate from disease symptoms that can be used to map multiple genes of small effect [202]. The use of endophenotypes also allows researchers to model certain aspects of psychiatric disorders in animals, using quantifiable measures that can be related to the human condition [203]. While there are no universally accepted risk loci for major psychiatric syndromes [204, 205], researchers are concentrating efforts on modeling loci which have been identified through cytogenetics, copy number variants, association, linkage or gene expression studies as candidate susceptibility genes. Multiple mouse models of psychiatric syndromes now exist, though this section will concentrate on examples of Disc1 animal models.

Koike et al (2006) [206] reported that a 25bp deletion variant in exon six of Disc1 in the 129S6/SvEv mouse strain, results in a lack of full length Disc1 protein and impairment of working memory in a C57BL6 backcross (>98% C57BL6 genotype). Gross brain morphology of these animals was reported to be normal, however in a delayed non-match space test (which measures special working memory) there was a clear working memory impairment in both heterozygous and homozygous mutant animals, with a dosage dependant effect. This same model was used by Kvajo et al [207] when studying neuronal architecture and cognition. They confirmed the previous working memory impairment and found alterations in the organisation of neurons in the dentate gyrus, and
reduction of short term potentiation of the CA3/CA1 hippocampal synapses. The strength of synaptic transmission, release probability and long term potentiation were not significantly different in the mutant animal. Working memory deficits are considered fundamental cognitive symptoms of schizophrenia [208] and impaired dendritic growth and misorientation of dendrites have previously been described in human schizophrenia subjects [209]. A third study of this mouse strain, which used all available DISC1 antibodies, found that all isoforms of DISC1 were still expressed, suggesting bypassing of the deletion through exon skipping [210]. It is yet to be confirmed whether the original working memory deficits reported by Koike et al are indeed a result of the Disc1 deletion or what other factors may be in play.

Another mouse model, carrying a dominant-negative Disc1 mutation on a CamKII promoter, shows enlargement of the lateral ventricles and a selective reduction in the immunoreactivity of parvalbumin in the cortex, indicating a deficit in interneurons, leading to cortical asynchrony [193]. Enlargement of the lateral ventricles has long been associated with schizophrenia [211] and is usually present at disease diagnosis. In addition, this model also displays disturbed sensorymotor gating, increased hyperactivity and measures of anhedonia, all behavioural attributes that have previously been reported in patients with schizophrenia. It should be noted that these behavioural anomalies are not restricted to schizophrenia, and have been associated with multiple psychiatric syndromes, but the combination of phenotypes and the known genetic connection between DISC1 and schizophrenia indicate that this model may mimic some aspects of the human schizophrenia phenotype.

In 2008 Pletnikov et al [212] generated a transgenic mouse model with inducible expression of mutant human DISC1. Expression of the transgene was limited to the forebrain regions to counteract any effect of the mutation elsewhere. No neurodevelopmental abnormalities were observed, but there was mild enlargement of the lateral ventricles and primary cortical neurons derived from mutant DISC1 mice produced significantly less elaborate neurite outgrowth compared to those from wild-type controls. Some gender specific behavioural phenotypes were observed which were consistent with
other Disc1 mutant mouse models. Inducible Disc1 male mice show increased horizontal activity in the open field similar to the dominant-negative Disc1 model [193] and the L100P ENU mouse model [197]. Males also display reduced social interaction consistent with the Q31L ENU mouse model [197], and increased aggression. Females display a reduction in spacial learning and memory, which has not previously been shown. Further studies indicate that prenatal over-expression of N-terminal human DISC1 results in a reduction of parvalbumin-positive and dopamine-positive neurons in adult mouse cortex. This was indistinguishable to that observed in animals with continuous expression of N-terminal human DISC1, indicating that there may be a critical period for DISC1 expression [213].

In an effort to closely mimic the genetics of the Scottish (t1:11) translocation family, Shen et al (2008) developed a transgenic Disc1 mouse that expressed a truncated form of Disc1 comprising of exons 1-8 [214]. Truncated Disc1 is expressed in the cerebellum, cerebral cortex and hippocampus of these mice. Enlargement of lateral ventricles and a reduction in cerebral cortex was observed along with reduced neurite outgrowth in culture consistent with observations from Pletnikov et al in the inducible expression model [212]. Similar to the phenotype of the dominant-negative mouse model [193] and recent work on the inducible expression mouse model [213], these animals also display a reduction in parvalbumin neurons in the cortex. This model also displays many novel phenotypes not previously reported in a Disc1 mouse model, including reduced parvalbumin-positive neurons in the hippocampus, thinning of layer II/III in the cortex and selective decrease of neural proliferation in the cortex. Some aspects of the behavioural phenotype are consistent with those observed in both the dominant-negative mouse model [193] and the ENU mouse model (Clapcote et al 2007, discussed next), with increased immobility in depression related tests and reduced latent inhibition (table 1.1).
Table 1.1: Phenotypes of DISC1/Disc1 mouse models (adapted from Shen et al 2008). Column 1 shows the phenotype and columns 2-7 the mouse models. NS = not significant, - = unknown, ↑ = increased, ↓ = reduced, +, ++, +++ = positive effect with p<0.05, 0.01 and 0.001 respectively.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Phenotype</td>
<td>129 mice</td>
<td>DBN DISC1</td>
<td>Inducible DISC1</td>
<td>Truncated DISC1</td>
<td>Q31L L100P</td>
</tr>
<tr>
<td>Brain volumetry</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>↑ MALE</td>
<td>↓ 6%</td>
</tr>
<tr>
<td>Lateral ventricles</td>
<td>-</td>
<td>-</td>
<td>↑</td>
<td>-</td>
<td>↓ 13%</td>
</tr>
<tr>
<td>Cerebral cortex</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>Neural proliferation</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>↓</td>
<td>-</td>
</tr>
<tr>
<td>Neurite outgrowth</td>
<td>-</td>
<td>-</td>
<td>↑</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Corpus callosum</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Penialgic in MEP cortex</td>
<td>-</td>
<td>↓</td>
<td>↓ (Ayan et al 2010)</td>
<td>↓</td>
<td>-</td>
</tr>
<tr>
<td>Penialgic in hippocampus</td>
<td>-</td>
<td>-</td>
<td>↓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Open field horizontal activity</td>
<td>NS</td>
<td>↑</td>
<td>↑ MALE</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Anxiety</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>-</td>
<td>NS</td>
</tr>
<tr>
<td>Aggregation</td>
<td>-</td>
<td>-</td>
<td>↑ MALE</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Stereotypy</td>
<td>-</td>
<td>NS</td>
<td>↑ MALE</td>
<td>-</td>
<td>↑ NS</td>
</tr>
<tr>
<td>Spatial learning and memory</td>
<td>NS</td>
<td>NS</td>
<td>↑ FEMALE</td>
<td>-</td>
<td>NS</td>
</tr>
<tr>
<td>Working memory</td>
<td>↓</td>
<td>-</td>
<td>-</td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>Propulsive inhibition (FPI)</td>
<td>NS</td>
<td>↓</td>
<td>NS</td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>Latent inhibition (LI)</td>
<td>-</td>
<td>-</td>
<td>↑ FEMALE</td>
<td>-</td>
<td>↑ NS</td>
</tr>
<tr>
<td>Immobility in forced swim test</td>
<td>-</td>
<td>↑</td>
<td>-</td>
<td>↑</td>
<td>↑ NS</td>
</tr>
<tr>
<td>Immobility in tail suspension test</td>
<td>-</td>
<td>-</td>
<td>↑</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Stress calls</td>
<td>-</td>
<td>-</td>
<td>↑ MALE</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Rotarod in FPI</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>NS</td>
<td>+++</td>
</tr>
<tr>
<td>Bagrination in FPI</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+ + +</td>
<td>NS</td>
</tr>
<tr>
<td>Clearance in LI</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+ + +</td>
<td>+</td>
</tr>
<tr>
<td>Clearance in horizontal activity</td>
<td>-</td>
<td>-</td>
<td>NS</td>
<td>+ + +</td>
<td>+ + +</td>
</tr>
<tr>
<td>Bagrination in PST</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+ + +</td>
<td>+</td>
</tr>
</tbody>
</table>

Our collaborators Clapcote and Roder have developed two independent Disc1 mouse mutants at the University of Toronto each displaying a different phenotype of psychiatric illness [197]. N-ethyl-N-nitrosourea (ENU) is an alkylating agent that when injected into male mice, induces mutations at a frequency of one per locus in every 700 gametes (Davis et al 1998). Clapcote and Roder screened the RIKEN ENU-based gene-driven mutagenesis system (RGDMS) for Disc1 mutants and found to two independent point mutations, Q31L and L100P, in exon 2 of the mouse Disc1 gene (figure 1.1) in C57BL6/J sperm, which was then used to impregnate DBA female mice. First generation individuals carrying either mutation were backcrossed with C57BL6/J mice for 6 generations until a predominantly C57BL6/J background line (average 98.5% at N6) was intercrossed to generate homozygous and heterozygous mutation carriers.

The L100P and Q31L point mutations change the amino acid sequence and are predicted to alter the structure and therefore the function of the protein. The evidence suggests that in adult mice the gene is expressed at the same level as in wild types but that it may have
different functional properties. In the Q31L mutant the mutation from glutamine, a
hydrophilic amino acid normally expected to be on the outer surface of the protein, to
leucine, a hydrophobic amino acid expected to be found on the protein inner surface
would be expected to cause a distinct conformation change. The Q31L mutation also
disrupts a known pde4b binding site in the disc1 head domain. Cell line studies have
shown that binding of pde4b is significantly reduced in both the Disc1 mutated lines. In
Q31L mice, the amount of pde4b protein is not altered from the wild type level, but
activity is reduced by up to 50% [197]. The change from leucine to proline in the L100P
line would also be expected to result in distinct structural alterations of the disc1 protein
due to the ring structure of proline, which is known to cause a sharp transition in
polypeptide chain direction. It is highly likely both mutations result in altered Disc1
proteins at the structural level.
Figure 1.1: The Disc1 Protein. (A) Amino acid changes in Disc1 resulting from ENU mutagenesis of C57B6/J sperm cells (B) Adapted from Clapcote et al. 2007. Amino acid sequence of the Disc1 protein with the two ENU mutations highlighted. Conserved amino acids between mouse and human are shaded in grey. (C) Diagram of the Disc1 protein with known regions of interaction. The area directly affected by the ENU mutations is shown by two red lines in the head domain. NLS = nuclear localisation sequence, LZ = leucine zipper.
Prepulse inhibition and latent inhibition are common methods used to quantify information processing (sensory-motor gating) deficits in schizophrenia [107]. Mice homozygous for L100P and Q31L had lower prepulse inhibition (degree to which the startle response is reduced when startle eliciting stimulus is preceded by non-startle eliciting stimulus) than wild type controls, and latent inhibition (the degree to which exposure to a conditioned stimulus decreases the salience of a paired unconditioned stimulus) was impaired in both lines. Behaviours associated with depression were observed using the forced swim test, social interaction test and reward response test in Q31L mice only (Table 1.2). Homozygote Q31L mice displayed high immobility in forced swim tests and avoided social interaction and reward responsiveness [197].

<table>
<thead>
<tr>
<th>Behaviour</th>
<th>Change compared to wild type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety</td>
<td>=</td>
</tr>
<tr>
<td>Startle reactivity</td>
<td>= reduced</td>
</tr>
<tr>
<td>Spatial learning and memory</td>
<td>=</td>
</tr>
<tr>
<td>Taste Sensitivity</td>
<td>= n/a</td>
</tr>
<tr>
<td>Olfactory Function</td>
<td>=</td>
</tr>
<tr>
<td>Prepulse Inhibition</td>
<td>reduced</td>
</tr>
<tr>
<td>Latent inhibition</td>
<td>greatly reduced</td>
</tr>
<tr>
<td>Social interaction</td>
<td>reduced</td>
</tr>
<tr>
<td>Reward Responsiveness</td>
<td>reduced</td>
</tr>
<tr>
<td>Immobility during forced swim test</td>
<td>increased</td>
</tr>
</tbody>
</table>

Table 1.2: Common behavioural traits and the effects of the ENU missense mutation on phenotype when compared to the C57B6/J parent line = no significant change. Adapted from Clapcote et al 2007

Widely used antipsychotic and mood stabilising drugs were shown to reverse these behavioural anomalies, almost to the levels of non drug treated wild type controls (Table 3). Interestingly, treatment with the antipsychotics clozapine and haloperidol was effective in correcting the behavioural abnormalities in the L100P mutants, as was the PDE4 inhibitor rolipram, while these treatments had no effect on correcting the majority of the deficits observed in the Q31L line. Conversely the antidepressant biogenic amine
reuptake inhibitor Bupropion was effective in rescuing behaviour in Q31L but not L100P mice [197].

<table>
<thead>
<tr>
<th>Drug Treatment</th>
<th>Q31L hom</th>
<th>L100P hom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-pulse inhibition</td>
<td>Citalopram</td>
<td>=</td>
</tr>
<tr>
<td></td>
<td>Haloperidol</td>
<td>=</td>
</tr>
<tr>
<td></td>
<td>Rilpiram</td>
<td>=</td>
</tr>
<tr>
<td></td>
<td>Bupropion</td>
<td>positive p&lt;0.001</td>
</tr>
<tr>
<td>Latent inhibition</td>
<td>Citalopram</td>
<td>=</td>
</tr>
<tr>
<td></td>
<td>Haloperidol</td>
<td>positive p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Bupropion</td>
<td>positive p&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Rilpiram</td>
<td>=</td>
</tr>
</tbody>
</table>

Table 1.3: The effects of drug treatment on behavioural rescue in Q31L and L100P missense mutant mouse lines. Adapted from Clapcote et al 2007

From the pharmacological responses and distinct behavioural phenotypes observed Clapcote et al [197] categorized the Q31L mutation as depressive-like and the L100P mutation as schizophrenia-like. They also observed reductions in brain size coupled with tissue shrinkage in cortex, thalamus and cerebellum. These areas correspond well to known neuroanatomical features of schizophrenia. This could be due to the predicted altered function of DISC1 in these mutants during brain development, but prior to this time a developmental profile of whole genome gene expression has not been obtained for the ENU missense mutant lines.
1.8 Thesis Aims

There were four main aims of this thesis.

1. to determine the effect of the two missense mutations on whole genome gene expression through development and into adulthood
2. to determine the effects of certain drug treatments on gene expression in adult mice ultimately allowing us to determine potential target genes, or pathways for targeting by drug treatment.
3. to determine gene expression levels throughout embryonic and early post-natal development of genes identified in the initial study.
4. to determine the effects of the disc1 mutations on protein expression in the adult mouse and mature neuronal cultures

My hypothesis was that the expression levels of genes involved in brain function and development would be altered in the mutant mouse lines. I proposed that these alterations would occur from early embryonic stages, correlating with previous reports of delays in neuronal migration and maturation in Disc1 mouse models. My secondary hypothesis would suggest that administration of drugs previously shown to correct aberrant patterns of behaviour in this mouse model may correct the expression patterns of these genes in the tissues tested. This would correlate with the changes in endophenotypes and patterns of behaviour observed in previous studies. It would be hypothesized that if this is the case it is a secondary effect (due to the action of the drug binding to receptors and preventing neurotransmitter release) and not the primary action of the drug, but may give some indication of the drugs mode of action and future development potential.
Chapter 2
Materials and Methods
2.1 Collection of tissue samples

In order to generate a wild type Disc1 developmental profile, C57BL6 female mice were obtained from Harlan, UK and bred to males of the same strain held in-house at the Western General Hospital Biomedical Research Unit. Females were taken at key time points throughout gestation based on previous results from Shurov et al [134], and sacrificed by cervical dislocation. Pups taken for post-natal developmental stages were sacrificed by decapitation (if under 2 weeks) or cervical dislocation. C57BL/6 is an inbred laboratory mouse strain widely used in research as a wild type or background strain as it has a permissive background for maximal expression of most mutations (www.jax.org). They are long lived and breed well.

Missense mutant mice for the microarray study were obtained from Steve Clapcote and John Roder at the Samuel Lunenfeld Research Institute, Toronto, Canada. Wild type C57BL6J mice were obtained from Jackson Laboratories (www.jax.org). Where possible, matings were set up between homozygous individuals to remove the need to genotype all animals. Adults of 12 weeks used for the drug treatment trial and developmental stages were obtained as above.

2.1.1 Use of Inbred Strains

The C57BL/6J inbred strain was created by Dr. CC Little (who went on to found the Jackson laboratories) in the 1950’s and all mice in this strain are derived from this original mating. All C57BL6 mice carry a mutation Cdh23\footnote{753A} which causes a frame skip of exon 7 in Cadherin 23 resulting in age related hearing loss at around 10 months of age. Subsequent splitting of the original C57BL6 line has resulted in the arrival of new spontaneous mutations specific to subpopulations. One such mutation results in the deletion of the alpha-synuclein locus (which has been implicated in Parkinsons disease) in C57BL6 mice from the subpopulation available from Harlen UK (www.harlan.com). While this is not thought to affect other genes in this population we wanted to ensure our background strain was as close to that of the missense mutants as possible.
The L100P and Q31L mutants were backcrossed for several generations with C57BL/6J mice from the Jackson Laboratories (which do not carry the alpha-synuclein deletion). Due to time restrictions and problems achieving successful matings while in Toronto it was deemed necessary to complete the collection of C57BL6 embryonic stages back in the UK. The animals immediately available to us were C57BL6 mice from the Harlan subpopulation, so to ensure continuity it was necessary to order new breeders from Charles River (www.criver.com), the only UK supplier of the Jackson C57BL/6J substrain. This done, it meant we could confidently make direct comparison to those samples collected in Toronto. It would however be interesting to compare gene expression between the Harlan and Jackson subpopulations to determine if the spontaneous mutation would have had an effect on the experimental outcome.

All necessary measures were taken to ensure minimal stress to animals prior to sacrifice.

2.1.2 Extraction of Genomic DNA from adult tissue
Mice were earmarked for identification and the ear notches used to extract DNA for genotyping. 300µl 50mM NaOH was added to the ear notches at room temperature and then incubated at 95°C for 15-20 minutes on a heat block to break down the tissues. Samples were mixed by vortex and pulse spun in a centrifuge to collect any condensation. 50µl 1M Tris-HCl was added and samples again mixed by vortex for 15 seconds. Samples were spun in a centrifuge at 13,000rpm for 6 minutes and the supernatant collected for genotyping. Samples were stored at -20°C until required.

2.1.3 Extraction of Genomic DNA from embryonic tissue
Genomic DNA extraction was carried out on embryonic samples from E10 and E13 pups, and from those animals for which genotype had to be ascertained. Extractions were carried out using the DNeasy blood and tissue kit (Qiagen) as per the manufacturer’s protocol. Tissue (either a leg for embryonic samples or a tail clipping for postnatal samples) was cut into small pieces and placed in a 1.5ml eppendorff with 180µl buffer
ATL and 20µl proteinase K. The sample was mixed by vortexing and incubated at 56°C for 4 hours to ensure complete lysis of the tissues. Samples were vortexed and 200µl buffer AL added. 200µl of 100% ethanol was added immediately and the sample mixed by vortexing. The mixture was transferred to a DNeasy mini column in a 2ml collection tube and centrifuged at 8000rpm for 1 minute. Flow through was discarded and 500µl of wash buffer AW1 added to the column. This was then centrifuged for 1 minute at 8000rpm and flow through was discarded and 500µl of wash buffer AW2 added to the column. This was centrifuged for 3 minutes at 12,000rpm to dry the spin column and flow through and collection tube discarded. The spin column was placed in a new 1.5ml eppendorff and 200µl buffer AE added directly to the membrane. The column was incubated for 1 minute at room temperature and then centrifuged for 1 minute at 10,000rpm to elute the DNA. DNA was stored at -20°C until required.

2.1.4 Genotyping by sequencing reaction

2.1.4.1 Polymerase Chain Reaction (PCR)

A master mix was prepared on ice consisting of 1X Sigma PCR buffer (SIGMA), 0.1mM of each forward and reverse primer (Invitrogen), 0.6µM dNTPs and 1U Sigma Taq DNA polymerase (SIGMA). Typically 0.5µg gDNA was used for a 20µl reaction volume and the master mix aliquoted into the tubes. PCR reactions were run on the Peltier Thermal Cycler-225 (MJ Research) and heated lids used to minimise evaporation. The initial PCR was run under cycle conditions; 94°C 10mins, 94°C 1min, 55°C 1min, 72°C 1min (return to stage 2 and repeat 30times), 72°C 5mins, hold at 4°C. Samples were run on a 1.5% agarose gel to check for product.

2.1.4.2 Exosapit Clean-up

2µl product from the PCR was added to 2µl H2O and 1µl Exosapit for a 5µl reaction mix. For each sample there were two reactions, to allow the use of forward and reverse primers in the next stage. The plate was run on the Peltier Thermal Cycler-225 (MJ Research) at 37°C 1hour, 80°C 20mins and held at 4°C.
2.1.4.3 Sequencing reaction

A master mix was prepared consisting of 0.5µl BD v.3.1, 1.75µl Sequencing buffer, 1µl 3.2pmol Primer, 2.75µl H2O to a total volume of 6µl per sample. This was added to each sample in the plate used for the previous reactions and run on the Peltier Thermal Cycler-225 (MJ Research) under cycle conditions; 96ºC 1min, 96ºC 10secs, 50ºC 4secs, 60ºC 4mins (return to stage 2 and repeat 30times), hold at 4ºC. After PCR the samples were stored at -20ºC for ethanol/EDTA precipitation the following day.

2.1.4.4 Ethanol/EDTA Precipitation

Sample plate was allowed to equilibrate to room temperature and then pulse spun in a centrifuge to collect any condensation. 2.5µl 125mM EDTA was added to each well along with 30µl 100% EtOH and hybrid sealing mat replaced. Reactions were mixed by inversion 4times and incubated at room temperature for 15mins. Plates were spun in a Jouan centrifuge at 3000rpm for 30mins and seal removed before plate was inverted over a paper towel to remove most of the EtOH. The paper towel was placed in the bottom of a centrifuge bucket and plate spun inverted up to 1000rpm then stopped. 40µl freshly diluted 70% EtOH was added to each well and the seal replaced before mixing by inversion 4times. The plate was spun in Jouan centrifuge at 3000rpm for 15mins and seal removed before plate was inverted over a paper towel to remove most of the EtOH. The paper towel was placed in the bottom of a centrifuge bucket and plate spun inverted for 30secs then stopped. Wells were air dried for 15mins protected from light and the plate sealed with adhesive film before being sent to the Medical Research Council sequencing service to be sequenced by Agnes Gall.

2.1.5 Genotyping using the Transnetyx sequencing service

Ear notches were collected from mice and placed in the 96-well plate provided by Transnetyx services (www.transnetyx.com). Positive and Negative control samples from mice with known genotypes were included in the first run. Samples were sent by courier to Transnetyx to be analyzed using primers designed from sequence submitted by myself. Results were returned within 3 days.
2.1.6 Genotyping Q31L het x hom pups

Genotyping of Q31L pups from het x hom matings was carried out by Edward Weiss at the Toronto Sick Kids Hospital Research laboratories.

2.1.7 Dissection of embryos and pups

Pregnant females were sacrificed by cervical dislocation and observed until all movement had ceased. An incision was made into the females abdominal cavity and the skin pulled back to reveal the embryonic sacs. Embryos were removed and placed in chilled PBS on ice. Under a standard dissecting microscope the heads were removed and the top of the skull removed to expose the brain. The brain was extracted whole and placed in prechilled RNAlater (Ambion) at 4°C to stabilize the RNA for further analysis, and kept at 4°C for up to 2 weeks before required. For post-natal samples the heads were also removed and brains extracted in the same manner as the embryonic stages.

2.1.8 Dissection of adult brain samples

Adults were decapitated post mortem and the skin cut from the back of the head to the nose down the midline. The tissue was then peeled back to reveal the top of the skull. This was removed by making an incision around the circumference of the skull starting at the foramen magnum and proceeding round the top of the skull at the level of the occipital openings. The top of the skull was lifted off the brain and the whole brain removed intact from the skull and placed dorsal side up on filter paper moistened with PBS. The cerebellum was removed by making a coronal cut perpendicular to the paper just behind the inferior colliculi, and placed in pre-chilled RNAlater (Ambion). A sagittal incision was made down the midline between the two hemispheres to one third from the anterior end. Using two paintbrushes the hemispheres were separated exposing the hippocampal groove. One brush was slid into the groove, separating the hippocampus from the cortex, and the hippocampus gently rolled out. Once separated from the rest of the brain the hippocampus was placed in prechilled RNAlater prior to further analysis. This was repeated for the other hemisphere also. The cortex was gently peeled away and
stored in RINAlater. In some cases only the hippocampus was removed in which case the cortex and remaining brain tissues were stored in the same eppendorf as the cerebellum. Where no further dissection was being carried out the brain was removed from the skull whole and placed directly in pre-chilled RINAlater.

2.2 Sample Preparation

2.2.1 Extraction of total RNA from tissue samples

Total RNA was isolated from selected tissue in accordance with the protocol supplied with the RNeasy mini kit for animal tissues (Qiagen). Tissues were homogenized in 600µl buffer RLT using a Precellys 24 tissue homogenizer (Bertin Technologies) at maximum speed for two 45 second cycles. Lysate was centrifuged for 3 minutes at 13,000rpm and the supernatant transferred to a fresh micro centrifuge tube. 1 volume of 70% ethanol was added to the lysate and mixed by pipetting and 700µl of this sample transferred to an RNeasy spin column. After centrifugation at 10,000rpm for 15seconds, flow through was retained for protein analysis and 700µl buffer RW1 added to the spin column. Columns were again centrifuged under the same conditions and flow through retained and combined with that from the previous step. The spin column membrane was washed by addition of 500µl buffer RPE and spun for 15seconds at 10,000rpm. Flow through was again retained and the step repeated with a 2 minute spin time to dry the membrane. RNA was eluted by the application of 40 µl sterile H₂O directly to the membrane and centrifuged for 1 minute at 10,000rpm. Pooled flow through was stored at -20°C until required for protein analysis. Concentration and quality of total RNA samples (RIN) was determined using a spectrum analyzer (Agilent Bioanalyser ; see section 2.2.5) and samples stored at -20°C until required for cDNA synthesis.

The same method was followed for RNA extraction of the samples collected at the Samuel Lunenfeld Research Institute however due to lack of access to the Precellys 24 tissue homogenizer, all tissues were homogenized using a standard mortar and pestle. These samples also underwent on column DNA digest (see section 2.2.3).
2.2.2 PARIS procedure

The PARIS procedure from Qiagen allows simultaneous extraction of RNA and protein from the same sample tissue. This enables us to use the same animal and sample section for subsequent protein analysis after the initial RNA/DNA analysis is complete. The PARIS procedure was used on animals for the second round qRT-PCR Taqman analysis.

Lysate is prepared by placing tissue in a homogenizer tube with 600µl ice-cold Cell disruption buffer (provided with kit) and homogenizing in a rotor homogenizer for 30secs. The lysate was split into two eppendorf tubes of equal volume. One tube was stored at -80°C for later protein analysis while 300µl 2X lysis/binding solution was added to the other tube and mixed by pipetting. One volume 100% EtOH was added and the sample was mixed by pipetting to ensure the EtOH reached the bottom of the tube. The sample mix was transferred into a filter cartridge and spun at 10,000rpm for 1min in a centrifuge. Flow-through was discarded and 350µl was solution 1 added to the cartridge. The sample was spun at 10,000rpm for 30secs and flow-through again discarded. 500µl wash solution 2/3 was added to the cartridge and the sample was spun at 10,000 rpm for 1min. Flow-through was discarded and this step repeated a second time. The empty cartridge was spun at 10,000rpm for 15secs to remove any residual wash solution and transferred to a fresh collection tube. 40µl pre-heated (to 95ºC) elution solution was added to the cartridge membrane and the sample was spun at 10,000rpm for 30secs. This was repeated so a total of 80µl RNA was collected. Any genomic contamination was removed using the DNA-free kit (Ambion) outlined in section 2.2.3.

2.2.3 Post extraction DNase Treatment of RNA

RNA samples were treated with the DNA-free kit (Ambion) prior to cDNA synthesis to remove any traces of genomic contaminant. Treatment was carried out as per the protocol supplied with the kit. 0.1 volumes of 10x DNase1 buffer and 1µl rDNase 1 was added to the samples and mixed gently before incubation in a 37ºC water bath for 30minutes. 0.1 volume of DNase inactivation reagent was added to the samples and incubated at room temperature for 2 minutes ensuring the samples were continually mixed over the incubation period to re-disperse the inactivation reagent. After centrifugation at
13,000rpm for 2 minutes the supernatant containing the RNA was transferred to a fresh tube to be used for cDNA synthesis.

2.2.4 On column DNase Treatment of RNA

On column DNA digest was carried after the preliminary wash stage of RNA extraction using the RNase-Free DNase set (Qiagen) by adding 70µl Buffer RDD to 10µl DNase 1 and applying this mix directly to the spin column membrane. The column was then incubated at room temperature for 15-20 minutes to allow complete DNA digest to occur. The membrane was washed again with 350µl buffer RW1 and RNA extraction continued.

2.2.5 Assessing RNA quality using the Agilent Bioanalyser

The Agilent 2100 Bioanalyser uses electrophoresis to separate fragments of the RNA sample by size. Quantification of each fragment allows quality control to be carried out at the same time as checking quantity. Samples collected at the Samuel Lunenfeld Research Institute in Toronto were analyzed by me on my return to Edinburgh. All other samples, including those collected in-house at the Molecular Medicine Centre and those which required vacuum drying prior to amplification and labeling for the microarray study, were analyzed by Alison Condie at the Welcome Trust Clinical Research Facility, Western General Hospital, Edinburgh.

The RNA 6000 Nano kit was placed on the bench for 30mins prior to loading the chip to allow the dye concentrate to equilibrate to room temperature. 550µl of RNA 6000Gel matrix was pipetted into a spin column and centrifuged for 10 minutes at 1500g. Gel was split into 65µl aliquots to be used within 4 weeks. Once at room temperature the RNA 6000Nano dye concentrate was placed on a 10second pulse spin and 1µl added to a 65µl aliquot of filtered gel. The solution was vortexed for 15seconds to mix and centrifuged for 10mins at 13000g. RNA samples were denatured for 2minutes at 72ºC and placed on ice ready for loading onto the chip.
The RNA 6000 Nano chip was placed in the chip priming station and 9µl gel-dye mix added to the well marked and the priming station closed. The plunger was pressed until held in the clip and released after 30 seconds. After 5 seconds the plunger was pulled back to the 1ml position and the station opened. 9µl of gel-dye mix was pipetted into the wells marked and the remaining gel mix discarded. The RNA 6000 Nano marker was loaded into all 12 sample wells and the ladder well. 1µl of the RNA 6000 Nano ladder was added to the appropriatted well marked and 1µl of sample into each of the 12 sample wells. The chip was placed horizontally in the IKA vortex and vortexed for 1 minute at 2400rpm. The chip was placed in the Agilent 2100 Bioanalyser and run within 5 minutes of vortexing. The programme selected quantifies total RNA in the sample.

After the bioanalyser had completed its programme the chip was discarded and the bioanalyser pins cleaned by immersing in 500µl RNA zap for 2mins then 500µl distilled H₂O for 2 mins before being air dried for 10 seconds. Total RNA results were analyzed using the Agilent 2100Expert software. Samples with a RNA Integrity Number (RIN) of <5 were deemed unsuitable for use in the microarray study due to problems with amplifying low quality RNA (Figure 2.1).
Of all samples collected (both in house at the MMC and at the Samuel Lunenfeld Research Institute) 96.91% met the RIN number criteria for progressing forward for the microarray. Of those groups chosen for the microarray study 95.10% had suitable RNA for amplification.

### 2.2.6 Vacuum Drying RNA samples

While the majority of samples (97%) collected for the microarray study contained enough total RNA to go forward for amplification and labeling, a few had very low concentrations and would require drying before amplification. This was achieved using a vacuum centrifuge.

The centrifuge was cleaned with RNAzap to prevent cross contamination and was set to heat to 60°C for 10 minutes prior to loading the samples. Samples were loaded in 1.5ml eppendorffs with the lids removed and the vacuum pump and centrifuge set running. Progress was checked every 5 minutes to ensure the samples were not over-dried and the
running time was adjusted according to the volume left in the tubes at any time throughout the process. All samples were dried from 45µl down to around 15µl over a period of around 35 minutes. An aliquot of the concentrated sample was stored for quantification on the Agilent Bioanalyser and the remaining sample stored at -70ºC until required for amplification. Some samples returned poor RIN numbers post drying and so could not be used for the microarray study. The other samples, with RIN’s above 5, were deemed acceptable for amplification and would be checked after this to assess suitability.

**2.2.7 Synthesis of cDNA**

Single-stranded cDNA was synthesized from DNase treated total RNA using the core cDNA synthesis kit (Roche). 2µg total RNA was diluted in sterile H2O to a total volume of 16.4µl and denatured at 65°C for 15 minutes before being placed on ice for 5 minutes. This was then incubated at 47°C for 60 minutes in the presence of 10x buffer, 25mM MgCl2, 5mM dNTPs, OligoT, RNase inhibitor and AMV reverse transcriptase. Controls underwent the same conditions in the absence of AMV reverse transcriptase or RNA. The reaction was terminated by heat inactivation at 99°C for 5 minutes. Single stranded cDNA was stored at -20°C until required for PCR amplification.

**2.2.8 Synthesis of cDNA from low quantity RNA**

Synthesis of single stranded cDNA from low quantity RNA was carried out using the Transcriptor cDNA synthesis kit from Roche. 2µg total RNA was diluted in sterile H2O to a total volume of 22µl and added to 4µl random primers. The mix was denatured at 65°C for 15 minutes before being placed on ice for 5 minutes. This was then run of a cycle of 25°C for 10 minutes, 55°C for 30 minutes and 85°C for 10 minutes in the presence of 5x buffer, 5mM dNTPs, RNase inhibitor and AMV reverse transcriptase. After completion of the cycle the samples were held on ice for 5 minutes and transferred to -20°C until required for further experiments.
2.2.9 Determining cDNA concentration using the Quant-iT™ PicoGreen® dsDNA kit

Concentration of cDNA preparations was determined using the Quant-iT™ PicoGreen® dsDNA kit (Invitrogen). The kit contains the Quant-iT™ PicoGreen® dsDNA reagent which stains double stranded DNA in solution and fluoresces when excited at 480nm in a Cytofluor Fluorimeter (at the Cancer Research Facility, Western General Hospital, Edinburgh). Intensity of fluorescence emission of samples and standards was measured at 520nm.

The 100µg/ml lambda DNA standard supplied with the kit was diluted 50 fold in TE to give a 2µg/ml DNA stock for preparing standards. Standards were prepared at 0, 1, 10, 100 and 1000ng/ml DNA and DNA samples to be measured were diluted in duplicate in TE (5µl sample + 95µl TE) in a DynaTech 96 well plate. As the predicted yield of the cDNA reaction is 30% all samples should fall within the range of the standard curve.

The PicoGreen reagent was diluted 200 fold in TE and covered in foil to protect from light. The reagent must be used within a few hours to be most reliable so only enough working solution for one run was prepared allowing 100µl for each sample and standard.

The fluorimeter was set to excite at 480nm and measure emission at 520nm with a gain of 45 so the highest DNA standard would be close to the fluorimeter’s maximum. The Picogreen working solution was added to the samples and readings taken immediately. Emission readings of the DNA standards were plotted and concentration of cDNA samples calculated using the regression equation of the standard curve.
**2.2.10 Determining cDNA or RNA quantity on the NanoDrop**

The Thermo Scientific NanoDrop 1000 spectrophotometer is a highly accurate full spectrum spectrophotometer for measuring concentration and quality of nucleic acids and proteins. As little as 1µl of sample can be used to determine concentrations, reducing the amount of sample required compared to traditional methods. Sample concentration is determined using the Beer-Lambert equation ($A = e \cdot b \cdot c$) where $A$ is the sample absorbance measurement (-log[sample intensity/blank intensity]), $e$ is the wavelength-dependent extinction coefficient (liter/mol-cm), $b$ is the path length in centimeters, and $c$ is the analyte molarity (M). On start up the nucleic acid option was chosen and for cDNA the ‘ssDNA33’ programme was selected, and for RNA the ‘RNA-40’ programme was selected. Selection of the correct programme is essential for accurate calculation of concentrations.

The machine was blanked using 1µl dH2O and this is stored as reference for calculating the subsequent samples. Each sample was tested on the NanoDrop ensuring the pedestal was cleaned after every sample. Concentrations were saved in the report and absorbance ratios and curves checked for quality. The 260/280 ratio gives a comparative value of
absorbance at 260 and 280nm and is used to assess quality of RNA and DNA. A 260/280 ratio of 1.8 is accepted as ‘pure’ for DNA and a ratio of 2.0 ‘pure’ for RNA. Deviations from these values would suggest contamination of the sample. Any samples run which were not sufficiently close to these ratios were discarded from further study.

2.3 PCR
Polymerase chain reactions (PCR) allow the rapid amplification of a selected region of DNA sequence from a complex mixture. Specificity is provided by oligonucleotide primers designed to be complimentary to the 5´ region of sequence to be amplified. The PCR reaction is a three stage process. First the DNA is denatured at a high temperature, and then the temperature is reduced to allowing annealing of the oligonucleotides to the specific DNA sequence. Finally the temperature is adjusted again to allow elongation and amplification by a thermo-stable DNA polymerase. Multiple cycling of these stages results in exponential amplification of the desired sequence.

2.3.1 PCR Reagents
As only RNA expressed at certain time points was of interest, reverse transcriptase PCR (RT-PCR) was used. A master mix was prepared on ice consisting of 1X Sigma PCR buffer (SIGMA), 0.1mM of each forward and reverse primer (Invitrogen), 0.6µM dNTPs and 1U Sigma Taq DNA polymerase (SIGMA). Typically 0.1µg cDNA was used for a 20µl reaction volume and the master mix aliquoted into the tubes. PCR reactions were run on the Peltier Thermal Cycler-225 (MJ Research) and heated lids used to minimise evaporation.

