Understanding How Genetics and Environments Shape the Development of Schizophrenia

Choi Shing Wan

A thesis submitted in partial fulfillment of the requirements for
the Degree of Doctor of Philosophy

Department of Psychiatry
University of Hong Kong
Hong Kong
January 3, 2016

Abstract

Schizophrenia (SCZ) is a detrimental disorder affecting approximately 1% of the population worldwide. To fully understand the disease mechanism for the development of proper treatments, it is important not only to examine how certain genetic polymorphisms can predispose individuals to the disease development, but also how environmental factors triggers the disorder in apparently healthy individuals.

Genome Wide Association Study (GWAS) is now a standard approach for investigating associations of common genetic variations (mainly Single Nucleotide Polymorphisms (SNPs)) with SCZ. A recent meta-analysis of GWAS of SCZ has identified 108 loci significantly associated with SCZ. However, due to the limitation of sample size and the moderate-to-small effect size of an unknown number of causal loci, many SNPs associated with SCZ may be left undetected and a much larger sample size of GWAS may be required. However, it is also possible that these 108 loci have already contained all or near most of the SNPs associated with the disease. So estimating the contribution of these common SNPs to SCZ has important implications for future research strategy.

In this thesis, we proposed an alternative approach for estimating the contribution of SNPs to SCZ (SNP-heritability) from GWAS summary statistics, called the SNP HeRitability Estimation Kit (SHREK). Our simulation results suggested that when compared to the existing method (LD SCore regression (LDSC)), SHREK provided a more robust estimate for oligogentic traits and in case-control designs in which no confounding variables was present. Using the summary statistics from the latest

meta-analysis of GWAS of SCZ, we estimated that SCZ has a SNP-heritability of 0.174 (SD=0.00453), which is similar to the estimate of 0.197 (SD=0.0058) by our competitor LDSC. The result indicated that common SNPs have relatively less contribution to the genetic predisposition of individuals to SCZ as measured by the heritability estimated. Also, it suggested that alternative strategies like whole genome sequencing would be more efficient for identifying additional SCZ genes, compared to GWAS.

On the other hand, prenatal infection has been identified as the single largest environmental risk factor of SCZ. It was observed that a wide variety of infections are associated with the increased SCZ risk in the offspring. This suggests that maternal immune activation (MIA) during prenatal development may have a negative impact on fetal brain functions as well as behaviors. So it is important to understand how MIA triggers the disorder by examining the molecular events that take place in the cerebellum using established animal models, such as those involving the viral RNA mimic polyriboinosinic-polyribocytidilic acid (PolyI:C).

As a result, we also performed a RNA-sequencing study for the MIA on the change in global gene expressions in the fetal cerebellum in PolyI:C-treated pregnant mice. We found that several pathways related to neural functioning and calcium ion signaling were likely to be disrupted by MIA in the cerebellum. In addition, we investigated how a n-3 polyunsaturated fatty acid (PUFA) rich diet can help to reduce the SCZ-like phenotype in mice exposed to early MIA insults. We found that *Sgk1*, a gene that regulates the glutamatergic system, is potentially affected by the n-3 PUFA rich diet in the PolyI:C exposed mice. In conclusion, our results suggested that genes related to neural function or calcium ion signaling, as well as glutamate-related genes such as *Sgk1*, are potential targets for future SCZ research.

(550 words)

Declaration

I declare that this thesis represents my own work, except where due acknowledgments is made, and that it has not been previously included in a thesis, dissertation or report submitted to this University or to any other institution for a degree, diploma or other qualification.

Signed.....

Choi Shing Wan

Acknowledgements

I would like to express my deepest gratitude to Professor Pak Sham. I am eternally grateful for his trust, supervision, patience and support in the course of my study. I would also like to thanks Dr Stacey Cherny and Dr Wanling Yeung for giving me valuable advice for my projects. My special thanks go to Dr Johnny Kwan. He has provided critical advices on my projects and has taught me a great deal in the field of statistic.

The past 4 years has been a blast and I really enjoy my time in this department. This is only possible because of all the great people here. Thank you Beatrice Wu, Dr Li Qi, Tomy Hui, Vicki Lin, Nick Lin, John Wong, Dr Clara Tang, Dr Amy Butler, Dr Emily Wong, Dr Allen Gui, Dr Sylvia Lam, Yung Tse Choi, Oi Chi Chan, Pui King Wong and Dr Miaoxin Li, without you everything will be much different. I will forever cherish the time I spent with you.

Words alone cannot express my gratitude to Beatrice Wu and my family. Their support and encouragement have been my greatest source of energy and have helped me to continue on with my study.

THANK YOU!

Abbreviations

bp	base pair.
DEG	differentially expressed gene.
ECM	extracellular matrix.
EGF	epidermal growth factor.
ERCC	External RNA Controls Consortium.
FGF	fibroblast growth factor.
GD	Gestation Day.
GWAS	Genome Wide Association Study.
IL-6	Interleukin-6.
kb	kilobase.
LD	Linkage Disequilibrium.
LDSC	LD SCore regression.
LRT	likelihood ratio test.
maf	minor allele frequency.
MAPK	mitogen-activated protein kinase.
MIA	maternal immune activation.
MMP	matrix metalloproteinase.
MSigDB	Molecular Signatures Database.
NGS	next generation sequencing.
PC	Principle Component.
PCA	principle component analysis.
PET	positron emission tomography.
PGC	Psychiatric Genomics Consortium.
PI3K	phosphatidylinositol 3-kinase.
PolyI:C	polyriboinosinic-polyribocytidilic acid.
PUFA	polyunsaturated fatty acid.
QC	quality control.

RIN	RNA integrity number.
rt-PCR	real time PCR.
SCZ	schizophrenia.
SE	standard error.
SHREK	SNP HeRitability Estimation Kit.
SNP	Single Nucleotide Polymorphism.

Contents

List of Figures

List of Tables

3 n-3 Polyunsaturated Fatty Acid Rich Diet in Schizophrenia

3.1 Introduction

In the previous chapter, we have found that the Single Nucleotide Polymorphism (SNP) heritability of schizophrenia is estimated to be at most 20% which is much lower than expected in schizophrenia. This suggests that other factors such as rare variants such, epigenetic factors and gene-environmental interaction ($G \times E$) might contribute to the "missing" heritability of schizophrenia.

Previous studies have reported the possibility of interaction between prenatal infection and genetic variation in risk of developing schizophrenia (Tienari et al., 2004; Clarke et al., 2009). It has been suggested that the effect of prenatal infection was mainly mediated by maternal immune response, instead of the specific type of infection (Brown and Derkits, 2010). Therefore it is likely that the perturbation induced by maternal immune activation (MIA) interacts with genetic variations in the development of schizophrenia.

Slowly but steadily, progress has been made in the research of schizophrenia. Converging evidence from Genome Wide Association Study (GWAS), copy number variation (CNV) and sequencing studies suggest that rare and common

variants in genes related to postsynaptic density (PSD) (Purcell et al., 2014; T. N. Consortium and Psychiatric Genomics, 2015) and calcium ion channels (Purcell et al., 2014; Ripke et al., 2014; Szatkiewicz et al., 2014) contribute to the etiology of schizophrenia. Given these, it is possible for the effect of prenatal infection to also act upon the same functional gene sets during the development of schizophrenia.

Additionally, with the development of LD SCore regression (LDSC), partitioning of SNP heritability can now be performed using summary statistics from GWAS. It is therefore possible to not only identify gene sets or pathways that are associated with schizophrenia, but also estimate their relative contribution to the heritability of schizophrenia.

Furthermore, a number of studies have reported the potential of n-3 polyunsaturated fatty acid (PUFA) in the treatment of schizophrenia (Li, Leung, et al., 2015; Trebble et al., 2003). In mouse, it was found that n-3 PUFA can inhibits the production of Interleukin-6 (IL-6) (Trebble et al., 2003) - a major mediator in MIA model (Smith et al., 2007). Apart from its anti-inflammatory property, n-3 PUFA such as docosahexaenoic acid (DHA) also plays a critical role in the development of central nervous system (Clandinin, 1999; Kitajka et al., 2002). Given its strong implication in neuronal functioning, it is possible that n-3 PUFA rich diet may reduce the symptoms of schizophrenia, as reported by a recent study (Li, Leung, et al., 2015).

Herein, we conduct a hypothesis-driven study to investigate the gene expression changes induced by early MIA exposure in the brain of the adult offspring, and also expression changes induced by n-3 PUFA rich diet using RNA Sequencing. Another goal of the current study is to investigate whether the effects of MIA or diet act upon the same functional gene sets as the genetic variants associated with schizophrenia in the development of the disease. Finally, the relative contribution of the candidate gene sets to the heritability of schizophrenia was also estimated

using LDSC.

Although hippocampus (Velakoulis et al., 2006; Nugent et al., 2007) and prefrontal cortex (Knable and Weinberger, 1997; Perlstein et al., 2001) are the two brain regions that have been extensively studied in schizophrenia, cerebellum dysfunction has also been reported in schizophrenia (Yeganeh-Doost et al., 2011; Andreasen and Pierson, 2008). Specifically, positron emission tomography (PET) studies have shown that a dysfunction in the cortico-cerebellar-thalamic-cortical neuronal circuit, which contributes to "cognitive dysmetria", e.g. impaired cognition, and other symptoms of schizophrenia (Yeganeh-Doost et al., 2011). Taken together, cerebellum might plays an important part in the etiology of schizophrenia and are therefore selected for the current study.

