Genetic and Environmental risk factors 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

December 12, 2015

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...

Acknowledgements

Abbreviations

bp base pair. 123, 124

CEU Northern Europeans from Utah. 51, 55, 60, 64, 69

CI confidence interval. 16

CNS central nervous system. 30

CNV copy number variation. 19, 22, 23, 28

DEG differentially expressed gene. 127, 130

DSM Diagnostic and Statistical Manual of Mental Disorders. 3

DZ dizygotic. 15

GC Genomic Control. 26, 28

GCTA Genome-wide Complex Trait Analysis. 25, 36, 58, 61, 63, 65–67, 71–84, 86, 87, 89, 90, 92, 96, 98, 101, 102, 107, 114

GD Gestation Day. 119, 122

GO Gene Ontology. 130

GRM Genetic Relationship Matrix. 24, 25

GWAS Genome Wide Association Study. 19, 20, 22–29, 36–38, 46, 51, 58–60, 68, 80, 85, 92, 98, 100, 105, 106, 117, 137–139

IBD identity by descent. 67

IL-6 Interleukin-6. 4, 120

IQ intelligence quotient. 9

LD Linkage Disequilibrium. 19, 25–27, 29, 30, 37–39, 43, 47, 48, 50–57, 60, 61, 68–70, 92–95, 97, 99, 103, 105–107, 138

LDSC LD SCore regression. 26–28, 30, 32, 35, 36, 52, 53, 58, 61, 63, 65–68, 71–85, 87, 89–93, 95–99, 101–110, 112–117, 137–139

LRT likelihood ratio test. 126, 130

maf Minor Allele Frequency. 9, 19, 55, 57, 60–62, 64–66, 68, 103, 139

mb megabase. 52, 59, 68

MHC major histocompatibility complex. 22

MIA maternal immune activation. 4, 5, 7, 22, 118–121, 126, 133–135, 137

MLM mixed linear model. 24, 25

MSE mean squared error. 79, 80, 85, 86, 90, 95–97

MZ monozygotic. 15, 17

NCP non-centrality parameter. 45

NGS next generation sequencing. 139

PCA principle component analysis. 138

PCGC Phenotype correlation - genotype correlation regression. 92

PET positron emission tomography. 121

PGC Psychiatric Genomics Consortium. 20, 23, 28, 59, 67, 68, 91, 105, 106, 117, 137

PGS Polygenic Risk Score. 91, 104

PolyI:C polyriboinosinic-polyribocytidilic acid. 4–6, 122, 124, 129, 130

PUFA polyunsaturated fatty acid. 120–122, 124, 126, 129, 133, 135, 137

QC quality control. 123

REML restricted maximum likelihood. 24, 25

RIN RNA integrity number. 123

rt-PCR real time PCR. 135

SCZ schizophrenia. 7, 91, 105, 106

SE standard error. 25, 28, 44, 45, 58, 75, 79, 85

SHREK SNP Heritability and Risk Estimation Kit. 68–87, 89–99, 101, 102, 104–107, 110, 112–117, 137–139

SNP Single Nucleotide Polymorphism. 18–20, 23–26, 28–30, 35, 37–40, 43, 50–53, 55–68, 70–86, 88, 93–97, 99, 102, 103, 105–116, 137–139

SVD Singular Value Decomposition. 48, 51, 52

tSVD Truncated Singular Value Decomposition. 48–51, 106

WHO World Health Organization. 1, 2

YLD years lost due to disability. 1, 2

Contents

List of Figures

List of Tables

Chapter 3

n-3 Polyunsaturated Fatty Acid Rich Diet in Schizophrenia

3.1 Introduction

As research in schizophrenia progress, we start to identify an increasing amount of variants. There are now 108 genomic loci identified to be associated with schizophrenia (Ripke et al., 2013). Using LD SCore regression (LDSC) and SNP Heritability and Risk Estimation Kit (SHREK), we estimated that Psychiatric Genomics Consortium (PGC) schizophrenia Genome Wide Association Study (GWAS) only accounts for no more than 20% of the heritability, much least than what was estimated from the twin studies (Lichtenstein et al., 2009; Sullivan, Kendler, and Neale, 2003). Although the PGC schizophrenia GWAS bring great promises to the field of schizophrenia genetics, there is still a long way before one can translate the findings from the PGC schizophrenia GWAS into clinical applications.

