Clinically proven drug targets differentially expressed in the prefrontal cortex of schizophrenia patients

Running Title: Gene expression in schizophrenia

Joanne Voisey, Ph.D[a]*., Divya Mehta, PhD[a]., Robert McLeay, Ph.D[a]., Charles P Morris, Ph.D[a] ., Leesa F Wockner, Ph.D[b]., Ernest P Noble, Ph.D[c]., Bruce R Lawford, M.D[a]., Ross McD Young, Ph.D[a].

[a]Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia.

[b]Queensland Institute of Medical Research, Brisbane, Queensland, Australia.

[c]Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, California, USA.

[*]**Corresponding author**: Joanne Voisey

Institute of Health and Biomedical Innovation,

Queensland University of Technology,

2 George St., Brisbane, Queensland, 4000, Australia.

Ph: 61 7 31386261

Fax: 61 7 31386030

Email: j.voisey@qut.edu.au

Keywords: Brain; schizophrenia; gene expression; RNA-Seq; methylation; immune response

Manuscript word count: 3631 Abstract word count: 226

Abstract

Background: Due to the heterogeneous nature of schizophrenia, understanding the genetic risk for the disease is a complex task. Gene expression studies have proven to be more reliable than association studies as they are consistently replicated in a tissue specific manner.

Methods: Using RNA-Seq we analysed gene expression in the frontal cortex of 24 individuals with schizophrenia and 25 unaffected controls.

Results: We identified 1146 genes that were differentially expressed in schizophrenia, approximately 60% of which were up-regulated and 366 of 1146 (32%) also have aberrant DNA methylation ($p = 2.46 \times 10^{-39}$). The differentially expressed genes were significantly overrepresented in several pathways including inflammatory ($p = 8.7 \times 10^{-3}$) and nitric oxide pathways ($p = 9.2 \times 10^{-4}$). Moreover, these genes were significantly enriched for those with a druggable genome ($p = 0.04$). We identified a number of genes that are significantly up-regulated in schizophrenia as confirmed in other gene expression studies using different brain tissues. Of the 349 genes associated with schizophrenia from the Psychiatric Genomics Consortium we identified 16 genes that are significant from our list of differentially expressed genes.

Conclusions: Our results identified biological functional genes that are differentially expressed in schizophrenia. A subset of these genes are clinically proven drug targets. We also found a strong pattern of differentially expressed immune response genes that may reflect an underlying defect in schizophrenia.

Introduction

Schizophrenia is a debilitating disorder that is still poorly understood due to the complexity of the relationship between genetic and environmental risk factors [1]. Schizophrenia has a high genetic heritability and has been researched extensively with the hope of producing improved treatment options. Due to the heterogeneous nature of the disorder and the hundreds of small effect genetic variants associated with schizophrenia, a large proportion of the genetic risk is unknown. Gene expression studies performed in functional brain tissue show much stronger schizophrenia association even when performed with limited sample sizes [2]. Identifying the molecular mechanism of the functional pathways involved and the associated genetic markers may provide more clues to a more complete understanding of schizophrenia.

With the advent of next generation sequencing and high throughput techniques, new pathways have been identified including those involved in the immune system. A study of the prefrontal cortex of individuals with schizophrenia revealed an increase in inflammatory mRNA expression and identified a number of differentially expressed cytokines and immune modulators including *IL-6*, *IL-8*, *IL-1β* and *SERPINA3* [3]. There has also been schizophrenia RNA-Seq studies performed in other regions of the brain including the hippocampus. Like the study performed in the frontal cortex, genes were over-represented from pathways involved in immune/inflammation response [4]. Genetic association studies have also identified the human leukocyte antigen (HLA) locus and interleukin -1 gene complex as risk factors for schizophrenia [5, 6]. A recent molecular pathway analysis study identified genes in the toll-like receptor family and innate immunity to play a significant role in a number of pathways that are disrupted in schizophrenia [7]. Epigenetic mechanisms such as DNA methylation also effect the expression of genes and therefore it is important to investigate both gene expression and epigenetics if we are to understand the molecular mechanism of

schizophrenia. In a whole genome methylation analysis of schizophrenia we identified a number of differentially methylated genes that were previously associated with schizophrenia including *NOS1*, *AKT1*, *DTNBP1*, *DNMT1*, *PPP3CC* and *SOX10* [8]. In a further schizophrenia analysis we identified seven regions that were consistently differentially methylated in three separate cortex data sets including regions near genes *CERS3*, *DPPA5*, *PRDM9, DDX43, REC8*, *LY6G5C* and a region on chromosome ten [9]. Lymphocyte antigen 6 complex (*LY6G5C*) belongs to a cluster of leukocyte antigen-6 (LY6) genes located in the major histocompatibility complex (MHC) class III region [10]. MHC encodes 400 genes critical to immune system function and is strongly associated with schizophrenia susceptibility [11].