2.3.2 Design and Synthesis of Oligonucleotide Primer pairs
Oligonucleotides were designed using the Primer3 software (primer3 www.cgi v 0.2) with in-house default features. All oligonucleotides were 18-25bp long with an average
50% GC content and a melting temperature \( (T_m) \) of 60-65°C. No primers chosen were predicted to dimerise or hairpin when used as a pair. All primer pairs were commercially synthesised by Invitrogen at 25nM scale with standard purification.

### 2.3.3 RT-PCR Cycling

The PCR consisted of an initial denaturing step of 94°C for 2 mins followed by 30 cycles of denaturation, annealing and extension. The denaturation step lasted 30 seconds at 94°C. The annealing temperature was determined by the predicted Tm of the oligonucleotide primers. Primers were typically designed with a Tm of 64°C so the annealing step was carried out 5°C below this at 59°C for 30 seconds. The extension step was carried out at 72°C, the optimum temperature for Taq DNA polymerase, and the length of stage determined by final predicted product size. Typically 1 minute per 1Kb of sequence was used. After the final cycle a further elongation step of 94°C for 5 minutes was used to ensure fragments were complete.

The ‘Touch-down’ PCR method was used as standard as it increases specificity of the PCR reaction. The initial annealing temperature used is 10 °C above the final annealing temperature required, and drops by 1°C with every cycle until after ten cycles the final annealing temperature is reached. The normal 30 cycle reaction at the desired annealing temperature can than proceed. This method results in highly specific primer annealing in the first few cycles selecting for the correct product in the subsequent cycling sets.

### 2.3.4 Visualisation of PCR Products

PCR products were separated by electrophoreses on agarose gel of between 1 and 2% (w/v) depending on the size of the expected bands (gel used is noted on figure legends). Multi-purpose agarose was boiled in 0.5x TBE buffer (45 mM Tris base, 45 mM boric acid, 1 mM EDTA pH 8.0) and SYBRsafe (Invitrogen) added 1:50 (v/v). Gel was left to set in the dark to prevent any bleaching of the SYBR safe. 2µl loading dye was added to the PCR product prior to loading on the gel and 400ng of 1Kb ladder (Invitrogen) run.
alongside the samples. Products were visualized on a UV transilluminator and photographed with a digital camera.

2.4 Protein Preparations
2.4.1 Determining Protein Sample Concentrations
Concentrations of protein samples obtained during the PARIS procedure were determined using the BCA Protein Assay Kit (Pierce Biotechnologies) test tube procedure as per the manufacturers’ instructions. This relies on the reduction of copper ions by protein in an alkaline medium which is detected by kit reagent A. Protein standards of known concentrations from 0-2000µg were initially prepared by dilution of 2mg/ml stock Albumin. BCA working reagent was prepared by mixing 50 parts BCA reagent A with 1 part BCA reagent B. Total volume required was calculated as:

\[(\text{#standards + #unknowns}) \times (\text{#replicates}) \times 2 = \text{Total volume working reagent required}\]

0.1ml of each standard and unknown sample replicate was pipetted into an appropriately labeled tube and 2ml working reagent added to each. Tubes were incubated at 37°C in a waterbath for 30 minutes and then allowed to cool to room temperature. As the BCA reaction is not an end-point colour change, the colour will continue to develop after cooling to room temperature. However, the rate of colour change is so low at room temperature that no significant error will be recorded if all samples are measured within 10 minutes of each other. Sample absorbance was measured at 562nm on a spectrophotometer. The absorbance reading for the blank sample was subtracted from subsequent standard and unknown sample measurements. These blank correlated BSA standards were plotted on a standard curve against their known concentrations and this was used to determine the protein concentrations of the unknown samples.
2.5 Pre-Pulse Inhibition

Pre-pulse inhibition is a common method used to quantify information processing deficits in schizophrenia with reasonable validity. It can be used with both animal and human experiments. PPI measures the degree to which the acoustic startle response is reduced when a startle eliciting stimulus is preceded by a brief low-intensity stimulus which does not elicit a startle response.

2.5.1 Calibration

Animals from 11.5-14 weeks of age were weighed and the chamber platform of the prepulse box was calibrated to take into account the weight of the animal.

Figure 2.3: Image of a PrePulse inhibition chamber. Platform, cradle, amplifier and speaker are labeled.
2.5.2 Drug Administration

Drugs were administered 30 minutes before the commencement of behavioural testing by intra-peritoneal injection in a volume of 10ml/kg by Tatiana Lipina at the Samuel Lunenfeld Research Institute. Clozapine (3mg/kg; Tocris) was dissolved in saline (0.9% NaCl) containing 0.3% Tween (Biorad). Bupropion (4mg/kg; Sigma) was dissolved in distilled water and Rolipram (0.5mg/kg; Sigma) in saline containing 10% DMSO. Saline was used as a control for drug naïve animals. All drug doses were selected from previous studies carried out at this lab (Clapcote et al 2007 [197]).

Each genotype was split into groups for drug treatment. For the L100P mutant line and the C57BL6 line there were 12 mice in each drug treatment group of equal numbers male and female, including a control group administered with saline. For the Q31L mutant line 8 mice were used for each drug treatment with equal numbers of male and females. Rolipram was not administered to the Q31L line as it had been deemed unsuccessful at behavioural rescue in previous studies [197]. In all cases animals were sacrificed 1 hour after drug administration.

2.5.3 Behavioural Testing

Animals were placed in a pre-pulse chamber 15 minutes after administration of drug treatment. The pre-pulse program begins with a 15 minute habituation period so actual testing did not begin until 30 minutes post drug administration. Animals are then played a series of five ‘startles’, and movement is detected by the platform at the base of the chamber. This is followed by 50 trials of startle, startle preceded by pre-pulse or no stimulus. The final stage is a further series of 5 pure startle stimuli.

Animals were sacrificed immediately at the end of the pre-pulse testing by cervical dislocation, and whole brain was extracted and stored in RNAlater before further dissection.
2.5.4 Analysis of Behavioural Testing Results

Analysis of the amplitude of acoustic startle response in the Pre-pulse inhibition trials was carried out in Microsoft Excel and Statistica for Windows analysis packages. Repeated measures analysis of variance was used to determine any significant differences between groups with appropriate post-hoc tests applied and a standard p-value cut-off of 0.05 applied.

2.6 Microarray Analysis of ENU mutant gene expression

This section describes the methods used to carry out the microarray study from amplification onward. Preparation and extraction of RNA was as described in section 2.2.

2.6.1 cRNA amplification and Biotinylation

RNA samples were gathered as previously described in sections 2.2.1-2.2.6. Biotinylated, amplified cRNA was generated using the Illumina® TotalPrep RNA amplification kit as per the manufacturers instructions.

100ng total RNA was used as starting material and master mixes were prepared as recommended using the Illumina® TotalPrep™ RNA Amplification Master Mix Calculator available online (www.ambion.com/techlab/mm_calcs/illumina_rna_totalprep_amp_calc.php). The two 2-hour incubation cycles (II.C and II.D) were performed in a Peltier Thermal Cycler and hybridisation (step II.F) was carried out with overnight incubation for 14 hours at 37°C in a hybridisation chamber. The eppendorf’s were wrapped in parafilm for this step to prevent loss of material through condensation. 1.5µl of the final elute, containing amplified, biotinylated cRNA, was set aside for quantification and quality control and the remainder stored at -70°C until required.
2.6.2 Quantification and Quality control of cRNA

Samples were loaded onto the Agilent RNA nano chip as described in section 2.2.5. The chip was loaded into the bioanalyser and run under the ‘RNA nano-cRNA’ programme, designed to measure the quantity of cRNA in the sample. The read-out obtained from the bioanalyser gives a measure of concentration and contamination (as a percentage). Samples with high levels of contamination would not be amplified.

2.6.3 Preparation of Pooled cRNA

Due to the number of animals used for the study it was decided samples would be pooled before being hybridised to the array. Each pool contained 3 animals of the same sex from the same mouse line. In most cases there were 2 pools for each group and sex (table 2.1). After determining the concentration of each individual sample using the Agilent 2100 bioanalyser, equal amounts of each sample were combined to a total concentration of 150ng/µl.

<table>
<thead>
<tr>
<th>Array</th>
<th>Beadchip</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>C57BL6 Sal F-1</td>
</tr>
<tr>
<td>C</td>
<td>L100P Cloz F-2</td>
</tr>
<tr>
<td>D</td>
<td>L100P Rol M-2</td>
</tr>
<tr>
<td>E</td>
<td>Q31L E13 M-1</td>
</tr>
<tr>
<td>F</td>
<td>Q31L E13 M-2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Key</th>
</tr>
</thead>
<tbody>
<tr>
<td>C57BL6</td>
</tr>
<tr>
<td>L100P</td>
</tr>
<tr>
<td>Q31L</td>
</tr>
<tr>
<td>Sal</td>
</tr>
<tr>
<td>Cloz</td>
</tr>
<tr>
<td>Rol</td>
</tr>
<tr>
<td>Bugro</td>
</tr>
<tr>
<td>E13</td>
</tr>
<tr>
<td>F#</td>
</tr>
<tr>
<td>M#</td>
</tr>
</tbody>
</table>

Table 2.1: BeadChip layout with key. Numbers represent each individual beadchip and letters represent the position on the chip.
2.6.4 Hybridisation and scanning of Illumina BeadChips at the Wellcome Trust Clinical Research Facility

The Illumina MouseWG-6 v2.0 expression beadchip was used to target 45281 genes from the RefSeq database release 22 and RIKEN FANTOM across six pooled samples per chip (http://www.illumina.com/downloads/GX_Mousev2.0_DataSheet.pdf). Each chip carried one identical sample to be used as a between chip control, and one sample set was amplified, pooled and hybridised 3 times to control for batch differences.

10 microlitres of each 150ng/µl sample was submitted to Alison Condie at the Genetics Core of the WTCRF where the subsequent hybridisation, washing, blocking and streptavidin-Cy3 staining was performed in line with manufacturers guidelines. The Illumina BeadArray™ Reader was used to scan the BeadChips and Illumina BeadScan software performed specific image processing steps to determine bead intensities from raw image data (as described in Kuhn et al 2004 [215]). Illumina BeadStudio 2.0 Gene Expression software was used to read the image data and generate signal data for subsequent analysis. Probe information including raw signal intensity, number of beads, standard deviation of raw bead intensity and detection \( P \)-value were exported to text file and returned to myself from the WTCRF. Two sets of data were initially returned, one with no normalisation and no grouping, and the other with no normalisation but with samples grouped according to line and treatment. This second sample set returned group detection \( P \)-values which would be used for data filtered later in analysis.

Unfortunately during the hybridisation stage one chip (chip 5) showed signs of drying out while in the chamber which resulted in some skewed intensity values of controls. It was decided that all samples on this chip should be re-run before further analysis. The chip 5 repeat gave control intensities comparable with previous runs and intensities from this were taken forward for further analysis. Results from both chip 5 runs are shown in Chapter 3.
2.6.5 Data Analysis
General descriptive statistics and many quality control methods were carried out using Microsoft® Office Excel with various add-ins, and Statistica for Windows. Venn diagrams were created using a combination of Microsoft® Office Excel and Microsoft Office Powerpoint. Most downstream analyses were performed using the free software environment R (www.r-project.org) and the R-based open source software for bioinformatics, Bioconductor (www.bioconductor.org).

2.6.6 Microarray Quality Control Measures
This section describes the methods used to analyse the inbuilt controls, within the Illumina BeadArray Platform, technical replicates and between chip controls outwith the standard chip controls. Cluster analysis has also been performed to assess the success for the experiment and quality of the data generated.

2.6.6.1 Illumina direct hybridisation assay controls
The Illumina BeadArray™ platform has six control categories inbuilt to assess the quality and reproducibility of sample intensities obtained as follows: 1) housekeeping controls assess the intactness of the biological sample; 2) negative controls provide an estimate of background noise; 3) hybridisation controls determine success of hybridisation; 4) biotin controls assess the success of biotin labelling and signal generation; 5) high stringency and 6) low stringency hybridisation controls assess the stringency of hybridisation. Controls 1, 2 and 4 are dependant on the quality of the sample, while the others are not. It was at this stage the initial failing of the first chip 5 run was noticed and samples re-submitted.

2.6.6.2 Technical Replicates and Between Chip Controls
One sample (C57BL6Sal-F2) made up of 3 drug naive female C57BL6 mouse samples was loaded onto every chip analysed to assess BeadChip-BeadChip reproducibility. Another sample, prepared from the same total RNA but different cRNA preparations (C57BL6Sal-F1) was included in triplicate on one BeadChip as a measure of
experimental within chip variability. The reproducibility was assessed using the Pearsons correlation coefficient ($r^2$).

### 2.6.6.3 Cluster analysis for detection of experimental artifacts

Hierarchical cluster analysis was performed to determine whether samples clustered together through experimental variables such as batch effect or BeadChip. Various methods exist differing in measures of distance between pairs of samples and clusters. As this type of analysis had already been carried out by Andrea Christoforou within our group, the clustering method she used was followed as this had previously been shown to be suitable for similar sample sets. The R package `pvclust` ([www.is.titech.ac.jp/~shimo/prog/pvclust/](http://www.is.titech.ac.jp/~shimo/prog/pvclust/)) was used to create a final dendogram and to assess the results via multisample bootstrap resampling. One thousand bootstraps were performed on each of three bootstrap sample sizes ($r=0.5,1.0,1.5$ where $r$ is the size of the bootstrap sample relative to the actual dataset).

### 2.6.6.4 Power Calculations

Power calculations were performed using the `power.multi` function for multiple treatment designs from the Bioconductor `sizepower` package. No software package available at this time was able to produce power calculations for multiway ANOVA analysis so the calculations were carried out on the basis of 33 student t-tests on the smallest group size, giving a very conservative estimate of the experimental power. Curves were plotted for a range of standard deviations and effect sizes.

### 2.6.6.5 Pre-processing

Pre-processing is the steps taken to prepare data for analysis, and involved 3 stages: 1) variance stabilisation and transformation; 2) normalisation and 3) data filtering. The Bioconductor `lumi` package ([www.bioconductor.org/packages/2.1/bioc/vignettes/lumi/inst/doc/lumi.pdf](http://www.bioconductor.org/packages/2.1/bioc/vignettes/lumi/inst/doc/lumi.pdf)) which was created specifically for use with the Illumina BeadArray™ platform, was used for pre-processing. The first stage, variance stabilisation, aims to reduce the heteroskedacity (that is, make the variance more constant) across the sample set. This was carried out using the
The final stage of pre-processing was to filter the data. Data was filtered in two ways: First, any probes not expressed above background in at least one sample group were removed from further analysis. This was determined using the Illumina BeadStudio detection $P$-value, which is calculated for every probe in every sample using a non-parametric method that ranks the probe against that of a negative control. It is interpreted as the probability the observed signal is no greater than the background given the amount of noise in the data. Secondly, the probes were filtered based on fold change. Pairwise comparisons were made between samples grouped by strain and treatment and only probes with at least a 1.3 fold change (either above or below that of the reference group) were kept. The fold change cut off was chosen as this is the level of sensitivity offered by the Illumina platform (www.illumina.com/downloads/GX_Mousev2.0_DataSheet.pdf). Fold change was calculated by taking the inverse/anti-log$_2$ of the difference between the average normalised values. Cluster analysis was carried out after both filtering stages to check the effect on data distribution.
2.7 Testing For Differential Expression

This section describes the methods used to determine genes with significant differential expression in mutant strains and genes of interest for follow-up studies.

2.7.1 Differential expression analysis

The R package MicroArray Analysis of Variance model (MAANOVA) was used to test the probes for differential expression. The $F_t$ statistic, which makes no assumptions about the distribution of variance but borrows information across the probes tested to determine the variance, was used as the test statistic, with a nominal $P$-value threshold of 0.05. The false discovery rate (FDR) for each probe was also determined with the MAANOVA \textit{adjPval} function. Probes with a fold change of 1.3 (+/-) $p<0.05$ compared to the appropriate reference set were classed as being differentially expressed.

2.7.2 Functional Analysis (Including IPA Pathway analysis)

GO terms for each gene were collected from Mouse Genome Informatics (http://www.informatics.jax.org/) and Ensembl Genome Browser (http://www.ensembl.org/index.html) where available.

Pathway analysis was carried out using the Ingenuity Systems pathway analysis programme (http://www.ingenuity.com/) held on licence at the University of Edinburgh. All probes expressed above background for each strain were initially loaded into the system regardless of fold change or $p$-value. These two factors were then noted as observations and appropriate cut-offs (as mentioned in section 2.8.1) set before running the analysis. This resulted in the previously determined differentially expressed genes being highlighted and those expressed above background shaded grey indicating their presence but not significance. Any genes not expressed above background that appeared within the networks were white.

Ingenuity® works by accessing a database of gene interactions and comparing the input gene list with what is known by the database. A significance score is assigned for each
network and displayed as the negative log of the networks p-value. This indicates the likelihood that the assembly of focus genes within a network could be explained by random chance. A score of 3 indicates a 1 in 1000 chance that the focus genes are together by random chance so scores greater than this have at least 99.9% confidence that the genes are not grouped by random chance alone. Pathway analysis was carried out for both adult mouse mutants, and embryonic stages.

2.7.3 GOTree Analysis of differentially expressed genes

GOTreeMachine (GOTM) (http://bioinfo.vanderbilt.edu/gotm) is an online open resource for identifying clusters of genes by ontology. By comparing a list of genes of interest with a reference gene list it is possible to identify statistically enriched gene ontology categories by hypergeometric testing. This gives a list of genes which may have functional significance within the dataset and as a result within the disease group. I entered the complete list of differentially expressed genes from the microarray study as one data set and split by genotype and drug treatment to test if there was an overall over enrichment of specific gene sets within the list as a whole or within one group in particular.

2.7.4 Testing for overlaps and gene expression corrections from drug treatments

The resulting differentially expressed genes for each strain were compared using the countif function in Microsoft Office Excel. The adults were compared to each other and to their drug treated counterparts, and the embryonic stages were compared to each other resulting in 5 comparisons in total. Any overlaps in expression changes were noted as these may signify genes of interest for follow-up studies.
2.7.5 Comparing data with previously published work

The differential expression data was compared with previously published works on schizophrenia genetics to determine if any overlaps were present within our data. Gene lists from Walsh et al 2008 [83], Camargo et al 2007 [132] and Kirov et al 2008 [217] which have been implicated in schizophrenia were compared to my own lists generated by the differential expression analysis. Again any genes of overlap were noted for possible follow-up.

2.8 Follow up analysis of the microarray study

On completion of the microarray study it was necessary to validate the gene list using real time PCR. This was carried out over two stages to give a set of robust genes which would then be tested further using protein analysis.

2.8.1 Identifying the genes to take forward for validation

As previously mentioned the genes which were shown to be differentially expressed by the statistical analysis of the microarray data were then subject to functional analysis through IPA and GOTree, and were tested for overlap with previous key genetic studies of major mental illness. Literary searches were also completed to identify previous publications concerning these genes and whether they were significant to the study of major mental illness. Each gene was given a score based on the evidence available from the above analyses and its fold change and p value as noted from the ANOVA of the microarray. For each group the top 10-15% of genes after scoring were chosen for follow up by real time PCR.
2.8.2 Preparation of Samples for two phase follow up

As the follow up study was two phase there were two sets of samples to be prepared. The first phase involved the samples which had been used on the array. RNA from these mice was already stored at -80°C and cDNA synthesis carried out as outlined in section 2.2.6.

The second phase was to be carried out on an independent sample set of mice bred at the Biomedical Research Facility at the Western General Hospital, Edinburgh. These second phase mice were descendents of animals that had been shipped over from the Samuel Lunenfield Institute after the collection of the first phase samples. This second phase collection allowed me to use wild-type littermates, which had not been possible for the first round, and determine how useful the C57BL6/J was as a wild-type control in the initial study. Samples were processed in the same way as the first phase samples with the exception of the RNA extraction method. The PARIS procedure (section 2.2.2) was carried out on phase two mice to allow later protein analysis to be carried out on samples from the same mice used for the DNA study.

2.8.3 Identification of suitable probes for Taqman real time PCR

Real time PCR was to be carried out using the ABI taqman bioanalyser at the Wellcome Trust Clinical Research Facility. Once genes for follow up had been identified probes were ordered from Applied Biosystems. When choosing probes it was important to take into account the possible isoforms of any one gene and map the Taqman probe as closely to the original array probe as possible. The microarray probe sequences were provided with the output data from the CRF and I used these to map to the area of the gene targeted by the microarray. I then compared this to the area targeted by the various Taqman probes and picked the probe with the best overlap to ensure as much as possible the same isoforms of the gene were being detected. It was impossible to have direct overlap of both probes as Applied biosystems do not make the probe sequences available but identify gene regions targeted. I also tried to have probes with a _m1 product code.
ending as this signified a probe which spans an exon junction and so would be more reliable with cDNA and cut out background from any possible genomic contamination.

Two control genes were required to allow for accurate comparisons to be made. The most commonly used control genes for mouse Taqman studies are \textit{GAP-DH} and \textit{Actin}. As \textit{DISC1} is believed to play a role in the Actin cytoskeleton this was considered inappropriate as a control gene. \textit{Hprt1} was chosen instead. This had previously been used successfully by the Cystic Fibrosis group within the centre and was known to be expressed in brain tissues.

2.8.4 Verification of genes by ABI Taqman qRT-PCR

Samples were analysed using the ABI Taqman machine at the Wellcome Trust Clinical Research Facility, Western General Hospital, Edinburgh. Samples were transported on ice and Taqman master mix was kept at 4°C within the CRF facility for my use.

The analysis was run on 396 well clear plates with two control genes and \(x\) genes of interest (\(x\) determined by how many samples were being run that day). A standard curve from pooled cDNA was run for each gene on each plate as an indicator of how well the reaction was working. A 1:200 calibrator was also run on each plate, allowing comparison between plates and within plates as this was used in the statistical analysis as a correctional value. Each well contained 5\(\mu\)l Taqman mastermix, 0.5\(\mu\)l appropriate probe and 4.5\(\mu\)l cDNA. This was sealed with a plastic sheet using a heat sealer and centrifuged for 30 seconds before being placed on the ABI Taqman platform. After completing the plate layout on screen the platform was set to run standard real time PCR cycle lasting 1.5 hours. On completion of the cycle the samples were removed from the platform and discarded, and the data analysed.
2.8.5 Statistical Analysis of ABI Taqman output

Samples were initially analysed in the ABI taqman screen to determine the efficiency of the reaction. For the replicated samples to be efficient the mean CT value for each replicate should be within 1 cycle. Data was scanned to determine if this was the case and any samples which did not fall within these criteria were omitted. Results were then exported as a text file to be opened in excel for further analysis.

In excel the exported data was summarised to quantity mean and standard deviation of each sample. The normalisation factor was calculated by taking the geometric mean of the two control genes per sample and dividing by the average of the geometric mean, giving a unique normalisation factor for each sample. This was also done for the standard deviation using the calculation:

\[
NF(sd) = NF^{*}\left(\frac{(control1 \ sd/2^*(control1 \ average)) \ W^{2} + (control2 \ sd/2^*(control2 \ average)}{W^{2}}\right)^{0.5}
\]

For each gene of interest this normalisation factor was used to create a normalised quantity mean for each sample by dividing the sample quantity mean by the normalisation factor. The normalised standard deviation was then calculated using the equation:

\[
\text{Norm sd} = \text{Normalised quantity mean}^{*(N(Fsd/\text{NF}) \ W^{2} + (\text{Gene of Interest sd/\text{Gene of Interest quantity mean}^{W})^{0.5}}
\]

This normalised quantity mean and standard deviation was then rescaled to correct for the 1:200 calibrator by dividing the normalised quantity mean of the sample by the normalised quantity mean of the calibrator, and similarly with the standard deviations. It was these ‘rescaled’ values that would be used for further analysis. Outliers were detected using the inter-quartile range detection method and discarded from further analysis.

The group average, standard deviation and number were calculated using excel basic statistics and these values were transferred into a GraphPad Prism file for further analysis (www.graphpad.com). GraphPad Prism is a statistical analysis and graphing programme
on licence within the University. ‘Raw’ rescaled values were also transferred and used to create a scatterplot and carry out statistical analysis. For each gene of interest comparisons were made between appropriate groups using both unpaired t-tests and Mann-Whitney U tests. Which test was most appropriate for which gene was determined by drawing normal distribution curves for each gene of interest. Genes with normal distribution in samples were analysed by t-test, and non-normal distribution by Mann-Whitney U test. However, in most cases both tests gave similar results so the need to choose between them was arbitrary. Also with such relatively small sample sizes it is almost impossible to gain an accurate representation of normality distribution and this was used as more of a guide than a definitive distribution curve. Samples which maintained a p-value <0.05 were carried forward for further analysis.

Analysis of the Disc1 developmental profile was carried out in the same way but graphed using rescaled averages with standard error rather than scatterplots. Unpaired t-tests and Mann-Whitney U tests were still carried out to determine the difference in Disc1 expression at key time points between strains.

2.9 Protein analysis and Antibody Staining

Once the microarray validation was complete it was considered logical to look at protein expression both from protein lysates and cultured primary neurons.

2.9.1 Identification of suitable antibodies

Antibodies were ordered from Abcam UK and R and D systems, UK. Suitability for use was determined by searching the supplier website for antibodies which were known to work in western blotting and immunocytochemistry (ICC), and that had been shown to work in mouse samples.
2.9.2 Western Blotting

Composition of all buffers can be found in section 2.10.

Sample lysate (20µg in 10µl) was added to 10µl protein sample buffer and 1µl DTT and boiled for 5 minutes to denature. Due to the small product sizes all westerns were run on 7% polyacrilamide gels. Samples and marker were added to the gel in the presence of running buffer and a positive control of SHSY5Y cell lysate which was used in the initial antibody work up. The block was topped up with running buffer and set at a constant voltage of 150V for 90 minutes. Sponges were soaked in transfer buffer along with PVDF membrane which had previously been dipped in 100% methanol. Once the acrilamide gel had run for 90 minutes it was removed from the block and the case cracked open to allow access to the gel. The PVDF membrane was applied to the gel and filter paper placed either side. A transfer casket was loaded with 2 soaked sponges, then the filter paper – gel – membrane – filter paper sandwich, and finally another 2 sponges. The casket was loaded into the block and filled with transfer buffer while the surround was filled with 650ml cold dH2O to prevent overheating during transfer. Blotter was set on constant 30V for 60 minutes. The membrane was removed and washed twice with dH2O and then in Ponceau stain to visualise protein. The membrane was washed again with dH2O and put on a shaking plate at low rpm until protein was visible before transferring to 40ml blocking buffer for 1hour on a shaking plate.

Primary antibodies were diluted in 4ml blocking buffer and applied to the membrane. Antibodies were diluted as per manufacturers instructions and a full list of conditions can be found in Appendix 1. Membranes were incubated at room temperature for 1 hour, or at 4°C overnight, on a low rpm shaking plate. Membranes were washed in wash buffer as previously described and the appropriate secondary antibody added to the membrane, diluted in 4ml wash buffer, for 25 minutes on a shaking plate. Membranes were washed in wash buffer as previously described and dabbed dry with tissue paper. ECL PLUS mixture (4.8ml reagent A + 120µl reagent B) was applied to the membrane and left to incubate for 5 minutes at room temperature. Membranes were dabbed dry and wrapped in clingfilm, ensuring there were no air bubbles in the film, before being placed in a cassette.
to prevent bleaching. Light sensitive film was placed over the membrane in the cassette and sealed for the appropriate exposure time before being developed in a BioRad developer.

Films were kept for further analysis and the membranes stripped by placing in 10ml stripping buffer for 20 minutes on a shaking plate at room temperature. Membranes were then re-blocked and stained for GAP-DH using the same protocol as above.

2.9.3 Quantification of Westerns using Image J

Films were scanned into a PC in high resolution and imported into Image J for analysis. Image J reads the signal intensity of the bands and background within a chosen frame and returns a numerical output for each band. The signal intensities were corrected for loading differences by dividing the numerical output from the gene of interest by the signal intensity of GAPDH on the same membrane. The mean signal intensity and standard error for each group was calculated in Microsoft Excel and imported into GraphPad prism for further statistical analysis and graph production.

2.9.4 Culture of Primary Neurons

Coverslips were prepared by incubating overnight in 100% ethanol before washing with distilled water, heat-dry sterilisation and incubating in Poly-D-Lysine solution.

Adult female mice were sacrificed 18 days post conception and embryos removed for neuron culture. The E18 embryo brains were removed and placed in dissection buffer containing HBSS (+CaCl₂ +MgCl₂), L-glutamate and HEPES (1M, pH 7.3-7.5) and kept on ice through the dissection process. The hippocampus was removed and placed in fresh dissecting buffer under a standard dissection microscope. Once all hippocampi had been collected, the dissection buffer was removed and the tissue chopped up using a scalpel. Trypsin was added at 0.1% to aid protein digestion and incubated at 37°C for 45mins. The suspension was passed through a wide Pasteur pipette 5x to remove clumps and
centrifuged at 1500rpm for 5mins. Trypsin was removed and the pellet resuspended in 10ml DMEM with foetal bovine serum to deactivate the trypsin. The process was repeated to ensure all traces of trypsin were removed and then the suspension was passed through a wide necked Pasteur pipette 20x and a narrow necked Pasteur pipette 10x. Suspension was centrifuged at 1500rpm for 5 mins, supernatant removed and pellet resuspended in 10mls DMEM (no FBS). The suspension was passed through a 40µM cell filter and cells counted before being centrifuged and resuspended in Neurobasal +++ medium (Neurobasal medium, B-27 supplement, GlutaMAX-1 supplement, Pen/Strep solution) at a concentration of 2x10^5 in a 12 well dish previously prepared with Poly-D-Lysine coverslips.

Cells were incubated at 37°C for 21days before being used for ICC. The neurobasal +++ medium was changed every 7days and cells checked to ensure health and growth.

2.9.5 ICC of Primary Neurons

The neurobasal medium was removed from the plate wells with an aspirator and coverslips washed twice with PBS at 4°C. Ice cold methanol was added to the coverslips and left to incubate at room temperature for 10minutes to fix the cells. Methanol was removed by aspiration and coverslips washed a further two times in PBS. Cells were blocked in 3% BSA for 20 minutes on a shaking plate at low rpm. Cells were then given two quick washes in PBS followed by 3 5 minute washes in PBS to remove all of the BSA. Primary antibodies were diluted in 3% BSA to a total volume of 200µl. After numerous trials a dilution of 1:50 was deemed the most appropriate for all primary antibodies of the genes of interest. Cells were also co-stained for alpha-tubilin and PSD-95. The antibodies were added to the coverslips and left on a shaking plate at room temperature for 1 hour to incubate. The primary antibody was removed by aspiration and coverslips washed as after the blocking treatment. Appropriate fluorescently labelled secondary antibodies were chosen and diluted in 3% BSA to a total volume of 200 µl. The colour of the fluorescent probe determined the dilutions. Red was diluted 1:800, green 1:500 and blue 1:1000. Secondary antibodies were added to the coverslips and left
to incubate on a shaking plate at room temperature for 1 hour. To prevent bleaching of the secondary antibodies the plate was wrapped in tinfoil to protect from light. The coverslips were washed as after the primary antibody treatment. A drop of mounting medium containing a Dapi stain was placed onto fresh slides and the coverslips mounted onto these with the cells facing downwards. Air bubbles were removed and slides stored away from light at 4°C until viewed with the confocal microscope.

2.10 Buffers

2.10.1 Western Blotting

Protein Sample Buffer (PSB)

- 6.25ml 1M Tris pH 6.8
- 10ml glycerol
- 10ml 20% SDS
- 13.75ml dH₂O
- 0.5mg Bromophenol Blue

Ponceau Stain

- 0.5g ponceau S
- 2ml Acetic acid
- 98ml dH₂O

Wash Buffer

- 100ml 10x PBS
- 900ml dH₂O
- 2ml tween

Running Buffer

- 20ml 20x NuPage Tris Acetate SDS running buffer
• 380ml dH₂O

Transfer Buffer
• 25ml NuPage transfer buffer
• 50ml methanol
• 425ml dH₂O

Blocking Buffer
• 2.5g Marvel
• 50ml PBS
• 100µl tween

2.10.2 Primary Neuron Production

500 µg/ml Stock Poly-D-Lysine Solution
• 5mg Poly-D-Lysine
• 10ml distilled water
• Filter sterilised and stored at -20°C

20 µg/ml Poly-D-Lysine Solution
• 25ml Borate buffer pH 8.5
• 22.5ml Distilled H₂O
• 2.5ml Stock Poly-D-Lysine

Dissection Buffer
• 500ml HBSS (+CaCl₂ + MgCl₂)
• 5ml L-glutamate
• 3.5ml HEPES (1M, pH7.3-7.5)

PBS
• PBS – (no CaCl$_2$ + MgCl$_2$)

0.1% Trypsin

• 20ml Trypsin/EDTA 0.25%
• 30ml dissection buffer

Neurobasal Medium

• 500ml Neurobasal medium
• 10ml B-27 supplement
• 5ml GlutaMAX-1 supplement
• 2.5ml Pen/Strep solution

DMEM

• 500ml DMEM
• 50ml foetal bovine serum
• 2.5ml Pen/Strep solution
Chapter 3

Sample collection and processing of a genome wide microarray analysis of the Disc1 ENU mouse mutants
3. Sample collection and processing of a genome wide microarray analysis of the Disc1 ENU mouse mutants

3.1 Introduction

DNA microarrays are widely considered the gold standard in genome wide expression studies of psychiatric genomics [99]. The complex multistep process allows extensive analysis of genome wide gene expression levels from relatively small sample numbers. This chapter describes the experimental preparation and preliminary analysis of a genome-wide microarray study comparing the Disc1 ENU mutant mouse lines described by Clapcote et al (2007)[197], with C57BL/6J control animals.

3.1.1 Background and motivation

Recently studies of ENU mutant mice have shown deficits in prepulse inhibition connected to genetic mutations, which can be partially reversed by antipsychotic medications[197]. Clapcote et al (2007) identified two missense mutant mouse lines with mutations within exon 2 of the Disc1 gene, in which they showed PPI deficits that were partially reversible with antipsychotic and antidepressant medication. These two mouse strains, L100P and Q31L, were described as 'schizophrenic-like' and 'depressive-like' by virtue of their behavioural profile and physiological observations and pharmacological responses that were considered comparable to human disease [218] (Table 3.1).

The two exon-2 mutants display two slightly different phenotypes corresponding to a schizophrenic-like and a depressive-like phenotype. The more severe schizophrenic-like phenotype occurs in the L100P mouse mutant. Here a point mutation has altered amino acid 100 from a leucine to a proline residue in the mouse Disc1. These mice have a marked reduction in PPI and overall startle response. The pronounced PPI deficit is partially rescued by administration of the typical antipsychotic haloperidol and the atypical antipsychotic clozapine. Interestingly, the anti-depressant and PDE4B inhibitor, rolipram also had a significant effect in rescuing PPI in the L100P mutant mice. Disc1 is known to bind to particular PDE4 isoforms at specific binding sites in the protein’s head
domain[104]. Two of these binding sites are disrupted by the mutations in both mouse lines suggesting that binding to PDE4B is likely to be impaired. The Q31L mutant mouse also displayed a reduction in PPI, although not to the extent of the L100P mutant. Interestingly, the deficit displayed by the Q31L “depressive-like” mice was not reversible by treatment with antipsychotic medications or with rolipram. Consistent with the lack of response to rolipram treatment the Q31L mice also exhibit decreased PDE4B activity compared to wild-type littermates and L100P “schizophrenic-like” mutant mice [197]. The antidepressant bupropion abolished the PPI deficit in these animals.

Both mouse lines displayed a disruption in Latent Inhibition (LI), which is a phenomenon by which prior exposure to a stimulus that holds no reward decreases the response to that stimulus when it is paired with an unconditioned stimulus. Again, administration of clozapine abolished this effect in the L100P mouse mutant, but not in the Q31L line. The Q31L mouse mutant, but not the L100P mouse mutant, displayed increased immobility during a forced swim test, a measure of behavioural despair, and reduced social interaction and reward responsiveness (measures of withdrawal and anhedonia respectively). The Q31L and L100P mouse models had overall reductions in brain size of 6 and 13% respectively, with tissue contraction mainly of the cortex, entorhinal cortex, thalamus and cerebellum.

The aim of the work described in this and the following chapter was to partially replicate the study by Clapcote and Roder to confirm the presence of PPI deficits in the Q31L and L100P mouse lines when compared to my control group of C57BL/6J mice bought in from Jackson laboratories that were to be used in my microarray study, and to identify genes whose expression was altered as a result of these mutations. In the original behavioural study, administration of selected antipsychotic and antidepressant drugs partially rescued the behavioural phenotype of the two lines.

In this chapter I will describe an attempt to replicate the behavioural study and describe the sample collection and primary analysis of a genome wide gene expression study in the ENU mutant mouse lines. By identifying other genes dysregulated by these mutations
it may be possible to identify pathways involved in major mental illness and possible new targets for treatment. Millar et al previously described a dynamic interaction between DISC1 and PDE4B that served to regulate cAMP signalling (Millar 2005)[131]. Disruption of this interaction is predicted to alter modulation of cAMP signalling and may result in abnormal transcription of genes with cAMP response elements in regulatory regions. DISC1 also interacts with the cAMP response element-dependent transcription factor ATF4, and possibly binds to chromatin remodelling factors such as SMARCE1 [132]. These interactions with transcription factors, together with the fact that DISC1 localises to the nucleus [172, 173] is consistent with a role for DISC1 in transcriptional regulation. Thus the view was that examination of the gene expression profile, by whole-genome microarray analysis, in the developing and adult brain of the two Disc1 mutant mouse lines was a key step towards understanding the factors underlying the behavioural and anatomical features of this mouse model.