The work in this chapter were done in collaboration with my colleagues who have kindly provide their support and knowledges to make this piece of work possible. Dr Li Qi and Dr Basil Paul were responsible for generating the animal model and providing the sample for our study; Dr Li Qi and Dr Desmond Campbell helped with the experimental design; Vicki Lin has helped with the RNA extraction; Tikky Leung for her high quality sequencing service; Nick Lin for his help in tackling problems encountered during sequencing quality control; Dr Johnny Kwan, Dr Desmond Campbell, Dr Timothy Mak and Professor Sham for their guidance in the statistical analysis.

3.2 Methodology

3.2.1 Sample Preparation

Female and male C57BL6/N mice were bred and mated by The University of Hong Kong, Laboratory Animal Unit. Timed-pregnant mice were held in a normal

light–dark cycle (light on at 0700 hours), and temperature and humidity-controlled animal vivarium. All animal procedures were approved by the Committee on the Use of Live Animals in Teaching and Research (CULATR) at The University of Hong Kong.

The MIA model was generated following procedures previously reported (Li, Cheung, et al., 2009). A dose of 5mg kg^{-1} polyriboinosinic-polyribocytidilic acid (PolyI:C) in an injection volume 5ml kg^{-1}, prepared on the day of injection was administered to pregnant mice on Gestation Day (GD) 9 via the tail vein under mild physical constraint. Control animals received an injection of 5ml kg^{-1} 0.9% saline. The animals were returned to the home cage after the injection and were not disturbed, except for weekly cage cleaning. The resulting offspring were weaned and sexed at postnatal day 21. The pups were weighed and littermates of the same sex were caged separately, with three to four animal per cage. Half of the animal were fed on diets enriched with n-3 PUFAs and half were fed a standard lab diet until the end of the study. The latter 'n-6 PUFA' control diet had the same calorific value and total fat content as the n-3 PUFA diet. The diets were custom prepared and supplied by Harlan Laboratories (Madison, WI, USA). The n-6 and n-3 PUFA were derived from corn oil or menhaden fish oil, respectively. The n-6 PUFA control diet, was based on the standard AIN-93G rodent laboratory diet (Reeves, Nielsen, and Fahey, 1993), and contained 65 g kg^{-1} corn oil and 5 g kg^{-1} fish oil with an approximate (n6)/(n3) ratio of 13:1. The n-3 PUFA diet contained 35 g kg^{-1} corn oil and 35 g kg^{-1} fish oil with an approximate (n6)/(n3) ratio of 1:1 (Olivo and Hilakivi-Clarke, 2005). To avoid being confounded by sex difference, we only use the male offspring for our analysis. The male offspring were sacrificed by cervical dislocation on postnatal week 12, which roughly correspond to adulthood in human, and the cerebellum was extracted and stored in -80°C until RNA extraction.

SampleID	Litter	Diet	Condition	Lane	Batch	Rin
B1	3	O3	POL	1	B	7.7
B2	6	O3	POL	2	B	7.7
F1	4	O3	POL	1	F	7.6
F4	1	O3	SAL	2	F	8.1
B4	5	O3	SAL	1	B	7.8
B5	14	O3	SAL	2	B	7.7
F2	2	O6	POL	1	F	7.5
E3	11	O6	POL	2	E	7.8
C2	7	O6	POL	2	C	7.9
B6	13	O6	SAL	2	B	7.4
E6	14	O6	SAL	1	E	8
C6	1	O6	SAL	1	C	7.8

Table 3.1: Sample information. O3 = n-3 PUFA diet; O6 = n-6 PUFA diet; POL = PolyI:C exposed; SAL = Saline exposed. We have tried to separate the samples into different lane and batch to control for the lane and batch effect. Samples from different litters were also used with the exception of F4 and C6 which came from the same litter but were given a different diet.

3.2.2 RNA Extraction, Quality Control and Sequencing

Total RNA was extracted from each cerebellum tissue using RNeasy midi kit (Qiagen) following the manufacturer's instructions. RNA quality was assayed using the Agilent 2100 Bioanalyzer and RNA was quantified using Qubit 1.0 Flurometer. Samples with RNA integrity number (RIN) < 7 were not included in our study as the RNA are most likely degraded. As a hypothesis generation study, we select a minimum of 3 samples per group and each samples must come from a different litter to control for littering effect. The RNA Sequencing library was performed at the Centre for Genomic Sciences, the University of Hong Kong, using the KAPA Stranded mRNA-Seq Kit. All samples were sequenced using Illumina HiSeq 1500 at 2 lanes (2×101 base pair (bp) paired end reads). We distribute the samples such that each lane contain roughly the same amount of samples from different conditions.

3.2.3 Sequencing Quality Control

Quality control (QC) of the RNA Sequencing read data was assessed by FastQC (Andrews, n.d.), which reports the overall quality of the high throughput sequence, and allow the identification of any potential problems and biases.

From the FastQC report, it was noted that some adapter sequences remained in the final sequence. By using trim_glore, a wrapper for cutadapt (version 1.9.1) (Martin, 2011), the adapter sequences were removed from the sequence reads and only reads that were at least 75 bp long were retained for subsequent alignment.

3.2.4 Alignment

In a recent review by Engstrom et al. (2013), it was demonstrated that STAR (Dobin et al., 2013) has the best performance in term of accuracy and speed among all the aligners investigated. Thus STAR aligner was used in our study. The RNA sequencing reads were mapped to the *Mus musculus* reference genome (mm10, Ensembl GRCm38.82) using the STAR aligner (version 2.5.0a) (Dobin et al., 2013). And the quantification of the gene expression levels were conducted using featureCounts (version 1.5.0) (Liao, Smyth, and Shi, 2014).

3.2.5 Data Quality Assessment

Data quality assessment and quality control are essential steps of any data analysis. In order to assess the quality of the count data, unsupervised clustering was performed. Sample with abnormal count data was removed from the analysis.

3.2.6 Differential Expression Analysis

There are many statistical tools available for the differential gene expression analysis. Based on the review of Seyednasrollah, Laiho, and Elo (2015), it was suggested that DESeq2 and limma are the most robust statistical packages for analyzing RNA Sequencing data. As the authors of DESeq2 are very active in providing supports for the package, DESeq2 (version 2.1.4.5) (Love, Huber, and Anders, 2014) was used as the statistic package for the differential gene expression analysis.

One of the most controversial RNA sequencing study in RNA Sequencing was the mouse ENCODE study by Yue et al. (2014), where most of the findings reported were found to be confounded by lane and batch effect Gilad and Mizrahi-Man (2015). This highlights the importance of lane and batch effect in the design of RNA Sequencing. To avoid batch and lane effect, the whole sampling collection procedure and sequencing was performed in a way where we minimize the batch and lane difference between conditions (table 3.1). However, because of the sample quality differed across different batches, we were unable to fully balance out the batch effect. Therefore, it was necessary to control for batch effect in the analyzes.

The following statistical comparisons were performed:

1. Saline exposed samples with n-3 PUFA rich diet vs Saline exposed samples with n-6 PUFA rich diet
2. PolyI:C exposed samples with n-3 PUFA rich diet vs PolyI:C exposed samples with n-6 PUFA rich diet
3. Saline exposed samples with n-6 PUFA rich diet vs PolyI:C exposed samples with n-6 PUFA rich diet

We used $\sim Batch + Condition + Diet + Condition : Diet$ as our model of statistical analysis where Condition is the MIA exposure status. RIN was not included in the

statistical model as suggested by the author.

To further investigate whether batch effect may can lead to false positives, we performed the likelihood ratio test (LRT) to investigate the effect of batch on our result. The LRT examines two models for the counts, a full model with a certain number of terms and a reduced model, in which some of the terms of the full model are removed. The test determines if the increased likelihood of the data using the extra terms in the full model is more than expected if those extra terms are truly zero. Thus we compared the full model $\sim Batch + Condition + Diet + Condition : Diet$ with $\sim Condition + Diet + Condition : Diet$ to understand the effect of batch on our data.

In our analysis, genes with base mean count < 10 were removed to reduce noise associated with low expression and the Benjamini and Hochberg method was then used to correct for multiple testing.

3.2.7 Gene Set Analysis

The main goal of the current study is to investigate whether the effect of MIA or diet act upon the same functional gene sets as the genetic variants associated with schizophrenia in the development of the disease. Specifically, as genes related to PSD (Purcell et al., 2014; T. N. Consortium and Psychiatric Genomics, 2015) and calcium ion channel (Purcell et al., 2014; Ripke et al., 2014; Szatkiewicz et al., 2014) has been implicated to be involved in the etiology of schizophrenia, it is interesting to investigate whether these gene sets were also enriched by genes perturbed by MIA or diet.

To compile a list of relevant gene-sets, significant gene sets from Purcell et al. (2014) and T. N. Consortium and Psychiatric Genomics (2015) were retrieved. Gene sets and pathway related to PSD and calcium ion channel were also retrieved

from Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000) and Gene Ontology (GO) (T. G. O. Consortium, 2015), both have been widely used for systematic genetic analysis.

The Wilcoxon Rank Sum test was performed to assess whether the gene sets were enriched by genes affected by either MIA or diet. Pathways with adjusted p-value < 0.05 (using Benjamini and Hochberg adjustment) were considered as significant.

3.2.8 Partitioning of Heritability

In order to identify the relative contribution of the significant gene sets to the heritability of schizophrenia, partitioning of heritability of schizophrenia was performed using LDSC.