In this study high throughput next-generation sequencing (RNA-Seq) was used to analyse the brain transcriptome of schizophrenia patients and controls. We aimed to identify candidate genes that were differentially expressed and identify genes that are both differentially expressed and differentially methylated in the same sample set. We also verified the differentially expressed genes in an independent sample set from brain cortex using PD-NGSAtlas. A functional annotation and gene set enrichment was performed to identify statistically significantly enriched biological processes in the differentially expressed genes. Finally, we further examined the association of our differentially expressed genes with GWAS data from the Schizophrenia Working Group of the Psychiatric Genomics Consortium. By understanding the genetic and epigenetic regulatory mechanisms involved in schizophrenia we hope to gain a more complete understanding of the relationship between brain function and schizophrenia.

Methods and Materials

Samples

Frontal cortex post-mortem brain tissue from individuals with Diagnostic and Statistical Manual of Mental Disorders, 4th Edition-diagnosed schizophrenia (n=24) and controls (n=25) was provided by the Human Brain and Spinal Fluid Resource Centre, California (courtesy of James Riehl). Four samples failed quality control (QC) following RNA extraction and or library preparation with 45 samples remaining for final analysis (22 schizophrenia and 23 controls). Each sample consisted of a coronal section (7 mm thick) that had been quick frozen and a section of frontal cortex was dissected from each frozen section sample weighing (0.4-1.0 gram). 38 samples were taken from Brodmann's area 10 and 11 samples were taken from Brodmann's area 46. Demographic data including age, post-mortem interval (PMI) and gender are summarized in Supplementary Table 1. There were significant differences in age and PMI between cases and controls. These known confounds as well as unknown varibales were adjusted for in the analysis. PMI in our study is defined as the time between death and when the brain section is quick frozen. The mean (±SD) time between death and post-mortem refrigeration was 4.48 ± 3.86 hours. All but two of the participants with schizophrenia were known to be receiving antipsychotic medication at time of death. The cause of death of five schizophrenia patients was suicide.

Ethics approval for the project was obtained from the Human Research Ethics Committee of the Queensland University of Technology.

RNA Sequencing

Extraction of RNA was performed at the UCLA Clinical Microarray Core Laboratory using the Roche MagNa Pure Compact.

Quality of RNA samples was assessed by electrophoresis using the Agilent Bioanalyzer RNA 6000 Nano assay. The percentage of RNA fragments >200 nucleotides was assessed to determine the appropriate input. Samples with <30% of RNA >200 nucleotides were excluded. Forty-six samples passed the initial QC. Following sample QC, samples were processed with the TruSeq RNA Access Sample Preparation Kit as per the manufacturer's instructions (Illumina, San Diego, CA, USA). Briefly, the protocol follows the following steps: cDNA Synthesis; DNA library preparation; library validation (libraries were assessed using by electrophoresis using Agilent TapeStation D1K TapeScreen assay and quantified by PicoGreen fluorometry); coding exon enrichment; captured library validation (one sample failed to generate enough library to proceed to capture hybridisation and was excluded from the final analysis).

Two captured (eight samples) libraries were pooled for sequencing. Each pool of libraries was clustered on the Illumina cBot system using HiSeq PE Cluster Kit v4 reagents followed by sequencing on the Illumina HiSeq 2500 system with HiSeq SBS Kit v4 reagents with 159 cycles (75 base pair pared end reads). Illumina RTA 1.18.61 software was used for base calling and bcl2fastq pipeline 1.8.4 was used for quality scoring, de-multiplexing and FASTQ file generation.

Statistical Analysis and Read Normalisation

Reads were mapped to the NCBI hg38 reference genome using Tophat 2.0.13, following the protocol described by Trapnell et al. [12]. The tool featureCounts from the SubRead package 1.4.6p5 was used to assign raw read counts to exons and genes [13].

Non-normalised read counts were used for the edgeR package 3.12.0 [14, 15] to perform differential gene expression analysis after quality control and normalisation. The edgeR package normalises read counts on each gene using the TMM method [16] and then tests for

differential expression by fitting a model to the negative binomial distribution. Surrogate variable analyses (SVA) was used to correct for known confounds and hidden variables. The significance of differential gene expression was tested using a likelihood-ratio statistic.

Using the same tissue samples used in this study, 19 582 differentially methylated regions, associated with 7833 genes were identified from a previous methylation study by our group [8]. Set intersection was performed and the hypergeometric test was to test for the enrichment, this was performed in R.