Another direction of schizophrenia research was to investigate how different environmental risk factors contribute to the etiology of schizophrenia. Of all the non-genetic risk factors, prenatal infection has the largest effect size (Sullivan, 2005). Because of its important in schizophrenia, prenatal infection has been extensively studied.

Early studies of prenatal infection in schizophrenia mainly relies on ecological data such as influenza epidemics in the population to define the exposure status (A S Brown and Derkits, 2010). The problem of these studies was that the exposure status was based solely on whether an individual was in gestation at the time fo the epidemic without any confirmation of maternal infection during pregnancy. This leads to difficulties in replication of the findings. Subsequently, researchers uses birth cohorts where infection was documented using different biomarkers during pregnancies to provide a better labeling of the exposure status (A S Brown and Derkits, 2010). Through these rigorous studies it was found that the risk of schizophrenia increases as long as an individual's mother was infected by different form of infectious agents such as influenza, HSV-2 and *T.gondii* during gestation (A S Brown and Derkits, 2010). As different infectious agents all increase the risk of schizophrenia, it leads to the hypothesis of maternal immune activation (MIA) (A S Brown and Derkits, 2010) where it was hypothesized that instead of a particular infectious agents, it was the maternal immune response that disrupt the brain development in the offspring, thus leading to an elevated risk of schizophrenia.

To really understand how MIA increase the risk of schizophrenia, it is important to understand the molecular mechanism. A great challenge in the study of MIA was that one cannot carry out empirical experimental design in human samples due to ethical issues. Thus a popular alternative is to employ rodent models. However, unlike physiological traits, psychiatric disorder such as that of schizophrenia often contain symptoms related to higher level functioning such as hallucinations, delusion, disorganized speech etc (American Psychiatric Association, 2013). This raise challenge in diagnosing whether if the rodent has demonstrated the symptoms of schizophrenia for not only it was difficult to check whether if the high level functioning of the rodent is disrupted, there were no available biomarkers for schizophrenia. Thus instead of labeling whether if the rodent is “schizophrenic” or “normal”, one would rather consider whether if the rodent demonstrate any “schizophrenia-like” behaviours such as impaired prepulse inhibition, impaired working memory and reduced social interaction (Meyer, Yee, and Feldon, 2007). An important point to note here is that as autism and schizophrenia shares most of these phenotypes, and that risk of autism is also increased in patients whose mother were exposed to infections during gestation (Alan S Brown, 2012), studies using these rodent models to study effect of prenatal infection were usually non-specific to schizophrenia or autism. Rather, they should be considered together. For simplicity and focus of the current thesis, we would limit our discussion to schizophrenia.

Recent studies of global gene expression patterns in MIA-exposed rodent fetal brains (Oskviga et al., 2012; Garbett et al., 2012) suggest that the post-pubertal onset of schizophrenic and other psychosis-related phenotypes might stem from attempts of the brain to counteract the environmental stress induced by MIA during its early development (Garbett et al., 2012). To date, all these studies have focused on the changes elicited by a mid-to-late gestation exposure (e.g. Gestation Day (GD) 12.5 for mouse, or GD for rat). However, although Meyer, Yee, and Feldon (2007), Li, C. Cheung, Wei, Hui, et al. (2009), and Li, C. Cheung, Wei, V. Cheung, et al. (2010) have reported that MIA early in gestation event might exert a more extensive impact on the phenotype of offspring, the effect of early MIA on gene expression in brain of adult offspring have not been examined. It is therefore interesting to study the gene expression changes in adult offspring who were exposed to MIA during early gestation.

Ultimately, one would like to identify treatments / cures for schizophrenia thus help to boost the quality of life of the schizophrenia patients. One candidate is the n-3 PUFA which can inhibits the production of Interleukin-6 (IL-6) (Trebble et al., 2003), a major mediator in MIA (Smith et al., 2007). n-3 PUFA is also plays a critical role in the development of central nervous system (Clandinin, 1999) and it has robust anti-inflammatory properties (Trebble et al., 2003). Previous study from our lab suggested that a n-3 PUFA rich diet can help to reduce the schizophrenia-like phenotype in mouse exposed to early MIA insults (Li, Leung, et al., 2015). Thus we would also like assess the effect of an n-3 PUFA rich diet on the gene expression pattern in the brain of the adult offspring.