Firstly, SNPs were assigned to genes based on human genome hg19 positions if they lay within 35 kilobase (kb) upstream or 10 kb downstream of the gene. If SNPs mapped within more than one gene, they were assigned to all such genes, following the procedure employed by T. N. Consortium and Psychiatric Genomics (2015)

Then, the partitioning of heritability was performed using LDSC (Bulik-Sullivan et al., 2015) --annot and --overlap-annot options, with window size of 1000kb window size and the Linkage Disequilibrium (LD) score generated in section 2.2.8. The major histocompatibility complex (MHC) region (chr6:25,000,000-35,000,000) was removed from the analysis due to its unusual LD and genetic architecture (Finucane et al., 2015).

3.2.9 Designing the Replication Study

The sample size of the current study is relatively small and therefore only serves as a pilot study. It is therefore important to utilize the information from the current study to design a more powerful follow-up study.

In order to estimate the required sample size for the follow-up studies, power estimation was performed using Scotty (Busby et al., 2013). Based on the current count data, Scotty can estimate the required sample size of the follow up study in order to detect at least 90% of the differentially expressed genes with least 2× difference, and for at least 80% of genes to reach 80% of the maximum power.

3.3 Results

3.3.1 Sample Quality

On average, 87 million reads were generated for each sample of which more than 90% of the read bases has quality score > 30. More than 97% of the sequence reads remains after adapter trimming was performed. Over 90% of the trimmed reads were uniquely mapped to the *Mus musculus* reference genome (mm10, Ensembl GRCm38.82) using the STAR aligner (version 2.5.0a) (Dobin et al., 2013).

Unsupervised clustering was performed to assess the count data quality. It is observed that none of the samples are clustered by lane or by batch (fig. 3.1). However, one of the sample in the n3-PolyI:C group is found to be substantially different from all other samples. It is unclear whether the difference was a result of sample contamination from other sources, or was a result of sample mis-label. The sample was therefore excluded from subsequent analyses.

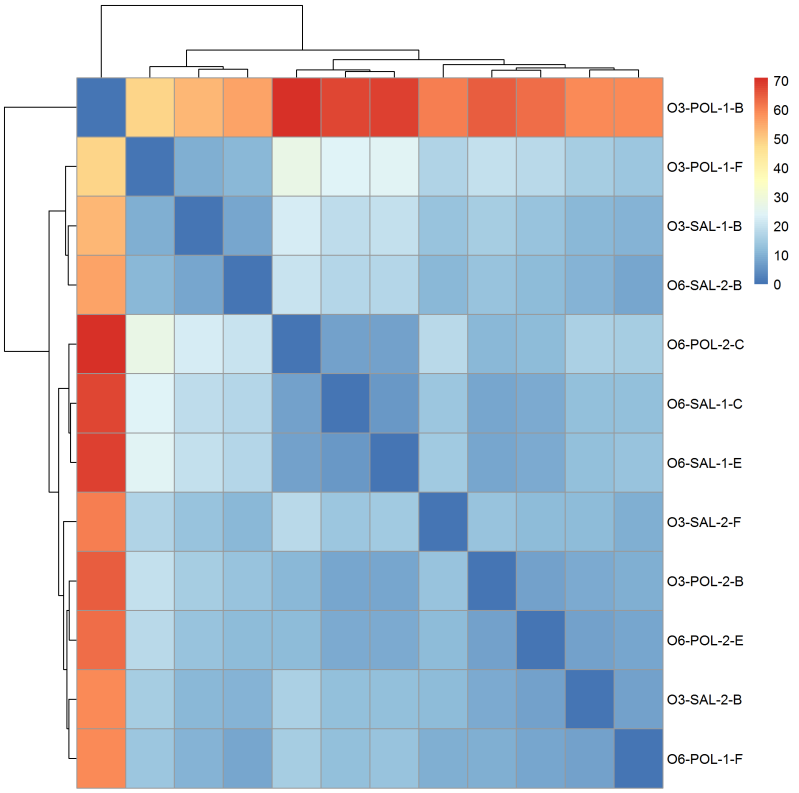


Figure 3.1: Sample Clustering results. Samples were labeled as <Diet>-<Condition>-<Lane>-<Batch> where O3 = n-3 PUFA rich diet; O6 = n-6 PUFA rich diet; POL = PolyI:C; SAL = Saline.
No clear clustering for lane or batch effects are observed. However, one sample from the n3-PUFA-PolyI:C group is found to be substantially different from all other samples. It is unclear whether the difference is due to sample contaminations or sample mis-label. To avoid problems in down-stream analysis, we excluded this sample from subsequent analyses

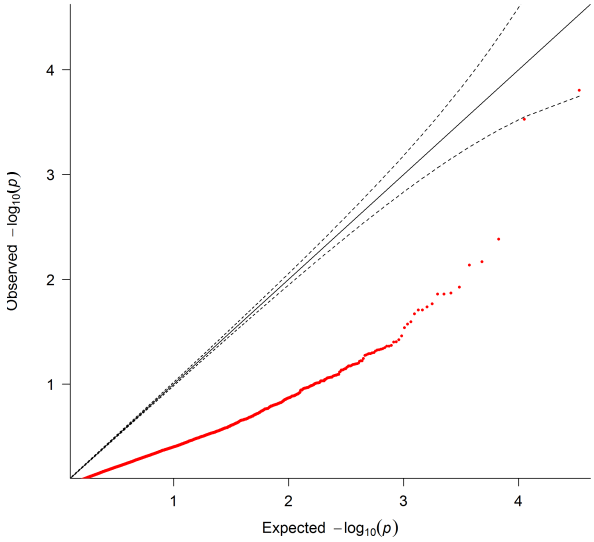
3.3.2 Differential Expression Analysis

DESeq2 analysis was performed after excluding the problematic sample. Of the 16,747 genes that passed through quality control, only *Sgk1* (p-adjusted=0.00186) was found to be significantly differentially when comparing the effect of n-3 PUFA rich diet in PolyI:C exposed mice (fig. 3.2c). On the other hand, no significant differentiation is observed in all other comparisons (figs. 3.2a and 3.2b).

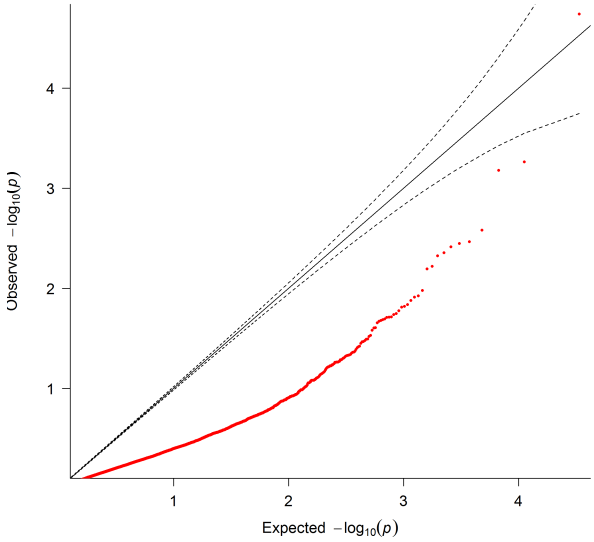
LRT was performed to test the goodness of fit of model with and without the batch effect include. It is observed that when "Batch" was not included in the model, 178 genes are found to be significant (fig. 3.2d). This indicates that by including "Batch" in the statistic model, a significant better fit can be obtained.

3.3.3 Gene Set Analysis

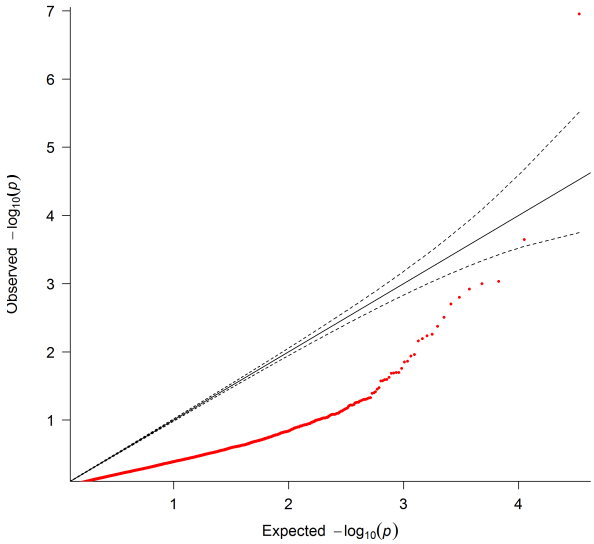
In total, 7 gene sets were included for the gene set analysis (table 3.2). Of the 7 gene sets tested, 6 are significantly enriched in MIA, whereas only the PSD gene set from GO are significantly enriched in PolyI:C exposed mice given the n-3 PUFA rich diet. None of the gene sets are significant in Saline exposed mice given the n-3 PUFA rich diet.



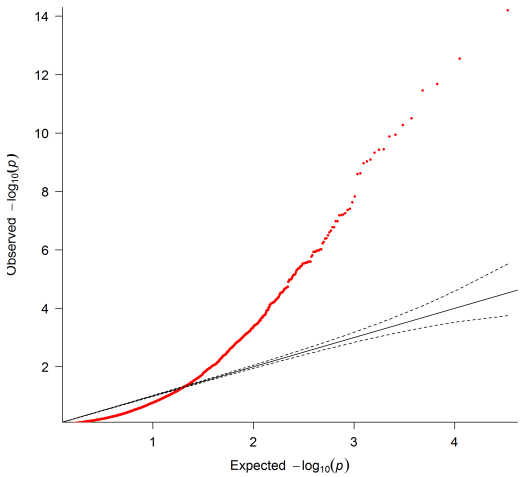
(a) O6-Saline mice vs O6-PolyI:C mice



(b) O6-Saline mice vs O3-Saline mice



(c) O6-PolyI:C mice vs O3-PolyI:C mice



(d) Batch Effect

Figure 3.2: QQ Plot of statistic results. From the quantile-quantile Plot (QQ-plot), it is observed that most of the observed p-values are less than expected. Because the sample size is relatively small, it is likely for the current study to lack detection power, therefore leads to an deflation in p-values.
Meanwhile, the results from LRT suggested that the full model, which adjusted for batch effect, might provide a better fit to our data. Therefore it is important to adjust for the batch effect.