Validation of RNA–Seq using PD-NGSAtlas, GWAS Loci, and Existing Disease Annotations

Using the PD-NGSAtlas web application, we downloaded differential gene analysis results in brain region BA9 between 5 schizophrenia and 6 control patients. We selected "edgeR, TMM method" as the closest equivalent of our statistical analysis. For Hwang et al and Qin Wu et al, we compared their differentially expressed lists with our list.

108 loci associated with schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium, [17] were used as further validation. All 349 genes identified by the study authors as associated with these loci (within ± 20 kb, or the nearest gene within 500 kb otherwise) were used as a validation set.

VarElect (http://varelect.genecards.org/) was used to determine which of our genes had literature annotations for "schizophrenia". VarElect reports both directly-associated genes and genes indirectly-associated genes (genes that form part of a pathway with directly-associated genes). Set intersection was performed and the hypergeometric test was to test for the enrichment, this was performed in R.

Functional Annotation

Genomatix Pathway System (GePS) that uses information extracted from public and proprietary databases to display canonical pathways or to create and extend networks based on literature data was used (http://www.genomatix.de/). Signal Transduction Pathway Associations were obtained based on all available PubMed abstracts (http://pubmed.com). Individual gene to pathway associations found on sentence level in the scientific literature were filtered for significance to avoid random matches. The significant associations were used for pathway annotations within large gene sets. Gene associations with pharmacological substances based on Genomatix literature data mining algorithm were used. Gene to pharmacological substance associations found on sentence level in the scientific literature (i.e. PubMed abstracts) were filtered for significance to avoid random matches. The significant associations were used for pharmacological substance annotations within large gene sets. An adjusted *p*-value from the results of 1,000 simulated null hypothesis queries was estimated via the GeneRanker and all reported pathways had at least 20 genes within the pathways and were significant after adjustment (p-adjusted < 0.05).

The Drug Gene Interaction database (DGIdb) was used to query for drug-gene interactions among the differentially expressed genes (http://dgidb.genome.wustl.edu/api) for a) clinically validated drug targets and b) genes with potential druggable genome (potential drug targets). The hypergeometric test was to test for the enrichment, this was performed in R.

Results

Differential Gene Expression

Following RNA extraction and library preparation 45 out of 49 post-mortem samples passed QC and library preparation. RNA-Seq data was generated from the frontal cortex of 22 individuals diagnosed with schizophrenia and 23 unaffected controls. Whole genome expression profiling was then carried out to identify differentially expressed genes. Reads

were then mapped to the human genome (hg38). For each sample, an average of 18.7 million reads were mapped, 97.7% uniquely (Supplementary Table 1). We then assigned mapped reads to each gene and identified differentially expressed genes.

SVA revealed 7 significant SVA vectors which were used as covariates in the model to correct for technical artifacts and hidden confounds [18].

A total of 1146 genes were identified as differentially expressed between schizophrenia and unaffected controls (FDR ≤ 0.05; Supplementary Table 2). A total of 659 genes were up-regulated, and 487 were down-regulated in the schizophrenia patients. The top 50 differentially expressed autosomal genes are displayed in Table 1.

Using the 7832 genes previously identified as associated with a differentially methylated site in these tissue samples [9], we tested for enrichment with differentially expressed genes. In total 366 genes were found to be in both sets ($p = 2.46 \times 10^{-39}$) Supplementary Table 3.