Herein, we introduce a pilot study aiming to study 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 - an approach considered to be more accurate and reliable compared to conventional microarrays (Z. Wang, Gerstein, and Snyder, 2009). Moreover, RNA Sequencing are more flexible when compared to microarrays in that it can also detect alternative splicings and novel transcripts. Although we don't have sufficient sample size for such analysis in our pilot study, the use of RNA Sequencing allow subsequent replication studies to incorporate the pilot samples for such analysis, thus potentially reducing the cost of experiment.

Brain is a complex organ in that it is subdivided into multiple regions, each with their own responsibility. Thus it is expected that the gene expression pattern differs from region to regions. It is then important for us to select a region of interest for our analysis. As a pilot study, we have exploit the samples from our previous study on MIA and effect of n-3 PUFA (Li, Leung, et al., 2015) where only the cerebellum was available. Although hippocampus (Velakoulis et al., 2006; Nugent et al., 2007) and prefrontal cortex (Knable and Weinberger, 1997; Perlstein et al., 2001) were the two most studied region in schizophrenia, the cerebellum has also been reported to be related to schizophrenia (Yeganeh-Doost et al., 2011; Andreasen and Pierson, 2008). Moreover, the cerebellum plays a central role in the cortico-cerebellar-thalamic-cortical neuronal circuit. Positron emission tomography (PET) studies show a dysfunction in this circuit can contribute to "cognitive dysmetria", e.g. impaired cognition and other symptoms of schizophrenia (Yeganeh-Doost et al., 2011). Altogether, this makes the cerebellum an interesting target to investigate.

To summarize, in this chapter, we conducted a pilot study on the effect of early MIA and n-3 PUFA rich diet on the gene expression pattern of the cerebellum of mouse using RNA Sequencing.

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 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, C. Cheung, Wei, Hui, 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 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 and the cerebellum was extracted and stored in -80°C until RNA extraction.

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 pilot 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 Strannder 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 were rather standardized where FastQC (Andrews, n.d.) is the most widely adopted tools. It can generate the required per base QC and provide a general picture of how well the sequencing were done.

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) (Mar-

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 1M_2 and 1M_3 which came from the same litter but were given a different diet.

tin, 2011), we trim the adapter sequences from the sequence and only retain reads that were at least 75 bp long for subsequent alignment.

3.2.4 Alignment

When aligning RNA Sequencing reads, one can either directly align the reads to the transcriptome or to the genome. However, when aligning to the transcriptome, multiple isoforms can share part of the sequence, thus leads to high level of multiple alignment, having an negative impact to the downstream analysis especially if one were only interested in the gene base expression. On the other hand, when directly aligning the reads to the genome, one need to use splicing aware aligners to handle the splicing. Aligners such as TopHat2 (Kim et al., 2013), STAR (Dobin et al., 2013) and MapSplice (K. Wang et al., 2010) are some of the popular aligners that are capable to align RNA

Sequencing reads to the genome by considering possible splicing. In a recent review by Engstrom et al. (2013), it was demonstrated that STAR has the best performance of all the aligners tested taking into account of accuracy and speed. 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, Gordon K Smyth, and Shi, 2014).

3.2.5 Differential Expression Analysis

Early RNA Sequencing experiment assumes the gene expression counts follows the Poisson distribution (Marioni et al., 2008) where the variance is assumed to be equal to the mean of the expression. However, it was found that the assumption of Poisson distribution is too restrictive where an over-dispersion was typically observed in RNA Sequencing data (S Anders and W Huber, 2010). Taking into account of the over-dispersion, modern RNA Sequencing statistical package usually models the RNA Sequencing counts using the negative binomial distribution (S Anders and W Huber, 2010; Robinson, McCarthy, and G K Smyth, 2010) or the beta negative binomial distribution (Trapnell et al., 2012). 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. Considering that the authors of DESeq2 were very active in providing supports for the package, we selected DESeq2 (version 2.1.4.5) (Love, Wolfgang Huber, and Simon Anders, 2014) as the statistic package for the differential gene expression analysis.