Table 3.2: Results of gene set analysis. In total, 7 gene sets were retrieved from Purcell et al. (2014), KEGG and GO. Firstly, Wilcoxon Rank sum test was performed. Except for the PSD gene set obtained from Purcell et al. (2014), all pathways are enriched in MIA. On the other hand, the PSD gene set obtained from GO is the only gene set that are significantly enriched in PolyI:C exposed mice receiving the n-3 PUFA rich diet, whereas none of the gene sets are significantly enriched in Saline exposed mice receiving the n-3 PUFA rich diet.

Gene Set	Source	ID	Category	Diet in PolyIC Mice	MIA Effect	Diet in Saline Mice	Proportion of h^2 explained	Enrichment P-value
Calcium Ion Signaling Pathway	KEGG	hsa04020	Calcium Ion	0.0402	4.40×10^{-7}	0.231	0.0151	0.397
Glutamatergic synapse	KEGG	hsa04724	PSD	0.118	0.00490	0.123	0.0142	0.335
Voltage-Gated Calcium Channel Activity	GO	GO:05245	Calcium Ion	0.0262	3.45×10^{-6}	0.137	0.0121	0.577
Calcium Channel Activity	GO	GO:05262	Calcium Ion	0.0942	0.00209	0.0880	0.0121	0.577
PSD	GO	GO:14069	PSD	4.86×10^{-3}	6.31×10^{-9}	0.0383	0.0364	0.00325
PSD	Purcell		PSD	0.113	0.328	0.977	0.0448	0.340
GWAS	Purcell		GWAS	0.3048	6.91×10^{-3}	0.551	0.101	1.52×10^{-7}

3.3.4 Designing the Replication Study

Using Scotty (Busby et al., 2013), it is estimated that a minimal of 10 samples per group are required for the follow-up study in order to obtain the desirable power.

3.4 Discussion

3.4.1 Serine/threonine-protein kinase

Our results demonstrated that the expression of Serine/threonine-protein kinase *Sgk1* in the cerebellum of PolyI:C exposed mice might have been affected by n-3 PUFA rich diet. *Sgk1* is a serine/threonine kinase activated by phosphatidylinositol 3-kinase (PI3K) signals,. Studies have reported that the expression of *Sgk1* is associated with spatial learning, fear-conditioning learning and recognition learning in rat (Tsai et al., 2002; Lee et al., 2003). For example, Tsai et al. (2002) observed a 4 fold increase of *Sgk1* in the hippocampus of fast learners when compared to slow learners. Furthermore, the transfection of *Sgk1* mutant DNA impairs the water maze performance in rat (Tsai et al., 2002).

On the other hand, it was found that *Sgk1* can regulates the AMPA and kainate glutamate receptors, especially GluR6, which is encoded by *Grik2* (Lang, Böhmer, et al., 2006; Lang, Strutz-Seebohm, et al., 2010). The kainate receptors contribute to the excitatory postsynaptic current and are important to the synaptic transmission and plasticity in the hippocampus (Lang, Böhmer, et al., 2006). The upregulation of AMPA and kainate receptors are therefore expected to enhance the excitatory effects of glutamate (Lang, Strutz-Seebohm, et al., 2010).

Furthermore, *Sgk1* can up-regulates the glutamate transporters such as EAAT4 (Bohmer et al., 2004), which are vital for the clearance of glutamate from

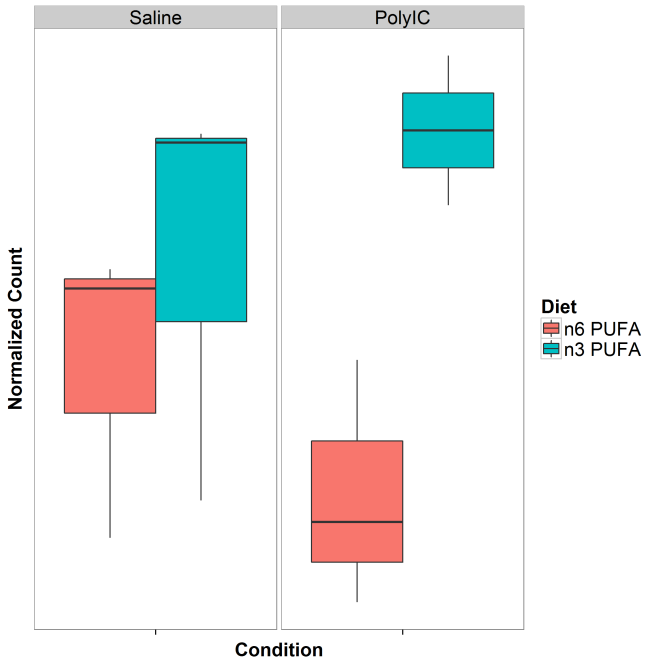


Figure 3.3: Normalized Expression of *Sgk1*. It was observed that the expression level of *Sgk1* increases after the mice was given a n3-PUFA rich diet where a significant increase was observed in mice exposed to PolyI:C.

the synaptic cleft. This prevents excessive glutamate accumulation, thus help to prevent the neurotoxic effects of glutamate (Lang, Strutz-Seebohm, et al., 2010). Considering the complexity of the glutamatergic system and the conflicting role of *Sgk1*, it is likely for more genes to play a role in the regulation of the glutamatergic system. Nonetheless, the disruption of *Sgk1* affect the normal functioning of the glutamatergic system.

In our study, it is observed that the expression of *Sgk1* is lower in the PolyI:C exposed samples when compared to the Saline exposed samples (fig. 3.3). Above all, the expression of *Sgk1* is significantly higher in PolyI:C exposed samples received the n-3 PUFA rich diet. Based on this observation, we hypothesize that the higher expression of *Sgk1* might have "rescued" the effect of MIA on the gene expression in cerebellum, therefore helping to reduce the schizophrenia-like phenotype in the PolyI:C exposed mice.

However, it is unclear whether the increased expression of *Sgk1* is a direct

effect of n-3 PUFA rich diet or just a secondary change. Additionally, although the expression of *Sgk1* is relatively lower in the PolyI:C exposed sample when compared to the Saline exposed samples, the difference was not significant (unadjusted p-value=0.0254, q-value = 0.999). Therefore, it is uncertain whether there is a deficients

Although no direct evidence can be Although we were unable to provide direct connection between the expression of *Sgk1* and the improve functioning of the PolyI:C mice given n-3 PUFA diet, our results do suggest a possible effect of the n-3 PUFA rich diet in the expression of genes related to the PI3K/Akt pathway and might affect the expression of *Sgk1*. Further studies are therefore required to understand whether the change in expression of *Sgk1* can account for the improved functioning of the PolyI:C mice. A possible design will be to induce the expression of *Sgk1* in PolyI:C mice through transfection and examine whether the PolyI:C mice with higher expression of *Sgk1* display a reduction in schizophrenia-like behaviours.

However, previous research of *Sgk1* has been focusing on the hippocampus instead of the cerebellum. It is uncertain whether *Sgk1* has the same function in the cerebellum. Therefore, further researches are required to investigate the role of *Sgk1* in the regulation of development of cerebellum. By studying *Sgk1*, it might be possible to understand how n-3 PUFA rich diet help to reduce the schizophrenia-like behavior in PolyI:C exposed mice, thus provide insight to possible treatment for schizophrenia.

3.4.2 Functional Annotations

When examine the expression change in mice exposed to PolyI:C, none of the genes were significantly differentiated. However, we do observe 12 pathways that contains genes that were more significant than genes not within the pathway (**??**). Interest-

ingly, of the 12 significant pathways, 5 pathways were related to neuronal functions such as neuroactive ligand-receptor interaction (padj=1.27×10^{-3}), genes involved in neuronal system (padj=0.00153) and genes involved in transmission across chemical synapses (padj=0.00401). It has long been developed that the neuronal system and the neurotransmitter regulation plays a critical role in schizophrenia. For example, the disruption of the GABAergic and glutamtergic neuronal system might leads to excitation/inhibition imbalance which might ultimately lead to schizophrenia (Wassef, Baker, and Kochan, 2003). Moreover, the alteration in balance betweem excitation and inhibition can distort the connectivity patterns between different brain regions, thus leads to developmental and behavioral deficits (Cline, 2005).

Additionally, it was found that the calcium signaling pathway was significant when comparing the effect of MIA. The association of the calcium signaling pathway with schizophrenia was not a new finding (Lidow, 2003; Purcell et al., 2014; Ripke et al., 2014). Previous exome sequencing study of schizophrenia by Purcell et al. (2014) has already report the enrichment of non-synonymous variants within the voltage gate calcium ion channel genes in the schizophrenia cases and the Psychiatric Genomics Consortium (PGC) schizophrenia GWAS has also found association between genes encoding the calcium channel subunits with schizophrenia. As calcium signaling pathway is the key component of the mechanism responsible for regulating neuronal excitability (Berridge, 2014), the disruption of the calcium signaling pathway is likely to have a profound effect on the neural function. Together, our results suggest that MIA might have disrupted the normal functioning of the neural system in the cerebellum, thus lead to schizophrenia-like behaviours in the adult mice yet follow up studies are required to validate our findings.