Table 1: Top 50 differentially expressed autosomal genes

Gene	log FC	*P* value	FDR	Chromosome
SNORD3B-1	2.745672	5.59E-214	4.16e-210	17
HBA2	-2.16737	1.30E-140	6.45e-137	16
SNORD3C	-2.94286	2.83E-117	1.05e-113	17
HBA1	-1.74136	3.21E-110	9.54e-107	16
FOS	1.640928	5.28E-077	1.31e-73	14
HBB	-1.9139	1.81E-076	3.85e-73	11
SNORD116-5	-1.55088	6.76E-030	1.26e-26	15
CCL2	1.399308	1.84E-024	3.04e-21	17
CD163	-0.86852	7.13E-024	1.06e-20	12
EGR1	0.932135	1.29E-023	1.74e-20	5
TXNIP	-0.7016	2.82E-023	3.49e-20	1
SNORD3D	-1.47839	4.16E-023	4.75e-20	17
CP	-0.87183	1.17E-022	1.24e-19	3
SNORA22	0.695145	7.10E-021	7.04e-18	7
DIO2	0.935332	9.67E-020	8.98e-17	14
CYR61	0.824907	5.48E-019	4.56e-16	1
RN7SK	-0.6192	5.52E-019	4.56e-16	6
HLA-DRB5	-0.81587	8.97E-019	7.02e-16	6
IGFN1	0.839495	6.77E-018	5.03e-15	1
CARTPT	-0.66133	1.41E-017	9.95e-15	5
VEGFA	0.673787	1.89E-016	1.23e-13	6
SCD	0.615229	1.90E-016	1.23e-13	10
LOC100288778	0.714972	2.42E-016	1.5e-13	12
SNORD13	0.895811	2.76E-016	1.64e-13	8
SST	0.746259	2.66E-015	1.52e-12	3
HSPB1	-0.52068	5.62E-014	2.98e-11	7
PDK4	-0.54891	2.20E-013	1.13e-10	7
NR4A1	0.77781	2.56E-013	1.27e-10	12
ROS1	-0.61903	7.42E-013	3.56e-10	6
RN7SL1	-0.526	1.55E-012	7.18e-10	14
MRC1	-0.53699	1.83E-012	8.24e-10	10
F3	0.537702	2.30E-012	9.79e-10	1
EGR2	1.087177	2.30E-012	9.79e-10	10
BCYRN1	-0.5929	3.10E-012	1.28e-09	2
OVOS	1.561954	4.87E-012	1.95e-09	12
SNORD116-3	-0.67094	9.92E-012	3.88e-09	15
COL11A1	0.492673	1.04E-011	3.97e-09	1
SNORA7B	0.577825	1.07E-011	3.97e-09	3
FAM189A2	0.640753	1.15E-011	4.19e-09	9
NOMO3	-0.59631	1.40E-011	4.95e-09	16
HIST1H4D	-0.58824	1.98E-011	6.57e-09	6
PHLDB1	0.471543	1.98E-011	6.57e-09	11
RYR3	0.485461	1.99E-011	6.57e-09	15
IL1B	0.89181	2.18E-011	7.04e-09	2
WIF1	0.577042	2.23E-011	7.06e-09	12
GRIN2C	0.563355	2.61E-011	8.08e-09	17
FKBP5	-0.50593	3.36E-011	1.02e-08	6
NR4A2	0.58417	6.24E-011	1.85e-08	2
FADS2	0.460321	7.69E-011	2.24e-08	11
SREBF1	0.528543	7.84E-011	2.24e-08	17

Functional Annotation of Differentially Expressed Genes

The Drug Gene Interaction database (DGIdb) was used to query for drug-gene interactions among the differentially expressed genes. Our gene list was significantly enriched for genes with a druggable genome (n = 217, p = 0.04) (Supplementary Table 10). The Drug Gene Interaction database looks for interactions between known drugs and their target genes, therefore the 217 genes identified by this database depict genes that have a potentially druggable genome. Additionally 23 of the significantly differentially expressed genes matched clinically proven drug targets. These genes include *ROS1*, *HIST1H4B*, *STAT4*, *ETV5*, *FGFR3*, *TNFAIP3*, *HIST1H3A*, *FGFR2*, *EPCAM*, *PDGFRA*, *FAM46C*, *MET*, *SHQ1*, *PDPK1*, *MDM4*, *IDH2*, *DOT1L*, *SOX9*, *EWSR1*, *SPOP*, *CDKN1A*, *EPHB1* and *AXL*.

Finally, functional annotation using the Genomatix literature data mining algorithm identified significant pathways enriched for our differentially expressed genes. Antipsychotic agents was the top enriched category within the pharmacological substances in genomatix (n = 20 genes p = 6.7e-5) (Figure 1a). Top signal transduction pathways enriched within the differentially expressed genes are listed in Table 2. As reported in previous studies [3, 4, 19], inflammatory genes showed significant up-regulation in schizophrenia patients, and were significantly enriched (n = 42 genes, p = 0.00876, Figure 1b). Other relevant pathways include Mitogen active protein kinase (MAPK) and Nitric oxide Figure 1c) (which have established roles in schizophrenia development. Our functional annotation validates a previous study that identified genes enriched in pathways related to immune response and MAPK cascade [7]. A list of all the genes within these pathways is listed in Supplementary Table 4.

Table 2. Top pathways enriched within the differentially expressed genes

Signal Transduction Pathway (Genomatix)	Number of genes in pathway	*P*-value
Mitogen activated protein kinase	66	1.14E-04
Nitric oxide	22	9.20E-04
Epidermal growth factor receptor	43	2.07E-03
Interleukin 1	26	8.24E-03
Inflammatory	42	8.76E-03

Validation of RNA-Seq using PD-NGSAtlas, GWAS Loci, other Brain Regions and VarElect

We subsequently validated the differentially expressed genes using data from the PD-NGSAtlas [20]. The PD-NGSAtlas is a publically available database of next-generation sequencing data from different brain regions in schizophrenic, bipolar, and control individuals. We selected brain region Brodman Area 9 (BA9), as it was the largest frontal cortex region available in the PD-NGSAtlas. Gene expression data in the BA9 region were available for 5 schizophrenia patients and 6 unaffected controls (Supplementary Table 5). Comparing our RNA-Seq results (n= 951 present in the PD-NGSAtlas) to differentially expressed genes in this region in the PD-NGSAtlas, 141 were significant (see Supplementary Table 6), this was greater than expected by chance (p = 1.984e-09).