3.2 Partial Replication of Clapcote et al’s Behavioural Analysis of the Disc1 ENU mouse mutants

Prepulse inhibition (PPI) is a neurological phenomenon in which the reaction to a startle stimulus (pulse) is reduced when it is preceded by a weaker prestimulus (prepulse) within a short time frame of under 500ms (figure 3.1). The reduction in startle response reflects the ability of the nervous system to adapt temporarily to a strong stimulus when a warning stimulus is present. It is generally accepted that this reflects the ability to filter out unnecessary information and so is a good measure of sensorimotor gating, a neurological function known to be deficient in schizophrenia[219].
In some healthy individuals prepulse inhibition is as great as 70%. Individuals with major mental disorders such as Alzheimer’s, schizophrenia, and bipolar depression however have significantly reduced prepulse inhibition compared to healthy controls[220]. As PPI deficits are observed in many mental disorders they are not diagnostic, but may indicate deficits in mental pathways common to these disorders. In particular, the inability to filter out unnecessary information could explain the symptoms of hallucinations and delusions associated with these disorders. In human patients there is also a noted gender difference in prepulse inhibition with males having higher PPI than females under all conditions[221]. Atypical antipsychotic medications such as risperidone and olanzapine are particularly effective in increasing PPI in individuals where a deficit is present[218].

Prepulse inhibition is not a purely human characteristic and has been widely observed in other mammals including rodents [219]. These animals are tested in a startle chamber (figure 2.4) with sensors detecting whole body movement as a measure of startle response. Pulse and pre-pulse are tones fed into the chamber via speakers mounted in the wall and controlled by computer programme. Pulse-alone results are compared to prepulse-plus-pulse, and the percentage of the reduction in the startle reflex represents prepulse inhibition. The baseline activity of each animal is also measured and subtracted.
from the startle response measurements. Baseline activity does not appear to affect the prepulse inhibition of an individual[222].

Table 3.1: Behavioural Phenotypes of the L100P and Q31L Disc1 mutant mice. Table displays result from multiple behavioural tests carried out by Clapcote et al (2007) on the L100P and Q31L mutant mice. Reduction in social interaction, reward responsiveness and increased immobility during a forced swim test indicated a ‘depressive-like’ phenotype in the Q31L mutant mouse. Greatly reduced PPI and LI in the L100P mutant mouse indicate a ‘schizophrenic-like’ phenotype.

Groups of mutant and control mice were also treated with drugs that had previously been successful in rescuing PPI in these mouse lines.
3.2.1 Effect of Gender on PPI in non-drug treated animals

Mice were tested as outlined in section 2.4.1. Twelve mice per group (8 Q31L adult) were tested on consecutive days. Analysis was carried out on startle response (120dB) and prepulse inhibition at 3 decibel levels (69, 73 and 81dB) and combined. The effects of gender, drugs and genotype were investigated.

There was no gender effect on total PPI (F=1.393, p =0.226) or acoustic startle response at 120dB minus pulse stimulus (F=2.357, p=0.069) in any group. As no gender differences were observed, further analyses were carried out grouping animals purely by genotype and drug treatment.

3.2.2 Effect of Genotype on PrePulse Inhibition and Startle Response in Mice

I found no difference in total prepulse inhibition (F=1.788, p=0.185) or acoustic startle response at 120dB minus pulse stimulus (F=2.091, p=0.142) by genotype in non-drug treated animals (Figure 3.2). It should be noted at this point, however, that comparison with previously published data showed that the mutant mice were displaying comparable levels of PPI and startle response as previously described, but that the wild type controls were displaying lower levels than previously reported. When the total prepulse inhibition for the C57BL6/J mice used in my study was compared to that of the wild type littermate controls in the previously published study they were significantly different (t=4.07, p=0.0003, df=31). This is also true for the acoustic startle response (t=2.63, p=0.014, df=26). If I substitute the previously reported wild type littermate data for my C57BL6/J controls in my comparisons, I find no significant differences except in one comparison, total PPI in the Q31L mutant (t=2.17, p=0.029, df=27).
Clapcote et al had previously reported that the level of prepulse may be significant in the manifestation of PPI in the mutant mouse lines. Analysis of variance was carried out on the data from the individual prepulse conditions (69dB, 73dB and 81dB) to determine if this was the case with my mice. While the both the L100P and Q31L mouse lines displayed a trend with reduced PPI at 69dB which increased through 73dB and 81dB there was again no significant differences between groups (F=1.478, p=0.844).

<table>
<thead>
<tr>
<th>% PrePulse Inhibition non-drug treated mice</th>
<th>L100P</th>
<th>Q31L</th>
</tr>
</thead>
<tbody>
<tr>
<td>C57BL6 69 dB</td>
<td>25.538 +/- 10.9</td>
<td>35.597 +/- 9.9</td>
</tr>
<tr>
<td>(n=12)</td>
<td>(n=12)</td>
<td>(n=8)</td>
</tr>
<tr>
<td>73 dB</td>
<td>21.146 +/- 11.5</td>
<td>48.231 +/- 9.6</td>
</tr>
<tr>
<td>(n=12)</td>
<td>(n=12)</td>
<td>(n=8)</td>
</tr>
<tr>
<td>81 dB</td>
<td>22.322 +/- 9.3</td>
<td>56.255 +/- 6.45</td>
</tr>
<tr>
<td>(n=12)</td>
<td>(n=12)</td>
<td>(n=8)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>23.002 +/- 9.9</td>
<td>46.354 +/- 8.3</td>
</tr>
<tr>
<td>(n=12)</td>
<td>(n=12)</td>
<td>(n=8)</td>
</tr>
</tbody>
</table>

Table 3.2: Mean results on PPI at all 3 levels tested, plus total. Pre-pulse inhibition was tested at 3 decibel levels, and averaged across all levels to produce a mean PPI total. Values shown are startle response and standard error.
3.2.3 Effect of Drug treatment on PrePulse Inhibition and Acoustic Startle Response in Mice

Prior to testing, 12 mice selected at random (8 Q31L) within each group had been treated with clozapine, bupropion or rolipram to determine the effects of these compounds on PPI and startle response within and between lines. Mice not receiving drug treatment received a saline injection.

<table>
<thead>
<tr>
<th>Acoustic Startle Response of Drug Treated Mice</th>
<th>Prepulse Inhibition of Drug Treated Mice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L100P</td>
</tr>
<tr>
<td>Saline</td>
<td>619.175 ± 67.7</td>
</tr>
<tr>
<td></td>
<td>(n=12)</td>
</tr>
<tr>
<td>Clozapine</td>
<td>713.575 ± 96.6</td>
</tr>
<tr>
<td></td>
<td>(n=12)</td>
</tr>
<tr>
<td>Rolipram</td>
<td>603.175 ± 90.1</td>
</tr>
<tr>
<td></td>
<td>(n=12)</td>
</tr>
<tr>
<td>Bupropion</td>
<td>610.533 ± 101.7</td>
</tr>
<tr>
<td></td>
<td>(n=12)</td>
</tr>
</tbody>
</table>

Mean acoustic startle response at 120dB stimulus, and average PPI over 69, 73 and 81 dB ranges

Table 3.3: Mean acoustic startle response and average PPI over all ranges tested. Table shows the total mean startle response and PPI by genotype and drug treatment. Values displayed are mean and standard error.

The effects of the drug treatment by genotype were analysed using a one-way ANOVA and post-hoc Tukey test. Total PPI was found to be significantly different in L100P drug treated mutants (F = 4.67, p = 0.019) and post-hoc analysis revealed clozapine treatment significantly decreased PPI, while bupropion treatment increased PPI in these mice. Acoustic startle response was not altered significantly in these comparisons. In the Q31L mutant no significant drug effect on PPI was found (F=0.118, p=0.889), however treatment with clozapine was found to significantly lower the acoustic startle response in these mice (p=0.02).
3.3 Whole Genome Gene Expression Study Experimental Design

3.3.1 Sample size and power

The RNA used in the microarray study was extracted from hippocampal brain tissues taken from the ENU mouse mutants immediately post behavioural testing. The result was 57 samples across six adult groups, as outlined in table 3.4.

<table>
<thead>
<tr>
<th>Biological Group</th>
<th>Group Symbol</th>
<th>Number of Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>C57BL/6J untreated</td>
<td>C57BL/6J</td>
<td>10</td>
</tr>
<tr>
<td>L100P untreated</td>
<td>L100Ptal</td>
<td>9</td>
</tr>
<tr>
<td>C31L untreated</td>
<td>C31Ltal</td>
<td>10</td>
</tr>
<tr>
<td>L100P rolipram treated</td>
<td>L100Proli</td>
<td>9</td>
</tr>
<tr>
<td>L100P clozapine treated</td>
<td>L100Pcloz</td>
<td>9</td>
</tr>
<tr>
<td>C31L bupropion treated</td>
<td>C31Lbupro</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 3.4 Summary of Adult Mouse Groups used in the genome-wide expression analysis. The groups are based on both genotype and presence or absence of drug treatment prior to behavioural testing. Group symbol is the annotation used throughout the rest of this thesis and the number of samples per group is shown in the final column.

Additionally, to test the effect of the Disc1 ENU mutations on genome-wide expression levels during development, 29 samples from three further groups were added from mouse Embryonic day 13.5 as outlined in table 3.5. Previously published data by Shurov et al (2004) [134] had shown Disc1 to have a peak in expression at this developmental time point, which coincides with neural determination and differentiation after a peak in neuronal migration at timepoint E12.5[223].
Chapter 3

Table 3.5 Summary of developmental mouse groups used in the genome wide expression analysis. Groups are based on genotype alone, no drug treatments were used in developmental samples. Group symbol is the annotation used throughout the rest of this thesis and the number of samples per group is shown in the final column.

<table>
<thead>
<tr>
<th>Biological Group</th>
<th>Group Symbol</th>
<th>Number of Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>C57BL6J Embryonic day 13.5</td>
<td>C57BL6.E13</td>
<td>12</td>
</tr>
<tr>
<td>L100P Embryonic day 13.5</td>
<td>L100P.E13</td>
<td>7</td>
</tr>
<tr>
<td>G31L Embryonic day 13.5</td>
<td>G31L.E13</td>
<td>10</td>
</tr>
</tbody>
</table>

The aim of this study was to determine which genes are dysregulated on the basis of genotype and drug treatment by comparing each of the ENU groups to their appropriate control groups. The C57BL/6J wildtype group was used as the control for testing the effect of genotype while the effect of drug treatment compared the drug treated group to their respective saline treated counterparts.

Power calculations were performed using the R/BioConductor (www.bioconductor.org) freeware with the `power.multi` function within the `sizepower` package[224]. As the true variability among individuals and actual effect size were unknown, a range of effect sizes and variance parameters were used to determine the ability of this experimental design to detect genes with differential expression. Natural variation among mice in inbred strains is thought to be very low (around 0.1-0.2 SD) so should result in low noise in the experiment and reasonable power to detect significant differential expression. As the mice used in this study had been backcrossed on to C57BL/6J (Jackson Laboratories) for 10 generations they could be considered an inbred strain and thus should have similar genetic variation. Standard deviations tested ranged from 0.1-0.5. It was, however, considered highly likely the natural variation between mice would be below the 0.2 SD range. No adjustments were made for multiple testing and five percent Type 1 error rate was assumed for each gene.

The power curves for the full range of SDs tested are shown in Figure 3.3. A sample size of 7 (as graphed) gives 80% power to detect a 1.3 fold change, the lower limit of the
illumina array platform, given an SD of 0.15. This sample size also provides 80% power to detect a two-fold change in more variable genes with SD of up to 0.4. This was the smallest group used, with most groups having sample sizes of 10-12 animals.

Figure 3.3: Power Curves. Power calculations were performed in the R/BioConductor (www.bioconductor.org) package sizepower using the multi.power function. Power curves were generated using the GraphPad software Prism 4.0 on license to the University of Edinburgh. The power (y-axis) was determined for the minimum number of samples and plotted against a range of standard deviation values (x-axis, range 0.1-0.5) using pairwise comparisons (NumTrt=2) with 7 samples per group (N=7, the smallest group size used) assuming a Type 1 error rate (α) of 0.05. Each coloured line represents the effect size expressed as fold change.
3.3.2 Sample Layout and Technical Replication

In total there were 86 samples to be hybridized to Illumina Mouse Ref-6 Beadchips. Due to the large number of samples available, and potential costs involved in running each sample individually, it was decided to pool samples based on genotype, drug treatment and gender to reduce the number of bead array chips required for the study. Pooling of samples resulted in 33 groups to be hybridized to the bead chip (figure 3.4) plus 9 replicate samples to be used for quality control. Pools from each biological group were evenly dispersed throughout the 7 BeadChips used to ensure that systematic bias or failure of one BeadChip did not sacrifice a whole biological group.

The nine replicate samples previously mentioned were used to assess the reliability between-BeadChip and within-BeadChip with different amplification and labelling preparations. This was an essential quality control method as comparisons would be made of data both between and within BeadChips. Firstly the non-drug treated C57BL/6J adult mouse female pool 2 (C57BL/6J-Sal F2) was hybridized to each BeadChip in the array in position A (figure 3.4) to allow assessment of the reproducibility between BeadChips. Secondly the same mouse RNA was amplified and labeled in three separate reactions (C57BL/6J-Sal F2 (1-3)) and hybridized to the same BeadChip in three different positions to assess the within-BeadChip experimental reproducibility. According to the BeadChip specifications (www.illumina.com), the expected array-to-array coefficient of variation (the ratio of the standard deviation of the mean expressed as a percentage and an indicator of within-BeadChip reliability) is <10%. This has been further corroborated by the Microarray Quality Control Consortium[225] through an intra-platform reproducibility study. As this has shown that within BeadChip reproducibility is reliable the main function of the technical triplicate was to determine the reproducibility and reliability of the amplification process and eliminate technical errors.
### Figure 3.4: Layout of the seven Illumina mouse ref-6 version 2 BeadChips.

Each column (1-7) represents a BeadChip and each row (A-F) represents an array on the BeadChip. Samples are named by their genotype, drug treatment, gender and pool number (as indicated in the key above). Samples in bold italics are the same sample on each BeadChip acting as between-BeadChip replicate controls. Samples in italics are from the same source RNA, amplified in three separate batches and hybridized to the array as within-BeadChip experimental reproducibility controls.

<table>
<thead>
<tr>
<th>Array</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>CS7BL6</td>
<td>CS7BL6</td>
<td>CS7BL6</td>
<td>CS7BL6</td>
<td>CS7BL6</td>
<td>CS7BL6</td>
<td>CS7BL6</td>
</tr>
<tr>
<td>B</td>
<td>CS7BL6</td>
<td>L100P</td>
<td>CS1L</td>
<td>CS7BL6</td>
<td>L100P</td>
<td>CS7BL6</td>
<td>CS7BL6</td>
</tr>
<tr>
<td></td>
<td>Sal F-1</td>
<td>Cloz F-1</td>
<td>E13 F-2</td>
<td>Sal M-2</td>
<td>Cloz F-1</td>
<td>Sal M-2</td>
<td>Sal F-2</td>
</tr>
<tr>
<td>C</td>
<td>L100P</td>
<td>L100P</td>
<td>L100P</td>
<td>CS7BL6</td>
<td>L100P</td>
<td>CS7BL6</td>
<td>CS7BL6</td>
</tr>
<tr>
<td></td>
<td>Cloz F-2</td>
<td>Sal F-2</td>
<td>Cloz F-2</td>
<td>Sal M-1</td>
<td>Cloz F-2</td>
<td>E13 F-2</td>
<td>E13 M-2</td>
</tr>
<tr>
<td>D</td>
<td>L100P</td>
<td>Q31L</td>
<td>L100P</td>
<td>L100P</td>
<td>Q31L</td>
<td>L100P</td>
<td>L100P</td>
</tr>
<tr>
<td></td>
<td>Roll M-2</td>
<td>Sal F-1</td>
<td>Cloz M-1</td>
<td>Sal M-2</td>
<td>E13 F-2</td>
<td>Sal F-2</td>
<td>E13 M-2</td>
</tr>
<tr>
<td>E</td>
<td>Q31L</td>
<td>Q31L</td>
<td>Q31L</td>
<td>Q31L</td>
<td>Q31L</td>
<td>L100P</td>
<td>L100P</td>
</tr>
<tr>
<td></td>
<td>E13 M-2</td>
<td>E13 M-1</td>
<td>E13 F-1</td>
<td>E13 F-1</td>
<td>E13 F-2</td>
<td>Roll F-1</td>
<td>E13 M-2</td>
</tr>
<tr>
<td>F</td>
<td>Q31L</td>
<td>Q31L</td>
<td>Q31L</td>
<td>Q31L</td>
<td>Q31L</td>
<td>Q31L</td>
<td>Q31L</td>
</tr>
<tr>
<td></td>
<td>L100P</td>
<td>Roll M-1</td>
<td>L100P</td>
<td>L100P</td>
<td>E13 F-1</td>
<td>L100P</td>
<td>E13 F-2</td>
</tr>
<tr>
<td></td>
<td>E13 F-1</td>
<td>E13 F-1</td>
<td>L100P</td>
<td>E13 F-1</td>
<td>E13 F-2</td>
<td>L100P</td>
<td>E13 F-2</td>
</tr>
<tr>
<td></td>
<td>E13 F-1</td>
<td>E13 F-1</td>
<td>L100P</td>
<td>E13 F-1</td>
<td>E13 F-2</td>
<td>L100P</td>
<td>E13 F-2</td>
</tr>
<tr>
<td></td>
<td>E13 F-1</td>
<td>E13 F-1</td>
<td>L100P</td>
<td>E13 F-1</td>
<td>E13 F-2</td>
<td>L100P</td>
<td>E13 F-2</td>
</tr>
<tr>
<td></td>
<td>E13 F-1</td>
<td>E13 F-1</td>
<td>L100P</td>
<td>E13 F-1</td>
<td>E13 F-2</td>
<td>L100P</td>
<td>E13 F-2</td>
</tr>
<tr>
<td></td>
<td>E13 F-1</td>
<td>E13 F-1</td>
<td>L100P</td>
<td>E13 F-1</td>
<td>E13 F-2</td>
<td>L100P</td>
<td>E13 F-2</td>
</tr>
<tr>
<td></td>
<td>E13 F-1</td>
<td>E13 F-1</td>
<td>L100P</td>
<td>E13 F-1</td>
<td>E13 F-2</td>
<td>L100P</td>
<td>E13 F-2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Key</th>
<th>CS7BL5</th>
<th>L100P</th>
<th>Q31L</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wild type non-termanate control</td>
<td>ENU mouse mutant with L100P Disc-1 mutation; &quot;schizophrenic-like&quot; phenotype</td>
<td>ENU mouse mutant with Q31L Disc-1 mutation; &quot;depressive-like&quot; phenotype</td>
</tr>
<tr>
<td>Sal</td>
<td>Non-drug treated</td>
<td>E13</td>
<td>Embryonic day 13</td>
</tr>
<tr>
<td>Cloz</td>
<td>Clozapine treated</td>
<td>F#</td>
<td>Female pool #</td>
</tr>
<tr>
<td>Roll</td>
<td>Rolipram treated</td>
<td>M#</td>
<td>Male pool #</td>
</tr>
<tr>
<td>Eupro</td>
<td>Euproprizine treated</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.4 Sample Preparation and Quality

Mice from the previously described behavioural study were sacrificed immediately post PrePulse inhibition testing by cervical dislocation. The brain was removed and stored in RNAlater for one week at 4°C to stabilize the tissues. Hippocampi were removed and RNA extracted. The time and date of tissue and RNA extraction was logged to allow identification of any batch effects. Total RNA was isolated and treated with DNase while on column to remove any contamination from genomic DNA.
3.4.1 Integrity of total RNA

Previous studies have shown that the integrity of total RNA is critical for successful microarray analysis. The Agilent 2100 Bioanalyser (Nano-LabChip) with 2100 expert software provides a standardized measure of RNA quality, the RNA Integrity Number (RIN) [226] which is based on the curve of the electrophoregraph produced by the analyzer software package. RIN numbers range from 1 (totally degraded RNA) to 10 (totally intact RNA). A high RNA RIN number is obviously desirable however previous publications have advised that consistency of RIN values across samples is more important than high values[227]. A comparison between RNA integrity and success of GeneChip (Affymetrix) microarray analysis [228] suggests that RIN>5.5 gives sufficient quality of expression data. With both these points in mind, samples with RIN numbers below 5.5 were not discarded immediately but were retained to be used if the other samples within that group were of consistently low value. Table 3.6 lists the identification numbers of mice used in this study alongside the date of RNA preparation and the RIN number returned from the Agilent Bioanalyser. RIN values ranged from 5.2 to 10.0, with an overall median RIN of 9.2 and a Mode RIN of 10.0. Median RINs from the three embryonic groups were comparable (C57BL6 E13, 9.85; L100P E13 9.80; Q31L E13 9.55), within the L100P mutant group the rolipram treated pools stood out, with a lower RIN than the others (L100P Sal 9.05; L100P Cloz 9.55; L100P Roli 7.65) and within the Q31L mutant group the bupropion treated pools stood out, with a lower RIN than the others (A31L Sal 9.30; Q31L Bupro 6.50). The C57BL6 Sal pools had a comparable median RIN to the two mutant untreated groups (C57BL6 Sal 9.30). Representative electrophoregraphs are shown in figure 3.5.
Figure 3.5: Examples of Agilent 2100 Bioanalyser output Electropherographs. (A) Total RNA with RIN 10 – intact with no contamination. (B) Total RNA with RIN 2.1 – low quality, high contamination and little intact RNA. Samples with RIN’s this low would not be used.
## Table 3.6: Mouse samples used in the microarray analysis

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Contained in Pool</th>
<th>Mutation</th>
<th>Drug treatment</th>
<th>Sex</th>
<th>RIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>C57-15-24</td>
<td>C57Bl/6J Sal-F1</td>
<td>Nil</td>
<td>SAINE</td>
<td>F</td>
<td>9.40</td>
</tr>
<tr>
<td>C57-14-24</td>
<td>C57Bl/6J Sal-F1</td>
<td>Nil</td>
<td>SAINE</td>
<td>F</td>
<td>9.40</td>
</tr>
<tr>
<td>C57-14-12</td>
<td>C57Bl/6J Sal-F1</td>
<td>Nil</td>
<td>SAINE</td>
<td>F</td>
<td>9.40</td>
</tr>
<tr>
<td>C57-13-24</td>
<td>C57Bl/6J Sal-F1</td>
<td>Nil</td>
<td>SAINE</td>
<td>F</td>
<td>9.40</td>
</tr>
<tr>
<td>C57-12-24</td>
<td>C57Bl/6J Sal-F1</td>
<td>Nil</td>
<td>SAINE</td>
<td>F</td>
<td>9.40</td>
</tr>
<tr>
<td>C57-12-12</td>
<td>C57Bl/6J Sal-F1</td>
<td>Nil</td>
<td>SAINE</td>
<td>F</td>
<td>9.40</td>
</tr>
<tr>
<td>C273-12-26</td>
<td>C57Bl/6J Sal-F2</td>
<td>Nil</td>
<td>SAINE</td>
<td>F</td>
<td>10.00</td>
</tr>
<tr>
<td>C273-12-12</td>
<td>C57Bl/6J Sal-F2</td>
<td>Nil</td>
<td>SAINE</td>
<td>F</td>
<td>10.00</td>
</tr>
<tr>
<td>C273-11-12</td>
<td>C57Bl/6J Sal-F2</td>
<td>Nil</td>
<td>SAINE</td>
<td>F</td>
<td>10.00</td>
</tr>
<tr>
<td>C273-10-12</td>
<td>C57Bl/6J Sal-F2</td>
<td>Nil</td>
<td>SAINE</td>
<td>F</td>
<td>10.00</td>
</tr>
<tr>
<td>C273-9-12</td>
<td>C57Bl/6J Sal-M2</td>
<td>Nil</td>
<td>SAINE</td>
<td>M</td>
<td>6.40</td>
</tr>
<tr>
<td>C273-8-12</td>
<td>C57Bl/6J Sal-M2</td>
<td>Nil</td>
<td>SAINE</td>
<td>M</td>
<td>6.40</td>
</tr>
<tr>
<td>C273-7-12</td>
<td>C57Bl/6J Sal-M2</td>
<td>Nil</td>
<td>SAINE</td>
<td>M</td>
<td>6.40</td>
</tr>
<tr>
<td>C273-6-12</td>
<td>C57Bl/6J Sal-M2</td>
<td>Nil</td>
<td>SAINE</td>
<td>M</td>
<td>6.40</td>
</tr>
<tr>
<td>C273-5-12</td>
<td>C57Bl/6J Sal-M2</td>
<td>Nil</td>
<td>SAINE</td>
<td>M</td>
<td>6.40</td>
</tr>
<tr>
<td>311E-1-30</td>
<td>C3H/101Sal-F1</td>
<td>3H101</td>
<td>SAINE</td>
<td>F</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-24</td>
<td>C3H/101Sal-F1</td>
<td>3H101</td>
<td>SAINE</td>
<td>F</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-18</td>
<td>C3H/101Sal-F1</td>
<td>3H101</td>
<td>SAINE</td>
<td>F</td>
<td>10.00</td>
</tr>
<tr>
<td>311E-1-12</td>
<td>C3H/101Sal-F1</td>
<td>3H101</td>
<td>SAINE</td>
<td>F</td>
<td>10.00</td>
</tr>
<tr>
<td>311E-1-18</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-12</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-12</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-18</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-24</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-28</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-30</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-18</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-12</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-18</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-24</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-28</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-30</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-18</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-12</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-18</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-24</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-28</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-30</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-18</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-12</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-18</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-24</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-28</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-30</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-16</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-12</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-18</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-12</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-18</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-24</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-28</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-30</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
<tr>
<td>311E-1-18</td>
<td>C3H/101Sal-M1</td>
<td>3H101</td>
<td>SAINE</td>
<td>M</td>
<td>9.60</td>
</tr>
</tbody>
</table>

**Table 3.6: Mouse samples used in the microarray analysis.** Column 1 gives the animal ID, column 2 the genotype and drug treatment, plus pool the animal was assigned to (F1 = female pool 1 etc), column 3 the mutation, column 4 the drug treatment, column 5 the gender and column 6 the RNA RIN number of the sample as an assessment of quality.
3.4.2 Integrity of cRNA

The RNA from the mouse samples in table 3.6 was used to produce biotinylated amplified cRNA (antisense mRNA). Labelled samples were produced within the two weeks prior to the microarray being run so no samples were stored over long periods. Quality and concentration of the samples were obtained using the Agilent Bioanalyser (Nano-LabChip). The Agilent Bioanalyser software does not provide a metric measure of quality for mRNA but rather reports on contamination levels by ribosomal RNA (rRNA) and provides an electropherograph of peaks in RNA size distribution that must be examined manually (figure 3.6). Evidence of a shift on RNA size distribution towards lower sized fragments would indicate degradation of RNA, which was not seen in any of the electropherographs analysed (figure 3.6). There were, however, some cases where contamination with rRNA was reported by the Agilent Bioanalyser which was not evident on manual examination of the traces. After communication with Agilent technical support, it was agreed the contamination reported was an artifact of the Agilent 2100 Expert software caused by high concentration levels close to the limits of detection for the Agilent Bioanalyser (Agilent RNA 6000 Nano chip = 25-250ng/µl mRNA).

Dilution of highly concentrated samples and re-quantification gave concentration measurements with a CV of less than 10% compared to the original undiluted measurement (Christoforou, personal communication). The Agilent Bioanalyser product specifications expect a CV of less than 10% for any reproduced quantification measurement making this well within the expected reproducibility of sample quantification. For this reason it was decided not to re-quantify these samples and to use the original quantification results provided the electropherograph traces showed no evidence on contamination peaks at 18S and 28S (the fragments which make up total RNA). This allowed me to be confident that the samples submitted to the Wellcome Trust Clinical Research Facility, Western General Hospital, Edinburgh, for hybridization to the array chips were of sufficient quality and accurately quantified.
Figure 3.6: Agilent Bioanalyser output of cRNA Integrity. Figure shows examples of two samples that were analysed using the Agilent Bioanalyser 2100 Expert software. Fluorescent units (FU) are plotted against size of RNA fragment (nucleotides). The RNA marker is indicated in the trace and on the gel image on the right hand side of each trace. RNA concentration and contamination are indicated under each graph. Trace A shows a sample with good concentration and no estimated contamination while trace B shows good concentration but predicted 18.2% contamination. As seen on the trace there is no evidence of a peak at 18S or 28S suggesting this is an artifact rather than true contamination.

3.5 Data Generation using the Illumina BeadArray Platform

Hybridisation of the cRNA samples was carried out at the Wellcome Trust Clinical Research Facility (WTCRF) at the Western General Hospital in Edinburgh. Samples were hybridized to seven Illumina Mouse-6 version 2 BeadChips on the Illumina BeadArray Platform. The chips were scanned using the Illumina BeadArray reader software, and image processing steps to determine the bead signal data from the raw images was automatically performed [229]. The Illumina BeadArray platform uses multiple bead copies for each probe allowing each probe signal intensity to be estimated from a number of beads reducing the likelihood of complete probe failure. Each bead copy is randomly positioned within the Beadchip to negate regional bias, and on average each probe signal intensity is estimated from readings from 30-40 bead copies with a low probability of being calculated from less than five beads [229]. Outlying beads (those with a median deviation more than 3) were removed and the average intensity of the remaining beads used to calculate the probe signal. Across the seven chips used in this experiment the
average number of beads per probe was 37.8 with one probe intensity being calculated from less than 5 beads (table 3.7). Illumina BeadStudio software was used to generate the probe signal intensity from multiple bead readings, along with the standard deviation of bead intensities. The detection p-value (which determines whether a probe is present in the given sample) was also calculated using this software.

Of the 48,000 probes on the BeadChip, 46,644 probes were successful (corresponding to 34,492 individual genes). For the purposes of quality control, preprocessing and differential expression analysis, each probe was identified not only by probe ID but by gene name, meaning some genes appear more than once in the initial stages or analysis. This is purely because some genes had multiple probes and as such could not be identified by gene name alone.

Table 3.6 shows the estimated average intensity and intensity range for each array as well as the average number of beads and minimum number of beads used to calculate the probe intensities. Across the whole experiment the mean probe intensity was 551.7 with a range of intensities from 10.1 to 48574. Interquartile range analysis was used to confirm no BeadChip probe intensity was considered an outlier and a one-way ANOVA (with 41 degrees-of-freedom) was performed on the raw, log2-transformed data to account for variability in the data. As expected the ANOVA showed statistically significant variation among the 42 arrays (F=5.23, p=0.0006), due to the effect of non-biological factors such as RNA integrity, hybridization and scanning, validating the need to normalize the data prior to further analysis.
<table>
<thead>
<tr>
<th>Sample ID (BeadChip)</th>
<th>Average # Beads (min)</th>
<th>Average signal (range)</th>
<th>Background (noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C57BL6Sal F2 (1)</td>
<td>37.8 (9)</td>
<td>518 (35.7-45324)</td>
<td>67.9 (20.3)</td>
</tr>
<tr>
<td>C57BL6Sal F1 (1)</td>
<td>38.0 (10)</td>
<td>531 (31.1-40870)</td>
<td>66.3 (19.5)</td>
</tr>
<tr>
<td>L100CPcloz F2 (1)</td>
<td>36.1 (6)</td>
<td>634 (36.6-45103)</td>
<td>73.0 (22.0)</td>
</tr>
<tr>
<td>L100CPRol M2 (1)</td>
<td>37.3 (5)</td>
<td>604 (36.4-44599)</td>
<td>74.5 (22.5)</td>
</tr>
<tr>
<td>Q31LBupr M1 (1)</td>
<td>37.4 (6)</td>
<td>540 (37.1-41067)</td>
<td>68.7 (19.8)</td>
</tr>
<tr>
<td>Q31LE13 M2 (1)</td>
<td>37.9 (9)</td>
<td>606 (37.5-43062)</td>
<td>71.6 (24.1)</td>
</tr>
<tr>
<td>C57BL6Sal F2 (2)</td>
<td>37.9 (11)</td>
<td>615 (42.4-48574)</td>
<td>75.0 (21.2)</td>
</tr>
<tr>
<td>L100CPcloz F1 (2)</td>
<td>37.6 (9)</td>
<td>727 (36.3-46952)</td>
<td>91.9 (28.0)</td>
</tr>
<tr>
<td>L100CPSal F1 (2)</td>
<td>36.4 (7)</td>
<td>487 (34.9-44619)</td>
<td>75.3 (18.2)</td>
</tr>
<tr>
<td>Q31LSal F1 (2)</td>
<td>36.2 (6)</td>
<td>575 (32.3-45038)</td>
<td>69.5 (19.9)</td>
</tr>
<tr>
<td>Q31LE13 M1 (2)</td>
<td>38.1 (7)</td>
<td>601 (35.6-45445)</td>
<td>71.8 (20.5)</td>
</tr>
<tr>
<td>Q31LSal M2 (2)</td>
<td>37.5 (9)</td>
<td>603 (35.6-43607)</td>
<td>71.5 (21.3)</td>
</tr>
<tr>
<td>C57BL6Sal F2 (3)</td>
<td>37.0 (9)</td>
<td>607 (37.0-41721)</td>
<td>75.2 (22.6)</td>
</tr>
<tr>
<td>Q31LE13 F2 (3)</td>
<td>37.7 (9)</td>
<td>563 (36.3-42270)</td>
<td>68.9 (17.7)</td>
</tr>
<tr>
<td>C57BL6Sal M1 (5)</td>
<td>38.1 (10)</td>
<td>643 (34.4-46216)</td>
<td>67.8 (18.5)</td>
</tr>
<tr>
<td>L100CPcloz M1 (3)</td>
<td>36.1 (8)</td>
<td>537 (35.5-40947)</td>
<td>67.9 (19.3)</td>
</tr>
<tr>
<td>Q31LBupr F1 (3)</td>
<td>36.1 (8)</td>
<td>467 (34.9-39263)</td>
<td>63.7 (17.7)</td>
</tr>
<tr>
<td>L100CPRol M1 (3)</td>
<td>37.5 (9)</td>
<td>534 (33.2-42965)</td>
<td>67.2 (19.1)</td>
</tr>
<tr>
<td>C57BL6Sal F2 (4)</td>
<td>37.7 (7)</td>
<td>574 (34.8-48172)</td>
<td>68.8 (19.6)</td>
</tr>
<tr>
<td>C57BL6Sal (4)</td>
<td>37.9 (8)</td>
<td>581 (26.3-41363)</td>
<td>60.0 (18.3)</td>
</tr>
<tr>
<td>L100CPcloz M2 (4)</td>
<td>37.8 (8)</td>
<td>506 (33.3-45196)</td>
<td>61.8 (15.6)</td>
</tr>
<tr>
<td>L100CPSal M2 (4)</td>
<td>38.2 (9)</td>
<td>622 (32.6-42080)</td>
<td>68.1 (19.5)</td>
</tr>
<tr>
<td>Q31LSal F2 (4)</td>
<td>38.1 (7)</td>
<td>571 (31.6-43520)</td>
<td>53.8 (18.2)</td>
</tr>
<tr>
<td>L100CPFe13 F1 (4)</td>
<td>37.9 (9)</td>
<td>528 (34.4-42920)</td>
<td>53.3 (16.9)</td>
</tr>
<tr>
<td>C57BL6Sal F2 (5)</td>
<td>37.6 (6)</td>
<td>524 (37.3-40258)</td>
<td>69.2 (22.5)</td>
</tr>
<tr>
<td>L100CPRol F1 (5)</td>
<td>36.1 (6)</td>
<td>427 (28.8-39550)</td>
<td>67.3 (16.7)</td>
</tr>
<tr>
<td>C57BL6El3 F2 (5)</td>
<td>37.7 (5)</td>
<td>433 (37.0-39070)</td>
<td>66.4 (19.7)</td>
</tr>
<tr>
<td>Q31LBupr F2 (6)</td>
<td>37.8 (4)</td>
<td>484 (34.0-38169)</td>
<td>67.6 (21.2)</td>
</tr>
<tr>
<td>C57BL6El3 F1 (5)</td>
<td>37.5 (8)</td>
<td>412 (32.4-39884)</td>
<td>63.5 (21.2)</td>
</tr>
<tr>
<td>L100CPFe13 F2 (5)</td>
<td>37.1 (6)</td>
<td>477 (30.0-40004)</td>
<td>66.4 (21.7)</td>
</tr>
<tr>
<td>C57BL6Sal F2 (6)</td>
<td>37.5 (10)</td>
<td>692 (10.1-44156)</td>
<td>77.2 (20.1)</td>
</tr>
<tr>
<td>L100CPSal F2 (6)</td>
<td>38.1 (12)</td>
<td>627 (12.3-47937)</td>
<td>76.2 (18.3)</td>
</tr>
<tr>
<td>C57BL6El3 M2 (6)</td>
<td>36.2 (7)</td>
<td>536 (14.6-47484)</td>
<td>69.1 (15.5)</td>
</tr>
<tr>
<td>L100CPSal M1 (6)</td>
<td>36.2 (8)</td>
<td>589 (26.3-46734)</td>
<td>71.1 (17.5)</td>
</tr>
<tr>
<td>L100CPRol F2 (6)</td>
<td>38.2 (9)</td>
<td>589 (25.7-43120)</td>
<td>71.8 (18.2)</td>
</tr>
<tr>
<td>Q31LE13 F1 (6)</td>
<td>37.9 (9)</td>
<td>607 (27.9-46465)</td>
<td>72.4 (18.8)</td>
</tr>
<tr>
<td>C57BL6Sal F1 (7)</td>
<td>37.2 (9)</td>
<td>514 (37.4-40968)</td>
<td>73.5 (15.5)</td>
</tr>
<tr>
<td>C57BL6Sal F1 (7)</td>
<td>38.1 (8)</td>
<td>513 (35.8-43560)</td>
<td>67.9 (15.5)</td>
</tr>
<tr>
<td>C57BL6Sal F1 (7)</td>
<td>37.9 (8)</td>
<td>551 (34.9-41587)</td>
<td>68.3 (17.1)</td>
</tr>
<tr>
<td>L100CPFe13 M1 (7)</td>
<td>36.2 (7)</td>
<td>505 (33.7-40032)</td>
<td>67.0 (14.7)</td>
</tr>
<tr>
<td>C57BL6El3 M1 (7)</td>
<td>36.4 (9)</td>
<td>477 (33.7-44449)</td>
<td>64.7 (16.9)</td>
</tr>
<tr>
<td>Q31LSal M1 (7)</td>
<td>36.9 (8)</td>
<td>508 (33.5-41563)</td>
<td>64.8 (16.1)</td>
</tr>
</tbody>
</table>

Table 3.7: Summary of raw expression data. Table includes pool ID and beadchip number (as in figure 3.4); average number of beads on each probe plus the minimum number of beads used to calculate probe intensity; the average probe intensity for each array plus the range of intensities recorded; and the background and noise recorded for each array. Replicate samples are shown in italics.
3.6 Quality Control

Multiple methods were employed to determine the quality of the microarray experiment, including both platform specific and sample specific controls as outlined below.