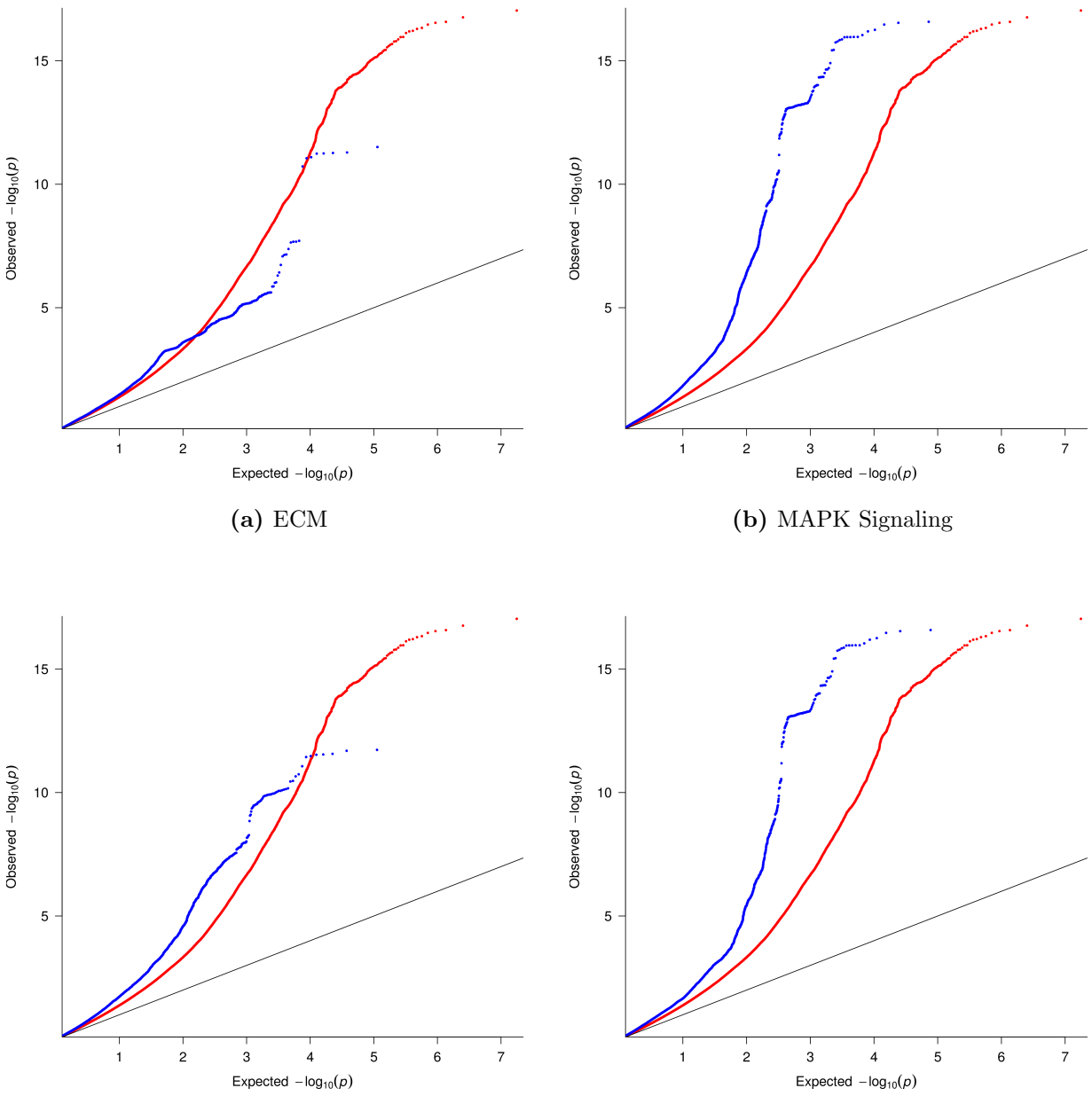
Moreover, we performed the partitioning of heritability hoping to see whether if the significant pathways have contributes disproportionately to the SNP heritability of schizophrenia. Interestingly, all 4 significant pathways that were found to be

contributing a significantly higher portion to the SNP heritability were affected by MIA. To assess whether the significance of these pathways were driven by a small number of very significant genes, we compared the quantile-quantile Plot (QQ-plot) of SNPs within the pathway and all the SNPs included in the PGC GWAS (fig. 3.4). It is observed that for most of the pathways, there is a general inflation of summary statistics when compared to the full set, suggest that the significance was not driven by a single significant gene. However, for the extracellular matrix (ECM) related pathway, only a small inflation was observed. This is therefore likely that the significance was driven by a small number of significant genes.

The "super pathway" containing all the genes participating in the MIA related pathways were also found to be significant (**??**) suggest that the differential gene expression in the cerebellum induced by early MIA events and the genetic variants might act upon similar pathways in the development of schizophrenia.

Interestingly, among the 4 pathways found to be significantly contributes to the SNP heritability, only the pathway related to the assembly of core ECM molecules such as the ECM glycoproteins, were also found to be significantly affected by the n-3 PUFA rich diet in the PolyI:C exposed mouse. Although the QQ-plot suggest that a small number of genes might have driven the significance of this pathway, it is nonetheless an interesting candidate.

Emerging evidences suggest that the ECM abnormality might be associated with schizophrenia (Berretta, 2012). The ECM glycoprotein Reelin has been reported to have a decreased expression in the cerebellum of schizophrenia patients (Maloku et al., 2010) and were found to be accompanied by decreased expression of glutamic acid decarboxylase 67 (Costa et al., 2001). Studies also suggested that Reelin might have important role in corticogenesis and synaptic maturation and stabilization (Berretta, 2012). Moreover, another ECM molecule, Semaphorin 3A has been reported to be increased in the cerebellum of subjects with schizophre-



(a) ECM

(b) MAPK Signaling

(c) Neuronal System

(d) Calcium Signaling

Figure 3.4: Comparing the QQ-plot with PGC SNPs. SNPs found within the pathways were colored in blue whereas the all the full set of SNPs from the PGC GWAS was coded in red. It is observed that for most of the pathways, there is a general inflation of summary statistics when compared to the full set, suggest that the significance was not driven by a single significant gene. However, for the ECM related pathway, only a small inflation was observed. This is therefore likely that the significance was driven by a small number of significant genes.

nia (Eastwood et al., 2003). The Semaphroin 3A protein was found to regulates axonal guidance and has a critical role in the regulation of tangential migration of cortical GABAergic interneurons (Zimmer et al., 2010). It was also reported that the elevated Semaphorin 3A is associated with down-regulation of genes involved in synaptic formation and maintenance (Eastwood et al., 2003). Together, these evidence suggest that the ECM molecules might have critical role in the development of schizophrenia.

It has been reported that the n-3 PUFA diet can modulate the matrix metalloproteinase (MMP) (Derosa et al., 2009; Kavazos et al., 2015) which can regulates the ECM composition (Stamenkovic, 2003). Therefore it is possible that the n-3 PUFA diet has exerted its effect to the ECM through MMP. However, from the QQ-plot, it was noted that only a modest inflation was observed (fig. 3.4a). This suggest that the significance of the ECM pathway might have been driven by a small number of significant genes. Due to difficulties in delineating the individual SNPs effect, it is difficult for us to pin-point the "driver" genes of this pathway. Moreover, none of the ECM genes were found to be differentially expressed in either of our condition, therefore we urge that further studies are required to understand how the n-3 PUFA direct interacts with the ECM or the ECM related genes and the effect of such interaction in MIA exposed individuals.

Finally, it is important to note that the current study serves only as a hypothesis generation study and the sample size was modest. We therefore like to use the current results to provide an estimation of sample size required for a replication study. By using Scotty (Busby et al., 2013), we have estimated that the replication study should contain at least 10 samples for each group in order for us to detect at least 80% of genes has at least 80% of the maximum power. We have also demonstrated that the batch effect can have a big impact to the association (fig. 3.2d), therefore one should always control for the batch effect whenever possible.

Given the current resources, one of the preferred design for the follow up study are given in table 3.3.

3.4.3 Limitations

We first acknowledge that the sample size of the current study is small and are underpowered. This is reflected in the QQ-plots (fig. 3.2) where the observed p-values were generally smaller than would have expected. A better study design will include more samples yet we were limited by our budget. However, the importance of a pilot study is to identify potential targets for replications, hypothesis generation or to provide guidance for follow up studies. In this study, we have identified *Sgk1* as an interesting candidate gene that might have an important role in the effect of n-3 PUFA in PolyI:C exposed individuals. Our results provide support for a possible converging functional effect of differential expression induced by early MIA and genetic variations observed in schizophrenia. These provide interesting candidates for follow studies and we were able to estimate and design a better replication study based on the current data. Therefore we argue that as a hypothesis generation study, our study is successful.

Second, we examined only the male brains in the current study. The decision to direct experimental resources to males was made because there is evidence that the male fetus is more vulnerable to environmental exposures such as inflammation in prenatal life (Bergeron et al., 2013; Lein et al., 2007). We acknowledge that an interesting follow up study would be to investigate the gender difference in response to MIA and dietary change.