In 2014, the Schizophrenia Working Group of the Psychiatric Genomics Consortium (PGC) published a very large genome-wide association study (GWAS) of 36 989 cases and > 100 000 controls [17]. In this study, they identified only 108 loci and then mapped each locus to all genes within ±20 kb, or to the nearest gene within 500 kb. Intersecting these 349 genes with our differentially expressed genes, we found 16 in common. We report these genes in Table 3.

Table 3: Top differentially expressed genes in both QUT+ PGC GWAS

Gene	logFC	*P* Value	FDR
C4orf27	-0.25957	0.003502	0.046682
CHADL	0.400274	0.000944	0.018193
DGKI	-0.38078	1.79E-08	2.80E-06
EGR1	0.932135	1.29E-23	1.74E-20
GALNT10	0.372323	0.000968	0.018375
GOLGA6L4	0.397727	4.18E-05	0.001775
GRM3	0.224529	0.002745	0.039696
HAPLN4	-0.43547	0.000641	0.013703
ITIH4	0.422974	0.000246	0.006849
NCAN	0.248411	0.000376	0.009352
PCDHA10	0.279487	0.00336	0.045409
PLCH2	0.371213	9.81E-06	0.000595
PTN	0.244742	0.00202	0.032282
SREBF1	0.528543	7.84E-11	2.24E-08
YPEL3	-0.27907	0.001033	0.019272
YPEL4	-0.35367	0.000229	0.006576

We also found significant gene enrichment compared to schizophrenia studies performed in other brain regions. Hwang et al. [4] identified 144 differentially expressed genes in the hippocampus and 20 of our differentially expressed genes were included in this list (*p*-value = 0.006984) (Supplementary Table 7) Gene enrichment analysis of a further study performed in cortical grey matter identified 87 of their 772 genes within our list (*p*-value = 5.035e-06) (Supplementary Table 8) [21].

Finally, we used VarElect (http://varelect.genecards.org/about/) to identify whether the identified genes had previously shown direct or indirect association with schizophrenia or schizoaffective disorder. After excluding 52 gene identifiers not included within the GeneCards database (http://www.genecards.org/), we found that 196 of 1117 differentially expressed genes had direct literature evidence of schizophrenia association. An additional 869 differentially expressed genes were reported as having an indirect association, defined as being a member of a pathway or complex with one or more directly associated genes. We include the full list of associated genes in Supplementary Table 9.

Discussion

GWAS and genetic association studies are yet to identify a large component of the schizophrenia genetic risk. Therefore transcriptome studies that identify genes based on function may provide more promise. Using a functionally relevant tissue we employed RNA-Seq to analyse prefrontal cortex tissue from 24 individuals with schizophrenia and 25 unaffected controls. We identified 1146 genes that are differentially expressed in the prefrontal cortex of individuals with schizophrenia. Nearly 60 % of these genes were up-regulated in schizophrenia. A number of genes from our list have validated schizophrenia genetic and or mRNA expression association, some of which include serotonin receptors, *BDNF*, *FKBP5*, *GRM3*, *RELN*, *GRIN2C* and RGS9. Gene enrichment analysis of the differentially expressed genes identified 23 clinically proven drug targets as well as 20 antipsychotic agents which include *BDNF, RELN*, *HTR1A* and *SLC6A9*. Other potential drug targets include *RYR1* and *RYR3* which are ryanodine receptors that mediate the release of Ca(2+) from intracellular stores in neurons, and may thereby promote prolonged Ca(2+) signalling in the brain. Other interesting drug targets are *GRIN2C* and *GRIN3A* which are glutamate receptors that encode a subunit of the N-methyl-D-aspartate (NMDA) receptor. NMDA receptors are found in the central nervous system, are permeable to cations and have an important role in physiological processes such as learning, memory, and synaptic development. *GRIN2C* has been previously associated with differential cortical NMDA receptor expression in schizophrenia[22].

Immune function showed strong gene enrichment and up-regulation of the immune system in schizophrenia has been reported in other prefrontal cortex studies [3, 23] but also in studies performed in the hippocampus [4]. The latter study reported three genes *IFITM1, IFITM2* and *IFITM3* that were consistently up-regulated in the hippocampus of those with schizophrenia

and members of the *interferon inducible transmembrane protein* (IFITM) family. *IFITM2* was also significantly up-regulated in our schizophrenia study. Another gene of interest from our list of differentially expressed genes is IL-1β which not only encodes a cytokine which is a significant mediator in inflammatory response but it also confers increased schizophrenia risk [24] IL-1β also plays a crucial role in neurodevelopment [25] and may contribute to the pathophysiology of schizophrenia.