Perhaps one of the most controversial study in RNA Sequencing was the mouse ENCODE paper by Yue et al. (2014) where Gilad and Mizrahi-Man (2015) demonstrated that most of the findings from Yue et al. (2014) was confounded by lane and batch effect. 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 across different batches, we were unable to fully balance out the batch effect. Therefore, in our analysis, we must control for the batch effect. Moreover, we were interested in the following comparisons:

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

To obtain the desire comparison, and also control for batch effect, we used $\sim Batch + Condition + Diet + Condition : Diet$ as our model of statistical analysis where Condition is the MIA exposure status.

We would also like to see if the batch effect can leads to false positive results. Therefore we performed the likelihood ratio test (LRT). 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, we removed all genes with base mean count < 10 to reduce noise associated with low expression. The Benjamini and Hochberg method were then used to correct for multiple testing.

3.2.6 Functional Annotation

Usually, one would like to perform functional annotation of the differentially expressed genes (DEGs) which allow one to identify biological processes (e.g. pathways) that were disrupted (e.g. enriched by the DEGs). However, if only a small number of DEGs were identified, the common enrichment analysis, which usually were an inclusion / exclusion based, can be biased. An alternative method was to perform the Wilcoxon Rank Sum test to test whether if the p-value of genes within the gene set are than the p-value of genes outside the gene set. We downloaded the canonical pathways from the Molecular Signatures Database (MSigDB) (v5.0 updated April 2015) (Subramanian et al., 2005) as our reference pathways. To avoid testing overly narrow or broad functional pathways, we selected pathways that contains at least 10 and at most 300 genes. The Wilcoxon Rank Sum test was then performed for each pathway to test for significance. Pathways with adjusted p-value < 0.05 (using Benjamini and Hochberg adjustment) were considered as significant.

3.2.7 Designing the Replication Study

One of the most important goal of a pilot study is to provide information for further replication studies. In order to estimate the power and required samples for further studies, we performed the power estimation using Scotty (Busby et al., 2013). We provide the count data from our pilot samples to Scotty to estimate the minimal required samples for our replication study if we would like to detect at least 90% of the genes that are differentially expressed by a 2× fold change at $p < 0.01$ and that at least 80% of genes has at least 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 meaning that the probability of having an incorrect base call is less than 1 in 1,000. After removing the adapter sequences from the reads, more than 97% of the reads remains. Over 90% of the trimmed reads could be uniquely mapped to the *Mus musculus* reference genome (mm10, Ensembl GRCm38.82) using the STAR aligner (version 2.5.0a) (Dobin et al., 2013). To obtained the expression count, we used the featureCounts (version 1.5.0) (Liao, Gordon K Smyth, and Shi, 2014) to generate the count matrix required for downstream analysis.

Next, we are interested in whether if there are any series batch or lane effect. We perform unsupervised clustering on the sample count data. It was observed that none of the samples were clustered by lane or batch, suggesting that there were no serious batch or lane effect presented in our samples. However, one sample from the n3-PolyI:C group was found to be substantially different from all other samples (fig. 3.1). It was unclear whether if the difference was due to sample contaminations or was due to sample mis-label. To avoid problems in down-stream analysis, we excluded this sample