3.6.1 Illumina BeadChip Built in Controls

The Illumina BeadChip system has six in built control categories designed to monitor multiple aspects of the array experiment including sample labeling, hybridization to the array chip, quality of biological specimen and generation of probe signals. Quality control results from the Illumina built in controls in this experiment are shown in figure 3.7.
Figure 3.7: Signal Intensities for B. bernadensis Quality control measures. (A) The figure depicts the signal intensities for (A) the negative control probes; (B) the housekeeping control probes; (C) the high stringency and high control probes; (D) the low stringency control probes; (E) the hybridization control probes across the array experiment.
The negative controls (figure 3.7:A) consisted of 1,602 probe sequences which were thermodynamically comparable to the gene specific probe sequences but were made up of random sequence without specificity in the genome. Signal obtained from these probes was therefore an artifact of the imaging system background, non-specific binding or cross-hybridisation. The mean signal produced from these negative probes defined the system background (~ 69) while the standard deviation of the probes determines the noise in the array (~19) as outlined previously in table 3.7. This is significant to p<0.01. This value is then used in further analysis to establish the limits of detection within the array to allow genes with expression values lower than the negative to be discarded as they are not expressed above background.

The housekeeping gene controls (gene intensity), as shown in figure 3.7 (B), consist of 24 probes targeting seven housekeeping genes designed to monitor the integrity of the samples on the array. As these genes are expressed well in all intact samples it is expected that in a high quality sample the bead intensity of the housekeeping genes will be considerably higher than the average of all genes. The bead intensity of the housekeeping genes in this experiment was ~18,900 compared to an average targeted gene intensity of ~555, indicating that the samples used were of high quality (p<0.001).

The high stringency control and biotin-labelled control are both shown in figure 3.7:C. The high stringency control consists of two probes with high G+C content which define the upper bounds of stringency and are expected to hybridise regardless of the level of stringency observed. Signal from these probes in the absence of signal from the other hybridization controls would indicate that the stringency was too high. This was not the case in this experiment and it was determined that overall stringency was acceptable.

The low stringency control consists of sixteen probes which determined the lower bounds of stringency (figure 3.7: D). The probes each contain two mismatch bases (mm2) which is compared to their perfect-match counterparts (pm2). If the mm2 approaches the pm2 the hybridization stringency is considered to be too low. In this case there is a significant difference between the mm2 and the pm2 values (p<0.001) indicating sufficient stringency of hybridization.
The hybridization controls (figure 3.7: E) consist of twelve Cy3-labelled oligonucleotide probes complimentary to those present in the hybridization buffer. This control is independent of sample quality and determines whether the hybridization step was successful during array preparation. The signal produced should be proportional to the concentration of the probe to indicate successful hybridization. As this was the case with the samples tested it was concluded that hybridization was successful.

The biotin-labelled control determines the success of secondary staining and signal generation due to the presence of biotin labeled oligonucleotides in the hybridization buffer. The three probes gave a positive hybridisation signal indicating successful secondary staining.

These primary quality control measures indicated that the RNA samples were of sufficient quality and hybridization to the BeadChips was successful in this experiment. However, I also analysed the control data from each array separately to compare quality between chips (figure 3.9), and found one chip appeared to be significantly different to the others tested.
Figure 3.8: Signal Intensities for Illumina built-in Quality control measures by BeadChip. The figure depicts the signal intensities for (A) the negative control probes; (B) the housekeeping control probes; (C) the high stringency control probes; (D) the low stringency control probes; (E) the CY3 hybridisation control probes and (F) the biotin-labelled control probes for each individual BeadChip. BeadChip 5 has much lower probe signal intensities for low stringency hybridization, CY3 hybridisation and Biotin-labelled probes.

While BeadChip 5 passed the previous quality control measures when all chips were analysed together, further analysis revealed the CY3, low-stringency hybridization and biotin labeled control measures were considerably lower than those of the other BeadChips. After discussions with the Illumina support staff it was determined that part of the hybridization process for chip 5 was unsuccessful, probably due to the chip drying
out slightly during this step. As a result, samples that had been run on chip 5 were re-submitted to the WTCRF and hybridization of chip 5 was repeated. The chip-by-chip control profiles with the new chip 5 data are shown in figure 3.9.

![Figure 3.9: Signal Intensities for Illumina built-in Quality control measures by BeadChip post chip5 repeat.](image)

After the repeat run of BeadChip 5 the low stringency hybridization, CY3 hybridisation and biotin-labelled control probes produce comparable signal intensities to the other BeadChips, allowing more confident comparisons between chips to be made.

The repeat run of BeadChip 5 gave results comparable to the control values previously observed for the other six BeadChips. From this point forward all analysis was carried out on the data from the repeated BeadChip 5 run.
In summary, the initial analysis of the system controls did not indicate a problem with any BeadChip. Interquartile range analysis had not highlighted BeadChip5 as an outlier and primary quality control analysis was as expected in a successful array. Only the secondary analysis looking at each chip’s controls individually showed a potential problem with one chip. I would therefore conclude that this step is vital in ensuring all system quality control measures are accurate.

### 3.6.2 Technical Replicates

To assess the reproducibility of the BeadChip platform and RNA preparation in this experiment technical and experimental replicates were included in this study. The reproducibility of the BeadArray platform was determined by the inclusion of the C57BL6 F1 pool on all BeadChips while the joint within-BeadChip and RNA preparation reproducibility was determined by the inclusion of three RNA preparations of the same sample (C57BL6 F2) on BeadChip 7.

Pairwise comparisons of the raw, untransformed C57BL6 F1 pool data (figure 3.10) revealed high correlation between the replicate samples across chips (Pearson’s $r^2 \geq 0.972$). Similarly high pairwise comparisons were observed between the RNA preparations of C57BL6 F2 (figure 3.11) on BeadChip7 (Pearson’s $r^2 \geq 0.995$) indicating that variability due to batch and experimental effect were within acceptable ranges.
Figure 3.10: Pairwise Comparisons of raw data for C57BL6 F1 pool samples. Figure shows pairwise scatter plots of the log2-transformed intensities of the C57BL/6J chip replicates produced using the Bioconductor lumi package. Correlation estimates (Pearson’s correlation coefficient r²) were calculated using the cor function in R.
Figure 3.11: Pairwise comparisons of raw data for C57BL6 F2 pool samples. Figure shows pairwise scatter plots of the log2-transformed intensities of the C57BL/6J technical replicates produced using the Bioconductor lumi package. Correlation estimates (Pearson’s correlation coefficient r²) were calculated using the cor function in R.

3.6.3 Quality Control using Cluster Analysis

Cluster analysis is a basic statistical approach which assigns samples into groups based on similarity. Hierarchical cluster analysis starts with single-member groups which are merged based on how similar their expression profiles are resulting in a hierarchical tree known as a dendrogram. Multiple methods of calculating hierarchical clusters exist which differ in the way they calculate similarity, and give different clusters [230] highlighting
different aspects of the data. The data was analysed by ‘average correlation’ using the freeware R with the pvclust application, that assesses robustness of the clusters using multiscale bootstrapping (1000 bootstraps). The cluster analysis was carried out on raw expression data to determine if technical or experimental artifacts affected clustering. It was expected that samples would not cluster into their respective biological group prior to data pre-processing and would show random spread across the dendrogram.

Figure 3.12 shows the average correlation dendrogram for the raw expression data. The pairwise distance (or similarity) was measured using the correlation method (1 pearsons correlation) and clusters joined using ‘average linkage’ which measures correlation distances across clusters. The bootstrapping results are represented by the approximate unbiased (AU) probability values, which indicate the frequency at which that particular cluster was observed over the 1000 bootstraps. An AU value of 100 indicates the cluster was produced on every occasion during bootstrapping.

At this stage the embryonic and adult samples are clearly separated into two distinct branches with a Pearsons correlation of around 0.75. This would be expected as the expression profiles of the embryonic samples would be markedly different to the adult samples. Within the two main branches there was a maximum distance between two samples of ~0.025, indicating a Pearsons correlation of $\geq 0.975$ which suggests low overall variability even before pre-processing and data stabilization. The technical replicates (C57BL6 F1) cluster together within the adult branch with an AU value of 100, and the chip replicates (C57BL6 F2 hybridised to each chip) also cluster together with the exception of the sample on chip 6 which falls closer to the technical replicate clusters. All other groups appear evenly spaced throughout the dendrogram with no observable clustering due to chip number or gender (figure 3.12).
Figure 3.12: Hierarchical cluster dendrogram of all samples based on the raw dataset. The height of the dendrogram (y-axis) represents the distance between any two samples measured by the correlation distance method (equivalent to Pearson’s r). Clusters were joined using the average linkage method and AU (average unbiased probability) values generated by preclust are shown in red. The technical replicates are bordered in green with the chip replicates bordered in blue and the complete embryonic samples in purple.
3.7 Data Pre-processing

The Illumina microarray platform is unique in having a high redundancy of randomly positioned beads, which reduces positional bias, increases robustness of hybridization intensity and measurement error estimates across each probe. As multiple samples are hybridized to each chip, hybridization batch effects are also reduced. These features mean pre-processing of the data prior to differential expression analysis should be carefully considered to take advantage of this redundancy.

Three levels of pre-processing were carried out on the raw data in this study. Firstly the variance within the data was stabilized, secondly the stabilized data was normalized, and finally the stabilized and normalized data was filtered to first remove any probes not expressed above background and then to remove any probes that were not expressed at a level considered biologically significant.

3.7.1 Variance Stabilisation

Microarray data has a tendency to be highly variable, with high intensity data displaying larger variance. Inconsistency in variance (heteroskedacity) is a violation of many parametric statistic tests, such as analysis of variance (ANOVA), which assumes constant equal variance. To stabilize the variance in this dataset the variance stabilizing function (lumi-vst) within the Illumina specific Bioconductor lumi package (www.basic.northwestern.edu/projects/lumi) was used. Again, previous studies within the laboratory had shown this to be more reliable than standard log\_2-transformation, as the lumi-vst function directly models the mean-variance relationship in the array by taking account of the within-array bead redundancy to generate an optimal transformation parameter. After stabilization of the data using the lumi-vst function, a generalized log transformation was then performed to give appropriate complete stabilization of the data and more manageable data values for further analysis.
3.7.2 Normalisation

The second step in pre-processing is to normalize the data to remove the systematic variation caused by non-biological sources such as RNA quality, hybridization and scanning. Again the Bioconductor package lumi was utilized in this step using the lumi-rsn function which was originally created to be used in conjunction with the lumi-vst function from the previous step (www.bioconductor.org/packages/2.1/bioc/vignettes/lumi/inst/doc/Bioc2007_lumi_prese ntation.pdf). Due to the size of the dataset it was necessary to process each chip individually. This was possible as each array was normalized against a target array which was most similar to the mean of all samples. The mean reported signal intensity across the whole array was ~555. With a mean signal intensity of ~551, the technical replicate C57BL6 F1 (1) was the closest to the overall mean and was used as the target array for normalization. All samples on the array were normalized to this array using the entire probe-set.

Hierarchical cluster analysis was carried out on the stabilized and normalized data (figure 3.13) showing a reduction in overall variability across the array. At this stage the embryonic and adult samples are still clearly separated into two distinct branches with a Pearsons correlation of around 0.85. This would be expected as the expression profiles of the embryonic samples would be markedly different to the adult samples. Within the two main branches there was a maximum distance between two samples of ~0.015, indicating a Pearsons correlation of ≥0.985. The technical replicates (C57BL6 F1) cluster together within the adult branch with an AU value of 100, and the chip replicates (C57BL6 F2 hybridised to each chip) also now all cluster together. All other groups still appear evenly spaced throughout the dendrograph although with greater correlation than seen previously, with no observed clustering due to chip number or gender. Pairwise correlation of the technical replicates increased to $r^2 \sim 0.996$ and chip replicates to $\sim 0.991$. Formal testing of the data using analysis of variance (ANOVA) with 41 degrees-of-freedom showed no overall significant variability of the data indicating normalization was successful ($F=5.23$, $p=0.0006$).
Figure 3.13: Hierarchical cluster dendrogram of all samples based on the stabilised and normalised dataset. The height of the dendrogram (y-axis) represents the distance between any two samples measured by the correlation distance method (−1 Pearson’s r). Clusters were joined using the average linkage method and AU (average unbiased probability) values generated by preluce are shown in red. The technical replicates are bordered in aqua with the chip replicates bordered in pink. Clusters are much closer than for the raw data as would be expected for successful normalisation.
As global background correction would have resulted in negative signals which would have affected the pre-processing steps this was not performed on the datasets. Instead the data was filtered to include only the probes that were expressed above background in each sample as described in section 3.7.3 below.

### 3.7.3 Data Filtering

Data filtering was used to reduce the multiple testing burden and increase sensitivity of the microarray analysis. The data was filtered in two steps; the first removed probes that were not expressed above background, and the second selected only the probes with biologically relevant differential expression levels.

The detection limits of the array experiment are automatically generated from the background and noise measurements which are estimated from the signal intensities produced by the built-in negative controls. The detection P-value for each probe is calculated for individual and multiple samples using non-parametric methods that rank the probe signal relative to the negative control. This can be interpreted as the probability that the probe signal is no greater than background, relative to the noise in the data. The higher the intensity signal, the lower the probability that it is generated by non-specific sources. The detection P-value was calculated for each probe in each of the nine biological groups (C57BL6Sal, L100PSal, Q31LSal, L100PCloz, L100PRoli, Q31LBupro, C57BL6E13, L100PE13 and Q31LE13) and only those probes with a detection P-value of less than 0.05 in at least one of the nine groups were retained for further analysis. This allowed probes to be expressed in one group but not in another. Of the 46,644 successful probes (corresponding to 34,492 genes) on the Mouse Ref-6 BeadChip, 21,869 (~47%) probes met this threshold. Estimates by others state that most tissues only express 30-40% of genes [231] so this is slightly higher than would be expected. Figure 3.14 illustrates the final cluster dendrogram after the data had been filtered. There is very little difference between this and the previous dendrogram for the stabilized and normalized data. This was unexpected as filtering of probes not significantly expressed above background should increase the bootstrapping AU values.
indicating stronger support for the observed clusters. However, the previous cluster dendrogram gave AU values all close to or equal to the maximum of 100 which is likely to be the reason that no difference is seen in this new filtered dataset.

In the second filtering step only probes which showed potentially significant differential expression when compared to the relevant control group were retained. This has previously been shown that the most reproducible and reliable method of determining differential expression in microarray analysis is based on a rank of fold-change and non-stringent statistical significance (P-value) cut-offs [225]. As the Illumina platform offers the sensitivity to detect down to a 1.3 fold change (www.illumina.com/downloads/Mouse6-8v2Datasheet.pdf) this was used as the cut-off for this second filtering step.

Fold change was calculated by taking the anti-log$_2$ of the difference between the average normalized values of the two groups being compared at each probe. This fold-change filter reduced the number of probes to be tested from 21,869 to 1266 in the L100PSal vs C57BL6Sal comparison, 26 in the Q31LSal vs C57BL6Sal comparison, 996 in the L100PCloz vs L100P Sal comparison, 1313 in the L100PRoli vs L100PSal comparison, 251 in the Q31LBupro vs Q31LSal comparison, 2005 in the L100PE13 vs C57BL6E13 comparison and 622 in the Q31LE13 vs C57BL6E13 comparison (Figure 3.15).
Figure 3.14: Hierarchical cluster dendrogram of all samples based on the filtered, transformed and stabilised dataset. The height of the dendrogram (y-axis) represents the distance between any two samples measured by the correlation distance method (1 Pearson’s r). Clusters were joined using the average linkage method and AU (average unbiased probability) values generated by pheatmap are shown in red. The technical replicates are bordered in grey, with the chip replicates bordered in purple and the complete embryo samples in pink.
3.8 Conclusions

This chapter has described the experimental preparation and pre-processing analysis of a genome wide expression study carried out on two ENU mutant mouse lines which display distinct endophenotypes categorised as ‘schizophrenic-like’ and ‘depressive-like’ using the Illumina BeadArray platform.

The behavioural study carried out prior to the collection of samples for the microarray was designed to replicate the results obtained by Clapcote et al (2007) [197] to ensure the mutation continued to carry the behavioural phenotype previously described. As described, the animals used for this experiment were backcrossed for a further 3 generations than those used in the previously published study and the pool of mice available was substantially affected by a spontaneous mutation within the breeding
programme which required a mass cull of animals prior to the breeding of those to be used in this experiment. This resulted in a much reduced gene pool available, and the requirement to cross homozygous animals to generate the number of animals required, which further compounded the bottleneck effect. This also made it impossible to use wild-type littermates as a control as no heterozygous animals were used in the crosses. As the ENU mutant lines were established on a C57BL/6J genetic background it was determined that these animals would be the closest genetically to wild-type littermates.

The PrePulse inhibition analysis carried out by myself did not replicate the study previously published by Clapcote et al using the ENU mutant mouse lines. There are a number of reasons which may explain this. It is possible that the further backcrossing reduced the effect of the mutation on the behaviour phenotype previously observed. This is unlikely however, as additional, more recent, studies (Tatiana Lipina, personal communication) have replicated the results previously published, and personal observations of the animals brain physiology and general behaviour concur with those reported previously. Subsequent literature searches combined with results obtained in this experiment show the C57BL/6J mice may not have been the ideal control for this experiment. Studies of PPI in rodents have been going on for many years as researchers attempt to find the genetic basis for such behaviours. Paylor and Crawley (1997) [232] studied the differences in PPI in commercially available inbred mouse strains. One strain used in their study was the C57BL/6J strain from Jackson Laboratories as used in this study. They looked at mice between the ages and 9 and 14 weeks over 5 auditory prepulse stimuli. Interestingly, the C57BL/6J mice gave the lowest PPI of all the mice they tested[232]. Two of the decibel levels used by Paylor and Crawley [232] can be compared to those used in the current study and give an average of 20% PPI compared to the 23% PPI observed in the current study. It is also important to note that the original Clapcote [197] experiment grouped the wild-type littermates from both mutants in one ‘control’ group so any differences which may appear between wild-types littermates will be hidden within the one group. While the C57BL6 mice were specifically sourced from the Jax substrain for the behavioural study, to be as close as possible to the original backcross strain used when establishing the ENU mutant colonies, the results obtained would suggest that they were not an adequate control, and there may be non-target
mutations in the wild-type littermates which may be responsible for the differences observed. These could either be secondary ENU mutations that co-segregated with the Disc1 mutation, or sections of DBA genome that remained, despite backcrossing, around the Disc1 gene. This is discussed further in Chapter 8 and any future work should review the validity of non-littermate controls. It should also be noted that while comparison between PPI deficits in rodents and humans are, for the most part consistent, some drug effects have been shown to give differing results across species[220]. Prepulse inhibition is regarded as a reliable endophenotype of major mental illness (and schizophrenia in particular) but care must be taken over the conclusions drawn from our results if we wish to relate our mouse model to human disease.

Despite the limitations described above, it was determined that the full microarray experiment should continue as any genes which were ‘flagged’ as being differentially expressed in the comparisons with the C57BL/6J mice would be subsequently tested against wild-type littermates after the array as it would be possible to re-establish a complete ENU mutant mouse colony within this time.

The aim of the microarray experiment was to determine which genes were differentially expressed on the basis of genotype, drug treatment and developmental stage (embryonic and adult). A total of seven comparisons were made across nine biological groups. The tissues used for the array analysis were taken from the animals used in the behavioural study immediately after completion of the PrePulse inhibition cycle. All animals that underwent behavioural testing were sacrificed and RNA extracted from hippocampal tissues, however only the groups outlined previously were included in the array. Abnormalities in hippocampal structure, activation, organization of neurons and synapse function have been well documented in human schizophrenia patients and other mouse models [207, 214, 233, 234]. As there are working memory deficits often associated with schizophrenia [208], it is clear there is substantial hippocampal involvement in the disorder. The use of a single brain region also removed any confounding effects of compensation in other brain regions, which may have diluted the result. Financial constraints meant that only the drug treated groups which had shown a marked behavioural response in the previously published study were included, as these were
more likely to give a distinct genetic profile for comparison. The RNA from the other groups was retained to be used for future experiments and to determine the importance of some of the results obtained (as described in Chapter 5). The sample preparation, quality control and pre-processing methods used had been previously tested by other researchers within the laboratory and were considered of good standard with reliable output for differential expression analysis. Biotin labeling and hybridization controls indicated the RNA used was of good quality, increasing the likelihood of a reliable output from the array.

The main limitation of this study was the requirement to use the C57BL/6J mouse as a wild-type control, particularly with the resulting PrePulse inhibition result, and if the experiment were to be repeated it would be advisable to use wild-type littermates as controls. The use of these animals was however justified by the fact the ENU mutants had been backcrossed for 10 generations and would be considered as close to the C57BL/6J genotype as possible, and further analysis would include wild-type littermates to validate any results. As RNA was available for other biological groups not included on the array it would also have been interesting to compare the drug treated C57BL/6J animals with the drug treated ENU mutant samples to potentially determine whether the drug treatments affected the ENU mutants in a different manner to their ‘wild-type’ controls.

The next stage of analysis will describe differential expression profiles and potential biological significance of the results obtained from this primary analysis.
Chapter 4
Differential Expression Analysis of the Disc1 ENU Mutant Mouse
Microarray
4. Differential Expression Analysis, Primary Annotations and Biological Relevance

4.1 Introduction
Analysis of differential expression between mutant lines provides insight into potential genes and pathways involved in the phenotypes expressed in these animals. By comparing these genes to those previously published in human studies we can identify genes and pathways involved in major mental illness and potentially determine targets for further study or new drug treatments. This chapter describes the process of differential expression analysis in the ENU mouse mutants including pathway analysis and overlaps with previously published human studies.

4.2 Differentially Expressed Genes
The array probes that survived the fold-change filter described in chapter 3 were tested for differential expression using the Analysis of Variance model (ANOVA) written in the R package (MAANOVA) which is specifically designed to be used with microarray data [235]. The two main effects tested were genotype and drug treatment. Tests of differential expression were carried out using the MAANOVA `matest` function using the Fs test statistic, which makes no assumption about the distribution of variance but estimates it based on the information across all probes in the dataset. This has previously been shown to be more powerful than other F statistics available in the package for detecting differential expression [236] because of the employment of the James-Stein shrinkage concept. This concept uses the information across all probes, rather than probe specific estimators, to gain a better estimate of variance within the dataset. This increases the power of the test thus giving a higher reward in small sample sizes than standard F statistics. Table 4.1 shows the number of genes found to be differentially expressed in each group comparison (Fs p-value ≤ 0.05) based on genotype or drug treatment where appropriate.
For each probe tested I also calculated the false discovery rate (FDR), using the *adj* p-value function in MAANOVA, to determine the stringency of the 0.05 cut-off. The FDR value determines the expected proportion of false positives when determining significance of each probe. Table 4.1 shows the mean FDR for each comparison based on the *adj* p-value for each probe called as significant within that comparison. The maximum FDR for any comparison was 0.35 suggesting a moderate probability of any probe being a false positive. As genes were to be subjected to further rounds of validation this was deemed to be acceptable.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Probes</th>
<th>Significant (%)</th>
<th>Mean FDR (maximum)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L100Padult vs C57BL6adult</td>
<td>1266</td>
<td>531 (41.9)</td>
<td>0.26 (0.3)</td>
</tr>
<tr>
<td>G31Ladult vs C57BL6adult</td>
<td>26</td>
<td>9 (34.5)</td>
<td>0.21 (0.35)</td>
</tr>
<tr>
<td>L100PE13.5 vs C57BL6E13.5</td>
<td>2005</td>
<td>33 (1.6)</td>
<td>0.19 (0.22)</td>
</tr>
<tr>
<td>G31LE13.5 vs C57BL6E13.5</td>
<td>622</td>
<td>15 (2.4)</td>
<td>0.22 (0.24)</td>
</tr>
<tr>
<td>L100Padult vs L100PRol</td>
<td>1313</td>
<td>44 (3.4)</td>
<td>0.13 (0.18)</td>
</tr>
<tr>
<td>L100Padult vs L100PCln</td>
<td>996</td>
<td>3 (0.3)</td>
<td>0.20 (0.23)</td>
</tr>
<tr>
<td>G31Ladult vs G31LBupro</td>
<td>251</td>
<td>103 (41.0)</td>
<td>0.18 (0.21)</td>
</tr>
</tbody>
</table>

Table 4.1 Summary of probes differentially expressed by genotype of drug treatment. Column 1 shows the pair-wise comparison made while column 2 gives the number of probes that passed the 1.3 fold-change filter. Column 3 shows the number and percentage of probes that were significantly differentially expressed (*F* < p-value ≤0.05). The final column gives the mean and maximum false discovery rate (FDR) observed in the list of significantly differentially expressed probes.

Some genes showed differential expression in more than one pair-wise comparison (figure 4.1)
In addition, eight genes that were found to be differentially expressed in the mutant vs wild-type comparisons were also differentially expressed when the mutant, non-drug treated (saline) animals were compared to their drug treated counterparts. When the drug treated animals were compared to the C57BL/6J non drug treated adults for expression at these genes no significant difference was observed, suggesting the drug treatment has corrected the expression level back to that of the control animal. Seven of these genes were in the L100Padult (saline) vs L100PRoli (rolipram) comparison while one gene was in the Q31Ladult (saline) vs Q31LBupro (bupropion) comparison. No correction of differential expression was observed in the L100PCloz (clozapine) group.

### 4.2.2 Differentially expressed genes: primary annotation and gene ontology

Genes that were shown to be differentially expressed were annotated using data provided by Illumina. Each probe sequence was also run through NGBI Gene to determine its location on the mouse genome and to check whether this corresponded to the gene name provided by Illumina. The official gene name, symbol, accession number, GI number and gene ontology were noted for each gene within this search.
Gene ontology (GO) analysis was performed to determine if the lists of significantly differentially expressed genes showed enrichment for any particular GO category relating to molecular function, biological processes or cell components. This was carried out using the web based GOTree machine (GOTM) (http://bioinfo.vanderbilt.edu/gotm) by comparing the list of differentially expressed genes to those expressed above background in the relevant groups. GO categories are arranged in a hierarchical structure of common, controlled vocabulary used to describe the roles of genes and gene products of different species. Only GO enrichments that are statistically significant (P<0.01, as determined by the hypergeometric test) are reported by GOTM.

The list of statistically significantly differentially expressed genes was then compared to gene lists from recently published studies. Three of the studies used for comparison [83, 84, 217] described copy number variants in cases and controls of schizophrenia and other major mental illness. A further two, unpublished, microarray studies identified genes dysregulated in two large families with a history of mental illness. The first used lymphoblastoid cell lines from a large Scottish family with a balanced translocation (t1:11) that formed the original sample set used to identify the DISC1 mutation and the motivation for characterising the two Disc1 mutants used in this study [8, 56]. This family has a history of schizophrenia and schizoaffective disorders spanning multiple generations. The second used lymphoblastoid cell lines from the F22 family which has a history of major depressive disorder and provides a dataset of both affected and unaffected individuals with married in controls[237]. Lastly, the DISC1 interactome [132] was used to identify dysregulated genes in this dataset that are known to interact with Disc1.

Finally, pathway analysis was carried out on the lists of differentially expressed genes using the web-based software Ingenuity Pathway Analysis (IPA). All probes expressed above background in each comparison were loaded into the system regardless of fold-change or p-value. These were then set as the main observations for determining cut-offs (ie, only include genes with FC>1.3 and p-val<0.05) so any pathways which were created by the system would highlight those genes with significant differential expression and
would also show (by shading) those genes expressed above background in the sample group. Any genes in the pathway which were not expressed above background in the sample group would be left unshaded. IPA then assigns a significance score to each network (as described in section 2.7.2) which indicates the likelihood that the assembly of focus genes within a network could be explained by random chance.

After these stages had been carried out all genes identified by the array analysis were scored on the basis of gene ontology, overlaps with previous studies and presence or absence in high scoring networks, to determine which genes would be carried forward for further analysis. The following sections describe the results of this process for each pairwise comparison from the array analysis.

### 4.2.3 Differential Expression between the L100P adult ENU mutant and the C57BL/6J adult controls

Of the 1266 probes that passed the pre-processing filters in the L100P adult group, 531 showed differential expression compared to the C57BL/6J control group (table 4.3). 261 of these genes were over-expressed and 270 under-expressed in the L100P adult group. Four of these genes (*Akap9, Gnb1, Ppm1e and Smarce1*) are known interactors of Disc1 (Camargo et al 2006 [132]). With 127 genes in the Disc1 interactome, hypergeometric probability suggests this is a significant result \( p=0.003 \). *Akap9, Gnb1* and *Smarce1* were over-expressed and *Ppm1e* under-expressed in the L100P adult sample group relative to the C57BL/6J controls.
Table 4.3: Complete list of dysregulated genes in the L100P adult group. Cont next page...
Table 4.3: Complete list of dysregulated genes in the L100P adult group. Continued next page….
Table 4.3: Complete list of dysregulated genes in the L100P adult group. Cont next page...
Table 4.3: Complete list of dysregulated genes in the L100P adult group. Continued next page....
Table 4.3: Complete list of dysregulated genes in the L100P adult group. Continued next page....
Table 4.3: Complete list of dysregulated genes in the L100P adult group. Continued next page....
Table 4.3: Complete list of dysregulated genes in the L100P adult group. Cont next page……

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Log2 Fold Change</th>
<th>p-value</th>
<th>FDR</th>
<th>Description</th>
</tr>
</thead>
</table>

...
Table 4.3: Complete list of dysregulated genes in the L100P adult group. Cont next page……
Table 4.3: Complete list of dysregulated genes in the L100P adult group. Columns show Target (gene) ID, full gene name, accession number and probe ID, log transformed expression values for both comparable groups, difference between groups, fold-change, p-value and GeneOntology terms for each gene.
Nine genes showed overlap with the previously published schizophrenia CNV study by Walsh et al [83] and six of these findings have the same presumed directionality as the CNV study (ie over/under-expression corresponding to gain/loss of copy number). Two genes were found to overlap with a second CNV study [84] and again followed the same directionality (Table 4.2). Comparison with a microarray study of the F22 family described above (Christoforou et al, unpublished) also showed two genes of overlap with the current study following the same directionality.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Study</th>
<th>Gain/Loss</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARM10</td>
<td>Walsh et al (2005) Schizophrenia CNVs (cases)</td>
<td>Up-regulated, CNV duplication</td>
<td>regulation of cell growth</td>
</tr>
<tr>
<td>DLG2</td>
<td>Walsh et al (2005) Schizophrenia CNVs (cases)</td>
<td>Down-regulated, CNV deletion</td>
<td>protein binding, sensory perception of pain, synaptic transmission</td>
</tr>
<tr>
<td>EMID2</td>
<td>Walsh et al (2005) Schizophrenia CNVs (cases)</td>
<td>Up-regulated, CNV duplication</td>
<td>phosphate transport, proteinaceous extracellular matrix</td>
</tr>
<tr>
<td>HEP1</td>
<td>Walsh et al (2005) Schizophrenia CNVs (cases)</td>
<td>Up-regulated, CNV duplication</td>
<td>Wnt receptor signalling pathway, DNA-dependent regulation of transcription</td>
</tr>
<tr>
<td>MLL5</td>
<td>Walsh et al (2005) Schizophrenia CNVs (cases)</td>
<td>Up-regulated, CNV duplication</td>
<td>metal ion binding, regulation of transcription</td>
</tr>
<tr>
<td>PPP1R3B</td>
<td>Walsh et al (2005) Schizophrenia CNVs (controls)</td>
<td>Up-regulated, CNV deletion</td>
<td>carbohydrate/glycogen metabolic process</td>
</tr>
<tr>
<td>SLC1A3</td>
<td>Walsh et al (2005) Schizophrenia CNVs (cases)</td>
<td>Down-regulated, CNV deletion</td>
<td>Dicarboxylic acid transport, cell projection, transmembrane transport</td>
</tr>
<tr>
<td>SLC26A4</td>
<td>Walsh et al (2005) Schizophrenia CNVs (cases)</td>
<td>Down-regulated, CNV duplication</td>
<td>ion transport, integral to membrane, sensory perception of sound</td>
</tr>
<tr>
<td>FLXNE2</td>
<td>Christoforou et al (unpublished) F27 family affected vs MIC</td>
<td>Down-regulated, Down-regulated</td>
<td>negative regulation of transcription</td>
</tr>
<tr>
<td>SORT1</td>
<td>Christoforou et al (unpublished) F22 family affected vs MIC</td>
<td>Down-regulated, Down-regulated</td>
<td>implicated in modulation of dopamine signals</td>
</tr>
</tbody>
</table>

Table 4.2: Overlap with previous studies. This table shows the overlaps between genes from the current microarray study and large genome wide studies previously published from individuals with schizophrenia and other major mental illness. All genes with overlap were found within the L100P non-drug treated (saline) group. Column 1 shows the gene name, column 2 the comparative previously published evidence, column 3 the direction of change (with this study first and comparative study second) and column 4 the gene function/location.
GOTree analysis of the differentially expressed gene list did not find any categories to be overenriched in the sample dataset (table 4.2, figure 4.2).

![Bar graph of GOTree categories for L100Padult vs C57BL6adult.](image)

**Figure 4.2: Bar graph of GOTree categories for L100Padult vs C57BL6adult.** For each biological process GOTM calculates the expected number of genes within a dataset and graphs this against the observed genes within each category. A p-value is then calculated to determine the significance of over/under-enrichment for each category. For this dataset none of the categories were significantly overenriched.

Pathway analysis of the L100P dysregulated genes using IPA gave three statistically significant high scoring networks. The top network for the L100P mutant adults was “neurological disorder” (IPA score 45. Scores greater than three have at least 99.9%
confidence that the genes are not grouped by random chance alone. For full score definition see Chapter 2 section 7.2) It contained 32 of the L100P dysregulated genes and had a high network score, with low probability these genes clustered by chance (fig 4.3). Calcium signalling was cited as the top canonical pathway. A canonical pathway is a core pathway established for a given molecule in the cell, in which molecular interactions occur in a linear and stepwise manner. Network 2 (IPA score 43) “cell cycle” contained 28 L100P dysregulated genes while network 3 (IPA score 41) “developmental disorder”, also contained 28 L100P dysregulated genes, including one gene, Bex2, whose expression level was corrected toward the wild-type level by rolipram treatment (Figure 4.3).
Figure 4.3: Ingenuity Pathway Analysis networks for L100P aduldt dysregulated genes. Those in green are downregulated, and those in red upregulated. Genes shaded grey do not appear in the dysregulated gene list.

Panel A shows the top L100P adult network 'neurological disorder', Panel B the second network 'cell cycle' and panel C the third network 'developmental disorder'. Panel D shows the canonical pathways for the L100P dysregulated gene list. The yellow 'ratio' line shows the ratio of genes in network to total number of genes in canonical pathway.
4.2.4 Differential Expression between the Q31L adult ENU mutant and the C57BL/6J adult controls

The differentially expressed genes in the Q31L non-drug treated (saline) adult mice are shown in Table 4.4. For this comparison, the top network contained only two genes from the array study. This was, however, expected as fewer genes were differentially expressed in this group. The small number of genes was insufficient to create canonical pathways.

<table>
<thead>
<tr>
<th>TargetID</th>
<th>Full Gene Name</th>
<th>Accession Number</th>
<th>GeneID</th>
<th>Q31L Sal</th>
<th>C57BL/6J Sal</th>
<th>Difference</th>
<th>Fold change</th>
<th>p-value</th>
<th>GO Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>WSP71</td>
<td>14-0294-0002</td>
<td>2000-0002</td>
<td>8.167</td>
<td>6.707</td>
<td>0.3705</td>
<td>1.037</td>
<td>0.0286</td>
<td>0.945</td>
<td>binding protein</td>
</tr>
<tr>
<td>T1C2F</td>
<td>transmembrane</td>
<td>2000-0002</td>
<td>9.217</td>
<td>3.180</td>
<td>0.7018</td>
<td>0.565</td>
<td>0.0006</td>
<td>0.945</td>
<td>transcription, translation, protein binding</td>
</tr>
<tr>
<td>NER</td>
<td>necdin</td>
<td>2000-0002</td>
<td>9.242</td>
<td>3.030</td>
<td>0.7018</td>
<td>0.565</td>
<td>0.0006</td>
<td>0.945</td>
<td>protein binding</td>
</tr>
<tr>
<td>DIFBH1</td>
<td>G94140269.11</td>
<td>2000-0002</td>
<td>8.265</td>
<td>6.535</td>
<td>0.3705</td>
<td>1.037</td>
<td>0.0286</td>
<td>0.945</td>
<td>differentially expressed, embryonic region</td>
</tr>
<tr>
<td>G230H4.4H21.1</td>
<td>2000-0002</td>
<td>10.093</td>
<td>8.702</td>
<td>6.160</td>
<td>0.3705</td>
<td>1.037</td>
<td>0.0286</td>
<td>0.945</td>
<td>N/A</td>
</tr>
<tr>
<td>E1C01</td>
<td>G10017790.1</td>
<td>1201</td>
<td>10.392</td>
<td>10.392</td>
<td>0.3705</td>
<td>1.037</td>
<td>0.0286</td>
<td>0.945</td>
<td>biological processes, cellular component</td>
</tr>
<tr>
<td>E1G01</td>
<td>G10017790.1</td>
<td>1201</td>
<td>10.392</td>
<td>10.392</td>
<td>0.3705</td>
<td>1.037</td>
<td>0.0286</td>
<td>0.945</td>
<td>N/A</td>
</tr>
<tr>
<td>G10017790.1</td>
<td>2000-0002</td>
<td>1201</td>
<td>10.392</td>
<td>10.392</td>
<td>0.3705</td>
<td>1.037</td>
<td>0.0286</td>
<td>0.945</td>
<td>N/A</td>
</tr>
<tr>
<td>E1G01</td>
<td>G10017790.1</td>
<td>1201</td>
<td>10.392</td>
<td>10.392</td>
<td>0.3705</td>
<td>1.037</td>
<td>0.0286</td>
<td>0.945</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 4.4: Dysregulated genes in the Q31L adult group compared to C57BL/6J controls. Table shows the probe target (gene) ID, full name, accession and probe number, log transformed expression data for both Q31L and C57BL/6J, difference in expression, fold-change and p-value, and Gene Ontology terms for each target as available.