Using microarray data from combined studies of prefrontal cortices (n = 315 and nine separate studies) a recent transcriptome mega-analyses identified immune gene expression to be dysregulated in schizophrenia [26]. Out of the 92 genes found to be differentially expressed two genes reached significance after multiple correction including *RHOBTB3* and *ABCA1*. The ATP-binding cassette transporter A1 (*ABCA1*) was also significantly up-regulated in our list. This gene plays an essential role in the regulation of extracellular cholesterol levels in the cerebrospinal fluid [27]. A polymorphism in the *ABCA1* gene is also associated with increased schizophrenia susceptibility and lower grey matter volume in schizophrenia patients [28]. Given that cholesterol is crucial for myelination and synaptic development and *ABCA1* reaches significance in a large transcriptome meta-analysis [26] this may be a potential target for future research.

It is not known how increased levels of these immune-related genes contribute to the clinical symptoms of schizophrenia. However, it is now widely accepted that an immune challenge caused by a prenatal infection increases schizophrenia development by two fold (reviewed in [29]). Further research is required to understand whether the changes in the prefrontal cortex are a result of a chronic infection that leads to immune activation throughout the disorder or whether an early infection changes the immune signature in individuals that are genetically susceptible. However, an earlier study did confirm that *IFITM2* mRNA levels were increased in schizophrenia and we not just the result of an infection prior to death [30].

A high proportion of membrane related genes was also represented in our list, especially those involved in myelination and myelin metabolism. This observation strongly indicates the presence of a myelin disorder in the brains of these schizophrenia patients. An underlying defect in myelination in schizophrenia is further supported by the observation of genetic association with schizophrenia or altered mRNA expression in prefrontal cortex in schizophrenia for several genes involved in myelination (*CNP*, *MOBP* and *MOG* [31, 32 33]. When combined with the observation of a disturbance in immune function in schizophrenia prefrontal cortex, we believe that it is possible that there is an underlying immune function disorder in schizophrenia that results in a defect in myelination. In support of this, it is well known that multiple sclerosis is an autoimmune disorder which results in demyelination [34].

We had previously generated a whole genome methylation data set from these samples [8]. Using this methylation data set we were able to identify 366 genes that were both differentially methylated and differentially expressed in schizophrenia patients compared to controls. This list of differentially expressed genes that have correlated aberrant methylation patterns may prove promising candidates for identifying the underlying mechanisms of schizophrenia. Among this list of genes is NeuroD1 which is a transcription factor that plays an essential role in embryonic neurogenesis [35]. More recently it has been shown to reprogram other cell types into neurons including astrocytes into glutamatergic neurons [36]. This transcriptional regulation occurs through epigenetic memory even though NeuroD1 is transiently induced during neurogenesis [37].

Although RNA-Seq is a reliable method for gene expression profiling and is more sensitive compared to microarrays [38], we validated our list of differentially expressed genes using PD-NGSAtlas. We were able to confirm 141 out of the 951 genes identified by PD-NGSAtlas. We also identified 16 genes from our differentially expressed list that were found associated with schizophrenia in a large schizophrenia GWAS [17]. Of particular interest is glutamate

receptor, metabotropic 3 (*GRM3*) which plays a role in glutamatergic neurotransmission and is essential for normal brain function. This gene has a long established role in the pathogenesis of schizophrenia [39, 40]. A recent schizophrenia imaging and genetics study has also shown that *GRM3* is important in in prefrontal cortex activation [41].

There are limitations to this study including the use of frontal cortex tissue which has numerous cell types. Future studies that identify transcriptome changes in a cell specific manner will produce more specificity. Another limitation was patients were not free of antipsychotic medication which can influence gene expression profiling. This is a difficult limitation to address as those with chronic schizophrenia are almost certain to be treated with antipsychotic medication. Other confounds such as substance use were also not available in this study cohort. We acknowledge that it is likely that other factors might influence gene expression, nevertheless, we used SVA that takes into account hidden confounds when performing the analysis. Brain banks or biobanks that collect a comprehensive medical history of patients will result in future studies that can control for antipsychotic medication.

In conclusion our results identified biological functional genes that are differentially expressed in schizophrenia. We found a strong pattern of differentially expressed immune response and myelination genes that may reflect an underlying defect in schizophrenia. By integrating genetic, methylation and transcriptome data from functionally relevant tissues we are more likely to identify and validate schizophrenia susceptibility loci.