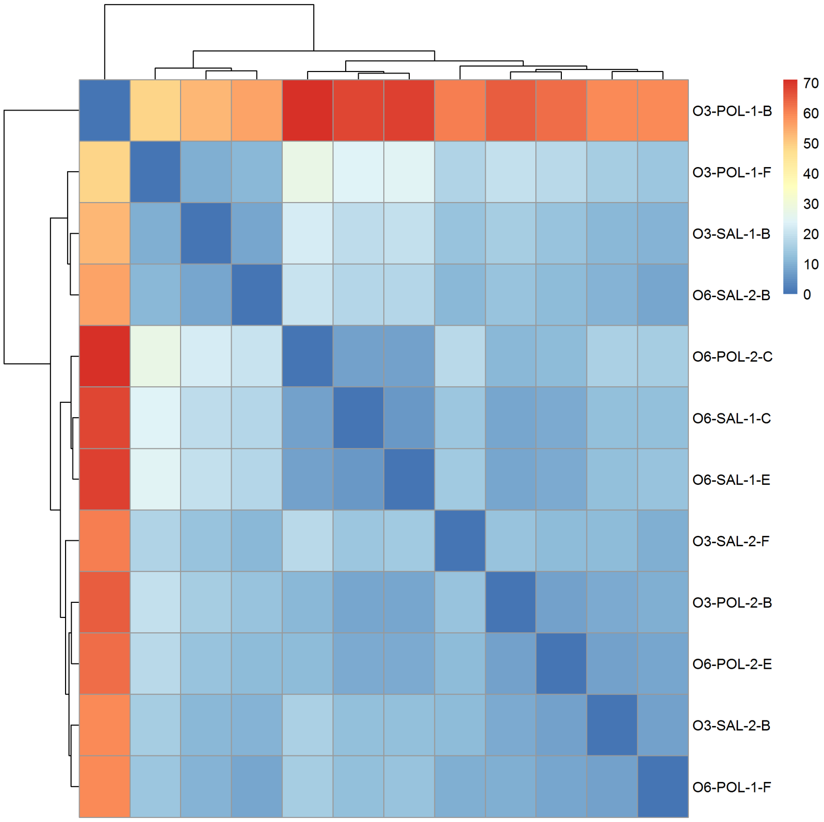
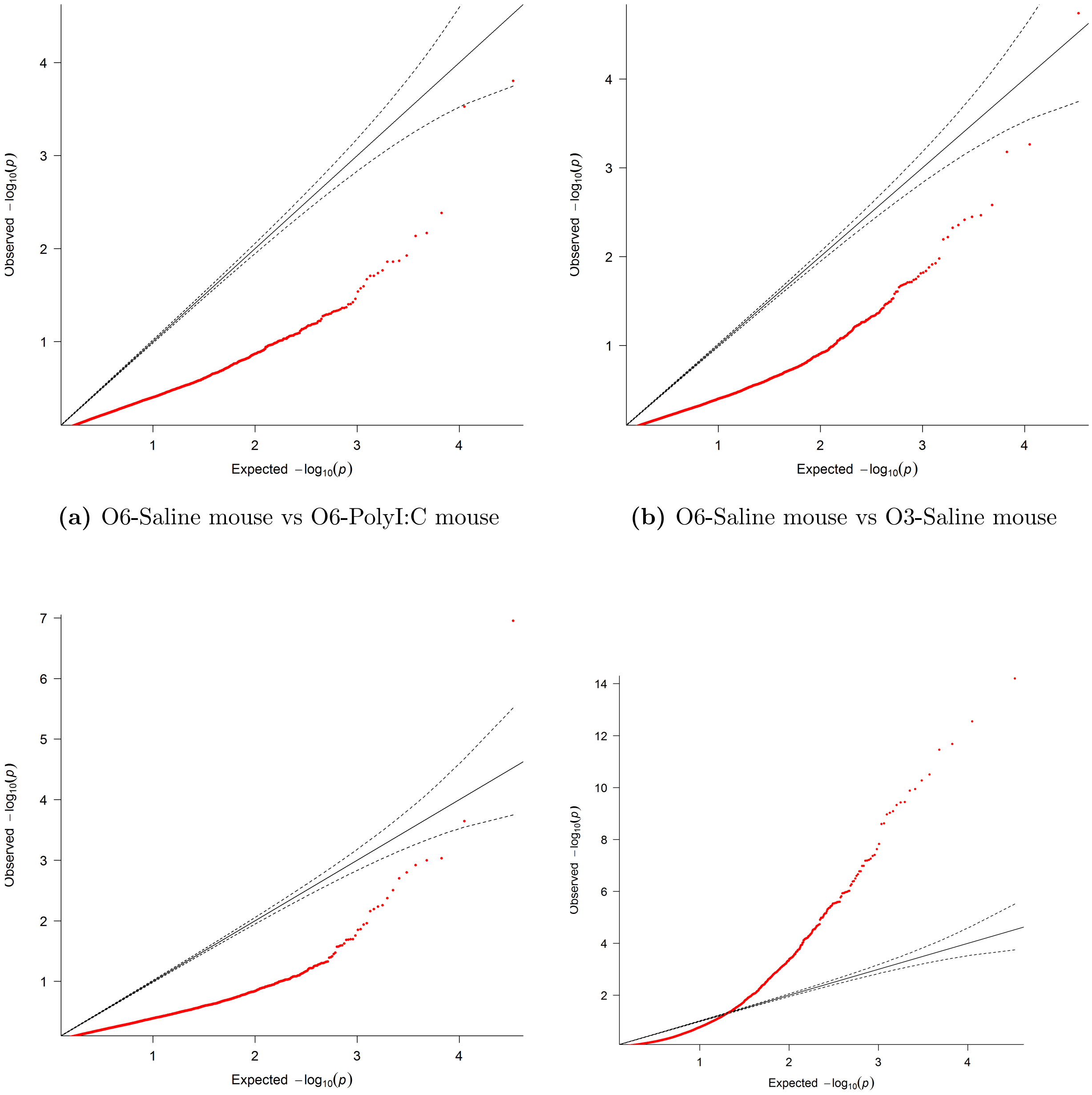