Third, although RNA Sequencing was performed, we did not performed any analysis on possible alternative splicing events or denovo transcript assembly. The reason behind such decision is that our sample size is simply too small. With-

out sufficient information, denovo transcript assembly can return noisy results. On the other hand, in order to investigate possible alternative splicing events, we would need to perform the analysis on transcript level instead of gene level. This increase the possible candidates from 47,400 genes to 114,083 transcripts. Combining with the difficulties of the quantification of different isoforms, a much larger power is required for the alternative splicing analysis. On top of that, the functional annotation of transcripts is another difficult aspect to tackle. While there are a lot of information for the annotation of genes, information on functional difference between isoforms of the same gene are generally lacking. The lack of annotation leads to difficulties in making sense of the data. Thus although we acknowledge the possible importance of alternative splicing and denovo transcripts, we did not perform any alternative splicing analysis or denovo transcripts assembly. Nonetheless, the use of RNA Sequencing allow us to easily perform these experiments once sufficient samples are obtained.

Forth, it is important to note that a high RNA expression level does not guarantee a high protein concentration (Vogel and Marcotte, 2012). Post transcriptional, translational and degradation regulation can all affect the rates of protein production and turnover, therefore contributes to the determination of protein concentrations, at least as much as transcription itself (Vogel and Marcotte, 2012). The RNA Sequencing thus only provide an approximation to the concentration of a particular protein in the samples. However, we do argues that RNA Sequencing can help to identify potential targets for protein assays where detail analysis can be performed on the protein level.

Finally, at the time of this thesis, we have yet completed any real time PCR (rt-PCR) or any functional studies to validate our findings. One of the most vital steps after any RNA Sequencing results is to validate the differential expression findings using the rt-PCR. Ideally, not only should one perform the rt-PCR on

the sequenced samples, one should also perform the rt-PCR on an independent set of samples. Moreover, the RNA Sequencing only helps to identify possible candidates that were "associated" with a particular trait. It does not provide any causal linkage between the phenotype and the differential expression. If one would like to establish a direct linkage between the phenotype and the expression of a gene, one will need to carry out functional studies such as knock-in knock-out mouse design. For example, in order to understand the functional impact of the differential expression of *Sgk1*, one might try to examine whether the pure up-regulation of *Sgk1* through transfection can reduce the schizophrenia-like behavior in PolyI:C exposed mice.

Currently, we are planning to perform the rt-PCR on *Sgk1* on all available samples. Shall the results be validated, we can then perform subsequent functional studies.

3.5 Supplementary

Litter	Condition	Diet	Cage	Batch	Lane
1	PolyIC	n-3 PUFA	1	1	1
1	PolyIC	n-6 PUFA	2	5	1
2	PolyIC	n-3 PUFA	3	4	2
2	PolyIC	n-6 PUFA	4	3	3
3	PolyIC	n-3 PUFA	5	2	4
3	PolyIC	n-6 PUFA	6	1	1
4	PolyIC	n-3 PUFA	7	5	1
4	PolyIC	n-6 PUFA	8	4	2
5	PolyIC	n-3 PUFA	9	3	3
5	PolyIC	n-6 PUFA	10	2	4
6	PolyIC	n-3 PUFA	1	2	1
6	PolyIC	n-6 PUFA	2	1	2
7	PolyIC	n-3 PUFA	3	5	2
7	PolyIC	n-6 PUFA	4	4	3
8	PolyIC	n-3 PUFA	5	3	4
8	PolyIC	n-6 PUFA	6	2	1
9	PolyIC	n-3 PUFA	7	1	2
9	PolyIC	n-6 PUFA	8	5	2
10	PolyIC	n-3 PUFA	9	4	3
10	PolyIC	n-6 PUFA	10	3	4
11	Saline	n-3 PUFA	1	3	1
11	Saline	n-6 PUFA	2	2	2
12	Saline	n-3 PUFA	3	1	3
12	Saline	n-6 PUFA	4	5	3

Continued

Litter	Condition	Diet	Cage	Batch	Lane
13	Saline	n-3 PUFA	5	4	4
13	Saline	n-6 PUFA	6	3	1
14	Saline	n-3 PUFA	7	2	2
14	Saline	n-6 PUFA	8	1	3
15	Saline	n-3 PUFA	9	5	3
15	Saline	n-6 PUFA	10	4	4
16	Saline	n-3 PUFA	1	4	1
16	Saline	n-6 PUFA	2	3	2
17	Saline	n-3 PUFA	3	2	3
17	Saline	n-6 PUFA	4	1	4
18	Saline	n-3 PUFA	5	5	4
18	Saline	n-6 PUFA	6	4	1
19	Saline	n-3 PUFA	7	3	2
19	Saline	n-6 PUFA	8	2	3
20	Saline	n-3 PUFA	9	1	4
20	Saline	n-6 PUFA	10	5	4

Table 3.3: Design for follow up study. This design will allow one to balanced out litter effect, cage effect, batch effect and lane effects such that the confounding effects were minimized. One can also include the External RNA Controls Consortium (ERCC) spike in control to serves as an internal standard for additional level of control (Jiang et al., 2011).

4 Conclusion

In this thesis, we presented SNP HeRitability Estimation Kit (SHREK), an robust algorithm for the estimation of Single Nucleotide Polymorphism (SNP) heritability using summary statistics from Genome Wide Association Study (GWAS), an alternative to LD SCore regression (LDSC). Through simulations, it was suggested that when compared to LDSC, SHREK can provide a more robust estimate for oligogentic traits and in case-control designs where no confounding variables was present. Using the latest GWAS summary statistics released by the Psychiatric Genomics Consortium (PGC), we estimated that schizophrenia has a SNP-heritability of 0.174 (SD=0.00453), which is similar to the estimate of 0.197 (SD=0.0058) by LDSC.

When compared to the heritability estimated from twin studies (81%) (Sullivan, Kendler, and Neale, 2003) and large scale population based study (64%) (Lichtenstein et al., 2009), the SNP heritability is much lower, suggesting that factors other than common SNPs might have accounted for the remaining heritability.

On the other hand, we also performed an RNA sequencing on the polyriboinosinic-polyribocytidilic acid (PolyI:C) maternal immune activation (MIA) mouse model to investigate if differential gene expression induced by MIA and genetic variations observed in schizophrenia were acting on the same functional pathway in the development of schizophrenia. We were able to identify a total of 12 pathways that might be perturbed by early MIA events in the cerebellum of the mouse, including calcium ion signaling and pathways related to neural or synaptic functioning. Using

LDSC, it was found that of the 12 significant pathways, 4 pathways related to the extracellular matrix (ECM), mitogen-activated protein kinase (MAPK) signaling, neuronal system and calcium signaling were all contributed disproportionately to the SNP heritability of schizophrenia, suggesting that the differential expression induced by early MIA and the genetic variants associated with schizophrenia might have act upon the same functional pathways in the development of schizophrenia.

Providing that recent study suggest a n-3 polyunsaturated fatty acid (PUFA) rich diet can help to reduce the schizophrenia-like behaviour in mouse exposed to early MIA events (Li, Leung, et al., 2015), we also investigated how the n-3 PUFA rich diet affect the gene expression pattern in the adult cerebellum. *Sgk1*, a gene that regulates the glutamatergic system, were found to be significant in PolyI:C exposed mouse given different diet. Moreover, we found that pathway related to ECM were affected not only by MIA, but also in PolyI:C samples given different diets. It is therefore possible that the ECM pathway or genes within the ECM pathway might have mediated the effect of n-3 PUFA diet on MIA exposed mouse, making them an important target for further research.

4.1 Schizophrenia: Future Perspectives

With the success of the PGC schizophrenia GWAS, research in schizophrenia genetics has finally entered an era of success. Through international collaboration, the PGC has finally identified 108 genetic loci that were associated with schizophrenia using GWAS approach (Ripke et al., 2014). However, the actual causal variants have not been identified. Functional analysis of these associated variants, and their contribution to the etiology of schizophrenia will become an important topic for further research in schizophrenia genetics.

On the other hand, when estimating the SNP-heritability of schizophrenia,

it was found that no more than 20% of the heritability has been accounted for by the current GWAS which is lower than the 81% estimated based on twin studies (Sullivan, Kendler, and Neale, 2003). This suggested that factors other than common SNP were contributing to the heritability of schizophrenia.

Clear evidences suggested that schizophrenia patients has a higher mortality than the general population (Saha, Chant, and Mcgrath, 2007). Given this strong selective pressure, it is likely that the causal variants of schizophrenia with large effect size will be selected against in the population. As a result of that, causal variants with large effect size are likely to be rare (fig. 4.1). With the technological advancement in next generation sequencing (NGS), we are now able to investigate the human genome at per base resolution using Exome Sequencing and even Whole Genome Sequencing technology. Recent study by Purcell et al. (2014) was able to identify gene sets enriched by rare variants that were associated with schizophrenia using Exome Sequencing. This demonstrate the power of the sequencing technology in the identification of possible risk variants. Moreover, there was overlaps observed between genes harboring rare risk variants and those within the PGC schizophrenia GWAS (Purcell et al., 2014), suggesting that the rare variants and common variants studies are complementing each other. As more resources are devoted in to sequencing the genome of schizophrenia patients, more rare variants associated with schizophrenia are expected to be identified.