4.2.5 Differential Expression between the Embryonic ENU mutant and the C57BL/6J embryonic controls

In total 33 genes were found to be differentially expressed in the L100P embryo, and 15 in the Q31L embryo (table 4.5).
### Table 4.5: Dysregulated genes in the Embryonic sample groups.

<table>
<thead>
<tr>
<th>Target ID</th>
<th>Full Gene Name</th>
<th>Accession Number</th>
<th>Protein ID</th>
<th>GI</th>
<th>Parameters</th>
<th>O/E (Mean)</th>
<th>O/E (SD)</th>
<th>Log</th>
<th>Fold change</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>L100P</td>
<td>Notch3</td>
<td>NM_001086</td>
<td>ENSP00000000000</td>
<td></td>
<td>NM_001086</td>
<td>1.029615</td>
<td>0.92349</td>
<td>0.8</td>
<td>1.298081</td>
<td>0.00072</td>
</tr>
<tr>
<td>Q31L</td>
<td>TGF-β1</td>
<td>NM_009692</td>
<td>ENSP00000000000</td>
<td></td>
<td>NM_009692</td>
<td>1.029615</td>
<td>0.92349</td>
<td>0.8</td>
<td>1.298081</td>
<td>0.00072</td>
</tr>
</tbody>
</table>

**Notes:**
- O/E (Mean): Overexpression (Mean) relative to control.
- O/E (SD): Overexpression (Standard Deviation) relative to control.
- Log: Log-transformed expression data.
- Fold change: Fold change relative to control.
- p-value: Statistical significance.

**Gene Ontology:**
- Gene Ontology terms were not available for the genes listed in this table.
For the L100P E13.5 embryos, there were nine genes in the top network (IPA score 22) and Notch signaling was the top canonical pathway. The Q31L E13.5 embryos had a lower but still significant top network (IPA score 10), with only four genes interconnecting. Three of these genes were also present in the L100P E13.5 embryo top network which added weight to taking them forward for further analysis. Consistent with this, the top canonical pathway for this group was also Notch signaling (figure 4.4).

**Figure 4.5: Ingenuity Pathway Analysis of Embryonic dysregulated gene lists.** Panels A and B show the top seven canonical pathways for L100P E13 and Q31L E13 respectively. In both cases Notch signaling is the top canonical pathway, and in the case of the L100P, the only one to meet the significance threshold. *Threshold* bar shows cut-off point of significance $P < 0.05$. The yellow ‘ratio’ line shows the ratio of genes in network to total number of genes in canonical pathway.

### 4.2.6 The effect of Drug Treatment on gene expression in adult ENU mutant mice

In total there were 150 genes dysregulated in the drug treatment groups compared to their drug naïve (saline) counterparts (table 4.6). While many of these genes were unique to the drug treatment groups, some had previously been seen to be differentially expressed in the non-drug treated (saline) groups compared to C57BL/6J animals. Treatment of the L100P adult with rolipram showed the most extensive gene expression rescue, with seven genes returning to levels not significantly different to wild-type (fold change +/-1.3,
p<0.05). Clozapine showed no correction of expression and the bupropion treated Q31L animals only had one gene show corrected expression.

<table>
<thead>
<tr>
<th>TargetID</th>
<th>Full Gene Name</th>
<th>Accession Number</th>
<th>ProbeID</th>
<th>GI number</th>
<th>UMIS/GoM</th>
<th>CMAP/GoM</th>
<th>Difference</th>
<th>Anti-log</th>
<th>Fold change</th>
<th>Pval</th>
<th>GO Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>THDMB</td>
<td>transaminase (acid) pyruvate</td>
<td>NM_028887</td>
<td>17048</td>
<td>193894204</td>
<td>11.0077</td>
<td>11.52681</td>
<td>-0.496520</td>
<td>0.794495</td>
<td>1.42644193</td>
<td>0.0000867</td>
<td>integral membrane protein</td>
</tr>
<tr>
<td>DDC</td>
<td>aspartate aminotransferase</td>
<td>NM_001593</td>
<td>619304</td>
<td>142262299</td>
<td>11.41932</td>
<td>11.90567</td>
<td>-0.496792</td>
<td>0.7962279</td>
<td>1.20264527</td>
<td>0.0009804</td>
<td>metabolic processing</td>
</tr>
<tr>
<td>MAO-C</td>
<td>monoamine oxidase, cytoplasmic</td>
<td>NM_023017748</td>
<td>420038</td>
<td>62945035</td>
<td>0.42091</td>
<td>0.1121111</td>
<td>-0.307913</td>
<td>1.2986095</td>
<td>1.23986847</td>
<td>0.0007497</td>
<td>negative regulation of transcription on DNA</td>
</tr>
</tbody>
</table>

Table 4.6.1: Dysregulated genes in Clozapine treated L100P adults. Table shows the genes that are dysregulated in L100P adult mice treated with anti-psychotic Clozapine, compared to L100P drug naïve adult mice. Table shows probe target ID (gene symbol), full gene name, accession number, Illumina probe ID, GI number, log transformed expression data, expression difference, anti-log of expression difference (to correct for negative values), fold-change, p-value and GeneOntology terms where available.
Table 4.6.2: Dysregulated genes in Rolipram treated L100P adults. Table shows the genes that are dysregulated in L100P adult mice treated with the PDE4 inhibitor and anti-depressant rolipram, compared to L100P drug naïve adult mice. Each table shows probe target ID (gene symbol), full gene name, accession number, Illumina probe ID, GI number, log transformed expression data, expression difference, anti-log of expression difference (to correct for negative values), fold-change, p-value and GeneOntology terms where available.
Table 4.6.3: Dysregulated genes in Bupropion treated Q31L adults. Continued on next page...
### Table 4.6.3: Dysregulated genes in Bupropion treated Q31L adults.

Table shows the genes that are dysregulated in Q31L adult mice treated with the anti-depressant Bupropion, compared to Q31L drug naïve adult mice. Each table shows probe target ID (gene symbol), full gene name, accession number, Illumina probe ID, GI number, log transformed expression data, expression difference, anti-log of expression difference (to correct for negative values), fold-change, p-value and GeneOntology terms where available.

<table>
<thead>
<tr>
<th>Target ID</th>
<th>Full gene name</th>
<th>Accession number</th>
<th>Illumina probe ID</th>
<th>GI number</th>
<th>log transformed expression</th>
<th>expression difference</th>
<th>anti-log expression difference</th>
<th>Fold change</th>
<th>p-value</th>
<th>GeneOntology terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSMUSG00000004886</td>
<td>NPY</td>
<td>ENSMUS000000000004886</td>
<td>ENSMUSP000000000004886</td>
<td>6370499</td>
<td>-0.06787</td>
<td>0.65</td>
<td>0.006</td>
<td>0.65</td>
<td>0.006</td>
<td>-1.54</td>
</tr>
<tr>
<td>ENSMUSG00000001065</td>
<td>Slc1a2</td>
<td>ENSMUS000000001065</td>
<td>ENSMUSP000000001065</td>
<td>6045224</td>
<td>0.95776</td>
<td>-0.95776</td>
<td>-0.95776</td>
<td>0.001</td>
<td>-0.95776</td>
<td>0.65</td>
</tr>
<tr>
<td>ENSMUSG00000003207</td>
<td>Bcl2</td>
<td>ENSMUS000000003207</td>
<td>ENSMUSP000000003207</td>
<td>4905662</td>
<td>0.93465</td>
<td>-0.93465</td>
<td>-0.93465</td>
<td>0.001</td>
<td>-0.93465</td>
<td>0.65</td>
</tr>
<tr>
<td>ENSMUSG00000006843</td>
<td>Fancd2</td>
<td>ENSMUS000000006843</td>
<td>ENSMUSP000000006843</td>
<td>4632819</td>
<td>0.95776</td>
<td>-0.95776</td>
<td>-0.95776</td>
<td>0.001</td>
<td>-0.95776</td>
<td>0.65</td>
</tr>
<tr>
<td>ENSMUSG00000009005</td>
<td>Tmem70a</td>
<td>ENSMUS000000009005</td>
<td>ENSMUSP000000009005</td>
<td>3776316</td>
<td>0.95776</td>
<td>-0.95776</td>
<td>-0.95776</td>
<td>0.001</td>
<td>-0.95776</td>
<td>0.65</td>
</tr>
<tr>
<td>ENSMUSG00000000540</td>
<td>Ctg2</td>
<td>ENSMUS00000000540</td>
<td>ENSMUSP00000000540</td>
<td>3518027</td>
<td>0.95776</td>
<td>-0.95776</td>
<td>-0.95776</td>
<td>0.001</td>
<td>-0.95776</td>
<td>0.65</td>
</tr>
<tr>
<td>ENSMUSG00000000871</td>
<td>Lcn2</td>
<td>ENSMUS00000000871</td>
<td>ENSMUSP00000000871</td>
<td>3453820</td>
<td>0.95776</td>
<td>-0.95776</td>
<td>-0.95776</td>
<td>0.001</td>
<td>-0.95776</td>
<td>0.65</td>
</tr>
<tr>
<td>ENSMUSG00000001618</td>
<td>Gfpm1</td>
<td>ENSMUS00000001618</td>
<td>ENSMUSP00000001618</td>
<td>3402913</td>
<td>0.95776</td>
<td>-0.95776</td>
<td>-0.95776</td>
<td>0.001</td>
<td>-0.95776</td>
<td>0.65</td>
</tr>
<tr>
<td>ENSMUSG00000002448</td>
<td>Atp1b1</td>
<td>ENSMUS00000002448</td>
<td>ENSMUSP00000002448</td>
<td>3363328</td>
<td>0.95776</td>
<td>-0.95776</td>
<td>-0.95776</td>
<td>0.001</td>
<td>-0.95776</td>
<td>0.65</td>
</tr>
<tr>
<td>ENSMUSG00000000862</td>
<td>Slc2a3</td>
<td>ENSMUS00000000862</td>
<td>ENSMUSP00000000862</td>
<td>3313731</td>
<td>0.95776</td>
<td>-0.95776</td>
<td>-0.95776</td>
<td>0.001</td>
<td>-0.95776</td>
<td>0.65</td>
</tr>
<tr>
<td>ENSMUSG00000002677</td>
<td>Cacna1d</td>
<td>ENSMUS00000002677</td>
<td>ENSMUSP00000002677</td>
<td>3274624</td>
<td>0.95776</td>
<td>-0.95776</td>
<td>-0.95776</td>
<td>0.001</td>
<td>-0.95776</td>
<td>0.65</td>
</tr>
<tr>
<td>ENSMUSG00000001614</td>
<td>Lcc</td>
<td>ENSMUS00000001614</td>
<td>ENSMUSP00000001614</td>
<td>3235527</td>
<td>0.95776</td>
<td>-0.95776</td>
<td>-0.95776</td>
<td>0.001</td>
<td>-0.95776</td>
<td>0.65</td>
</tr>
<tr>
<td>ENSMUSG00000002162</td>
<td>Srca2</td>
<td>ENSMUS00000002162</td>
<td>ENSMUSP00000002162</td>
<td>3196430</td>
<td>0.95776</td>
<td>-0.95776</td>
<td>-0.95776</td>
<td>0.001</td>
<td>-0.95776</td>
<td>0.65</td>
</tr>
<tr>
<td>ENSMUSG00000002807</td>
<td>Cav1.3</td>
<td>ENSMUS00000002807</td>
<td>ENSMUSP00000002807</td>
<td>3157333</td>
<td>0.95776</td>
<td>-0.95776</td>
<td>-0.95776</td>
<td>0.001</td>
<td>-0.95776</td>
<td>0.65</td>
</tr>
</tbody>
</table>

N/A
The potential importance of genes corrected by drug treatment was assessed by where they fell in the non-drug treated groups. For the L100P rolipram treated mice, Bex2 was sited in the 3rd network produced with a good network score and high probability that the grouping was not by chance. It also, interestingly, made indirect connections with NFKβ which has been implicated in major mental illness and forms a central connection in this network for many other genes. Pcbp1, Hnrpk and Wbp4 all appear in network 4 alongside Egr2, Egr3 and Egr4 which have been previously linked to major mental illness[238]. Egr2 and Wbp4 have a direct connection while Pcbp1 and Hnrpk connect to each other and an Actin complex (Figure 4.6)
Figure 4.5: Ingenuity Pathway analysis of untreated groups. Panel A shows the third highest L100P network with Bax2 (connected by red point). Panel B shows the fourth L100P network with PCEB1, HNRPA1 and HBF1 (all connected by red point) along with EGR2, 3 and 4, which are dysregulated in the L100P adult and have previously been associated with major mental illness.
4.3 Selection of Genes of Interest for Validation by QT RT PCR

The analysis outlined above resulted in a list of genes that were dysregulated in the Disc1 mutant mice compared to C57BL/6J wild-type controls, and genes which showed correction of expression in drug treated mutants. This list totalled 835 genes with some overlap between groups. Obviously it was not going to be possible to validate all of the genes highlighted so it was necessary to determine which genes most warranted follow up. This was done by scoring genes based on fold change, p-value, IPA analysis, over-enrichment of GO terms, presence in other groups from the array and overlaps with previously published studies on major mental illness. Genes were ranked within each pairwise comparison using a simple tally system, where each of the above selection criteria would add one tally to the gene ‘score’, and then the top 10% of genes were carried forward for validation. The list of genes taken forward for validation are shown in table 4.7.
Table 4.7: Genes for validation by qRT PCR. Columns 1 and 2 show the gene symbol and full name, column 3 shows the comparison in which it is dysregulated, columns 4 and 5 show probe ID and GI numbers for the gene, columns 7 and 8 show fold-change and p-value, column 9 gives GO terms, and column 10 gives further evidence to support validation.
4.4 Identification of L100P outlier

During the Taqman qRT-PCR, a conversation with a more experienced microarray statistician lead me to take a closer look at the normalised raw data for the L100P mutant mouse. As the number of genes classified as differentially expressed were very high it was thought that there may be an anomaly in the data skewing the result (Dr Freeman, personal communication).

I picked twenty genes at random through the microarray dataset and ran interquartile-range outlier analysis on the four L100P adult (saline) pools (3-4 mice per pool) which confirmed one pool, L100P Sal F2, as an outlier in 15 out of the 20 genes analysed. As there were 46,644 genes in the total dataset, it was not possible to test every gene to see whether this pool was an outlier therefore this pool was removed and the entire dataset reanalysed.

The qRT PCR validation that had been run so far (22 genes on each mouse sample) suggested that only one mouse in the pool was an outlier by inter-quartile range analysis. Of the outliers removed during analysis of the Taqman qRT-PCR data, 40% were mice from the L100P Sal F2 pool. Given that there were four L100P adult (saline) pools this is higher by ~15% than expected than random chance of 25%. 87.5% of the time the L100PSalF2 pool sample removed was B100-1-32 (14 times in 22 analyses). As the array data was run on the complete pool, all mice in the pool had to be removed from further analysis. The genes already validated by Taqman qRT-PCR would be reanalysed and will be described in chapter 5.

After identifying this pool I went back and looked at the behavioural data from the PrePulse Inhibition trial to determine if the mice in this pool had displayed behaviours which would be suggestive of different genetic expression (figure 4.7).
4.4.1 Reanalysis of array dataset

The identification of the outlier group necessitated complete re-analysis of the entire array dataset post the normalisation step (see chapter 3 section 7.2). Removal of the L100Psaline F2 pool left seven samples in the L100P saline group, which still gave 80% power to detect a 1.3 fold-change. Analysis was carried out as described initially in chapter 3, and chapter 4 section 2.

Removal of the L100Psaline pool altered the detection p-values for each probe and this gave rise to a completely new dataset for analysis. Of the 46,644 successful probes, 20,302 (43.5%) probes were expressed above background. This is slightly higher than expected [231]. As described in Chapter 3, the next filtering step was to remove all probes with less than 1.3 fold-change in the dataset. This reduced the number of probes in the L100Psaline set to 1272, in the Q31Lsaline group to 26, in the L100PEmbryo to
1344, in the Q31LEmbryo to 613, in the L100PRolipram to 456, in the L100PClozapine to 321 and in the Q31LBupropion to 262 (Table 4.8)

Finally, using the same methods as described in section 4.2, the probes that survived the fold-change filter were tested for differential expression using the Analysis of Variance model (ANOVA) (table 4.8).

<table>
<thead>
<tr>
<th>Group</th>
<th>Control</th>
<th>Number of Probes Expressed above Background</th>
<th>Number of Probes Surviving fold-change filter</th>
<th>Number of Probes Differentially Expressed Compared to Controls (% dataset)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L100P Saline</td>
<td>Q57SELEmbryo</td>
<td>19067</td>
<td>1272</td>
<td>35 (0.427)</td>
</tr>
<tr>
<td>Q57L Saline</td>
<td>Q57SELEmbryo</td>
<td>17700</td>
<td>26</td>
<td>19 (0.107)</td>
</tr>
<tr>
<td>L100P Embryo</td>
<td>Q57SELEmbryo</td>
<td>19075</td>
<td>1344</td>
<td>47 (0.262)</td>
</tr>
<tr>
<td>Q57L Embryo</td>
<td>Q57SELEmbryo</td>
<td>19097</td>
<td>615</td>
<td>26 (0.130)</td>
</tr>
<tr>
<td>L100P Rolipram</td>
<td>L100P Saline</td>
<td>17891</td>
<td>321</td>
<td>4 (0.103)</td>
</tr>
<tr>
<td>Q57L Bupropion</td>
<td>Q57L Saline</td>
<td>16871</td>
<td>262</td>
<td>18 (1.102)</td>
</tr>
</tbody>
</table>

Table 4.8 Summary of probes differentially expressed by genotype of drug treatment. Column 1 shows the group of interest and column 2 the control group for the comparison. Column 3 shows the number of genes expressed above background, column 4 shows the number of probes surviving the fold-change filter, and column 5 shows the number of probes, and percentage of those probes that were significantly differentially expressed (F s p-value ≤0.05).

In total 368 genes were found to be differentially expressed in this dataset (table 4.9). 66 of these genes were also present in the previous dataset, and overlap of only 18%.


Table 4.9: New list of dysregulated genes post outlier removal.

<table>
<thead>
<tr>
<th>Gene_ID</th>
<th>Protein ID</th>
<th>Differentially Fold change</th>
<th>Pool</th>
<th>Function (and site)</th>
<th>Localization</th>
<th>Feature Implications in Major Neurodegeneration/Disease Intervention</th>
<th>Downstream GOF/LoF</th>
<th>mRNA Pathways</th>
</tr>
</thead>
<tbody>
<tr>
<td>H2B</td>
<td>99552103</td>
<td>1.50647 ±0.16</td>
<td>0.036±0.01</td>
<td>protein binding</td>
<td>cytoplasmic complex</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HIF1A</td>
<td>1.004285</td>
<td>1.01457 ±0.03</td>
<td>0.000005 ±0.000005</td>
<td>protein binding, gsp accumulation</td>
<td>cytoplasmic concentrate</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SLC10A4</td>
<td>99552104</td>
<td>1.50647 ±0.16</td>
<td>0.036±0.01</td>
<td>protein binding, gsp accumulation</td>
<td>cytoplasmic concentrate</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CEBPA</td>
<td>99552105</td>
<td>1.50647 ±0.16</td>
<td>0.036±0.01</td>
<td>protein binding, gsp accumulation</td>
<td>cytoplasmic concentrate</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CRTAP</td>
<td>99552106</td>
<td>1.50647 ±0.16</td>
<td>0.036±0.01</td>
<td>protein binding, gsp accumulation</td>
<td>cytoplasmic concentrate</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TRPM2</td>
<td>99552107</td>
<td>1.50647 ±0.16</td>
<td>0.036±0.01</td>
<td>protein binding, gsp accumulation</td>
<td>cytoplasmic concentrate</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>LOC100803890</td>
<td>99552108</td>
<td>1.50647 ±0.16</td>
<td>0.036±0.01</td>
<td>protein binding, gsp accumulation</td>
<td>cytoplasmic concentrate</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 4.9: New list of dysregulated genes post outlier removal.
Continued next page…….
<table>
<thead>
<tr>
<th>Gene_ID</th>
<th>Protein</th>
<th>Function (and role)</th>
<th>Localization</th>
<th>Chromatin Interaction</th>
<th>Gene Ontology GO:Term</th>
<th>IPK Top Canonical Pathways</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUP175</td>
<td>25024093</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP133</td>
<td>25024108</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP214</td>
<td>25024113</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP153</td>
<td>25024128</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP358</td>
<td>25024133</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP374</td>
<td>25024148</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP375</td>
<td>25024153</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP84</td>
<td>25024168</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP85</td>
<td>25024173</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP133</td>
<td>25024188</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP214</td>
<td>25024193</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP153</td>
<td>25024208</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP358</td>
<td>25024213</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP374</td>
<td>25024228</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP375</td>
<td>25024233</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP84</td>
<td>25024248</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP85</td>
<td>25024253</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP133</td>
<td>25024268</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP214</td>
<td>25024273</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP153</td>
<td>25024288</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP358</td>
<td>25024293</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP374</td>
<td>25024308</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP375</td>
<td>25024313</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP84</td>
<td>25024328</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP85</td>
<td>25024333</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP133</td>
<td>25024348</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP214</td>
<td>25024353</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP153</td>
<td>25024368</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP358</td>
<td>25024373</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP374</td>
<td>25024388</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP375</td>
<td>25024393</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP84</td>
<td>25024408</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP85</td>
<td>25024413</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP133</td>
<td>25024428</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP214</td>
<td>25024433</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP153</td>
<td>25024448</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP358</td>
<td>25024453</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP374</td>
<td>25024468</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP375</td>
<td>25024473</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP84</td>
<td>25024488</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP85</td>
<td>25024493</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP133</td>
<td>25024508</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP214</td>
<td>25024513</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP153</td>
<td>25024528</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP358</td>
<td>25024533</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP374</td>
<td>25024548</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP375</td>
<td>25024553</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP84</td>
<td>25024568</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP85</td>
<td>25024573</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP133</td>
<td>25024588</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP214</td>
<td>25024593</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP153</td>
<td>25024608</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP358</td>
<td>25024613</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP374</td>
<td>25024628</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP375</td>
<td>25024633</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP84</td>
<td>25024648</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP85</td>
<td>25024653</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP133</td>
<td>25024668</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP214</td>
<td>25024673</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP153</td>
<td>25024688</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP358</td>
<td>25024693</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP374</td>
<td>25024708</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP375</td>
<td>25024713</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP84</td>
<td>25024728</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP85</td>
<td>25024733</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP133</td>
<td>25024748</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP214</td>
<td>25024753</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP153</td>
<td>25024768</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP358</td>
<td>25024773</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP374</td>
<td>25024788</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP375</td>
<td>25024793</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP84</td>
<td>25024808</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP85</td>
<td>25024813</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP133</td>
<td>25024828</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP214</td>
<td>25024833</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP153</td>
<td>25024848</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP358</td>
<td>25024853</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP374</td>
<td>25024868</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP375</td>
<td>25024873</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP84</td>
<td>25024888</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP85</td>
<td>25024893</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP133</td>
<td>25024908</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP214</td>
<td>25024913</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP153</td>
<td>25024928</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP358</td>
<td>25024933</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP374</td>
<td>25024948</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP375</td>
<td>25024953</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP84</td>
<td>25024968</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NUP85</td>
<td>25024973</td>
<td>nuclear transport</td>
<td>nucleus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
### Table 4.9: New list of dysregulated genes post outlier removal.

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Pearson ID</th>
<th>Compartments</th>
<th>Fold change</th>
<th>P-value</th>
<th>Function (and alias)</th>
<th>Localization</th>
<th>Previous Implication in Major Mental Disease (Prior to its Loss)</th>
<th>Overrepresentation GO-Test</th>
<th>IPA Top Canonical Pathways</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table Not Visible</td>
<td>Table Not Visible</td>
<td>Table Not Visible</td>
<td>Table Not Visible</td>
<td>Table Not Visible</td>
<td>Table Not Visible</td>
<td>Table Not Visible</td>
<td>Table Not Visible</td>
<td>Table Not Visible</td>
<td>Table Not Visible</td>
</tr>
</tbody>
</table>
Table 4.9: New list of dysregulated genes post outlier removal.

<table>
<thead>
<tr>
<th>Gene_ID</th>
<th>Gene Symbol</th>
<th>Function (and alias)</th>
<th>Localization</th>
<th>Protein Implications in Major Mental Disease</th>
<th>Overrepresentation GO:Term</th>
<th>IPA Top Canonical Pathway</th>
</tr>
</thead>
<tbody>
<tr>
<td>3X-01</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3X-02</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3X-03</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3X-04</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3X-05</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3X-06</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3X-07</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3X-08</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3X-09</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3X-10</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3X-11</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3X-12</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3X-13</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3X-14</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3X-15</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3X-16</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3X-17</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3X-18</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3X-19</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3X-20</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table continued on next page...
Table 4.9: New list of dysregulated genes post outlier removal.

<table>
<thead>
<tr>
<th>Gene_ID</th>
<th>Probe</th>
<th>Fold Change</th>
<th>P-Value</th>
<th>Function(s) and Others</th>
<th>Localization</th>
<th>Possible Implications in Human Disease</th>
<th>Overexpression GO Terms</th>
<th>IPA Top Canonical Pathways</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCC</td>
<td>3250</td>
<td>2.044E-07</td>
<td>0.000</td>
<td>transporter activity</td>
<td>integral to cognitive motoric in maintenance</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C20orf43</td>
<td>3250</td>
<td>2.044E-07</td>
<td>0.000</td>
<td>transporter activity</td>
<td>integral to cognitive motoric in maintenance</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PM001</td>
<td>3250</td>
<td>2.044E-07</td>
<td>0.000</td>
<td>transporter activity</td>
<td>integral to cognitive motoric in maintenance</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>LOC000069</td>
<td>3250</td>
<td>2.044E-07</td>
<td>0.000</td>
<td>transporter activity</td>
<td>integral to cognitive motoric in maintenance</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>LOC000069</td>
<td>3250</td>
<td>2.044E-07</td>
<td>0.000</td>
<td>transporter activity</td>
<td>integral to cognitive motoric in maintenance</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>EPF001</td>
<td>3250</td>
<td>2.044E-07</td>
<td>0.000</td>
<td>transporter activity</td>
<td>integral to cognitive motoric in maintenance</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>LOC000069</td>
<td>3250</td>
<td>2.044E-07</td>
<td>0.000</td>
<td>transporter activity</td>
<td>integral to cognitive motoric in maintenance</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>LOC000069</td>
<td>3250</td>
<td>2.044E-07</td>
<td>0.000</td>
<td>transporter activity</td>
<td>integral to cognitive motoric in maintenance</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>LOC000069</td>
<td>3250</td>
<td>2.044E-07</td>
<td>0.000</td>
<td>transporter activity</td>
<td>integral to cognitive motoric in maintenance</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Notes: Column 9 and 10 state whether the gene was overexpressed in GOTree analysis and/or IPA analysis.
Some probes were found to be differentially expressed in more than one comparison (figure 4.8)

![Venn diagram of significant probes in the genotype comparisons](image)

**Figure 4.8: Venn diagram of significant probes in the genotype comparisons.** Pane A shows significant probes in both adult comparisons while pane B shows significant probes in both embryonic comparisons. The total number of differentially expressed genes are shown. Within the circles of the Venn diagram are the number of genes unique to each comparison and, where the circles overlap, the number common to the comparisons.

In addition, there were eight genes that showed correction with drug treatment, and a further four genes that had borderline fold-change and p-value for correction with drug treatment. Bupropion treatment of the Q31L saline mouse corrected three genes back to the expression level of the wild type, clozapine treatment of the L100P mouse corrected one gene back to the expression level of the wild type, and rolipram treatment of the L100P mouse corrected five genes, and was close to correction of four genes back to a level not significantly different to that of the C57BL/6J mouse.
4.4.2 New Gene Ontology, Pathway analysis and Comparisons with previous study

Of the 88 genes shown to be differentially expressed in the L100PSaline group, 62 were upregulated and 26 were downregulated. GOTree analysis highlighted 11 biological processes over-enriched in the sample-set (table 4.10)

<table>
<thead>
<tr>
<th>Biological process</th>
<th>Observed</th>
<th>Expected</th>
<th>P-value</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>cell-cell signaling</td>
<td>5</td>
<td>0.61</td>
<td>0.00103</td>
<td>Eg2, Eg3, Sna2, Nmn, Syn2</td>
</tr>
<tr>
<td>transmission of nerve impulse</td>
<td>5</td>
<td>0.65</td>
<td>0.0000151</td>
<td>Eg2, Eg3, Sna2, Nmn, Syn2</td>
</tr>
<tr>
<td>synaptic transmission</td>
<td>4</td>
<td>0.48</td>
<td>0.0010831</td>
<td>Eg2, Sna2, Nmn, Syn2</td>
</tr>
<tr>
<td>neurotransmitter secretion</td>
<td>2</td>
<td>0.11</td>
<td>0.0056543</td>
<td>Nmn, Syn2</td>
</tr>
<tr>
<td>secretory pathway</td>
<td>7</td>
<td>0.64</td>
<td>0.0000013</td>
<td>Cplb2, Fodl1, Napb, Nmn, Exr2, Syn2, Snc24d</td>
</tr>
<tr>
<td>vesicle docking during synaptogenesis</td>
<td>2</td>
<td>0.05</td>
<td>0.001013</td>
<td>Cplb2, Exr2</td>
</tr>
<tr>
<td>regulated secretory pathway</td>
<td>3</td>
<td>0.13</td>
<td>0.0000234</td>
<td>Cplb2, Nmn, Syn2</td>
</tr>
<tr>
<td>vesicle-mediated transport</td>
<td>5</td>
<td>1.04</td>
<td>0.0034996</td>
<td>Cplb2, Sna1, Napb, Exr2, Snc24d</td>
</tr>
<tr>
<td>vesicle docking</td>
<td>2</td>
<td>0.05</td>
<td>0.0011376</td>
<td>Cplb2, Exr2</td>
</tr>
<tr>
<td>peripheral nervous system development</td>
<td>2</td>
<td>0.05</td>
<td>0.0011376</td>
<td>Eg2, Eg3</td>
</tr>
<tr>
<td>secretion</td>
<td>7</td>
<td>0.78</td>
<td>0.0000063</td>
<td>Cplb2, Fodl1, Napb, Nmn, Exr2, Syn2, Snc24d</td>
</tr>
</tbody>
</table>

Table 4.10: Table of biological processes over-enriched in the L100PSaline group. Column 1 shows the biological process, column 2 the number of genes in the dataset involved in the process, column 3 the number of genes expected in a dataset, column 4 the p-value of the finding, and column 5 the genes found.

None of the other groups showed overenrichment of GO categories. IPA analysis revealed a top network score of 48 “cellular development” for the L100P adult comparison, with 21 dysregulated genes appearing in the network (figure 4.9). The top canonical pathway for L100P was synaptic long-term depression (p= 0.005). There was now a significant canonical pathway for the Q31L analysis (p=0.01, figure 4.10 panel B), but no significant network of genes was identified. The L100P embryo pathway analysis changed very little, but the Q31L embryonic network analysis was greatly changed. For the L100P embryo the top network identified, “gene expression”, gave a score of 25, and for the Q31L embryo the top network “cell signalling” gave a score of 35. A score of 25 represents a 1.0E-24 probability of the genes within the network grouping by chance. While for the L100P embryo notch signalling remained the top canonical pathway (p=0.006), for the Q31L embryo this was replaced by caveolar-mediated endocytosis (p=0.001), pushing notch signalling into second spot (figure 4.10).
Figure 4.9: Ingenuity Pathway Analysis. Panel A shows top L100P adult network “cellular development”. Panel B shows the top L100P embryo network “gene expression” and Panel C the top Q31L embryo network “cell signalling”. Those in green are downregulated, and those in red upregulated. Genes shaded grey do not appear in the dysregulated gene, but are expressed above background in the dataset.
Figure 4.10: Canonical pathways for IPA analysis on new dysregulated gene list. Panel A shows the top six pathways in the L100P adult group, panel B the top six pathways in the Q31L adult group, panel C the top six pathways in the L100P embryonic group, and panel D the top six pathways in the Q31L embryonic group. Threshold bar shows cut-off point of significance $P < 0.05$. The yellow ‘ratio’ line shows the ratio of genes in network to total number of genes in canonical pathway.

There were also eight genes found to overlap with previous studies of major mental illness, with one gene being present in two of the studies included in the analysis (table 4.11). Probability distributions consider this a significant value ($p<0.001$). Studies used for comparison are described in section 4.2.3.
Table 4.11: Overlap with previous studies. This table shows the overlaps between genes from the current microarray study and large genome wide studies previously published from individuals with schizophrenia and other major mental illness. All genes with overlap were found within the L100P non-drug treated (saline) group. Column 1 shows the gene name, column 2 the comparative previously published evidence, column 3 the direction of change (with this study first and comparative study second) and column 4 the gene function/location.