Figure 3.1: Sample Clustering results. It was observed that there was no clear clustering for lane or batch effects. However, one sample from the n3-PUFA-PolyI:C group was found to be substantially different from all other samples.
It was unclear whether if the difference was due to sample contaminations or was due to sample mis-label. To avoid problems in down-stream analysis, we excluded this sample from subsequent analyses

from subsequent analyses

3.3.2 Differential Expression Analysis

After excluding the problematic samples, we performed the DESeq2 analysis. Of the 16,747 genes that passed through quality control, only one gene, *Sgk1* (p-adjusted=0.00186) was found to be significantly differentiated when comparing the expression in PolyI:C exposed mouse given different diet (fig. 3.2c). No genes were found to be significant for



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

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

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

(d) Batch Effect

Figure 3.2: QQ Plot of statistic results. From the QQ plot, it was observed that most of the observed p-value was less than what would have been expected. This is likely due to the small sample size of our study which leads to an under powered association. The only exception was the analysis of batch effect were a large amount of genes were found to be significant. This demonstrate the importance of adjusting for batch effect

the other two comparison (figs. 3.2a and 3.2b).

We also performed the LRT to compare test the effect of batch on our analysis. A total of 178 genes were found to be significant differentiated (fig. 3.2d), suggesting that the “Batch” is indeed an important factor to consider in our analysis.

3.3.3 Functional Annotation

It is common practice to try and perform functional annotation to the DEGs. However, in most of our analysis, there were either no DEG or only 1 DEG, making it difficult to perform functional annotation such as Gene Ontology (GO) enrichment analysis. We used the Wilcox rank sum test to analysis whether if a pathway contain genes that are more significant than genes not within the pathway.

None of the pathway were found to be significant when comparing the effect of the n-3 rich diet in Saline exposed mouse. On the contrary, 17 pathways were found to be more significant when comparing the effect of n-3 PUFA rich diet in PolyI:C exposed samples (table 3.2) where 4 pathways were related to growth factors such as fibroblast growth factor (FGF) or epidermal growth factor (EGF) and 4 others were related to kinases such a s phosphatidylinositol 3-kinase (PI3K) or mitogen-activated protein kinase (MAPK).

Finally, 12 pathways were found to significant when comparing Saline and PolyI:C exposed mouse given the n-6 PUFA rich diet (table 3.3) with pathways such as neuroactive ligand-receptor interaction (p-adj = 1.27×10^{-3}), calcium signaling pathway (p-adj = 2.79×10^{-3}) and genes involved in Neuronal System (p-adj=0.00153).