Currently, most of the focus in schizophrenia was directed to genetic variation yet it is possible that the heritability of schizophrenia is also transmitted in the form of epigenetic changes such as methylation. It was observed that the risk for individual born from a schizophrenic mother is larger than that from a schizophrenic father. This suggests that maternal specific elements, such as maternal imprinting and mitochondria might account for part of the risk of schizophrenia. Epigenetic studies in schizophrenia (Wockner et al., 2014; Nishioka et al., 2012) has identified

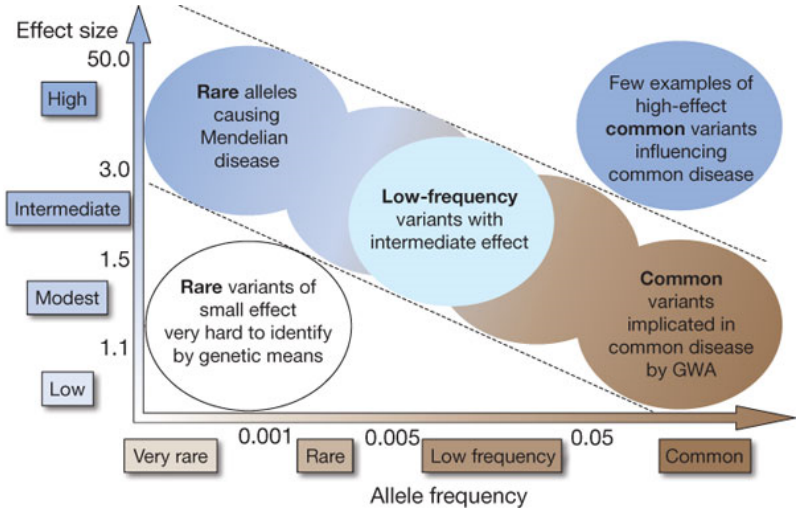


Figure 4.1: Relationship between effect size and allele frequency. It is expected that rare variants with large effect size were actively selected against in the population and therefore should be rare.

genes with differential DNA methylation patterns associated with schizophrenia, suggesting the importance of epigenetics in the etiology of schizophrenia.

As a highly heritable disorder, most of the research of schizophrenia has been focusing on the genetic factors. Although the genetic variation accounted for majority of the variations in schizophrenia, the environmental factors, especially prenatal infection is also an important factor to consider. It was estimated that prenatal infection accounts for roughly 33% of all schizophrenia cases (Brown and Derkits, 2010). The MIA rodent model has provide vital information on the possible interaction between the immune and neuronal system in the etiology of schizophrenia (Meyer, Yee, and Feldon, 2007). For example, Interleukin-6 (IL-6), a pro-inflammatory cytokine has been found to be an important mediator in generating the schizophrenia-like behaviour in rodent model (Smith et al., 2007). More importantly, there are evidence of the interaction between prenatal infection and genetic variation, supporting a mechanism of gene-environment interaction in the causation of schizophrenia (Clarke et al., 2009). As the SNP-heritability estimation does not take into account of the gene environmental interactions, it is possible that the "missing" heritability can be due to gene-environmental interactions. Efforts is

now made by the European network of national schizophrenia networks studying Gene-Environmental Interaction (EUGEI) to identify possible genetic and environmental interaction that contributes to the disease etiology of schizophrenia.

With the sophistication of technologies, we can now perform whole genome sequencing with the HiSeq X Ten system costing less than $1,000. Therefore, the largest challenge now resides in how to make sense of the data instead of data generation. For example, the alignment of sequence read to low complexity sequence or low-degeneracy repeats remains challenging and might be error prone, thus have a negative impact to the quality of the results(Sims et al., 2014). New sequencing technology such as Oxford Nanopore which can provide extra long-reads, might help to make alignment easier due to the extra information for each individual reads. However, the Oxford Nanopore is still under development and has a relatively high error rate (Mikheyev and Tin, 2014). Only until the error rate is dramatically decreased can the use of Oxford Nanopore system become feasible.

Even if the reads can be perfectly aligned to the genome, the functional annotation of variants remains challenging. When it comes to complex disease such as schizophrenia, there can be a lot of causal variants observed throughout the genome yet currently one can only provide estimates of the functional impact of variants on the exomic regions. The development of ENCODE project (ENCODE Project Consortium, 2012) and Genotype-Tissue Expression (GTEx) project (T. G. Consortium, 2015) have helped provide reference point for the annotation of genetic variations in the intergenic regions yet there are still many genetic variation in the genome where their function remains unknown. Only through the tireless effort of the molecular biologist can we gain sufficient information required to make sense of the sequencing data obtained.

In conclusion, we have only catch a glimpse of the etiology of schizophrenia and there are still a lot of questions left unanswered. It is expected that only by

combining the study of epigenetic, genomic variation, gene expressions, and gene environmental interaction can provide a deeper understanding of the complex disease mechanism of schizophrenia be obtained.

Bibliography

Andreasen, Nancy C and Ronald Pierson (2008). "The role of the cerebellum in schizophrenia." eng. In: *Biological psychiatry* 64.2, pp. 81–88 (cit. on p. 113).

Andrews, S. *FastQC A Quality Control tool for High Throughput Sequence Data* (cit. on p. 116).

Bergeron, J D et al. (2013). "White matter injury and autistic-like behavior predominantly affecting male rat offspring exposed to group B streptococcal maternal inflammation". eng. In: *Dev Neurosci* 35.6, pp. 504–515 (cit. on p. 132).

Berretta, S (2012). "Extracellular matrix abnormalities in schizophrenia". eng. In: *Neuropharmacology* 62.3, pp. 1584–1597 (cit. on p. 129).

Berridge, Michael J (2014). "Calcium signalling and psychiatric disease: bipolar disorder and schizophrenia." eng. In: *Cell and tissue research* 357.2, pp. 477–492 (cit. on p. 128).

Bohmer, Christoph et al. (2004). "Stimulation of the EAAT4 glutamate transporter by SGK protein kinase isoforms and PKB." eng. In: *Biochemical and biophysical research communications* 324.4, pp. 1242–1248 (cit. on p. 125).

Brown, A S and E J Derkits (2010). "Prenatal infection and schizophrenia: a review of epidemiologic and translational studies". eng. In: *Am J Psychiatry* 167.3, pp. 261–280 (cit. on pp. 111, 140).

Bulik-Sullivan, Brendan K et al. (2015). "LD Score regression distinguishes confounding from polygenicity in genome-wide association studies". In: *Nature Genetics* 47.3, pp. 291–295 (cit. on p. 119).

Busby, Michele A et al. (2013). "Scotty: a web tool for designing RNA-Seq experiments to measure differential gene expression". In: *Bioinformatics* 29.5, pp. 656–657 (cit. on pp. 120, 125, 131).

Clandinin, M T (1999). "Brain development and assessing the supply of polyunsaturated fatty acid." eng. In: *Lipids* 34.2, pp. 131–137 (cit. on p. 112).

Clarke, Mary C et al. (2009). "Evidence for an interaction between familial liability and prenatal exposure to infection in the causation of schizophrenia." eng. In: *The American journal of psychiatry* 166.9, pp. 1025–1030 (cit. on pp. 111, 140).

Cline, H (2005). "Synaptogenesis: a balancing act between excitation and inhibition". eng. In: *Curr Biol* 15.6, R203–5 (cit. on p. 128).

Consortium, The Gene Ontology (2015). "Gene Ontology Consortium: going forward". In: *Nucleic Acids Research* 43.D1, pp. D1049–D1056 (cit. on p. 119).

Consortium, The GTEx (2015). "The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans". In: *Science* 348.6235, pp. 648–660 (cit. on p. 141).

Consortium, The Network and Pathway Analysis Subgroup of the Psychiatric Genomics (2015). "Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways". In: *Nat Neurosci* 18.2, pp. 199–209 (cit. on pp. 112, 118, 119).

Costa, E et al. (2001). "Dendritic spine hypoplasticity and downregulation of reelin and GABAergic tone in schizophrenia vulnerability." eng. In: *Neurobiology of disease* 8.5, pp. 723–742 (cit. on p. 129).

Derosa, G et al. (2009). "Effects of long chain omega-3 fatty acids on metalloproteinases and their inhibitors in combined dyslipidemia patients." eng. In: *Expert opinion on pharmacotherapy* 10.8, pp. 1239–1247 (cit. on p. 131).

Dobin, A et al. (2013). "STAR: ultrafast universal RNA-seq aligner". eng. In: *Bioinformatics* 29.1, pp. 15–21 (cit. on pp. 116, 120).

Eastwood, S L et al. (2003). "The axonal chemorepellant semaphorin 3A is increased in the cerebellum in schizophrenia and may contribute to its synaptic pathology." eng. In: *Molecular psychiatry* 8.2, pp. 148–155 (cit. on p. 131).

ENCODE Project Consortium (2012). "An integrated encyclopedia of DNA elements in the human genome". In: *Nature* 489.7414, pp. 57–74 (cit. on p. 141).

Engstrom, Par G et al. (2013). "Systematic evaluation of spliced alignment programs for RNA-seq data". In: *Nat Meth* 10.12, pp. 1185–1191 (cit. on p. 116).

Finucane, Hilary K et al. (2015). "Partitioning heritability by functional annotation using genome-wide association summary statistics". In: *Nat Genet* advance online publication (cit. on p. 119).

Gilad, Yoav and Orna Mizrahi-Man (2015). "A reanalysis of mouse ENCODE comparative gene expression data." eng. In: *F1000Research* 4, p. 121 (cit. on p. 117).

Jiang, Lichun et al. (2011). "Synthetic spike-in standards for RNA-seq experiments". In: *Genome Research* 21.9, pp. 1543–1551 (cit. on p. 136).

Kanehisa, M and S Goto (2000). "KEGG: kyoto encyclopedia of genes and genomes". eng. In: *Nucleic Acids Res* 28.1, pp. 27–30 (cit. on p. 119).