Acknowledgements

This work was financially supported by the Nicol Foundation and the Institute of Health and Biomedical Innovation, QUT.

Financial Disclosures

There are no financial or other relationships that lead to a conflict of interest in the publication of this work. All authors named in this article state they have nothing to declare.

References

1. van Os J, Rutten BP, Myin-Germeys I, et al. Identifying gene-environment interactions in schizophrenia: contemporary challenges for integrated, large-scale investigations. *Schizophr Bull* 2014;40:729-736.

2. Horvath S, Mirnics K. Schizophrenia as a disorder of molecular pathways. *Biol Psychiatry* 2015;77:22-28.

3. Fillman SG, Cloonan N, Catts VS, Miller LC, Wong J, McCrossin T, Cairns M, Weickert CS. Increased inflammatory markers identified in the dorsolateral prefrontal cortex of individuals with schizophrenia. *Mol Psychiatry* 2013;18:206-214.

4. Hwang Y, Kim J, Shin JY, Kim JI, Seo JS, Webster MJ, Lee D, Kim S. Gene expression profiling by mRNA sequencing reveals increased expression of immune/inflammation-related genes in the hippocampus of individuals with schizophrenia. *Translational psychiatry* 2013;3:e321.

5. Ripke S, O'Dushlaine C, Chambert K, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet* 2013;45:1150-1159.

6. Xu M, He L. Convergent evidence shows a positive association of interleukin-1 gene complex locus with susceptibility to schizophrenia in the Caucasian population. *Schizophr Res* 2010;120:131-142.

7. Crisafulli C, Drago A, Calabro M, Spina E, Serretti A. A molecular pathway analysis informs the genetic background at risk for schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry* 2015;59:21-30.

8. Wockner LF, Noble EP, Lawford BR, Young RM, Morris CP, Whitehall VL, Voisey J. Genome-wide DNA methylation analysis of human brain tissue from schizophrenia patients. *Translational psychiatry* 2014;4:e339.

9. Wockner LF, Morris CP, Noble EP, Lawford BR, Whitehall VL, Young RM, Voisey J. Brain-specific epigenetic markers of schizophrenia. *Translational psychiatry* 2015;5:e680.

10. Mallya M, Campbell RD, Aguado B. Characterization of the five novel Ly-6 superfamily members encoded in the MHC, and detection of cells expressing their potential ligands. *Protein science : a publication of the Protein Society* 2006;15:2244-2256.

11. Corvin A, Morris DW. Genome-wide association studies: findings at the major histocompatibility complex locus in psychosis. *Biol Psychiatry* 2014;75:276-283.

12. Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 2012;7:562-578.

13. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;30:923-930.

14. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139-140.

15. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* 2012;40:4288-4297.

16. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 2010;11:R25.

17. Consortium SWGotPG. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014;511:421-427.

18. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS genetics* 2007;3:1724-1735.

19. Kim S, Hwang Y, Webster MJ, Lee D. Differential activation of immune/inflammatory response-related co-expression modules in the hippocampus across the major psychiatric disorders. *Mol Psychiatry* 2016;21:376-385.

20. Zhao Z, Li Y, Chen H, et al. PD_NGSAtlas: a reference database combining next-generation sequencing epigenomic and transcriptomic data for psychiatric disorders. *BMC medical genomics* 2014;7:71.

21. Wu JQ, Wang X, Beveridge NJ, Tooney PA, Scott RJ, Carr VJ, Cairns MJ. Transcriptome sequencing revealed significant alteration of cortical promoter usage and splicing in schizophrenia. *PLoS One* 2012;7:e36351.

22. Dean B, Gibbons AS, Boer S, Uezato A, Meador-Woodruff J, Scarr E, McCullumsmith RE. Changes in cortical N-methyl-D-aspartate receptors and post-synaptic density protein 95 in schizophrenia, mood disorders and suicide. *Aust N Z J Psychiatry* 2016;50:275-283.

23. Arion D, Unger T, Lewis DA, Levitt P, Mirnics K. Molecular evidence for increased expression of genes related to immune and chaperone function in the prefrontal cortex in schizophrenia. *Biol Psychiatry* 2007;62:711-721.

24. Lewis CM, Levinson DF, Wise LH, et al. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet* 2003;73:34-48.

25. Nawa H, Takahashi M, Patterson PH. Cytokine and growth factor involvement in schizophrenia--support for the developmental model. *Mol Psychiatry* 2000;5:594-603.

26. Hess JL, Tylee DS, Barve R, et al. Transcriptome-wide mega-analyses reveal joint dysregulation of immunologic genes and transcription regulators in brain and blood in schizophrenia. *Schizophr Res* 2016.