ID	Size	Source	Description	Adjusted P-Value
M508	78	REACTOME	Genes involved in Signaling by SCF-KIT	0.00671
M570	44	REACTOME	Genes involved in PI3K events in ERBB2 signaling	0.0242
M3008	196	NABA	Genes encoding structural ECM glycoproteins	0.0309
M1090	112	REACTOME	Genes involved in Signaling by FGFR	0.0309
M563	109	REACTOME	Genes involved in Signaling by EGFR in Cancer	0.0309
M17776	100	REACTOME	Genes involved in Downstream signaling of activated FGFR	0.0309
M1076	83	REACTOME	Genes involved in Amyloids	0.0309
M850	56	REACTOME	Genes involved in PI-3K cascade	0.0309
M10450	38	REACTOME	Genes involved in GAB1 signalosome	0.0309
M16227	24	REACTOME	Genes involved in Cholesterol biosynthesis	0.0309
M5872	17	KEGG	Steroid biosynthesis	0.0309
M16334	10	BIOCARTA	Eph Kinases and ephrins support platelet aggregation	0.0309
M5884	275	NABA	Ensemble of genes encoding core extracellular matrix including ECM glycoproteins, collagens and proteoglycans	0.0456
M635	127	REACTOME	Genes involved in Signaling by FGFR in disease	0.0456
M568	38	REACTOME	Genes involved in PI3K events in ERBB4 signaling	0.0456
M165	32	PID	Syndecan-4-mediated signaling events	0.0456
M1262	15	REACTOME	Genes involved in GRB2:SOS provides linkage to MAPK signaling for Intergrins	0.0456

Table 3.2: Significant Pathways when comparing effect of diet in PolyI:C exposed mouse. The pathway IDs are the systematic name from MSigDB. Most of the significant pathways were related to the kinase such as PI3K and MAPK or growth factors such as FGF and EGF.

ID	Size	Source	Description	Adjusted P-Value
M13380	272	KEGG	Neuroactive ligand-receptor interaction	1.27×10^{-3}
M2890	178	KEGG	Calcium signaling pathway	2.79×10^{-3}
M12289	188	REACTOME	Genes involved in Peptide ligand-binding receptors	0.00118
M5884	275	NABA	Ensemble of genes encoding core extracellular matrix including ECM glycoproteins, collagens and proteoglycans	0.00119
M735	279	REACTOME	Genes involved in Neuronal System	0.00153
M15514	186	REACTOME	Genes involved in Transmission across Chemical Synapses	0.00401
M4904	121	REACTOME	Genes involved in G alpha (s) signalling events	0.0127
M3008	196	NABA	Genes encoding structural ECM glycoproteins	0.0131
M752	137	REACTOME	Genes involved in Neurotransmitter Receptor Binding And Downstream Transmission In The Postsynaptic Cell	0.0131
M10792	267	KEGG	MAPK signaling pathway	0.0195
M17	59	PID	Notch signaling pathway	0.0406
M18437	184	REACTOME	Genes involved in G alpha (q) signalling events	0.0406

Table 3.3: Significant Pathways When Comparing Effect PolyI:C in Mouse Given n-6 PUFA Rich Diet. The pathway IDs are the systematic name from MSigDB. Interestingly, we observed a lot of neural related pathways and even got significant signal in the calcium signaling pathway, which was reported to be associated with schizophrenia (Purcell et al., 2014).

3.3.4 Designing the Replication Study

The main purpose of the current study is to serve as a pilot for subsequent replications. It is therefore vital for us to exploit the current data to estimate the number of samples required in order to have sufficient power for association. Using Scotty (Busby et al., 2013), given that we would like to detect at least 90% of the genes that are differentially expressed by a 2× fold change at $p < 0.01$ and that at least 80% of genes has at least 80% of the maximum power, we will need at least 10 samples per group in the replication study given the current sequencing depth.

3.4 Discussion

In this pilot study, we demonstrated that *Sgk1* might be affected by n-3 PUFA rich diet in the cerebellum of MIA exposed mouse.

When investigating the function of these pathways, it was found that most were related to either FGF receptor and EGF receptor, or were related to PI3K and MAPK. What was interested was that studies (Ojeda et al., 2011; Schlessinger, 2004) has shown that major canonical pathways activated by FGF and EGF share similar signals that includes MAPK and PI3K. More importantly, Ojeda et al. (2011) demonstrated that when cultured in EGF-containing media, the human fetal neural stem cells were drove away from the motor neuron differentiation path and instead differentiated toward glutamate and γ-aminobutyric acid (GABA) neuronal subtypes which were both found to be associated with schizophrenia (Wassef, Baker, and Kochan, 2003; Javitt, 2010; Nakazawa et al., 2012).