4.5 New list of genes for qRT PCR
Genes were selected for validation by qRT PCR based on fold-change, p-value, GOTree analysis, previous implications in mental illness research and overlaps with current studies of the human (t1:11) translocation carriers (Table 4.12). Initially, genes were sorted based on p-value and then fold change, and scored by where they fell within the list (ie those in the top 10% based on p-value were given a score of 10, top 11-20% scored 9 etc). Genes were then scored on whether they appeared in the ‘over-enriched’ functions based on GOTree analysis and/or Ingenuity pathway analysis. Finally, genes were scored based on whether they had previous implications in major mental illness or overlapped with the human translocation carrier data. Scores were totalled and the top 10% of genes from each comparison (now based on the total combined score from p-value, fold change, GOTree analysis, IPA and overlaps with other studies) were carried forward for further validation.
Table 4.12: List of genes to be analysed by QT RT PCR. Column 1 gives the gene name, column 2 the comparison in which it is dysregulated, column 3 the fold change from the array, column 4 the p-value from the array, column 5 the GeneOntology terms for that gene, and column 6 extra evidence for further analysis.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Comparisons</th>
<th>Array FC</th>
<th>Array Pval</th>
<th>GO Terms</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arc</td>
<td>L100P/CS7BL6</td>
<td>-1.37</td>
<td>0.0105</td>
<td>synaptic binding post synaptic membrane</td>
<td>Over-enriched Gtoret category</td>
</tr>
<tr>
<td>Atp5b</td>
<td>L100P/CS7BL6</td>
<td>1.24</td>
<td>0.045</td>
<td>ATP synthase complex, coupling factor F0</td>
<td>Over-enriched Gtoret category</td>
</tr>
<tr>
<td>Bmi2</td>
<td>L100P/CS7BL6</td>
<td>-2.17</td>
<td>0.0004</td>
<td>apoptosis, cell cycle</td>
<td>PA pathway</td>
</tr>
<tr>
<td>Cdh11</td>
<td>L100P/CS7BL6</td>
<td>1.38</td>
<td>0.0489</td>
<td>cell adhesion</td>
<td>PA pathway</td>
</tr>
<tr>
<td>CORO1B</td>
<td>L100P/CS7BL6</td>
<td>-1.33</td>
<td>0.0036</td>
<td>actin binding</td>
<td>PA pathway</td>
</tr>
<tr>
<td>Cyrb</td>
<td>L100P/CS7BL6</td>
<td>1.84</td>
<td>0.0022</td>
<td>neurotransmitter transport</td>
<td>Over-enriched Gtoret category</td>
</tr>
<tr>
<td>Ddx11</td>
<td>L100P/CS7BL6</td>
<td>0.25</td>
<td>0.0042</td>
<td>mRNA processing</td>
<td>PA pathway</td>
</tr>
<tr>
<td>Ddx9</td>
<td>L100P/CS7BL6</td>
<td>-2.08</td>
<td>0.018</td>
<td>MAP kinase tyrosine/threonine kinase activity</td>
<td>Over-enriched Gtoret category, PA pathway, Dysregulated in 11:11 translocation family (unpublished)</td>
</tr>
<tr>
<td>Egr2</td>
<td>L100P/CS7BL6</td>
<td>-1.47</td>
<td>0.0036</td>
<td>brain segmentation, motor axon guidance, schwann cell differentiation</td>
<td>Over-enriched Gtoret category, PA pathway, Yarada et al 2006</td>
</tr>
<tr>
<td>Egr3</td>
<td>L100P/CS7BL6</td>
<td>-1.38</td>
<td>1.59E-06</td>
<td>neuromuscular synaptic transmission</td>
<td>Over-enriched Gtoret category, PA pathway, Yarada et al 2007</td>
</tr>
<tr>
<td>Egr4</td>
<td>L100P/CS7BL6</td>
<td>-1.36</td>
<td>0.0056</td>
<td>regulation of transcription, DNA dependent</td>
<td>Over-enriched Gtoret category, PA pathway, Yarada et al 2006</td>
</tr>
<tr>
<td>Eno4</td>
<td>L100P/CS7BL6</td>
<td>1.26</td>
<td>9.2E-05</td>
<td>necroptosis during necroptosis</td>
<td>PA pathway</td>
</tr>
<tr>
<td>FBDD1A</td>
<td>L100P/CS7BL6</td>
<td>1.03</td>
<td>0.0003</td>
<td>ubiquitin dependent protein degradation process</td>
<td>Carnengo et al 2006</td>
</tr>
<tr>
<td>Gpr95</td>
<td>CS11U/CS11Bl6</td>
<td>-1.56</td>
<td>0.0005</td>
<td>G-protein coupled receptor signalling pathway</td>
<td>-</td>
</tr>
<tr>
<td>Gpr1a</td>
<td>L100P/CS7BL6</td>
<td>1.3</td>
<td>0.0004</td>
<td>receptor internalisation, ion transport</td>
<td>Over-enriched Gtoret category, PA pathway</td>
</tr>
<tr>
<td>Gpr2a</td>
<td>L100P/CS7BL6</td>
<td>1.27</td>
<td>0.01</td>
<td>synaptic transmission</td>
<td>Over-enriched Gtoret category, PA pathway</td>
</tr>
<tr>
<td>Htr3a</td>
<td>L100P/CS7BL6</td>
<td>-1.19</td>
<td>0.0011</td>
<td>protein amino acid dephosphorylation</td>
<td>-</td>
</tr>
<tr>
<td>Htr3b</td>
<td>L100P/CS7BL6</td>
<td>-1.28</td>
<td>0.0018</td>
<td>chromatin modification, cell cycle arrest</td>
<td>Over-enriched Gtoret category, Walsh et al 2009</td>
</tr>
<tr>
<td>Htr3c</td>
<td>L100P/CS7BL6</td>
<td>-1.44</td>
<td>0.0019</td>
<td>protein binding</td>
<td>-</td>
</tr>
<tr>
<td>Nppb</td>
<td>L100P/CS7BL6</td>
<td>1.86</td>
<td>0.0019</td>
<td>axon extension and fasciculation</td>
<td>PA pathway</td>
</tr>
<tr>
<td>Ntr2l3</td>
<td>L100P/CS7BL6</td>
<td>1.25</td>
<td>0.0022</td>
<td>calcium channel regulator activity</td>
<td>Novak et al 2009, dysregulated in 11:11 translocation (unpublished)</td>
</tr>
<tr>
<td>Numb</td>
<td>L100P/CS7BL6</td>
<td>-1.52</td>
<td>0.0033</td>
<td>axonogenesis, forebrain development</td>
<td>-</td>
</tr>
<tr>
<td>Pains</td>
<td>L100P/CS7BL6</td>
<td>1.56</td>
<td>0.0002</td>
<td>kinase activity</td>
<td>Over-enriched Gtoret category, PA pathway, Stemmer et al 2000, Donley et al 1996, Potok et al 2007</td>
</tr>
<tr>
<td>Pdeo1a</td>
<td>L100P/CS7BL6</td>
<td>1.37</td>
<td>0.0002</td>
<td>calcium and calmodulin-regulated TSS-cyclin-BMP phosphorylation kinase activity</td>
<td>Over-enriched Gtoret category</td>
</tr>
<tr>
<td>Pdeo4c</td>
<td>L100P/CS7BL6</td>
<td>1.84</td>
<td>0.0009</td>
<td>cAMP phosphodiesterase activity</td>
<td>Carnengo et al 2009, Miller et al 2005</td>
</tr>
<tr>
<td>Pgf4</td>
<td>L100P/CS7BL6</td>
<td>1.7</td>
<td>0.005</td>
<td>lnRNA processing/degradation</td>
<td>-</td>
</tr>
<tr>
<td>Pparoc1</td>
<td>L100P/CS7BL6</td>
<td>1.28</td>
<td>0.0011</td>
<td>positive regulation of transcription from RNA polymerase II promoter</td>
<td>Over-enriched Gtoret category, PA pathway</td>
</tr>
<tr>
<td>Prkaa1</td>
<td>L100P/CS7BL6</td>
<td>-1.52</td>
<td>0.0011</td>
<td>cell differentiation, mitotic chromosome condensation</td>
<td>-</td>
</tr>
<tr>
<td>Prmt6</td>
<td>L100P/CS7BL6</td>
<td>1.26</td>
<td>0.0004</td>
<td>methyltransferase activity</td>
<td>PA pathway</td>
</tr>
<tr>
<td>Syn1</td>
<td>C57BL/6J</td>
<td>1.5</td>
<td>0.0004</td>
<td>neurotransmitter secretion</td>
<td>-</td>
</tr>
<tr>
<td>Syn2</td>
<td>L100P/CS7BL6</td>
<td>1.3</td>
<td>0.0002</td>
<td>neurotransmitter secretion</td>
<td>Over-enriched Gtoret category</td>
</tr>
<tr>
<td>Wdfy4</td>
<td>C57BL/6J</td>
<td>1.5</td>
<td>3.9E-10</td>
<td>metal binding</td>
<td>Over-enriched Gtoret category</td>
</tr>
</tbody>
</table>

Chapter 4  161
In total 36 genes were selected for further analysis, 20 from the non-drug treated adult mice, ten that showed correction, or near to correction, with drug treatment, and six from the embryonic stages. qRT PCR analysis of these genes will be described in the following chapter.

4.6 Conclusions
This chapter describes the analysis and significance of the results obtained from a genome-wide expression study as described in chapter 3. Genes that were found to be differentially expressed were weighted on their functional significance (based on gene ontology and pathway analysis) and previous implications in major mental illness research. Genes which showed correction by drug treatment were also considered.

The initial analysis revealed a total of 835 genes that were differentially expressed across the sample set. As the one outlier pool gave erratic and unpredictable expression values, being neither consistently higher nor lower than the other samples, this likely explains why the initial analysis did not identify this pool as an outlier, as the effect was muted by combining all genes together. Thus when the average expression for each pool was compared no pool was initially identified as an outlier. This shows the need to take comparable biological groups and analyse the raw data of a subset of genes to ensure outliers are not missed in future.

Secondary analysis revealed a total of 368 genes differentially expressed across the sample set. In hindsight, secondary analysis could have been avoided by checking a random subset of genes from the outset, rather than after the analysis had been carried out.

The comparison between the L100P adult and the C57BL/6J adult mice identified 88 genes that were differentially expressed, 62 of which were over-expressed and 24 under-expressed in the L100P adult samples. The Q31L bupropion treated mouse showed the
highest levels of differential expression in the drug treated groups with 91 genes over-expressed and 95 genes under-expressed compared to non-drug treated adults. Only three of these genes were also found to be dysregulated in the Q31L non-drug treated mouse and it is likely the high numbers are a factor of the drug treatment and not the Q31L genotype. The L100P rolipram treated mouse gave the most genes which showed correction under pharmacological influence. Of the 32 genes identified, seven showed correction of the L100P non-drug treated disregulation. Correction of expression was defined as returning that gene to a level not significantly different (FC+/- 1.3, p<0.05) to that of C57BL/6J controls.

In the embryonic groups, L100P identified 45 differentially expressed genes, while Q31L identified 25. Four genes were identified in both groups, PRM3, Coro1B, Numb and PDE4D.

Selection of genes for further analysis was performed in multiple ways. Genes were first ranked in order of fold-change and p-value then GOTree analysis was performed to determine which biological processes, cellular components and molecular functions were or were not overrepresented in the differentially expressed gene lists generated from the microarray analysis. This not only facilitates direct comparison with other studies of major mental illness due to the simple and uniform characterizations, but also helped to “compartmentalize” the data and identify promising genes for further analysis. The GOTree analysis revealed over-enrichment of genes involved in cell-cell signaling, transmission of nerve impulses and neurotransmitter secretion in the L100P non-drug treated (saline) adult dataset. Among these genes are the two Neurexins (*Nrxn1* and *Nrxn3*), *Gria1* and *Gria2*, *Erg2*, *Egr3* and *Egr4*. Both *neurexin* genes and *Gria1* and *Gria2* were found to be over-expressed in the L100P adult samples, while the three *Egr* genes were found to be under-expressed in the sample set.

Ingenuity pathway analysis was also used to identify dysregulated genes that grouped together to form established networks and pathways. Nine genes showed both overenrichment in GO categories and were present in the top statistically significant IPA
network for the relevant group. One of the potential pitfalls of Ingenuity analysis for this study is the relevance of known networks to major mental illness. Much of the data currently held refers to networks established by studies of cancer genetics and diseases with simple genetic components. While these networks currently provide good points of reference and identify potential genes for further analysis, many connections are indirect. With continuing genetic studies being added to the existing datasets this is constantly improving but for this study IPA was used purely for identifying potential genes for further analysis and not to make direct claims to gene networks involved in major mental illness at this stage. Because the Ingenuity database is constantly being updated with new datasets, it is a dynamic tool and justifies repeat analysis in the future to determine whether new networks are identified with the same experimental dataset.

Comparing the dataset from this study to selected previously published studies allowed me to not only identify genes for further analysis, but to determine how this study sat in relation to previous work, and to gauge the relevance of these two ENU mouse models to major mental illness research. Recent work has investigated the role of copy number variants (CNVs) in neurodevelopmental disorders (reviewed in Kirov 2010) [82]. As the aetiology of schizophrenia and other major mental illness is so varied, with multiple genetic components identified, the general consensus is that multiple small components are likely to be responsible, and as CNVs account for a substantial proportion of human genetic variation, are likely to play an important role. With this in mind I compared my dataset with three recently published studies of CNVs in major mental illness to determine whether they were comparable. Only two genes were present in both the dysregulated gene list from this study and copy number variants in the previous studies. *Mll5*, a transcription regulator, was present as a duplication in schizophrenia cases from Walsh 2008 [83], and was upregulated in the L100P non-drug treated adult group. MLL5 has not been reported to interact directly with DISC1 [132], but DISC1 is well established as interacting with other kinases and modulating signaling pathways. For example, DISC1 interacts with PDE4B to dynamically regulate cAMP signaling, and also interacts with the cAMP response element-dependent transcription factor, ATF4, possibly binding to chromatin remodelling factors, such as SMARCE1 [132]. PDE4s are
orthologous to the Drosophila *Dunce*, which is involved in learning and memory and known to affect synaptic plasticity, which in turn requires alteration of gene expression profiles [170]. These interactions with transcription factors, together with the fact that DISC1 localises to the nucleus [172, 173] is consistent with a role for DISC1 in transcriptional regulation. *Nrxn1* was also found to be up-regulated in the L100P non-drug treated adult group, with a CNV deletion present in a schizophrenia case and affected sibling [217]. The deletion spans the promoter and first exon of the gene and partially overlaps with previous deletions identified in autism [76] and mental retardation [239].

I also had access to data from recently completed microarray studies on the DISC1 translocation (t1:11) family that allowed an almost direct comparison to be made between the mouse and human model. Five genes (*Dusp6, F5, Hook3, Nrxn1* and *Nrxn3*) were present in both studies, with *Hook3, Nrxn1* and *Nrxn3* dysregulated in the same direction in both arrays. These overlapping results between the mouse and human sample subjects is promising as it provides further evidence of the validity of the animal model to this system. One major flaw of this comparison is the two arrays used different tissues as the primary RNA source. While it was possible to use brain tissues from the ENU mutant mouse, the human samples were from a blood-derived cell line. Blood-derived (lymphoblastoid) cell lines do not express all genes and only a sub-set of those expressed in the brain[240]. This could obviously result in some genes not being present in the human dataset by virtue of the original tissues used, and not because they are not expressed, or potentially dysregulated in the human condition.

This study had obvious limitations and weaknesses, not least of all the presence of an outlier group in the initial analysis. As described previously, the Illumina built in controls did not identify this one pool as an outlier, highlighting the need to assess a subset of genes raw data across a group for complete confidence. A number of the genes identified as differentially expressed on the array were from a cDNA bank, with little or no functional information available. This made it difficult to determine the relevance of these predicted sequences to major mental illness and as such most were not chosen for
further analysis. It is possible that the lack of information available for many genes biased the selection of the subset for further analysis. While the methods used for microarray analysis described in this section have been widely used, it is important to note that other methods of data analysis are equally popular and different statistic procedures will give different lists of genes. The outcome of the analysis is obviously dependent on the starting data-set, thus testing for differential expression and then filtering on fold-change (rather than filtering by fold-change then potential differential expression as described here) would potentially lead to a different variance distribution (as more probes would be used to estimate the variance) and a different test statistic. This would potentially generate a different gene list from the same data. This has previously been addressed [241] with the conclusion that care should be taken when interpreting any results that rely on the structure of the starting dataset. It may be suggested that the dataset be reanalyzed in the future using the various methods available and only the genes which are consistant through all analyses be retained for further follow-up work. Non-parametric tests, Baysian models and t-tests (as used in this study) are all used to generate lists of differentially expressed genes from normalized microarray datasets [242]. Furthermore, once a list of genes is generated, there are multiple methods of describing changes in expression and identifying functional significance. Heirarchical clustering methods group genes by expression levels, but are of little use in determining functional significance [243]. Pathway analysis (as used here) is useful as it groups genes by GOTree classifications and so can be useful in identifying common functions that may be affected by gene expression changes [243]. Neither of these methods, however, address the issue of false-positives and erroneous artifacts, which may lead to genes being studied further with no viable results. Bertsch et al (2005) propose to narrow the list of genes using Convergent Functional Genomics (CFG), which uses Baysian methods to cross-validate datasets between arrays and linkage data, with the aim of reducing uncertainty and removing false positives [244].

The multiple stage selection process was employed to select the potentially most interesting and viable genes for further analysis, based on current knowledge of gene ontology, gene pathways and previous genetic work on schizophrenia and major
depression. However, there are multiple genes identified in the initial analysis which will not be investigated further at this stage. Ideally all genes identified would be validated using qRT-PCR but constraints on finances and time meant it was necessary to prioritise the most interesting and potentially viable genes for further study.

In summary, this study has identified several putative candidate genes in both embryonic and adult stages. Correction of gene expression through drug administration has also been shown and is worthy of further investigation.
Chapter 5
Quantitative real-time PCR analysis of candidate differentially expressed genes in the L100P and Q31L mutant mice
5. Quantitative real-time PCR analysis of candidate differentially expressed genes in the L100P and Q31L mutant mice

5.1 Introduction

For the past decade researchers have been using qRT-PCR as a robust and reliable measure of gene expression through levels of RNA species [245]. It has widely become the standard technique for validating results obtained by microarray analysis [246] and the recent commercial availability of standard fluorescent tagged probes negates the need for researchers to design multiple probes. This makes confirmation of microarray results quicker and easier. qRT-PCR does have its limitations. One of the biggest limitations of qRT-PCR for gene expression studies is the necessity to use RNA which is then reverse transcribed to cDNA. RNA isolation must be performed with the utmost care to minimise the risk of contaminants (such as genomic DNA) and maintain the integrity of the RNA itself (reviewed in [247]). Design and testing of primers must also be stringent to minimise the likelihood of non-specific binding. Compounds found within tissues may also inhibit PCR. A number of measures were used in this study to minimise error in the PCR reactions. All RNA samples were tested prior to reverse transcription to identify any loss of integrity or possible contamination (chapter 2). RNA was treated to remove genomic contamination and primers chosen which spanned exon boundaries, so should not bind to genomic DNA in the sample. Primers used were all commercially available and were assumed to have been stringently tested, however standard curves were obtained for all primers to ensure compatibility with the samples. While these measures were used to minimise error it is important to note that these limitations do exist and data should be viewed accordingly.

In this chapter I will present the results from qRT-PCR analysis of 36 genes that were identified as differentially expressed in the microarray study of the ENU mutant mice. It should be noted that this is only a subset (~10%) of the genes identified as being differentially expressed. These genes were chosen for further study based on the criteria outlines in chapter 4. I will also describe the reanalysis of 20 genes that were tested prior to outlier removal described in chapter 4, and analysis of 3 drug treatment genes which
failed to validate in the drug naïve (saline) L100P mice, but which showed persistent
differential expression in the drug treated group.

5.2 Validating the Disc1 microarray result
As described in Chapter 4, prior to the discovery of the outlier group qRT-PCR had been
carried out with some genes showing differential expression in the same direction seen in
the microarray experiment. When the array data was re-analysed these genes would not
have been selected for validation analysis (selection criteria outlined in chapter 4 section
6), and in most cases did not make the criteria for differential expression (FC +/- 1.3,
p<0.05). None-the-less, as these genes had already been tested the data was re-analysed
with the outlier pool removed, and any genes which continued to show differential
expression on qRT-PCR were tested on a second batch of mouse samples, alongside the
new gene list. Two batches of mouse RNA were used for qRT-PCR validation. Batch 1
was the RNA that went on the array, and batch 2 an independent sample set from a
colony derived from the original animals.

5.2.1 Initial Validations and Outlier removal of the L100P drug naïve adult
microarray
Twenty genes identified as differentially expressed in the comparison between the L100P
drug naïve (saline) animals and C57BL/6J animals had been tested by qRT-PCR prior to
the identification of the outlier pool (described in Chapter 4 section 4). When the 20 qRT-
PCR data-sets were reanalysed following outlier removal, five of the 20 genes remained
significant by qRT-PCR (FC +/- 1.3, p<0.05) (Table 5.1). Sort1 did not exceed the fold-
change cut-off imposed after the microarray, however, as this gene had previously been
implicated in major mental illness it was tested in a second batch of mouse RNA
alongside the five genes where the qRT-PCR result remained valid after outlier removal.
Of these genes Bex2 was also still differentially expressed at the array level. All genes
tested from the initial analysis are shown in table 5.1.
Nrgn1, Litaf, Idl and Syn1 had significant fold-change and p-values after outlier removal, however in the opposite direction to the microarray. It is already known that between platform consistency decreases significantly with different primers or increased sequence distance [246]. These genes have multiple transcripts which arise through alternate splicing which may account for the discrepancy in the results between platforms (www.ensembl.org).

Twenty one genes that showed differential gene array expression in the L100P drug naïve saline-condition, adult mouse, post outlier removal were tested by qRT-PCR in both the RNA that was hybridised to the array, and a second batch of RNA samples. The six genes previously tested in the initial validation effort were also tested in a second batch of RNA and are displayed in table 5.1.
Table 5.1: Table of genes differentially expressed in the L100P drug naïve adult that were tested by qRT-PCR. Column 1 gives gene name, column 2 and 3 the array fold-change and p-value to be validated, columns 4 and 5 the first round qRT-PCR fold-change, and p-value for Mann-Whitney U test. Columns 6 and 7 give these values for the second round of qRT-PCR on independent RNA samples. Significant p-values are denoted in red text for both qRT-PCR rounds. Genes in the bottom half of the table below the thick dividing line were carried over from the original analysis. Genes not tested in the second round are denoted with '-' in those columns.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Array FC</th>
<th>Array Pval</th>
<th>Taqman Fold change</th>
<th>Taqman MannWhitney</th>
<th>Taqman Fold change</th>
<th>Taqman MannWhitney</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nmri1</td>
<td>1.35</td>
<td>1.00E-03</td>
<td>1.34</td>
<td>0.041</td>
<td>1.29</td>
<td>0.020</td>
</tr>
<tr>
<td>Wdyl</td>
<td>1.46</td>
<td>2.05E-10</td>
<td>2.52</td>
<td>0.010</td>
<td>2.45</td>
<td>0.0022</td>
</tr>
<tr>
<td>Egfl4</td>
<td>-1.36</td>
<td>5.00E-03</td>
<td>-2.29</td>
<td>0.030</td>
<td>-2.95</td>
<td>0.0210</td>
</tr>
<tr>
<td>Nppa</td>
<td>1.39</td>
<td>4.07E-05</td>
<td>1.43</td>
<td>0.010</td>
<td>1.35</td>
<td>0.0325</td>
</tr>
<tr>
<td>Nnr1d3</td>
<td>1.25</td>
<td>2.13E-02</td>
<td>1.61</td>
<td>0.049</td>
<td>1.25</td>
<td>0.0270</td>
</tr>
<tr>
<td>Prx1</td>
<td>1.18</td>
<td>1.12E-01</td>
<td>1.43</td>
<td>0.020</td>
<td>1.20</td>
<td>0.0630</td>
</tr>
<tr>
<td>Evol1A</td>
<td>1.25</td>
<td>9.25E-05</td>
<td>1.64</td>
<td>0.072</td>
<td>1.16</td>
<td>0.1170</td>
</tr>
<tr>
<td>Gja1</td>
<td>1.30</td>
<td>4.00E-04</td>
<td>1.09</td>
<td>0.072</td>
<td>1.05</td>
<td>0.3690</td>
</tr>
<tr>
<td>s100b6</td>
<td>1.61</td>
<td>1.63E-07</td>
<td>1.97</td>
<td>0.117</td>
<td>1.63</td>
<td>0.0140</td>
</tr>
<tr>
<td>Mig</td>
<td>1.29</td>
<td>2.48E-03</td>
<td>1.64</td>
<td>0.009</td>
<td>1.11</td>
<td>0.0000</td>
</tr>
<tr>
<td>s100a8</td>
<td>1.28</td>
<td>2.35E-05</td>
<td>4.45</td>
<td>0.015</td>
<td>1.59</td>
<td>0.0000</td>
</tr>
<tr>
<td>Pde4a3</td>
<td>1.27</td>
<td>2.24E-03</td>
<td>1.23</td>
<td>0.100</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Arc</td>
<td>-1.35</td>
<td>1.00E-02</td>
<td>-1.07</td>
<td>0.140</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gria2</td>
<td>1.27</td>
<td>1.00E-02</td>
<td>1.14</td>
<td>0.157</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Nappb</td>
<td>1.29</td>
<td>6.32E-04</td>
<td>1.13</td>
<td>0.250</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Foxo41</td>
<td>1.33</td>
<td>2.92E-03</td>
<td>-1.18</td>
<td>0.270</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Dusp6</td>
<td>-1.28</td>
<td>1.80E-02</td>
<td>-1.01</td>
<td>0.300</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Egf3</td>
<td>-1.26</td>
<td>1.09E-05</td>
<td>-1.11</td>
<td>0.305</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Egf2</td>
<td>-1.26</td>
<td>1.09E-05</td>
<td>-1.31</td>
<td>0.377</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Smp2</td>
<td>1.30</td>
<td>2.00E-03</td>
<td>-1.01</td>
<td>0.409</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cplx</td>
<td>1.41</td>
<td>1.00E-02</td>
<td>-1.02</td>
<td>0.473</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sort1</td>
<td>-1.05</td>
<td>0.3965</td>
<td>-1.12</td>
<td>0.0029</td>
<td>-2.72</td>
<td>0.0410</td>
</tr>
<tr>
<td>Sh2ai3</td>
<td>-1.05</td>
<td>0.2543</td>
<td>-1.49</td>
<td>0.05</td>
<td>-2.84</td>
<td>0.0410</td>
</tr>
<tr>
<td>Syp</td>
<td>-1.11</td>
<td>0.0264</td>
<td>-1.36</td>
<td>0.0084</td>
<td>-1.39</td>
<td>0.0792</td>
</tr>
<tr>
<td>Ube3c</td>
<td>-1.06</td>
<td>0.3708</td>
<td>-1.29</td>
<td>0.055</td>
<td>-1.37</td>
<td>0.2454</td>
</tr>
<tr>
<td>Gsf1</td>
<td>-1.14</td>
<td>0.1363</td>
<td>2.04</td>
<td>0.0032</td>
<td>-1.39</td>
<td>0.0030</td>
</tr>
<tr>
<td>Smn1</td>
<td>-1.09</td>
<td>0.2266</td>
<td>2.281</td>
<td>0.0056</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PlkB</td>
<td>1.03</td>
<td>0.1734</td>
<td>19.21</td>
<td>0.0083</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ifil</td>
<td>1.06</td>
<td>0.1517</td>
<td>-1.29</td>
<td>0.0223</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Nnr1</td>
<td>-1.17</td>
<td>0.1361</td>
<td>2.61</td>
<td>0.028</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lifr</td>
<td>1.06</td>
<td>0.2759</td>
<td>-2.13</td>
<td>0.0327</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pebp1</td>
<td>1.15</td>
<td>0.1227</td>
<td>-3.02</td>
<td>0.0592</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Carnk2a</td>
<td>-1.12</td>
<td>0.0501</td>
<td>2.901</td>
<td>0.061</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gstn1</td>
<td>-1.08</td>
<td>0.1571</td>
<td>-1.36</td>
<td>0.0891</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Fhnb2</td>
<td>1.01</td>
<td>0.7463</td>
<td>-2.96</td>
<td>0.1129</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Dig2</td>
<td>-1.09</td>
<td>0.0419</td>
<td>1.129</td>
<td>0.164</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gp1ba</td>
<td>-1.19</td>
<td>0.0056</td>
<td>1.23</td>
<td>0.186</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mgat5</td>
<td>-1.04</td>
<td>0.3609</td>
<td>1.17</td>
<td>0.045</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gh10</td>
<td>1.10</td>
<td>0.2532</td>
<td>1.18</td>
<td>0.3916</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Enal2</td>
<td>1.01</td>
<td>0.3445</td>
<td>-1.04</td>
<td>0.4555</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Apo1</td>
<td>-1.04</td>
<td>0.5778</td>
<td>2.43</td>
<td>0.48</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Four genes (*wdfy1, Nrxn1, Nrxn3 and Egr4*) from the post outlier removal data set validated robustly through both rounds of analysis. *Rapgef5* failed to meet the fold-change cut-off in the second round of qRT-PCR, even though it gave a significant p-value in statistical testing. Of the genes carried forward from the pre-outlier removal list, *Sort1* and *Shank3* both gave significant valid results in the second round of qRT-PCR analysis.

### 5.2.2 Validation of the Q31L drug naïve adult microarray

Only two genes were selected from the Q31L drug naïve array for validation by qRT-PCR. *Wdfy1* and *Ndn* were both up-regulated in the Q31L adult compared to C57BL/6J wild-type controls. Differential expression of both genes was validated in the first qRT-PCR analysis, however both failed to replicate when tested in a second batch of independent RNAs using wild-type littermates as controls (table 5.2).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Array FC</th>
<th>Array Pval</th>
<th>Batch 1 Taqman Fold change</th>
<th>Batch 1 MannWhitney</th>
<th>Batch 2 Taqman Fold change</th>
<th>Batch 2 MannWhitney</th>
</tr>
</thead>
<tbody>
<tr>
<td>WDFY1</td>
<td>1.5</td>
<td>4.9E-11</td>
<td>2.75</td>
<td>0.0016</td>
<td>1.12</td>
<td>0.116</td>
</tr>
<tr>
<td>NDN</td>
<td>1.36</td>
<td>2.00E-02</td>
<td>1.9</td>
<td>0.0002</td>
<td>1.03</td>
<td>0.433</td>
</tr>
</tbody>
</table>

**Table 5.2**: Table of genes differentially expressed in the Q31L drug naïve adult that were tested by qRT-PCR. Column 1 gives gene name, column 2 and 3 the array fold-change and p-value to be validated, columns 4 and 5 the first round qRT-PCR fold-change, and p-value for Mann-Whitney U test. Columns 6 and 7 give these values for the second round of qRT-PCR on independent RNA samples. Significant p-values are denoted in red text for both qRT-PCR rounds.

A further three genes that showed correction with drug treatment were also tested in the Q31L mouse and will be discussed in section 5.2.4.
5.2.3 Validation of the L100P and Q31L embryonic microarray

Seven genes were analysed using qRT-PCR for the embryonic array samples. Three of these genes, \textit{Pde4d}, \textit{Coro1b} and \textit{Numb} were differentially expressed in both the L100P and Q31L embryonic samples in the array analysis (table 5.3).

![Table 5.3: Table of genes differentially expressed in the L100P and Q31L embryo that were tested by qRT-PCR.](image)

Of the seven genes tested, four met the criteria for significant differential expression in qRT-PCR analysis of the RNA that was hybridised to the array. \textit{Prm3} and \textit{Cdhl1} also validated in the second round of analysis in the L100P embryonic samples. \textit{Pde4d} validated in the first batch in the L100P embryo but failed to replicate in the independent sample, and did not validate in the Q31L embryonic samples at either stage of analysis. \textit{Coro1b} and \textit{Numb} unfortunately gave significant results in the opposite direction to those previously reported in the microarray analysis in both the L100P and Q31L embryonic samples. This negated them from further analysis at this stage.

5.2.4 Validation of the L100P and Q31L drug treatment microarray

As previously mentioned, there were a number of genes that were differentially expressed in the adult samples that showed correction of expression under drug treatment conditions. Correction of expression was defined as differential gene expression (FC +/-
1.3, p<0.05) in the drug treated group when compared to the homozygous mutant, but not when compared to the wild-type control. Of these, eight genes were corrected in the L100P Rolipram treated group and three in the Q31L Bupropion treated group. One gene corrected by Rolipram treatment was also corrected by Clozapine treatment according to the microarray result, although this did not validate by qRT-PCR. These genes were not tested as part of the L100P adult analysis as in some cases they were slightly below the fold-change cut-off, although became interesting for analysis when compared to the drug treated individuals. These eleven genes were analysed by qRT-PCR (Table 5.4) in both treatment groups (drug naïve vs wild-type control, and drug naïve vs drug treated) through two rounds where the first batch were consistent with the microarray data.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Comparison</th>
<th>Array FC</th>
<th>Array Pval</th>
<th>Batch 1 Taqman Fold change</th>
<th>Batch 1 Taqman MannWhitney</th>
<th>Batch 2 Taqman Fold change</th>
<th>Batch 2 Taqman MannWhitney</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aqp5A</td>
<td>L00P/L100P Rolipram</td>
<td>-1.37</td>
<td>0.04</td>
<td>-2.27</td>
<td>0.0002</td>
<td>-2.59</td>
<td>0.003</td>
</tr>
<tr>
<td>Bex2</td>
<td>L00P/L100P Rolipram</td>
<td>-1.3</td>
<td>0.0003</td>
<td>-3.45</td>
<td>0.0019</td>
<td>-4.11</td>
<td>0.015</td>
</tr>
<tr>
<td>Bex2</td>
<td>L00P/L100P Rolipram</td>
<td>1.22</td>
<td>0.06</td>
<td>2.66</td>
<td>0.0064</td>
<td>-1.99</td>
<td>0.013</td>
</tr>
<tr>
<td>Dmp2A</td>
<td>L00P/L100P Rolipram</td>
<td>-1.05</td>
<td>0.0141</td>
<td>-1.26</td>
<td>0.033</td>
<td>-1.04</td>
<td>0.413</td>
</tr>
<tr>
<td>Dmp2A</td>
<td>L00P/L100P Rolipram</td>
<td>1.52</td>
<td>0.00451</td>
<td>2.46</td>
<td>0.0006</td>
<td>2.66</td>
<td>0.009</td>
</tr>
<tr>
<td>Mtd3</td>
<td>L00P/L100P Rolipram</td>
<td>-1.19</td>
<td>0.001</td>
<td>-1.3</td>
<td>0.12</td>
<td>-1.07</td>
<td>0.095</td>
</tr>
<tr>
<td>Mtd3</td>
<td>L00P/L100P Rolipram</td>
<td>1.26</td>
<td>0.03</td>
<td>1.49</td>
<td>0.05</td>
<td>2.49</td>
<td>0.048</td>
</tr>
<tr>
<td>Ndtb</td>
<td>L00P/L100P Rolipram</td>
<td>1.45</td>
<td>0.009</td>
<td>1.91</td>
<td>0.0041</td>
<td>1.06</td>
<td>0.329</td>
</tr>
<tr>
<td>Ndtb</td>
<td>L00P/L100P Rolipram</td>
<td>-1.72</td>
<td>0.015</td>
<td>-3.02</td>
<td>0.0006</td>
<td>-5.11</td>
<td>0.005</td>
</tr>
<tr>
<td>Fak3</td>
<td>L00P/L100P Rolipram</td>
<td>1.36</td>
<td>0.0002</td>
<td>1.03</td>
<td>0.033</td>
<td>1.15</td>
<td>0.05</td>
</tr>
<tr>
<td>Fak3</td>
<td>L00P/L100P Rolipram</td>
<td>-1.96</td>
<td>0.048</td>
<td>-1.74</td>
<td>0.0003</td>
<td>-1.66</td>
<td>0.033</td>
</tr>
<tr>
<td>Pop4</td>
<td>L00P/L100P Rolipram</td>
<td>1.62</td>
<td>0.035</td>
<td>1.08</td>
<td>0.006</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pop4</td>
<td>L00P/L100P Rolipram</td>
<td>1.17</td>
<td>0.06</td>
<td>-1.14</td>
<td>0.007</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pop4</td>
<td>L00P/L100P Rolipram</td>
<td>1.73</td>
<td>0.05</td>
<td>1.02</td>
<td>0.42</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ranbp9</td>
<td>L00P/L100P Rolipram</td>
<td>1.14</td>
<td>0.07</td>
<td>-1.15</td>
<td>0.172</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ranbp9</td>
<td>L00P/L100P Rolipram</td>
<td>-1.93</td>
<td>0.04</td>
<td>-3.1</td>
<td>0.0057</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PRO15010Rik</td>
<td>L00P/L100P Rolipram</td>
<td>1.66</td>
<td>0.03</td>
<td>1.11</td>
<td>0.21</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PRO15010Rik</td>
<td>L00P/L100P Rolipram</td>
<td>1.66</td>
<td>0.045</td>
<td>-1.45</td>
<td>0.15</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gpr55A</td>
<td>L00P/L100P Rolipram</td>
<td>1.28</td>
<td>0.003</td>
<td>2.98</td>
<td>0.0043</td>
<td>-2.56</td>
<td>0.005</td>
</tr>
<tr>
<td>Gpr55A</td>
<td>L00P/L100P Rolipram</td>
<td>-1.52</td>
<td>0.03</td>
<td>-1.82</td>
<td>0.1</td>
<td>9.8</td>
<td>0.003</td>
</tr>
<tr>
<td>Syn1</td>
<td>L00P/L100P Rolipram</td>
<td>-1.15</td>
<td>0.04</td>
<td>2.7</td>
<td>&lt;0.0001</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Syn1</td>
<td>L00P/L100P Rolipram</td>
<td>1.32</td>
<td>0.004</td>
<td>1.65</td>
<td>0.0003</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 5.4: Table of genes differentially expressed in the L100P and Q31L drug treated adults that were tested by qRT-PCR. Column 1 gives gene name, column 2 comparison, column 3 and 4 the array fold-change and p-value to be validated, columns 5 and 6 the first round RT-qPCR fold-change, and p-value for Mann-Whitney U test. Columns 7 and 8 give these values for the second round of qRT-PCR on independent RNA samples. Significant p-values are denoted in red text for both qRT-PCR rounds. Genes not tested in the second round are denoted with '-' in the appropriate columns.

Ranbp9, Pop4, 1500015O10Rik and Syn1 gave results in the opposite direction to the array negating them from further study. More concerning was the result from Bex2 and
Gpr88, which replicated in the first round analysis but gave significant fold-changes in the opposite direction in the second round. As this was the same platform, probe incompatibility can be ruled out in these cases suggesting there was another factor involved. qRT-PCR was repeated for both these genes in the second batch of samples to check for experimental error and the result was found to be the same in both cases.

Pak3 gave a consistent result, with higher p-values and lower fold change in the L100P drug naïve comparison across both rounds of testing. Mast3, Atp5b, Ndfip and Dusp1 only replicated in the L100P drug naïve vs L100P Rolipram treated in the batch 2 analysis. To determine if this change in expression was due to genotype or drug treatment I drug treated eight C57BL/6J adult mice and ran the qRT-PCR analysis of these four genes on RNA samples from the C57BL/6J Rolipram treated and drug naïve mice (Table 5.5).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Comparison</th>
<th>Batch 2 - L100P vs L100P Rolipram</th>
<th>Batch 2 - C57BL6 vs C57BL6 Rolipram</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mast3</td>
<td>C57BL6/C57BL6 Rolipram</td>
<td>Taqman Fold change 2.49</td>
<td>Taqman MannWhitney 0.048</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Taqman Fold change 1.23</td>
<td>Taqman MannWhitney 0.433</td>
</tr>
<tr>
<td>Atp5b</td>
<td>C57BL6/C57BL6 Rolipram</td>
<td>Taqman Fold change -2.68</td>
<td>Taqman MannWhitney 0.0043</td>
</tr>
<tr>
<td>Ndfip</td>
<td>C57BL6/C57BL6 Rolipram</td>
<td>Taqman Fold change -1.51</td>
<td>Taqman MannWhitney 0.0037</td>
</tr>
<tr>
<td>Dusp1</td>
<td>C57BL6/C57BL6 Rolipram</td>
<td>Taqman Fold change 2.56</td>
<td>Taqman MannWhitney 0.0012</td>
</tr>
</tbody>
</table>

Table 5.5: Table of genes differentially expressed in the C57BL/6J drug treated adults that were tested by qRT-PCR. Column 1 gives gene name, column 2 comparison, columns 3 and 4 the second round qRT-PCR fold-change, and p-value for Mann-Whitney U test from the L100P drug naïve vs L100P Rolipram treatment. Columns 5 and 6 give these values for the second round of qRT-PCR on independent RNA samples from C57BL/6J mice treated with the same rolipram dose compared to drug naïve C57BL/6J mice. Significant p-values are denoted in red text for both qRT-PCR rounds.

Mast3 did not show significant differential expression in the C57BL/6J drug treated mice suggesting a possible effect of genotype on drug action. Atp5b, Ndfip and Dusp1 all gave significant results in the C57BL/6J Rolipram treated group which would suggest that the results seen in the L100P Rolipram treated mice were more a factor of drug treatment than genotype. However, the Atp5b differential expression in the C57BL/6J comparison is in the opposite direction to the L100P comparison, which may suggest some effect of
genotype on the drug action. This would require further study to determine if this is the case.

5.3 Conclusions
Of the 40 genes tested by qRT-PCR only seven gave robust and reliable results through both rounds of validation. Of these, four were from the L100P drug naïve adult group (\textit{Wdfy1}, \textit{Nrxn1}, \textit{Nrxn3} and \textit{Egr4}), two from the L100P embryonic group (\textit{Cdhl1}, \textit{Prm3}) and one from the L100P Rolipram treated group which showed correction (\textit{Pak3}). One (\textit{Rapgef5}) gave p-value<0.05 but did not reach the 1.3 fold-change cut-off implemented in earlier analysis. Of the five genes carried over from the pre-outlier removal gene list, two validated through both rounds of RT-qPCR with significant fold-change and p-value to meet the criteria imposed. While microarrays are powerful molecular biological tools, they have technical aspects which can produce results that erroneously under or over-represent specific genes. False negativity can result from low expression levels, inefficient priming of specific mRNAs resulting in transcript drop out, poor adhesion of DNA to the slide, and splice variants with sequences not included on the array. Conversely, sources of false positivity include repetitive nucleotide elements, sequence homology between functionally different transcripts and high background levels due to nonspecific binding of nucleotides [248]. This not only justifies the need to validate results through other platforms, but may help to explain why two genes that were not seen to be differentially expressed on the array platform post outlier removal, may be seen to be differentially expressed on the qRT-PCR platform.