27. Wahrle SE, Jiang H, Parsadanian M, Legleiter J, Han X, Fryer JD, Kowalewski T, Holtzman DM. ABCA1 is required for normal central nervous system ApoE levels and for lipidation of astrocyte-secreted apoE. *J Biol Chem* 2004;279:40987-40993.

28. Ota M, Fujii T, Nemoto K, et al. A polymorphism of the ABCA1 gene confers susceptibility to schizophrenia and related brain changes. *Prog Neuropsychopharmacol Biol Psychiatry* 2011;35:1877-1883.

29. Meyer U, Feldon J, Yee BK. A review of the fetal brain cytokine imbalance hypothesis of schizophrenia. *Schizophr Bull* 2009;35:959-972.

30. Saetre P, Emilsson L, Axelsson E, Kreuger J, Lindholm E, Jazin E. Inflammation-related genes up-regulated in schizophrenia brains. *BMC psychiatry* 2007;7:46.

31. Mitkus SN, Hyde TM, Vakkalanka R, Kolachana B, Weinberger DR, Kleinman JE, Lipska BK. Expression of oligodendrocyte-associated genes in dorsolateral prefrontal cortex of patients with schizophrenia. *Schizophr Res* 2008;98:129-138.

32. Cannon DM, Walshe M, Dempster E, Collier DA, Marshall N, Bramon E, Murray RM, McDonald C. The association of white matter volume in psychotic disorders with genotypic variation in NRG1, MOG and CNP: a voxel-based analysis in affected individuals and their unaffected relatives. *Translational psychiatry* 2012;2:e167.

33. Peirce TR, Bray NJ, Williams NM, et al. Convergent evidence for 2',3'-cyclic nucleotide 3'-phosphodiesterase as a possible susceptibility gene for schizophrenia. *Arch Gen Psychiatry* 2006;63:18-24.

34. McFarland HF, Martin R. Multiple sclerosis: a complicated picture of autoimmunity. *Nature immunology* 2007;8:913-919.

35. Hevner RF, Hodge RD, Daza RA, Englund C. Transcription factors in glutamatergic neurogenesis: conserved programs in neocortex, cerebellum, and adult hippocampus. *Neurosci Res* 2006;55:223-233.

36. Guo Z, Zhang L, Wu Z, Chen Y, Wang F, Chen G. In vivo direct reprogramming of reactive glial cells into functional neurons after brain injury and in an Alzheimer's disease model. *Cell stem cell* 2014;14:188-202.

37. Pataskar A, Jung J, Smialowski P, Noack F, Calegari F, Straub T, Tiwari VK. NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program. *EMBO J* 2016;35:24-45.

38. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews Genetics* 2009;10:57-63.

39. Egan MF, Straub RE, Goldberg TE, et al. Variation in GRM3 affects cognition, prefrontal glutamate, and risk for schizophrenia. *Proc Natl Acad Sci U S A* 2004;101:12604-12609.

40. Cherlyn SY, Woon PS, Liu JJ, Ong WY, Tsai GC, Sim K. Genetic association studies of glutamate, GABA and related genes in schizophrenia and bipolar disorder: a decade of advance. *Neurosci Biobehav Rev* 2010;34:958-977.

41. Kinoshita A, Takizawa R, Koike S, et al. Effect of metabotropic glutamate receptor-3 variants on prefrontal brain activity in schizophrenia: An imaging genetics study using multi-channel near-infrared spectroscopy. *Prog Neuropsychopharmacol Biol Psychiatry* 2015;62:14-21.

Figure Legends

Figure 1: Functional annotation of differentially expressed genes and cellular localisation.

1a. Antipsychotic agents.

Gene associations with pharmacological identified 20 genes as antipsychotic agents (p = 6.7e-5).

1b. Inflammation.

Inflammatory genes were significantly enriched (n = 42 genes, p = 8.76E-03).

1c. Nitric oxide.

Genes within the nitric oxide cascade were significantly enriched (n = 22 genes, p = 9.20E-04).

Supplementary Data:

Supplementary Table 1: Demographics and sequencing reads for control and schizophrenia patients

Supplementary Table 2: List of all differentially expressed genes.

Supplementary Table 3: Top differentially expressed genes in our RNA-Seq + Methylation data

Supplementary Table 4: Significant pathways and list of genes enriched in these pathways.

Supplementary Table 5: Demographics of replication data set (PD-NGSAtlas)

Supplementary Table 6: Validation of our genes with PD-NGSAtlas.

Supplementary Table 7: Validation of our genes with hippocampus data set.

Supplementary Table 8: Validations with cortical grey matter data set.

Supplementary Table 9: Genes with direct or indirect association with schizophrenia using VarElect.

Supplementary Table 10. Enrichment for genes using the Drug Gene Interaction database.