Kavazos, Kristyn et al. (2015). "Dietary supplementation with omega-3 polyunsaturated fatty acids modulate matrix metalloproteinase immunoreactivity in a mouse model of pre-abdominal aortic aneurysm." eng. In: *Heart, lung & circulation* 24.4, pp. 377–385 (cit. on p. 131).

Kitajka, Klára et al. (2002). "The role of n-3 polyunsaturated fatty acids in brain: Modulation of rat brain gene expression by dietary n-3 fatty acids". In: *Proceedings of the National Academy of Sciences of the United States of America* 99.5, pp. 2619–2624 (cit. on p. 112).

Knable, M B and D R Weinberger (1997). "Dopamine, the prefrontal cortex and schizophrenia." eng. In: *Journal of psychopharmacology (Oxford, England)* 11.2, pp. 123–131 (cit. on p. 113).

Lang, Florian, Christoph Böhmer, et al. (2006). "(Patho)physiological Significance of the Serum- and Glucocorticoid-Inducible Kinase Isoforms". In: *Physiological Reviews* 86.4, pp. 1151–1178 (cit. on p. 125).

Lang, Florian, Nathalie Strutz-Seebohm, et al. (2010). "Significance of SGK1 in the regulation of neuronal function". In: *The Journal of Physiology* 588.18, pp. 3349–3354 (cit. on pp. 125, 126).

Lee, Eminy H Y et al. (2003). "Enrichment enhances the expression of sgk, a glucocorticoid-induced gene, and facilitates spatial learning through glutamate AMPA receptor mediation." eng. In: *The European journal of neuroscience* 18.10, pp. 2842–2852 (cit. on p. 125).

Lein, E S et al. (2007). "Genome-wide atlas of gene expression in the adult mouse brain". eng. In: *Nature* 445.7124, pp. 168–176 (cit. on p. 132).

Li, Q, C Cheung, et al. (2009). "Prenatal immune challenge is an environmental risk factor for brain and behavior change relevant to schizophrenia: evidence from MRI in a mouse model". eng. In: *PLoS One* 4.7, e6354 (cit. on p. 114).

Li, Q, Y O Leung, et al. (2015). "Dietary supplementation with n-3 fatty acids from weaning limits brain biochemistry and behavioural changes elicited by prenatal exposure to maternal inflammation in the mouse model." eng. In: *Translational psychiatry* 5, e641 (cit. on pp. 112, 138).

Liao, Yang, Gordon K Smyth, and Wei Shi (2014). "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features." eng. In: *Bioinformatics (Oxford, England)* 30.7, pp. 923–930 (cit. on p. 116).

Lichtenstein, Paul et al. (2009). "Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study". In: *The Lancet* 373.9659, pp. 234–239 (cit. on p. 137).

Lidow, Michael S (2003). “Calcium signaling dysfunction in schizophrenia: a unifying approach.” eng. In: *Brain research. Brain research reviews* 43.1, pp. 70–84 (cit. on p. 128).

Love, Michael I, Wolfgang Huber, and Simon Anders (2014). “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.” eng. In: *Genome biology* 15.12, p. 550 (cit. on p. 117).

Maloku, Ekrem et al. (2010). “Lower number of cerebellar Purkinje neurons in psychosis is associated with reduced reelin expression”. In: *Proceedings of the National Academy of Sciences* 107.9, pp. 4407–4411 (cit. on p. 129).

Martin, Marcel (2011). “Cutadapt removes adapter sequences from high-throughput sequencing reads”. In: *EMBnet.journal; Vol 17, No 1: Next Generation Sequencing Data Analysis* (cit. on p. 116).

Meyer, U, B K Yee, and J Feldon (2007). “The neurodevelopmental impact of prenatal infections at different times of pregnancy: the earlier the worse?” eng. In: *Neuroscientist* 13.3, pp. 241–256 (cit. on p. 140).

Mikheyev, Alexander S and Mandy M Y Tin (2014). “A first look at the Oxford Nanopore MinION sequencer.” eng. In: *Molecular ecology resources* 14.6, pp. 1097–1102 (cit. on p. 141).

Nishioka, Masaki et al. (2012). “DNA methylation in schizophrenia: progress and challenges of epigenetic studies.” eng. In: *Genome medicine* 4.12, p. 96 (cit. on p. 139).

Nugent, Tom F. et al. (2007). “Dynamic mapping of hippocampal development in childhood onset schizophrenia”. In: *Schizophrenia Research* 90.1-3, pp. 62–70 (cit. on p. 113).

Olivo, Susan E and Leena Hilakivi-Clarke (2005). “Opposing effects of prepubertal low- and high-fat n-3 polyunsaturated fatty acid diets on rat mammary tumorigenesis.” eng. In: *Carcinogenesis* 26.9, pp. 1563–1572 (cit. on p. 114).

Perlstein, W M et al. (2001). "Relation of prefrontal cortex dysfunction to working memory and symptoms in schizophrenia." eng. In: *The American journal of psychiatry* 158.7, pp. 1105–1113 (cit. on p. 113).

Purcell, S M et al. (2014). "A polygenic burden of rare disruptive mutations in schizophrenia". eng. In: *Nature* 506.7487, pp. 185–190 (cit. on pp. 112, 118, 124, 128, 139).

Reeves, P G, F H Nielsen, and G C Jr Fahey (1993). *AIN-93 purified diets for laboratory rodents: final report of the American Institute of Nutrition ad hoc writing committee on the reformulation of the AIN-76A rodent diet.* eng (cit. on p. 114).

Ripke, Stephan et al. (2014). "Biological insights from 108 schizophrenia-associated genetic loci". In: *Nature* 511, pp. 421–427 (cit. on pp. 112, 118, 128, 138).

Saha, Sukanta, David Chant, and John Mcgrath (2007). "A Systematic Review of Mortality in Schizophrenia". In: *Archives of general psychiatry* 64.10, pp. 1123–1131 (cit. on p. 139).

Seyednasrollah, Fatemeh, Asta Laiho, and Laura L Elo (2015). "Comparison of software packages for detecting differential expression in RNA-seq studies". In: *Briefings in Bioinformatics* 16.1, pp. 59–70 (cit. on p. 117).

Sims, David et al. (2014). "Sequencing depth and coverage: key considerations in genomic analyses". In: *Nat Rev Genet* 15.2, pp. 121–132 (cit. on p. 141).

Smith, S E et al. (2007). "Maternal immune activation alters fetal brain development through interleukin-6". eng. In: *J Neurosci* 27.40, pp. 10695–10702 (cit. on pp. 112, 140).

Stamenkovic, Ivan (2003). "Extracellular matrix remodelling: the role of matrix metalloproteinases." eng. In: *The Journal of pathology* 200.4, pp. 448–464 (cit. on p. 131).

Sullivan, Patrick F, Kenneth S Kendler, and Michael C Neale (2003). "Schizophrenia as a Complex Trait". In: *Archives of general psychiatry* 60, pp. 1187–1192 (cit. on pp. 137, 139).

Szatkiewicz, J P et al. (2014). "Copy number variation in schizophrenia in Sweden". In: *Mol Psychiatry* 19.7, pp. 762–773 (cit. on pp. 112, 118).

Tienari, Pekka et al. (2004). "Genotype-environment interaction in schizophrenia-spectrum disorder". In: *The British Journal of Psychiatry* 184.3, pp. 216–222 (cit. on p. 111).

Trebble, Timothy et al. (2003). "Inhibition of tumour necrosis factor-alpha and interleukin 6 production by mononuclear cells following dietary fish-oil supplementation in healthy men and response to antioxidant co-supplementation." eng. In: *The British journal of nutrition* 90.2, pp. 405–412 (cit. on p. 112).

Tsai, Kuen J et al. (2002). "sgk, a primary glucocorticoid-induced gene, facilitates memory consolidation of spatial learning in rats." eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 99.6, pp. 3990–3995 (cit. on p. 125).

Velakoulis, Dennis et al. (2006). "Hippocampal and amygdala volumes according to psychosis stage and diagnosis". In: *Archives of general psychiatry* 63, pp. 139–149 (cit. on p. 113).

Vogel, Christine and Edward M Marcotte (2012). "Insights into the regulation of protein abundance from proteomic and transcriptomic analyses." eng. In: *Nature reviews. Genetics* 13.4, pp. 227–232 (cit. on p. 133).

Wassef, A, J Baker, and L D Kochan (2003). "GABA and schizophrenia: a review of basic science and clinical studies". eng. In: *J Clin Psychopharmacol* 23.6, pp. 601–640 (cit. on p. 128).

Wockner, L F et al. (2014). "Genome-wide DNA methylation analysis of human brain tissue from schizophrenia patients". In: *Transl Psychiatry* 4, e339 (cit. on p. 139).

Yeganeh-Doost, Peyman et al. (2011). “The role of the cerebellum in schizophrenia: from cognition to molecular pathways”. In: *Clinics* 66.Suppl 1, pp. 71–77 (cit. on p. 113).

Yue, Feng et al. (2014). “A comparative encyclopedia of DNA elements in the mouse genome.” eng. In: *Nature* 515.7527, pp. 355–364 (cit. on p. 117).

Zimmer, Geraldine et al. (2010). “Chondroitin sulfate acts in concert with semaphorin 3A to guide tangential migration of cortical interneurons in the ventral telencephalon.” eng. In: *Cerebral cortex (New York, N.Y. : 1991)* 20.10, pp. 2411–2422 (cit. on p. 131).