The neurexin family of proteins function as cell adhesion molecules and receptors. Neurexins 1-3 utilize two alternate promoters and include numerous alternatively spliced exons to generate thousands of distinct mRNA transcripts and protein isoforms. The majority of transcripts are produced from the upstream promoter and encode alpha-neurexin isoforms; a much smaller number of transcripts are produced from the downstream promoter and encode beta-neurexin isoforms. The numerous isoforms of
Neurexin 1 (NRXN1) is involved in maintenance of synaptic junctions and mediate intracellular signalling[249]. By making heterophillic connections with neuroligands it leads to maturation and differentiation of GABAergic and glutamatergic synapses through bi-directional signalling [250]. NRXN1 was implicated by Walsh et al (2008) [217] in an analysis of copy number variants in schizophrenia. Other studies on copy number variants have also implicated NRXN1 in schizophrenia pathology. Need et al (2009) [251] found deletions in the 3’ end of NRXN1 in patients with schizophrenia, but not in control groups. Data from Vrijenhoek et al (2008) [252] suggest that CNVs that affect the first few exons of NRXN1 confer greater risk of major mental illness, while that from Rujescu et al (2009) [253] implies that deletions that affect exons directly increase susceptibility. In addition, a missense mutation in exon 1 of human NRXN1 has also been linked to autism [254]. Neurexin 3 (NRXN3) is a membrane protein involved in cell adhesion, synaptic transmission and neurotransmitter secretion [249]. SNPs in NRXN3 have been associated with alcohol and nicotine dependence and linked with opiate dependence [255, 256].

The Early Growth Response Factors (EGRs) 2, 3 and 4 are synaptic activity inducible immediate early genes and all show nominal association with schizophrenia in a Japanese population. They activate transcription of genes required for mitogenesis and differentiation. EGR1, 2 and 3 are downregulated in the prefrontal cortex of brains from patients with schizophrenia [238], EGR4 expression was too low to measure disregulation. Treatment with the atypical antipsychotic and antidepressant Aripiprazole increases expression of Egr4 in rat frontal cortex[257].

Little is known about the function of WD repeat and FYVE domain-containing protein 1 (WDFY1 or FENS1) but it is known to be ribosomal and has been associated with alcohol consumption and preference [258]. Primary functional analysis identifies Wdfy1 as being involved in endosome trafficking [259].

Eight genes from the initial analysis showed correction of the L100P non-drug treated disregulation. Only one of these validated through qRT-PCR. Serine/threonine-protein
kinase 3 (PAK3), is involved in the Erb signalling pathway, axon guidance and focal adhesion. PAK proteins form an activated complex with GTP-bound RAS-like (P21), CDC2 and RAC1 proteins which then catalyzes a variety of targets. It is thought this protein may be necessary for dendritic development and for the rapid cytoskeletal reorganization in dendritic spines associated with synaptic plasticity. Mutations in PAK3 have been associated with X-linked nonsyndromic mental retardation [260-262] and suppression of PAK3 results in formation of abnormally elongated dendritic spines and a reduction of mature synapses [263]. These elongated spines fail to express post-synaptic densities or contact presynaptic terminals resulting in reduced spontaneous activity and defective LTP. Pak3 knockout mice have deficiencies in learning and memory and abnormalities in synaptic plasticity along with a reduction of transcription factor cAMP-responsive element-binding protein (CREB) suggesting a novel signalling mechanism with PAK3 and Rho signalling regulating synaptic function and cognition [264]. PAK genes also interact with the known Disc1 interactor Kalirin-7 which is thought to control multiple aspects of spinal plasticity [265].

Cadherin 11 (Cdh11) is a type 2 classical cadherin involved in mediation of calcium dependant cell adhesion. Type II (atypical) cadherins are defined based on their lack of a HAV cell adhesion recognition sequence, that is specific to type I cadherins. It comprises of a large N-terminal domain (extracellular), a single membrane spanning domain, and a small intracellular C-terminal domain. It was first discovered in rodent brain samples [266] with high expression in the developing olfactory system at E13-E17 [267] and has since been linked to multiple cancer types [268, 269].

The remaining three genes that have validated in the rolipram treated L100P adults but have not validated in the L100P adult are Atp5b, Ndfip and Dusp1. Two of these genes (Atp5b and Dusp1) are involved in phosphate metabolism and were classed as over-enriched in the sample set by GOtree analysis. This could be interesting on two counts; the action of Rolipram may target phosphate metabolism genes, which may or may not be significant in schizophrenia. Or phosphate metabolism is a key pathway in schizophrenia. When the list of genes differentially expressed in only the L100P adult group was run
through Gotree analysis the only over enriched genes were involved in synaptic plasticity, not phosphate metabolism, suggesting this is more a function of the drug treatment than the mouse phenotype. To confirm this I tested these genes in a C57BL6 drug naïve group vs C57BL6 rolipram treated group as this would allow me to determine if the effect was due to the drug treatment or the mutation. *Dusp1* was significantly upregulated in the C57BL6 rolipram treated group compared to the drug naïve group, as it had been in the L100P rolipram treated group compared to the L100P drug naïve group. *Ndfip* also behaved in the same way in the C57BL6 rolipram treated group as it had in the L100P drug treated group, being significantly down regulated compared to the drug naïve control. *Atp5b* was significantly down regulated in the C57BL6 rolipram treated group where it had been up regulated in the L100P rolipram treated group, but a significant difference in expression was observed suggesting it is an effect of the drug treatment as opposed to the mutation. As the changes in expression in these three genes appear to be an effect of the rolipram treatment rather than the mutation I feel they do not fit the criteria for further analysis within this project.

Finally *Sortilin* (*Sort1*) and *Shank3* which were carried over from the original gene list, validated in the second sample set. *Sortilin* (alias neurotensin receptor 3) has been implicated in the modulation of dopamine signals [270] and was found to be downregulated in a Scottish family with a history of bipolar disorder (Christoforou et al, manuscript in preparation). *Shank3* is a structural post-synaptic density protein which has previously been implicated in mental retardation [271]. Mutation of a single *Shank3* allele has also been found in Autism spectrum disorders [272].

As described previously, some genes tested had fold-changes in opposite directions across the two platforms. This has been documented previously [246] and is thought to be due to the platform probes amplifying different isoforms of the gene in question. While every effort was made to ensure overlap of the two probes, a perfect match could not be ascertained. In all cases both probes were mapped to the same exon but as the sequences of the qRT-PCR probes are not made available identical sequence alignment could not be achieved. A more concerning observation was that *Bex2* and *Gpr88*, both genes that
displayed correction through drug treatment, gave opposite fold-change directions between sample batches. As both batches were run on the same platform, this cannot be attributed to probe amplification discrepancy. Differences in the wild-type expression could be due to minor genetic differences between the C57BL/6J and the L100P wild-type. Although the L100P animals had been backcrossed for over 10 generations it is unlikely, but not impossible, that there were still some genetic differences.

The difference in the drug treated groups is more difficult to explain. It could be speculated that differences in environmental conditions (such as feed, housing etc) can affect the effect of pharmacologicals. As the batches of RNA were collected from animals bred in different facilities (though of the same background strain) this may account for the differences seen here. Also, the animals used in the first batch underwent behavioural testing just prior to being culled. It may be that this would have an effect on gene expression and thus drug efficacy.

Additional research would ideally focus on testing the other dysregulated genes present from the microarray study, which was outwith the scope of this project. As Disc1 has been shown to be developmentally regulated [134], establishing a developmental profile for both the Disc1 mutant, and the genes dysregulated by the mutation, would be an advantageous addition to the current data and is described in chapter 6. Protein analysis of those genes that showed robust disregulation is a logical next step for this project and is described in chapter 7.
Chapter 6

Establishing a developmental profile for the genes of interest in L100P mutant mice and wild type controls
6. Establishing a developmental profile for the genes of interest in L100P mutant mice and wild type controls

Since the 1940’s and 50’s scientists have considered the possibility that schizophrenia could be a developmental disease. In 1986 Weinberger proposed the developmental hypothesis of schizophrenia which suggested that small neurodevelopmental deficits in key circuits could lie dormant until puberty, when normal molecular changes could facilitate the onset of disease in affected individuals [189]. With this in mind I decided to establish a developmental profile of each of the validated genes of interest in the L100P mouse model. As Disc1 has been shown to be developmentally regulated [134] the aim was to determine if a mutation to Disc1 would alter the developmental profile and/or the genes found to be dysregulated in the mutants by the microarray analysis.

6.1 The Disc1 developmental profile

While collecting samples for the Disc1 microarray I had also collected a developmental profile panel from both the L100P and Q31L mice, to determine whether Disc1 expression was altered during development. A Disc1 developmental profile was previously published by Shurov et al (2004) [134] describing peaks in Disc1 protein expression at E13 and P35, corresponding to periods of neurogenesis in the developing whole brain and puberty in the adult mouse. The study used cortex from the adult mouse and whole brain from mouse embryos to determine protein expression through western blotting. However, Shurov et al did not specify the mouse strain used in their study. As each inbred strain is genetically different, I decided to establish a developmental profile for both ENU lines and wild-type counterparts to be used in this study, to determine whether the mutations affected the expression of Disc1 through development.

Mice were taken at eight stages throughout development from both ENU mutant lines, their wild-type littermates and C57BL/6J animals which were used as controls in the microarray experiment. Whole brain was used for embryonic samples and hippocampus from mature mice matching the stages used in the array experiment. There were six mice at each stage, for each line, giving a total of 240 samples to be run in triplicate. I used a
“pan” Disc1 probe, designed to hybridise at exons 4 and 5, thus covering all Disc1 isoforms bar one (the extreme short isoform Es: contains an alternatively spliced DISC1 exon 1a and terminates transcription two intronic codons after exon 3) to establish a developmental profile in these animals (Figure 6.1).

Figure 6.1 Developmental Profile of Disc1 in the mouse brain. Graph A shows the expression profile of Disc1 in the brain of L100P wild type littermates, L100P mutant and C57BL/6J mouse lines through 8 stages of development. Graph B show the expression profile of Disc1 in the brain of Q31L wild type littermates, Q31L mutant and C57BL/6J mouse lines through 8 stages of development.
I performed a one-way analysis of variance with a post-hoc Tuckey test to determine where the strains differed significantly in *Disc1* expression across the panel tested. There was a significant difference across the whole dataset (F=41.65, df=34, p<0.0001) with within stage significant differences arising at E15, E18 and P1.

With regard to *Disc1* expression, the homozygous L100P mutant is significantly different from the L100P wild type littermate and the C57BL/6J (p<0.001) at E13. At E18, the L100P mutant and the L100P wild-type are not significantly different than each other, but are significantly different (p<0.001) to the C57BL/6J. At P1, the L100P mutant has significantly higher expression than both the C57BL/6J and the L100P wild-type controls. As reported previously (Clapcote et al 2007) [197], there was no difference in *Disc1* expression levels in the adult mouse.

The homozygous Q31L mutant has significantly higher *Disc1* expression (p<0.001) than the C57BL/6J animals. At E18, the Q31L wild-type differs significantly from the Q31L homozygous mutant and the C57BL/6J (p<0.05). Perhaps the greatest difference was observed at P1, where the C57BL6 mouse is significantly different to all other groups (p<0.001) and the Q31L animals have significantly higher expression than their wild-type controls.

It can be speculated that the different developmental expression patterns of the Q31L and L100P mutant may contribute to the different phenotypes observed in behavioural testing. This result gives rise to two important questions; what neurobiological processes are occurring at these stages where Disc1 expression is altered and was the C57BL6 a suitable control animal for the microarray experiment given the results?

A review of the literature on mouse developmental stages gave very few results but one review article can shed some light on the processes involved at stages E15-P1 where the Disc1 expression differences have been observed. Matsuki et al (2005) [223] carried out a microarray experiment to determine which clusters of genes were most highly expressed at stages E12, E15, E18 and P1. Their data suggests that, based on genes with known
neural functions, the functional gene clusters most highly expressed at E15 are involved in neural determination and differentiation, and at E18 and P1 are involved in synapse formation and function, and survival and growth. Based on all genes differentially expressed at E18 and P1 the top functions are intra/inter cellular molecular transport, and signal transduction.

From the results obtained from the developmental panel I would suggest that the C57BL/6J mouse was an adequate control at the stages (E13 and adult) used in this study as their expression of Disc1 did not differ significantly from the wild-type littermates at these stages. However, the results suggest that either the ENU lines were not suitably backcrossed for the C57BL/6J to be an ideal control, or there was another mutation present in the ENU mice which was retained throughout the backcrosses in the ‘wild-type’ control animals. For this reason the remainder of the developmental panels were run with both ‘control’ groups for reference, although only the L100P wild-type animals would be used for comparisons.

6.2 Establishing developmental profiles for the genes of interest
As all genes that were successfully validated came from the L100P comparisons, the genes were only tested in the L100P developmental panel. Due to constraints on time and finances I pooled the samples at each developmental stage resulting in a total of 24 samples to be run in triplicate, compared to 144 if left unpooled.

6.2.1 The Disc1 developmental profile on pooled samples
As I had pooled the samples for subsequent gene panels I decided to run the Disc1 panel on pooled samples to check it was comparable to the previous result. Figure 6.2 shows the pooled samples compared to unpooled samples across all stages.
While absolute expression values in the pooled and unpooled samples were not identical, the trend of expression was the same and there was no statistically significant difference between the datasets (p>0.05) suggesting little effect of pooling. As a new dilution of calibrator DNA was used for the pooled samples, the absolute expression values are graphed as normalized to both calibrator samples. These data indicate that pooled samples should give comparable data to the individual samples for subsequent developmental panels.
When comparing the L100P mutant and wild-type littermates over the developmental panel, what is perhaps most evident, is the appearance of a shift in the peak of Disc1 expression during development. The L100P wild type expression values peak at E18 and the drop to level out after birth. The L100P mutant however does not show a peak in expression until P1, with a shallower decline to adulthood, with the two lines converging by P20. This delay in the L100P mutant mouse could have detrimental effects on neurodevelopment in the mutant animals. Personal observations and those of other researchers working with these animals show that the mutant individuals are physically smaller when born, which could signify a delayed development (Clapcote et al. personal communication). While the result of one gene expression profile is not enough to stipulate cause and affect it is worth noting that low birth weights have previously been associated with schizophrenia [273] and a latent shift in the expression of key developmental proteins could be a factor in this. Hikida et al [193] also note that there is a 3month delay in neuronal migration in a Disc1 mouse model, which could potentially be attributed to a change in Disc1 expression during development. While this chapter will only describe the developmental profile results of genes of interest from the microarray study I believe this altered Disc1 expression during development is worthy of further investigation in the future.
6.2.2 Developmental Profile for \textit{Egr4}

As mentioned in chapter 5, \textit{Egr4} was the gene with the largest fold change in the sample set and a good gene for further analysis due to its nominal association with schizophrenia in a Japanese association study [238].

As with the pooled \textit{Disc1} panel, the \textit{Egr4} probe was run against a pooled DNA panel across seven developmental stages from E13 to adult (figure 6.4). The \textit{Disc1} panel is being shown for comparison.

![Graph A: Developmental expression patterns of Disc1 in the brain of both the L100P wild-type and L100P mutant mouse.](image1)

![Graph B.1: Developmental expression pattern of EGR4 in the brain of the L100P wild-type, L100P mutant and C57BL/6J mouse as a bar chart and B.2 as a line graph.](image2)

\textbf{Figure 6.4: Developmental profile of EGR4 in cDNA pools from the L100P mutant mouse.} Graph A shows the developmental expression patterns of Disc1 in the brain of both the L100P wild-type and L100P mutant mouse. Graph B.1 shows the developmental expression pattern of EGR4 in the brain of the L100P wild-type, L100P mutant and C57BL/6J mouse as a bar chart and B.2 as a line graph. In all cases the X axis shows the developmental stage, and the Y axis the arbitrary expression value from Taqman RT-qPCR analysis.

The Taqman RT-qPCR analysis could not detect \textit{Egr4} in stages E13-E18. Levels of expression in all lines peaked at P20 with significant differences in expression between the L100P wild-type and L100P mutant samples at P7 (p=0.029) and P20 (p=0.0007), as well as the adult stage which was reported in chapter 5 (p=0.0465).
6.2.3 Developmental Profile for \textit{Nrxn1} and \textit{Nrxn3}

As mentioned previously, NRXN1 is involved in maintenance of synaptic junctions and is a mediator of intracellular signalling. A missense mutation in exon 1 of human \textit{NRXN1} has been linked to autism [254] and many studies have noted CNV’s in \textit{NRXN1} and \textit{NRXN3} linked to schizophrenia [83, 217, 251-253]. NRXN3 is a membrane protein involved in cell adhesion, synaptic transmission and neurotransmitter secretion. SNPs in \textit{NRXN3} have been associated with alcohol and nicotine dependence and linked with opiate dependence [255, 256]. The developmental panel for both \textit{Nrxn1} and \textit{Nrxn3} are shown in Figure 6.5.

![Figure 6.5: Developmental profile of Nrxn1 and Nrxn3 in cDNA pools from the L100P mutant mouse.](image)

Graph A shows the developmental expression patterns of Disc1 in the brain of both the L100P wild-type and L100P mutant mouse. Graph B.1 show the developmental expression pattern of Nrxn1 in the brain of the L100P wild-type, L100P mutant and C57BL/6J mouse as a bar chart and B.2 as a line graph. Graph C.1 shows the developmental expression pattern of Nrxn3 in the brain of the L100P wild-type, L100P mutant and C57BL/6J mouse as a bar graph and C.2 as a line graph. In all cases the X axis shows the developmental stage, and the Y axis the arbitrary expression value from Taqman RT-qPCR analysis.
The most obvious observation from both panels is the inverted trend of expression from E15 to P1 in both lines. That is, the L100P wild-type and L100P mutant show almost exact opposites in trend of expression at these stages. For \( Nrxn1 \) there is a peak in expression at E18 in the wild-type lines, while this stage represents a dip in expression in the mutant lines (\( Nrxn1 \) \( p = 0.01 \)). \( Nrxn1 \) expression in the L100P mutant peaks at P7 with significantly different expression to the wild-types (\( p = 0.029 \)) and shows significant over-expression in adulthood (\( p = 0.048 \)), confirming the previously reported result from chapter 5. \( Nrxn3 \) is significantly overexpressed at E15 (\( p = 0.026 \)), P7 (\( p = 0.0036 \)) and adulthood (\( p = 0.04 \)) and is significantly lower in the L100P mutants at E18 (\( p = 0.016 \)). The mutant lines show two peaks, at E15 and P7. At E15 neurogenesis occurring in the developing mouse brain and at P7 neurite outgrowth, myelination, synaptic pruning and apoptosis occur [274]. Given the later peak of \( Disc1 \) expression reported in the L100P mutants, and lower birth weights observed, it is possible there is a developmental delay in synaptogenesis in these animals but this cannot be confirmed without further physiological examination. As the predicted developmental function at E18 is the formation of synapses, the dip in expression of genes involved in synaptic maintenance and transmission in the mutant animals is also of key importance.

### 6.2.4 Developmental Profile of Pak3

As mentioned in chapter 5, Pak3 is involved in the Erb signalling pathway, axon guidance and focal adhesion. Mutations in \( PAK3 \) have been associated with X-linked nonsyndromic mental retardation [260, 263]. \( Pak3 \) knockout mice have deficiencies in learning and memory, and abnormalities in synaptic plasticity. They also show reduction of transcription factor cAMP-responsive element-binding protein suggesting a novel signalling mechanism with \( PAK3 \) and Rho, regulating synaptic function and cognition [264]. \( PAK \) genes also interact with the known DISC1 interactor Kalirin-7 which is thought to control multiple aspects of synaptic plasticity [265]. It is also the only drug corrected gene which gave robust and consistent results through all stages of validation.
Figure 6.6: Developmental profile of Pak3 in cDNA pools from the L100P mutant mouse. Graph A shows the developmental expression patterns of Disc1 in the brain of both the L100P wild-type and L100P mutant mouse. Graph B.1 show the developmental expression pattern of PAK3 in the brain of the L100P wild-type, L100P mutant and C57BL/6J mouse as a bar chart and graph B.2 as a line graph. In all cases the X axis shows the developmental stage, and the Y axis the arbitrary expression value from Taqman RT-qPCR analysis.

Expression of Pak3 in the L100P wild type littermate rises from E13 to E18. At E18 there is a significant difference between the expression levels of the L100P wild type and the L100P mutant (p = 0.049) as expression in the L100P mutant line starts to decline after E15. There is also a significant difference in expression between the wild-type and the mutants at P1 (p=0.0093). The L100P mutant peaks at E15 and falls away steadily to level out at P20 to adulthood. This is the only profile where the peak in expression of the mutants precedes that of the wild-type, although the level of the peak in the mutants is considerably lower than that of the wild-type, and no difference in expression value is observed at the E15 developmental stage. The previously reported over-expression of Pak3 in the adult mutant mice is not replicated here.
6.2.5 Developmental Profile of Cdh11

Cdh11 is a type II classical cadherin that mediates calcium dependant cell-cell adhesion. It was originally found to be differentially expressed in the embryonic samples run on the microarray. Enrichment of Cdh11 in future subplate neurons [275] has been linked to neuron extension and guidance in the developing mouse brain.

Throughout the developmental panel there is no significant difference in Cdh11 expression between the mouse lines, except at E13 as reported in chapter 5. At E13 the expression of Cdh11 is significantly higher (p=0.044) in the L100P mutant than in the wild type littermate. An increase in expression levels of Cdh11 at E13 may suggest excessive neurite extension in the developing brain of the L100P mutant mouse which would affect the positioning and effectiveness of neural connections in the developed animal. While the expression profile indicates enrichment at P7 this was not statistically significant (p=0.08).

Figure 6.7: Developmental profile of Cdh11 in cDNA pools from the L100P mutant mouse. Graph A shows the developmental expression patterns of Disc1 in the brain of both the L100P wild-type and L100P mutant mouse. Graph B.1 show the developmental expression pattern of Cdh11 in the brain of the L100P wild-type, L100P mutant and C57BL/6J mouse as a bar chart and B.2 as a line graph. In all cases the X axis shows the developmental stage, and the Y axis the arbitrary expression value from Taqman RT-qPCR analysis.
6.2.6 Developmental Profile of Sort1

Sort1 is one of the genes which was discarded after the identification of the outlier mouse during the microarray analysis. It did, however, validate as being dysregulated during Taqman RT-qPCR analysis (both pre and post outlier removal). For this reason, and because it was a gene of interest because it is dysregulated in a Scottish family with a history of mental illness, the developmental panel was run for Sort1 also (figure 6.8).

![Figure 6.8: Developmental profile of Sort1 in cDNA pools from the L100P mutant mouse. Graph A shows the developmental expression patterns of Disc1 in the brain of both the L100P wild-type and L100P mutant mouse. Graph B.1 show the developmental expression pattern of Sort1 in the brain of the L100P wild-type, L100P mutant and C57BL/6J mouse as a bar chart and B.2 as a line graph. In all cases the X axis shows the developmental stage, and the Y axis the arbitrary expression value from Taqman RT-qPCR analysis.

The L100P wild type showed peaks in Sort1 expression at E18 and P20 with the L100P mutant showing peaks at E15 and P20. Statistical analysis of the expression levels at each time point gave significant results at E18 (p=0.046) and adulthood (p=0.048). In both cases the expression of Sort1 was reduced in the L100P mutant animals.
6.3 Conclusions from the developmental profiles

The developmental profile of Disc1 highlighted some key stages of development where expression is altered in the mutant mouse line, and also confirmed the previous result that expression was not significantly different at either E13 or adulthood between the L100P mutant and wild type controls. The disparity between the developmental profile of Disc1 described in this chapter and that previously published by Shurov [134] highlights the importance of strain differences during development, and had this experiment been carried out prior to the microarray study it is likely a different developmental stage would have been chosen for the whole genome gene expression experiment.

The time shift observed in the peak of Disc1 expression could signify a delay in neural development of the L100P mutant, although this cannot be confirmed without further physiological analysis. Further weight is added to this suggestion by a similar time shift in the peak expression of both the neurexin genes and Pak3 (summary figure 6.9).
The lack of definable expression of $Egr4$ in the embryonic stages is of great interest and has not been previously reported. Its significantly different levels of expression at P20, just prior to when the animal is effectively going through puberty, is also of interest due to the significance of pubescent molecular changes in the brain in the developmental hypothesis of schizophrenia [189]. If this is the case then the altered expression of the
other genes studied during development could be responsible for deficits in key circuitry which are exposed at this stage facilitating the disease phenotype, as described in Weinbergers developmental hypothesis [189], in the L100P mutant model. It is my opinion that these findings warrant further investigation into the developmental process of the schizophrenia-like phenotype in this mouse model, which may shed light on the development of the human disease.
Chapter 7

Preliminary protein analysis of validated genes in L100P drug naïve (saline) adult mouse samples
7. Preliminary protein analysis of validated genes in L100P drug naïve adult mouse samples

In this chapter I will present the results from semi-quantitative western blotting and immunocytochemistry for five of the genes validated through qRT-PCR in the L100P adult mouse mutant. Wdfy1 and Cdh11 were not tested at the protein level as suitable antibodies were not available.

Validation of the changes at the protein level allows the investigation of alterations in isoform production and assessment of possible alterations in expression patterns. The data described here is not a full analysis but rather a first step in protein analysis for the genes identified in the array.

7.1 Protein analysis of Egr4

Egr4 had previously validated in an independent RNA sample set with a fold-change of -2.89 and a p-value of 0.013. Protein extracted from the same animals used to collect the independent RNA sample was used for western blotting. Hippocampal neurons cultured from a third batch of adult mice were used for immunocytochemistry after 21 days in culture (figure 7.1). Egr4 is expressed as a 50 kDa protein. Three replicates of each western blot were carried out and the results quantified using the Image J programme, which takes a scanned image of the blot and measures the intensity of each band. Two replicates and the quantification analysis were carried out by myself, and one replicate by Susan Anderson.
Figure 6.1: Western blotting and ICC for EGR4 in L100P ENU mutant mice. Panel A shows a representative western blot for Egr4 with the Gapdh membrane used for correction of loading differences below. Panel B shows the semi-quantitative analysis of the blot using Image J and a standard student t-test (p<0.01). Panel C shows the confocal microscope image of wild-type primary cultured hippocampal neurons (day 21) stained for Egr4 (green) and co-stained for β-actin (red). Panel D shows the mutant primary cultured hippocampal neurons stained for Egr4 (green) and co-stained for β-actin (red).

As can be seen from the figure, the disregulation of Egr4 previously described at the RNA level is maintained at the protein level. There appears to be no overall change in expression patterns of Egr4 in the cultured neuron between wild-type and mutant animals. In both cases there are few disperse puncta on the dendrites, with the majority of Egr4 expression contained within the soma. Disc1 has been shown to be expressed in similar puncta in the dendritic shaft with localisation to the ribosomes and microtubule structures in the soma [276]. It can be speculated that Egr4 has a similar, although not identical, expression pattern to Disc1 in the neuron, although further staining of specific cell components would be required to confirm this. Quantitative analysis of the neuron
cultures was not carried out so small changes not obvious to the naked eye cannot be ruled out, but this would require further analysis, which will be discussed later.

7.2 Protein analysis of Pak3

The mental retardation gene *Pak3* was found to be differentially expressed in the L100P drug naïve adults in the microarray analysis, and was corrected to a level not significantly different to wild type when the animals were treated with the PDE4 inhibitor rolipram. This differential expression was confirmed in two sample batches by qRT-PCR. Pak3 is expressed as a 75kDa protein (figure 7.2). Again, one of the western blot replicates was carried out by Susan Anderson.

![Figure 6.2: Western blotting and ICC for Pak3 in L100P ENU mutant mice. Panel A shows a representative western blot for Pak3 with Gapdh membrane used for correction below. Panel B shows the semi-quantitative analysis of the blot using image J and a standard student t-test (p<0.05). Panel C shows the confocal microscope image of wild-type primary cultured neurons stained for Pak3 (green) and co-stained for β-actin (red). Panel D shows the mutant primary cultured neurons stained for Psk3 (green) and co-stained for β-actin (red).](image)
There is a small but significant upregulation of expression at the protein level in the L100P drug naive adult mouse compared to wild-type littermates that is corrected by treatment with rolipram (p<0.05). This correlates well with the results obtained from the gene expression analysis. No global changes in protein localisation are observed from ICC of primary cultured hippocampal neurons, but as the protocol used is not quantitative this cannot be accurately defined without further analysis. Initial observations suggest there may be slightly more puncta in the dendrites of the mutant animal, but further analysis and staining would be required to confirm this. The overall expression pattern of Pak3 in the hippocampal neuron is similar to that of Egr4. Further work is again required to determine whether this signifies co-localisation with Disc1.

7.3 Protein analysis of NRXN1

*Nrxn1* has previously been identified as a schizophrenia candidate gene [83, 251-253] and was found to be upregulated in the L100P drug naive adult mouse, in both the array experiment and subsequent qRT-PCR analysis. Protein which had been extracted from the hippocampus of animals used for batch two of the qRT-PCR was used to determine differences in expression at the protein level (figure 7.3). *Nrxn1* is expressed as multiple isoforms due to alternate exon splicing. The two classes of isoform are determined by splice site and size. Large, or alpha, isoforms are the most common of the two. The most widely expressed *Nrxn1* isoform is around 150kDa and the antibody used in this study was designed to primarily detect this isoform, although it would also detect other large isoforms of *Nrxn1*. One replicate was carried out by Rosie Walker and the rest by myself.
Figure 7.3: Western blotting and ICC for Nrxn1 in L100P ENU mutant mice. Panel A shows a representative western blot for Nrxn1 with Gapdh membrane used for correction below. Panel B shows the semi-quantitative analysis of the blot using image J and a standard student t-test (p>0.05). Panel C shows the confocal microscope image of wild-type primary cultured hippocampal neurons stained for Nrxn1(red) and co-stained for β-actin (blue). Panel D shows the mutant primary cultured hippocampal neurons stained for Nrxn1(red) and co-stained for β-actin (blue).

The trend of expression of Nrxn1 protein between the L100P drug naïve adult mouse and the wild-type littermate is in the opposite direction to that observed in both the array and qRT-PCR analysis, however no significant difference was observed (p>0.05). Additionally, the datasheet for the antibody used stated the band size as 150kDa and not the 75kDa observed. After consulting the technical team at Abcam who supplied the antibody, they have stated the 75kDa band is a Nrxn1 isoform, but have removed the mouse from their ‘tested in’ species. In the future this experiment should be repeated with a better characterised antibody. This will be discussed further in section 7.6. Again no global changes in protein localisation are observed from ICC of primary cultured neurones but as the protocol used is not quantitative this cannot be accurately defined.
without further analysis. The pattern of expression suggests endoplasmic reticulum involvement but further staining of cellular components would be required to substantiate this.

7.4 Protein analysis of Nrxn3

Nrxn3 is involved in synaptic transmission and neurotransmitter secretion, and has previously been associated with alcohol and nicotine dependence [255]. It was found to be upregulated in the L100P drug naïve adult mouse, in both the array experiment and subsequent qRT-PCR analysis. Protein which had been extracted from the hippocampus of animals used for batch two of the qRT-PCR was used to determine differences in expression at the protein level (figure 7.4). Like Nrxn1, Nrxn3 also has multiple isoforms due to alternate splicing. The antibody used was designed to detect alpha isoforms of around 150kDa. One replicate was performed by Susan Anderson.
Figure 7.4: Western blotting and ICC for NRXN3 in L100P ENU mutant mice. Panel A shows a representative western blot for Nrxn3 with GAPDH membrane used for correction below. Panel B shows the semi-quantitative analysis of the blot using image J and a standard student t-test (p>0.05). Panel C shows the confocal microscope image of wild-type primary cultured hippocampal neurons stained for Nrxn3 (green) and co-stained for β-actin (red). Panel D shows the mutant primary cultured neurons hippocampal stained for Nrxn3 (green) and co-stained for β-actin (red).

While the trend of expression is consistent with the array and qRT-PCR analysis, no significant difference was observed at the protein level between the L100P drug naïve adult mouse and the wild-type littermate (p>0.05). The blot was considered rather dirty and it was difficult to obtain a clear result from this antibody. Another Nrxn3 antibody was tested but the result was even more ambiguous and this was considered the better of the two for analysis, however, these results should be viewed as preliminary. Again no global changes in protein localisation are observed from ICC of primary cultured neurones but as the protocol used is not quantitative this cannot be accurately defined without further analysis.
7.5 Protein analysis of Sort1

*Sortl* (neurotensin receptor 3) has previously been found to be down-regulated in sufferers of bipolar disorder (Christoforou et al, manuscript in preparation) and has been implicated in the modulation of dopamine signalling [270]. It was found to be down-regulated in the L100P drug naïve adult mouse pre outlier removal, and remained significant by qRT-PCR analysis post outlier removal. Sort 1 is expressed as a 95kDa protein (figure 7.5). One replicate was performed by Susan Anderson.

**Figure 7.5: Western blotting and ICC for Sort1 in L100P ENU mutant mice.** Panel A shows a representative western blot for Sort1 with gapdh membrane used for correction below. Panel B shows the semi-quantitative analysis of the blot using image J and a standard student t-test (p<0.01). Panel C shows the confocal microscope image of wild-type primary cultured hippocampal neurons stained for Sort1(green) and co-stained for β-actin (red). Panel D shows the mutant primary cultured hippocampal neurons stained for Sort1(green) and co-stained for β-actin (red).

There is a significant difference at the protein level between the L100P drug naïve adult mouse and the wild-type littermate (p<0.01). Images obtained from the ICC of Sort1
appeared blurry with bleaching of the Sort1 signal and the Actin signal. This occurred in every image taken and was consistent between wild-type and mutant, suggestive of an incompatibility between the two antibodies. Again no global changes in protein localisation are observed from ICC of primary cultured neurones but as the protocol used is not quantitative this cannot be accurately defined without further analysis.

7.6 Conclusions
While identification of dysregulated genes gives insight into the mechanisms of mental illness, correlation at the protein level signifies a lack of cellular correction mechanisms allowing consequential effects of the gene disregulation at the cellular level. In a review by Chuaqui et al (2002) [277] they state that a correlation between changes at the mRNA level and the protein level are accurate less than 50% of the time. This is due to post-translational modifications of the proteins, and cellular feedback mechanisms which maintain proteins at optimal levels for cellular function and development.

This study has identified three genes that are significantly differentially expressed at both the gene and protein level; *Egr4*, *Pak3* and *Sort1*. These three genes have all been previously implicated in major mental illness. Both *Nrxn1* and *Nrxn3* failed to replicate their differential expression at the protein level. This could be due to the mechanisms mentioned previously, however there were also problems involved with the NRXN antibodies used which could account for the lack of change observed. NRXN1 is present in multiple isoforms and non-specific binding could account for the problems encountered with this antibody. As mentioned in section 7.3, the NRXN1 antibody did not give a band corresponding to that suggested on the manufacturers’ datasheet. Consultation with the Abcam technical team revealed the antibody had not been tested in this species, and it could not be guaranteed that this was specific binding. A clear result could not be consistently obtained for the NRXN3 antibody. There are multiple isoforms of both NRXN1 and NRXN3 that could potentially bind the antibodies used and obtaining reliable results proved problematic.
Protein concentration was measured for all samples prior to loading using the Thermo Scientific NanoDrop 100 spectrophotometer. Despite loading what was believed to be the same amount of protein (after quantification and required dilutions) in each well, the GAPDH loading control did not give the same intensity band for each well. There was no consistency in the direction of difference (i.e., the mutant animals were not consistently higher than the wild-type with regards to protein concentration) nor was the position on the gel a factor. It is likely that these inconsistencies were due to errors with the initial quantification or inaccurate pipetting when loading the gel. It would be advisable to test the calibration of the NanoDrop for protein quantification by running samples on both the NanoDrop platform and a Bradford or a Lowry assay to check for inconsistencies. Due to these loading inconsistencies there was a greater margin for error when correcting the band intensity for overall protein concentration by densitometry. The outcome of western blot densitometry relies highly on the equipment used at every stage. In 2009, Gassmann et al. [278] discuss the pitfalls of current methods of densitometry and state that the same blots can give vastly different p-values (0.000013 to 0.76) depending on the method of digitization used. By comparing the densitometry results obtained from the same blot using a densitometry camera and a standard office scanner (as used in this experiment) they concluded that the office scanner was not a suitable tool for digitizing the image due to the inability to switch off the gain control. This resulted in a short optical density range and saturation at the higher levels was more frequent than with the densitometry camera. Lack of homogeneity across the scan area also means that the optical density of a band or spot on the gel is a function of its position, and shading correction that would correct this problem cannot be performed due to the gain control. In this experiment, the image J software used relied on the quality of scanner used when uploading the blot image to the PC, and on the blots being compared being of a similar exposure. Grassmann et al.’s results would suggest that a different digitization technique would have resulted in very different results from the same blots. As such the western blots displayed here may only be taken as an indication of differential expression and not an absolute value. Further attention to quantification methods may or may not confirm the reported result and it is suggested this be carried out using other suggested methods.
A similar statement could be made for the immunocytochemistry staining reported in this chapter. There were no observable differences between mutant and wild-type animals in relation to protein distribution in the neuron, however this was a very broad brush approach aiming to identify areas for further study. Further research should involve co-staining the protein of interest with a synaptic marker, such as PSD-95, to determine effects on synapse localization. Of interest would also be co-staining with markers for other cellular components (such as mitochondria etc) to give a greater idea of the effect of the dysregulated proteins on cellular function as a whole, and their localization in relation to Disc1.

The above results, however, provide a good grounding for further research into the expression of these candidate genes at the protein level, in particular the localization and expression in vitro.
Chapter 8
Discussion
8.1 Summary of findings

This thesis has described findings from genome wide expression analysis, qRT-PCR and protein analysis of an accepted animal model of major mental illness. I have reported the discovery of a number of potential candidate genes for further research. In this final chapter I will present a summary of the findings, how they relate to the field of mental illness research, and a possible direction for future work based on these findings.

8.1.1 Chapter 3: Sample collection and processing of a genome wide microarray analysis of the Disc1 ENU mouse mutants

One of the primary driving forces behind this study was the identification of the L100P and Q31L mouse mutants, and subsequent behavioural data classifying them as ‘schizophrenia-like’ and ‘depressive-like’ [197]. Both of these models carry missense mutations in exon 2 of the Disc1 gene which disrupt Pde4 binding sites. Expression of Disc1 protein is not altered in either animal, however Disc1-Pde4b binding is reduced in both models, and activity of Pde4b is reduced by up to 50% in the Q31L model [197].

I repeated the prepulse inhibition studies carried out by Clapcote et al (2007) [197] in an attempt to confirm the behavioural phenotype in the mouse model prior to genome wide expression analysis. The prepulse inhibition experiments I carried out did not give statistically significant results, but did follow the trend previously published. However, as mentioned in section 3.8, there were a number of caveats associated with these experiments. The identification of a de novo mutation in the mutant mouse lines meant it was necessary to cross only homozygous animals, making the use of wild-type littermates as controls impossible. As the mutation was on a C57BL6 background, this strain was used as a wild-type control. Studies of PPI in rodents have been going on for many years as researchers attempt to find the genetic basis for such behaviours. In 1997, Paylor and Crawley studied the differences in PPI in commercially available inbred mouse strains. Interestingly, the C57BL/6J mice gave the lowest PPI of all the mice they tested [232]. The PPI of the C57BL6 mice used in this study was not statistically different to the
results obtained by Paylor and Crawley, but was significantly lower than Clapcote et al’s wild-type controls. The result obtained for the mutant mouse lines are not significantly different to those obtained by Clapcote et al (2007) in the mutant lines, suggesting that, at least for the behavioural study, the C57BL6 mice are a less than ideal control animal. It was concluded that, due to the similarity between the levels of PPI I observed and those obtained by Clapcote et al (2007) [197], that the PPI deficit in the mutant lines was a true result and the behavioural phenotype was present in the mice. In addition to this, other researchers have confirmed the PPI deficit in the ENU mouse lines using wild-type controls (Steven Clapcote, personal communication).

Despite the limitations, the genome wide gene expression analysis was carried with C57BL6 control animals out due to time pressure and because validation of differential expression would be carried out using wild-type littermate controls once a colony had been established. The microarray analysis, therefore, compared drug naive adult mutant mice or adult mutant mice treated with selected antipsychotic or antidepressant drugs, with C57BL6 or drug naïve controls. Samples were also taken from embryonic day 13, which had previously been shown to be a peak in Disc1 protein expression [134]. Quality of RNA extraction from hippocampal tissue (or whole brain from the embryonic stage) was tested using Agilent technology. This showed that the RNA extracted was of sufficient quality and could therefore be used to look for differential gene expression.

8.1.2 Chapter 4: Differential Expression Analysis of the Disc1 ENU Mutant Mouse Microarray

Analysis of the genome wide expression data initially revealed a total of 835 genes that were differentially expressed across the sample set. However, subsequent analysis of individual gene expression using statistical methods and qRT-PCR identified one outlier pool that gave erratic and unpredictable expression values. Re-analysis of the whole genome data without this pool as expected revealed a reduced set, with 368 genes differentially expressed across the sample set. Of these genes, 62 were over-expressed
and 24 under-expressed in the L100P adult samples and 19 were found in the Q31L adult sample set. Three genes (\textit{wdfy1}, \textit{LOC674214} and \textit{2900040C04RIK}) were found to be differentially expressed in both the L100P and Q31L non-drug treated adult mice. 91 genes were over-expressed and 95 genes under-expressed in the Q31L bupropion treated group compared to non-drug treated adults. The L100P rolipram treated mouse gave the most genes which showed correction under pharmacological influence. Of the 32 genes identified, seven showed correction of the L100P non-drug treated disregulation. Correction of expression was defined as returning that gene to a level not significantly different (FC+/- 1.3, p<0.05) to that of C57BL/6J controls. 45 genes were found to be differentially expressed in the L100P embryo and 25 genes differentially expressed in the Q31L embryo. Four genes were identified in both groups, \textit{Prm3}, \textit{Coro1B}, \textit{Numb} and \textit{Pde4d}. With the exception of bupropion treatment, the L100P comparisons all produced a higher yield of dysregulated genes than the Q31L comparisons. The ‘schizophrenia-like’ phenotype of the L100P mouse is considered the more severe of the two, and so it could be expected that this would give a higher level of differential expression observed in this animal. Due to the high number of genes differentially expressed with bupropion treatment that were not correspondingly differentially expressed in the Q31L mutant without treatment, it was concluded that this was an effect of drug treatment and not related to the genotype. No differences were observed between male and female mice in gene expression. The small number of genes that were differentially expressed in the adult mutants and then corrected by drug treatment would suggest the drugs used were active primarily at the protein function level, and had minimal effect on gene expression.

GOTree analysis revealed over-enrichment of genes involved in the following categories: cell-cell signaling, transmission of nerve impulses/synaptic transmission. No overenrichment of gene categories was observed in any other group tested, possibly due to the small number of differentially expressed genes. Previous studies have identified clusters of genes involved in neural function [199] synaptogenesis and sensory perception [128] in schizophrenia. From the GOTree analysis, \textit{Egr2} and \textit{Egr3}, which have previously been associated with schizophrenia in a Japanese population, [238] were both significantly down-regulated in the L100P adult sample, and are involved in cell
signaling and transmission of nerve impulses. The excitatory glutamate receptor 2 (Gria2) was overexpressed in the L100P adult and is involved in cell signaling, transmission of nerve impulse and neurotransmitter secretion. It has also been identified as a potential candidate gene for schizophrenia in its own right [279]. Synapsin 2 was overexpressed in the L100P adult and was identified as being involved in cell signaling, neurotransmitter secretion, and transmission of nerve impulses. Syn2 has previously been implicated in conferring risk to schizophrenia in a Korean cohort [280]. Neurexin 1 (Nrxn1) was overexpressed in the L100P adult mouse. It is involved in maintenance of synaptic junctions and is a mediator of intracellular signalling [249]. NRXN1 was implicated by Kirov et al (2008) [217] in an analysis of copy number variants in schizophrenia. Other studies on copy number variants have also implicated NRXN1 in schizophrenia pathology. Need et al (2009) [251] found deletions in the 3’ end of NRXN1 in patients with schizophrenia, but not in control groups. Vrijenhoek et al (2008) [252] suggest that CNVs that affect the first few exons of NRXN1 confer greater risk of major mental illness, while Rujescu et al (2009) [253] suggest that deletions that affect exons directly increase susceptibility. In addition, a missense mutation in exon 1 of human NRXN1 has also been linked to autism [254] and recent studies have linked deletions in NRXN1 to a range of phenotypes, including autism spectrum disorder, mental retardation, hypotonia and language delays [281].

Lots of recent work has investigated the role of copy number variants (CNVs) in neurodevelopmental disorders (reviewed in Kirov 2010) [82]. Only two genes were present in both the dysregulated gene list from this study and copy number variants in the previous studies. MLL5, a transcription regulator, was present as a duplication in schizophrenia cases from Walsh et al 2008 [83], and was upregulated in the L100P non-drug treated adult group. As previously mentioned, NRXN1 was also found to be up-regulated in the L100P non-drug treated adult group. Walsh et al (2008) reported a NRXN1 deletion present in a schizophrenia case and affected sibling [83]. The deletion spans the promoter and first exon of the gene and partially overlaps with previous deletions co-segregating with autism [76] and mental retardation [239]. I also had access to data from recently completed microarray expression studies on LCLs from individuals
from the \textit{DISC1} translocation (t1:11) family that allowed an almost direct comparison to be made between the mouse and human model. Five genes (\textit{Dusp6, F5, Hook3, Nrxn1 and Nrxn3}) were differentially expressed in both studies, with \textit{Hook3, Nrxn1} and \textit{Nrxn3} dysregulated in the same direction in both arrays. These overlapping results between the mouse and human sample subjects is promising in terms of candidate gene identification and because it provides further evidence of the validity of the animal model to this system.

One major flaw of this comparison is the two arrays used different tissues as the primary RNA source. While it was possible to use brain tissues from the ENU mutant mouse, the human samples were from a blood-derived cell line. Blood-derived (lymphoblastoid) cell lines do not express all genes and in particular those expressed in the brain[240]. This could obviously result in some genes not being present in the human dataset by virtue of the original tissues used, and not because they are not expressed, or potentially dysregulated in the human condition.

Using Ingenuity pathway analysis, nine genes were identified that showed both overenrichment in GO categories and presence in the top statistically significant Ingenuity pathway analysis network in the L100P adult group. Ingenuity pathway analysis identified networks involved in cellular development, gene expression and cell signaling for the L100P adult, L100P embryo and Q31L embryo respectively. On the basis of fold-change, p-value, function (by GeneOntology classification) and overlap with previous studies, 40 genes from the microarray study were chosen for validation by qRT-PCR. Genes chosen were mainly involved in synaptic transmission, neurotransmitter transport and secretion, and central nervous system development. Genes whose differential expression was corrected by drug treatment were also carried forward for validation.

This study had obvious limitations and weaknesses, not least of all the presence of an outlier group in the initial analysis. In addition, a number of the genes that showed differential expression on the array were anonymous sequences from a cDNA bank with
little or no functional information attached. This made it difficult to determine the relevance of these predicted sequences to major mental illness and therefore, most were not chosen for further analysis. It is possible that the lack of information available for many genes biased the selection of the subset for further analysis.

8.1.3 Chapter 5: Quantitative real-time PCR analysis of candidate differentially expressed genes in the L100P and Q31L mutant mice

qRT-PCR was used for the validation of the microarray results. Two cDNA batches were used; one from the RNA that went on the array, and an independent batch extracted from mutant animals and wild-type littermate controls. Genes that failed to validate from the first batch of cDNA were not tested in the second batch. While every effort was made to ensure overlap of the microarray probes with the qRT-PCR probes, a perfect match could not be ascertained. In all cases both probes were derived from the same exon but, as the sequences of the qRT-PCR probes are not made publicly available, identical sequence alignment could not be achieved. Slight misalignment of probes could result in amplification of different isoforms of the gene in question. Also, false positivity in array data can occur by binding of repetitive nucleotide elements, sequence homology between functionally different transcripts and high background levels due to nonspecific binding of nucleotides [248]. For these reasons it was decided that genes must show consistent results across platforms to be considered valid, thus ensuring as much as possible a true result.

Of the 40 genes tested by qRT-PCR, 24 validated in the first round and only seven gave robust and reliable results through both rounds of validation. Correlation estimates predict a Spearman coefficient of 0.708 [282] between microarray data and qRT-PCR. The results obtained from this study are therefore slightly lower than would be expected, however lack of concordance between platforms at low fold-change (<1.4) has been previously reported [283]. What is more concerning is the lack of concordance between the two batches of RNA, suggesting differences in the wild-type expression could be due to minor genetic differences between the C57BL/6J and the L100P wild-type. Although
the L100P animals had been backcrossed for over 10 generations it is unlikely, but not impossible, that there were still some genetic differences present.

Of the seven genes validated, four were from the L100P drug naïve adult group (\textit{Nrxd1}, \textit{Nrxd3}, \textit{Egr4} and \textit{Wdfy1}), two from the L100P embryonic group (\textit{Cdh11}, \textit{Prm3}) and one from the L100P Rolipram treated group which showed correction (\textit{Pak3}). The neurexin genes (\textit{Nrxd1} and \textit{Nrxd3}) are involved in synaptic transmissions and the maintenance of synaptic junctions. \textit{Nrxd1} has been shown to be involved in the maturation and differentiation of GABAergic and glutamatergic synapses through bi-directional signalling [250]. DISC1 has previously been shown to interact with GSK3β, which suppress long-term potentiation and presynaptic release of excitatory glutamate in cortical neurons by inhibition [284]. As the glutamate hypothesis of schizophrenia predicts that decrease of NMDA receptor signaling during interneuron development could result in the behavioural phenotypes observed in schizophrenia [285], the potential interaction between \textit{DISC1} and another regulatory glutamate gene is of great interest. \textit{NRXN1} has also previously been identified as a candidate gene for schizophrenia [83, 217, 251-253] and Autism [254] while SNPs in \textit{NRXN3} have been associated with alcohol and nicotine dependence and linked with opiate dependence [255, 256]. The Early Growth Response Factors (\textit{EGRs}) 2, 3 and 4 are synaptic activity inducible immediate early genes and all show nominal association with schizophrenia in a Japanese population. \textit{EGR1}, 2 and 3 are downregulated in the prefrontal cortex of schizophrenic brains [238]. \textit{EGR4} expression was too low to measure disregulation however, treatment with the atypical antipsychotic and antidepressant Aripiprazole increases expression of \textit{EGR4} in rat frontal cortex[257]. Little is known about the function of WD repeat and FYVE domain-containing protein 1 (\textit{WDFY1} or \textit{FENS1}) but it is known to be ribosomal and has been associated with alcohol consumption and preference [258].

Cadherin 11 (\textit{Cdh11}) is a type 2 classical cadherin involved in mediation of calcium dependant cell adhesion. It was first discovered in rodent brain samples [266] with high expression in the developing olfactory system at E13-E17 [267] and has since been linked to multiple cancers [268, 269].
Pak3 was the only drug-corrected gene validated. Mutations in PAK3 have been associated with X-linked nonsyndromic mental retardation [260-262] and suppression of Pak3 results in formation of abnormally elongated dendritic spines and a reduction of mature synapses [263]. Pak3 knockout mice have deficiencies in learning and memory and abnormalities in synaptic plasticity along with a reduction of transcription factor cAMP-responsive element-binding protein (CREB) suggesting a novel signalling mechanism with PAK3 and Rho signalling regulating synaptic function and cognition [264]. PAK genes also interact with the known DISC1 interactor Kalirin-7 which is thought to control multiple aspects of spinal plasticity [265]. Kal-7 activates Rac-1 to control spine size. In the presence of DISC1, Kal-7 is anchored, regulating access to Rac-1 and controlling duration and intensity of Rac-1 activation. It is thought that constitutive activation of Rac-1 (due to Kal-7 not being anchored by DISC1) results in decreased spine size and may underlie the disturbances in glutamatergic neurotransmission observed in schizophrenia [286].

Five genes from the pre-outlier removal gene list (as described in chapter 5) were tested by qRT-PCR and two validated through both rounds, with significant fold-change and p-values that meet the criteria imposed. As previously mentioned, false positivity in array data can occur through multiple routes. In parallel, false negativity can occur through low expression levels, inefficient priming of specific mRNAs resulting in transcript drop out, poor adhesion of DNA to the slide, and splice variants with sequences not included on the array [248]. As the fold changes observed on the microarray for these genes were below the accepted sensitivity level of the array (http://www.switchtoi.com/pdf/GXHuman6-8v2Datasheet.pdf) but were subsequently found to be differentially expressed through qRT-PCR they may be considered as false negatives. This does, however, raise speculation to the criteria imposed when selecting genes for validation and this will be discussed in section 8.2.

Three additional genes (Dusp1, Ndfip and Atp5b) were found to be robustly dysregulated in the L100P rolipram treated group, but did not differ in the L100P adult vs wild-type controls. I, therefore, proposed that the disregulation of these genes was a function of drug treatment and not a result of the mutation. To confirm this I tested these genes in a
C57BL6 drug naïve group vs C57BL6 rolipram treated group. Dusp1 was significantly up regulated in the C57BL6 rolipram treated group compared to the drug naïve group, as it had been in the L100P rolipram treated group compared to the L100P drug naïve group. Ndfip also behaved in the same way in the C57BL6 rolipram treated group as it had in the L100P drug treated group, being significantly down regulated compared to the drug naïve control. Atp5b was significantly down regulated in the C57BL6 rolipram treated group where it had been up regulated in the L100P rolipram treated group, but a significant difference in expression was observed suggesting it is an effect of the drug treatment as opposed to the mutation. As the changes in expression in these three genes appear to be an effect of the rolipram treatment rather than the mutation I feel they do not fit the criteria for further analysis within this project. These data may, however, provide some information on the pathways through which these drugs act.

8.1.4 Chapter 6: Establishing a developmental profile for the genes of interest in L100P mutant mice and wild type controls

The developmental profile of Disc1 highlighted key stages of development where expression is altered in the mutant mouse line, and also confirmed my previous microarray result which showed that Disc1 expression was not significantly different at either E13 or adulthood between the L100P mutant and wild type controls. It was previously thought that the ENU mutations affected DISC1 protein function and not expression levels [197]. The data from these developmental profiles redefines our knowledge of this model and suggests that the behavioural phenotype observed is a combined effect of differences in DISC1 function and developmental expression.

In the L100P mutant mouse there is an obvious peak in Disc1 expression at P1, with a peak at E18 observed in the wild-type controls. This is similar to the pattern observed in the Q31L mutant mouse. In addition, the Q31L mutant animals have significantly higher Disc1 expression than their wild-type counterparts at E15. It can be speculated that the
different developmental expression patterns of the Q31L and L100P mutants may contribute to the different phenotypes observed in behavioural testing. A review of the literature on mouse developmental stages suggests that, based on genes with known neural functions, the functional gene clusters most highly expressed at E15 are involved in neural determination and differentiation, and at E18 and P1 are involved in synapse formation and function, and survival and growth [274]. Based on all genes differentially expressed at E18 and P1 the top functions are intra/inter cellular molecular transport, and signal transduction [274]. This delay in Discl expression in the L100P mutant mouse could have detrimental effects on neurodevelopment in the mutant animals. Personal observations and those of other researchers working with these animals show that the mutant individuals are physically smaller when born, which could signify a delayed development (Steven Clapcote personal communication). Activity at the three developmental time points highlighted here roughly correspond to human neurodevelopment at weeks 9-12, 12-15.5 and 13.5-17.5 in utero [287].

While the result of one gene expression profile is not enough to stipulate cause and effect it is worth noting that low birth weights have previously been associated with schizophrenia [273] and a latent shift in the expression of key developmental proteins could be a factor in this. Hikida et al [193] also note that there is a 3month delay in neuronal migration in a Discl mouse model, which could potentially be attributed to a change in Discl expression during development.

Developmental expression analysis of five differentially expressed genes highlighted some interesting developmental changes. Both neurexin genes showed a marked shift of peak expression in the L100P mutant compared to wild-type littermates. Nrxn1 peaks at E18 in wild-type, and P1 in L100P mutant animals. Conversely, Nrxn3 plateaus from E15-P1 in wild-tpe and peaks at E15 in mutant animals, with significantly lower expression at E18 and P1 than wild-type. As the predicted developmental function at E18 is the formation of synapses, the dip in expression of genes involved in synaptic maintenance and transmission in the mutant animals is considered of key importance. Pak3 developmental expression is identical in the wild-type and mutant animal until E15.
After this the wild-type expression continues to rise to peak at E18, while the mutant expression decreases, plateauing at P20. There is significant underexpression of \textit{Pak3} at E18 and P1 in the mutant mouse. As \textit{Pak3} was the only gene to show correction through drug treatment, it would be interesting to see if earlier intervention would facilitate complete rescue of the phenotype, or if treatment post hoc is sufficient. \textit{Egr4} displayed a distinct lack of expression in both wild-type and mutant until P1, which has previously not been reported, and significantly higher expression in the L100P mutant at P20, just prior to puberty in the mouse. According to the developmental hypothesis of schizophrenia, the molecular changes in the brain at the time of puberty are partially responsible for disease development in predisposed individuals\cite{189}. If this is the case then the altered expression of this and other genes studied during development could be responsible for deficits in key circuitry which are exposed at this stage, facilitating the disease phenotype, as described in Weinberger’s developmental hypothesis \cite{189}, in the L100P mutant model.

\textbf{8.1.5 Chapter 7: Preliminary protein analysis of validated genes in L100P drug naïve (saline) adult mouse samples}

Due to post-translational modifications of proteins, and cellular feedback mechanisms that maintain proteins at optimal levels for cellular function and development, it has been reported that a correlation between changes at the mRNA level and the protein level are accurate less than 50\% of the time \cite{277}. I identified three genes that are significantly differentially expressed at both the gene and protein level; \textit{Egr4}, \textit{Pak3} and \textit{Sort1} in the L100P adult mouse model. Both \textit{Nrxn1} and \textit{Nrxn3} failed to replicate their differential expression at the protein level. This could be due post translational modifications of the proteins, however there were also problems involved with the neurexin antibodies used which could account for the lack of change observed. The \textit{Nrxn1} antibody did not give a band corresponding to that suggested on the manufacturers’ datasheet and clear result could not be consistently obtained for the \textit{Nrxn3} antibody. There are multiple isoforms of both \textit{Nrxn1} and \textit{Nrxn3} that could potentially bind the antibodies used and obtaining reliable results proved problematic.
Immunocytochemistry was used to visualise the proteins of interest in cultured hippocampal neurons from L100P mutant and wild-type animals. No obvious overall changes in expression pattern were observed in any of the cultures, however the analysis was not quantitative and small changes would not be observed by the naked eye. Egr4 appears to be expressed in disperse puncta on the dendrites, with the majority of Egr4 expression contained within the soma. Disc1 has been shown to be expressed in similar puncta in the dendritic shaft with localisation to the ribosomes and microtubule structures in the soma [276]. It can be speculated that Egr4 has a similar, although not identical, expression pattern to Disc1 in the neuron, although further staining of specific cell components would be required to confirm this. The overall expression pattern of Pak3 in the neuron appears similar to that of Egr4. Initial observations suggest there may be slightly more puncta in the dendrites of the mutant animals but further quantitative analysis would be required to confirm this. Nrxn3 again appears localised to the soma, with few puncta in the dendrites. Staining of the hippocampal neurons for Nrxn1 gave a slightly different pattern of expression, suggestive of endoplasmic reticulum involvement. Disruption of the neurexin PDZ-binding domain motif in Nrxn1 knockout mice has been shown to result in diffuse distribution patterns with Nrxn retained in the endoplasmic reticulum and not packaged into vesicles for transport to the synapse [288]. It could be speculated that disruption of Nrxn1/PDZ binding by Disc1, or one of its interactors, prevents trafficking to the synapse. Further staining with synaptic markers would be required to determine if Nrxn1 is localising to the synapse or if it is being held in the ER. Images obtained from the ICC of Sort1 appeared blurry with bleaching of the Sort1 signal and the actin signal. This occurred in every image taken and was consistent between wild-type and mutant, suggestive of an incompatibility between the two antibodies.

While there were a number of limitations to this study (as described in more detail in section 7.6), these results provide the basis for further research in this area. They are to be considered preliminary findings, and it would of course be advantageous to continue these experiments, and in some cases repeat with more reliable antibodies. Further neuron
staining using cellular markers to compare co-localisation of proteins between mutant and wild-type animals would also be recommended.

8.1.6 Conclusions

This work has identified a number of genes dysregulated by missense mutations in Disc1 that are proposed to be involved in the schizophrenia endophenotype displayed by the mouse model. Most notably, the striking dysregulation of a number of known candidate genes for major mental illness in the L100P Disc1 mutant animals suggests a common network involved in the pathogenesis of schizophrenia and major mental illness. The altered expression of these genes during development in the mouse model also suggests a substantial neurodevelopmental component, supporting previous hypotheses from Weinberger (1986) [189] and Torkamani (2010) [199].

8.2 Caveats

No experiment is perfect, and it is important to record and analyse the potential caveats to improve future work. In the whole genome expression study, potentially the biggest question was the use of C57BL6 as a wild type control, particularly given its reduced PPI compared to the mutant wild-type littermates. The justification for this was two fold; firstly, the results from the array would be validated using wild-type littermate controls so any change attributed to the use of C57BL6 over wild-types would be identified at a later date. Secondly, the mutant mouse line had been backcrossed for 10 generations and would be expected to be genetically identical to at least 98.9%. C57BL6 mice were specifically sourced from the Charles River substrain to be as close as possible to the original backcross strain used when establishing the ENU mutant colonies, however, results obtained from both the behavioural study and comparison between the C57BL6
and wild-type littermate controls would suggest the presence of non-target mutations which may be responsible for the differences observed. These could either be secondary ENU mutations that co-segregated with the Disc1 mutation, or sections of DBA genome that remained, despite backcrossing, around the Disc1 gene. Comparative genotyping of the whole chromosome would allow this to be determined. In retrospect, the use of wild-type littermates for the array, as was the initial plan, would have reduced the number of false positives and false negatives, and provided a ‘cleaner’ result from the offset. However, due to circumstances outwith my control this was not possible, and it was necessary to continue with the closest viable control that was available at the time, the C57BL6 mouse. I would suggest that further experiments of this kind use wild-type littermate controls as standard, or that the C57BL6 be fully genotyped and compared to the mutant mouse line to fully establish and genetic differences that may be present.

Hippocampal tissue was used in the array study for a number of reasons. Abnormalities in hippocampal structure, activation, organization of neurons and synapse function have been well documented in human schizophrenia patients and other mouse models [207, 214, 233, 234]. Added to the working memory deficits often associated with schizophrenia [208], it is clear there is substantial hippocampal involvement in the disorder. The use of a single brain region also removed any confounding effects of compensation in other brain regions, which may have diluted the result. As such, this study can make direct inferences as to the gene expression in the hippocampus of the adult mutant mouse. Due to researcher inexperience, it was not possible to remove only hippocampal tissue from the embryonic mouse samples, and in these cases whole brain was used. This was not ideal and it would be suggested that analysis of distinct brain regions should be carried out on the embryonic mutant mouse, particularly with reference to the developmental panel. It would be very interesting to discover if the changes of gene expression over development observed in this study are confined to one brain region, or if they are a global effect.

Microarray technology has a number of inherent limitations that cannot be ignored. Genes with low expression levels, or genes with a small change in expression level, can often be overlooked due to the background noise of a heterogenous sample. It is highly
probable that this could occur in the sample set used for this experiment. The genetics of psychiatric disorders are known to be very complex, with multiple genes likely to be involved. Illumina state confidence in detecting a gene expression fold changes greater than 1.3, essentially meaning that any change in expression less than 30% from the norm cannot be reliably detected. Prior to filtering the data based on fold-change, 78% of genes with a significant \( p \)-value gave a fold-change between 1.01 and 1.25. These were obviously discarded from further analysis due to the detection cut-offs imposed by Illumina. This cut-off should however be considered arbitrary. Prior to the identification of the outlier 20 genes had been tested by qRT-PCR that did not make the 1.3 fold-change cut-off on reanalysis. Of these, two validated independently by qRT-PCR, suggestive that in these cases there was differential expression that was not apparent from the array analysis. The problem of tissue heterogeneity was in part addressed by using inbred mouse strains, however even within a small region of the brain (in this case hippocampus) there are multiple cell types with different expression patterns adding to the background noise of the array. This could be overcome using microdissection of specific cell types and it would be suggested this is carried out for future experiments.

The identification of an outlier group post analysis was indeed a blow to the study. Not least it meant the total reanalysis of the dataset and questioned the validity of previous quality control measures. As the outlier did not consistently show over or under expression, its effects were diluted when analysing the data as a whole. While this may be an isolated case, it highlights how important it is to analyse each group individually, by checking a random set of genes, as well as the group as a whole.

It was stated in chapter 4 that no corrections were made for multiple testing in this dataset. False discover rates (FDRs) were calculated for each pairwise comparison. As the maximum FDR was \(~30\%\) it could be predicted that 110 of the 368 genes found to be differentially expressed would be false positives. While this value is high it was decided that continuing without correction was viable as the data would be validated on an independent platform which should see any false positives from the array removed at that stage. If this experiment were to be repeated I would suggest either analyzing the dataset using ANOVA and posthoc tests, or, if carrying out multiple t-tests as was done here, that
some multiple testing corrections be carried out. This could potentially increase the number of false negatives from the data but would produce a dataset that was more likely to be reproducible on the independent platform.

Validation of the microarray result was two fold. In the first instance, genes of interest were analysed by Taqman real-time PCR using RNA from the animals that were used on the array. Of the subset of genes analysed, 51.5% gave reproducible results across both platforms. While every effort was made to ensure overlap of the two probes, a perfect match could not be ascertained. In all cases both probes were mapped to the same exon but as the sequences of the qRT-PCR probes are not made available identical sequence alignment could not be achieved. This may account for the low levels of reproducibility. Also, as mentioned previously, the array false discovery rate was high, so it could be expected that some of the genes identified and subsequently analysed on the Taqman platform were false positives. Wang et al (2006) [289] compared microarray platform and Taqman real-time PCR performance and found correlation ranges between 0.45 and 0.79 between platforms. They also noted that correlations were higher in genes with greater expression levels and larger fold changes. As the majority of genes selected for validation in this study had low fold-changes (classed as <2) it appears consistent with a smaller correlation between platforms. The second round of validation used RNA from an independent set of animals derived from the original line, with wild-type littermate controls. Only those genes that had reproduced successfully in the first round were tested in this sample set. Seven of the 19 genes tested reproduced in this second round. This could not be due to between platform differences, or probe alignment issues. It is therefore assumed that the issues discussed previously regarding the use of the C57BL6 as a control affected the result. Genotyping both ‘wild-type’ controls would be advised if this experiment was to be repeated, although the use of wild-type littermates from the offset would be likely to reduce this problem.

As stated previously, the selection criteria for genes for further study was, with retrospect, less than ideal. Small fold changes close to the detection sensitivity of the array could have resulted in a high number of false negatives, as demonstrated by the
validation of two genes by qRT-PCR that did not meet the selection criteria on the array. If the data were to be reanalyzed it would perhaps be prudent to select more on the basis of functional significance and $p$-value with less weight being left to fold-change values. While it is accepted that there must be a cut-off level, it is suggested that due to the complex nature of schizophrenia and other major mental illness, and the predicted multiple genes of small effect, that current array technology may not yet be sensitive enough to pick up the small changes in gene expression that may contribute to disease pathology.

The inclusion of the drug treated groups, while of interest, also increased the spread of the data and the propensity for error. Increasing the number of groups increases the multiple testing burden and reduces the statistical power. It is also likely that the primary effects of the drug treatments are not at the level of gene expression, but in post-translational protein modification and/or action on neurotransmitter systems [290, 291]. Previous studies have identified gene expression changes after clozapine treatment ([292, 293] and others), however these studies used chronic treatments and whole brain tissues, increasing the heterogeneity of the sample. I would conclude that while the drug treatments were interesting they were not necessarily an essential part of this study. Further work to investigate the actions of psychiatric drugs would be of interest, and the use of this mouse model an advantage for this, but it would be of more interest to investigate at the level of the synapse for change in neurotransmitter binding and secretion, than at the gene expression level.

8.3 Relevance of this study to the field

Since the discovery of the $DISC1$ gene in 2000 [56], a number of studies have identified variants within this locus as susceptibility factors for major mental illness. Many $DISC1$ binding partners have also been associated with major mental illness in their own right.
suggested a network of gene involvement in these synaptopathies.

While microarray analysis has been previously used to identify candidate genes for major mental illness, this is the first study to look at genome wide gene expression in a model system with mutations in Disc1, an already established candidate gene. The use of a well classified animal model, with mutations in Disc1 and behavioural characteristics consistent with a schizophrenic or depressive-like endophenotype, has allowed me to carry out this analysis on a uniform genetic background, with sufficient numbers for statistically relevant conclusions. Due to problems with tissue access and genetic heterogeneity, this would not have been possible using human samples.

I have identified a number of genes that are dysregulated as a result of the Disc1 mutation. Most of these genes have not previously been shown to interact with Disc1, suggesting either indirect interactions and/or the possibility of network involvement of these genes, which together contribute to the disease pathology. It may also suggest involvement of transcriptional control. As some of the genes identified and validated in this study have previously been associated with major mental illness in their own right, there is a convincing argument for network involvement in schizophrenia pathology, at least in this system.

It has previously been reported that genes related to CNS development, neuron guidance and neurotransmitter secretion were down-regulated with age (birth-20 years) in normal individuals [200]. Torkamani et al (2010) [199] found that while this downregulation continued in control subjects, it did not in individuals with schizophrenia, suggesting a progressive neurodevelopmental deficit. This is not a new idea, as Weinberger (1986) [189] proposed the neurodevelopmental hypothesis of schizophrenia on the basis that altered gene expression during development could be responsible for deficits in key circuitry, which are exposed when molecular changes occur in the brain during puberty, facilitating the disease phenotype. As schizophrenia is not normally diagnosed until the late teens at the earliest, gene expression measurements of individuals with schizophrenia during their early post-natal development is not possible, however, using the L100P
Disc1 mutant I have shown that there are developmental gene expression differences between the L100P mutant and wild-type littermates around the time of birth, and at P20. This is an exciting development as it not only redefines what we thought we knew about the mouse model, but can be related to an already established hypothesis in a way which may help to broaden our understanding of disease development. While only a small number of genes were tested in the L100P developmental profiles, this evidence suggests a solid neurodevelopmental basis for the schizophrenia-like phenotype exists in this model system.

8.4 Differential networking and other future work

As schizophrenia and other major mental illnesses are considered complex trait disorders, the identification of individual candidate genes only gives us a small insight into the disease cause. A more general, and perhaps more useful, tool is the identification of networks of genes involved in the disease pathology. I presented data from Ingenuity Pathway Analysis, identifying predicted gene networks on the basis of fold-change and p-value from the differential expression analysis of the microarray. Networks identified were cellular development, gene expression and cell signaling in the L100P adult, L100P embryo and Q31L embryo respectively. These proposed networks rely on the expression differences between affected and unaffected individuals (or in this case, mutant and wild-type animals). Disregulation of cell-cell signaling has previously been implicated in schizophrenia [295] and Disc1 has been shown to be involved in cellular development and neurite outgrowth ([160, 296] and others) so these networks are comparable with that data. What is more interesting, is the failure of all but two of the genes contained in these networks to validate by qRT-PCR in the mouse model, suggesting the analysis used may not be an ideal method for selecting genes for further study.

Differential networking proposes that co-expression of genes within disease groups is more important than differences in expression between groups. By performing pair-wise comparisons for co-expression it is possible to identify regulatory relationships between
genes and, eventually, extrapolate whole disease networks [297] and regulatory systems of disease. It would be interesting to carry out this analysis on the current L100P dataset to compare the two methods, and determine whether the networks generated could also be related to human disease.

The evidence for developmental disregulation of candidate genes in the L100P mutant mouse is intriguing, and further work to determine a clear developmental course for the phenotype is highly recommended. Hikida et al (2007) [193] previously showed a three month developmental delay in neuronal migration in a dominant-negative Disc1 mouse model. Kamiya et al (2005) [133] have also shown that expression of dominant-negative mutant Disc1 in the mouse leads to a delay in neuronal migration at P2, with a reduction in correctly oriented pyramidal cells in the cerebral cortex at P14. Neither experiment looked at the whole developmental time course, but selected key points in postnatal development. Experiments of this type on the L100P ENU mutant mouse would allow a direct comparison between the rate and success of neuronal migration in the developing brain, and the expression levels of Disc1 and differentially expressed genes. This would allow conclusions to be made as to the effect of gene expression levels on brain development and would provide insight into the development of the ‘schizophrenia-like’ phenotype in this model.

Once the developmental neuroanatomy of the L100P Disc1 mouse is established, it would be interesting to see if any deficits can be rescued with the use of drug treatment, and if this also rescues the behavioural phenotype. For example, the Fmr1 knockout mouse, a model of Fragile-X syndrome – the most common genetic cause of childhood cognitive impairment- shows an increase in dendritic spine maturation and improved behavioural performance when treated with minocycline, a tetracycline analogue [298] from an early age. Minocycline has also been found to have antipsychotic effects, and has been used to help treatment resistance schizophrenia [299]. It would be interesting to test the effect of this, and the other drugs used in this study, in early postnatal animals and adults. This would help us to determine to effect of treatment on behaviour and neuroanatomy, and whether early intervention can rescue both aspects of the phenotype.
8.4 Final comments

It is very clear from this and previous work that the genetics of major mental illness are incredibly complex. The Disc1 mutant mouse lines have opened the door for further studies into genetic variations in the DISC1 locus culminating in susceptibility to major mental illness. By increasing our knowledge of the knock-on effects of Disc1 mutations on down-stream gene functions and affected pathways we could be a step closer to determining the biological mechanism conferring common risk to major mental disorders. A better understanding of the aetiology of schizophrenia, bipolar disorder and other related conditions is crucial for developing rational treatments for the future.
Probe information for qRT-PCR and protein analysis. Column 1 gives gene target ID, Column 2 probe type, Column 3 supplier, Column 4 catalogue number, Column 5 concentration (where applicable) and Column 6 Secondary antibody used (where applicable).


References 235


83. Walsh T, McClellan J., McCarthy SE, Addington AM, Pierce SB, Cooper GM, Nord AS, Kusenda M, Malhotra D, Bhandari A *Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia.* Science, 2008. 320: p. 539-543.


References 239

References 240


Ekelund, J. Haplotype transmission analysis provides evidence of association for DISC1 to schizophrenia and suggests sex-dependent effects. Human Molecular Genetics, 2003. 12: p. 3151-3159.


Prata DP, Mechelli A., Fu CH, Picchioni M, Kane F, Kalidindi S, McDonald C, Kravariti E, Touloupetou T, Miorelli A, Murray R, Collier DA, McGuire PK,


References


163. K Miyoshi, Honda A, K Baba, M Taniguchi, K Oono, T Fujita, Kuroda, T Katayama, and M Tohyama, Disrupted-In-Schizophrenia 1, a candidate gene for


References


References 248

References


References 251


References


Post-analysis follow-up and validation of microarray experiments. nature genetics supplement, 2002. 32: p. 509-514.


Association study of polymorphisms in synaptic vesicle-associated genes, SYN2 and CPLX2, with schizophrenia Behavioral and Brain Functions 2005. 1(15).