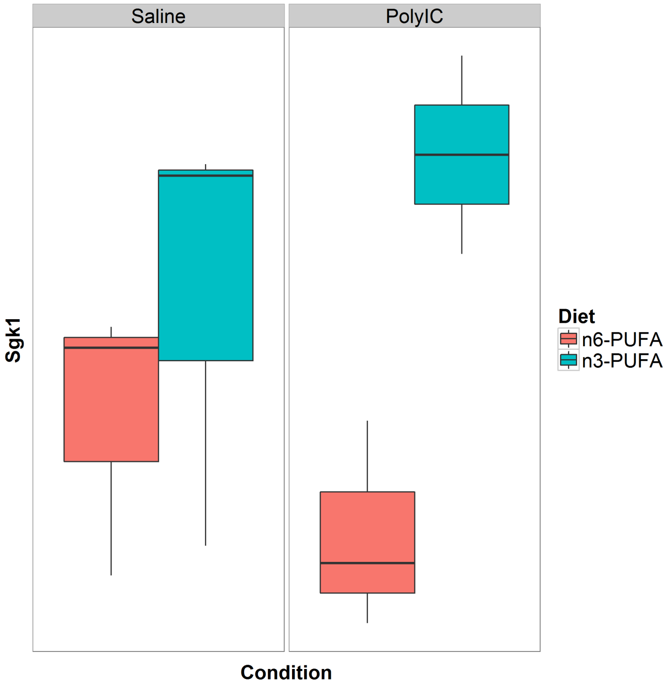


Figure 3.3: Normalized Expression of the DEGs. It was observed that the expression level of *Sgk1* increases after the mouse was given a n3-PUFA rich diet.

3.4.1 Limitation

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 and **??**) where the observed p-values were generally smaller than would have expected. A better study design will include more samples yet we are limited by our budget. However, the importance of a pilot study is to identify potential targets for replications or to provide guidance for follow up studies. With our data, we were able to identify two genes, the *Sgk1* and *Xbp1*, for follow up and were able to estimate the required number of samples in the follow up studies to achieve substantial power. Moreover, our study demonstrated the importance of controlling the batch effect which can severally confound the results. Therefore in subsequent follow up studies, one should always control for batch effect in the experimental design and statistical analysis.

Second, we examined only 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 have 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. Without 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. Combined 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 wealth of information on gene annotation, information on functional difference between isoforms of the same gene were generally lacking. The lack of annotation simply 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. Nonetheless, RNA Sequencing helps 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 performed 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 however 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 gene, one will need to carry out functional studies such as knock-in knock-out mouse design. So take for example, in order to understand how *Sgk1* interacts with MIA and the n-3 PUFA rich diet, we will need to investigate the effect of n-3 PUFA rich diet in MIA exposed *Sgk1* knock-out genes or the effect of co-ingestion of *Sgk1* inhibitor with n-3 PUFA diet in MIA exposed mouse.

Currently, we are planning to perform the rt-PCR on *Sgk1* and *Xbp1* genes on all available samples. Shall the results be validated, we can then perform subsequent functional studies.

Chapter 4

Conclusion

SHREK, an algorithm for the estimation of heritability using GWAS test statistics are reported in this thesis. To our knowledge, this is the only algorithm other than the LD Score regression that can perform heritability estimation using test statistics. In this thesis, we were able to demonstrate that SHREK can provided a more robust estimate in case-control designs when no confounding variables was present. By applying SHREK on the test statistic from the PGC schizophrenia GWAS, we estimated that schizophrenia has a Single Nucleotide Polymorphism (SNP)-heritability of 0.174 (SD=0.00453), which is similar to the estimate of 0.197 (SD=0.0058) by LDSC.

On the other hand, we report a pilot RNA Sequencing study aiming to investigate the effect of maternal immune activation (MIA) and n-3 PUFA rich diet on the gene expression pattern in the adult cerebellum.

4.1 Challenge in SNP-Heritability Estimation

Although now that we can estimates the SNP heritability based on the test statistic of large scale meta analysis, this is only the beginning for there are still a lot of questions left unanswered in the estimation of SNP heritability. One major problem of SHREK and LDSC is that they both heavily relies on the Linkage Disequilibrium (LD) structures from the reference panel. However, for any meta analysis, samples are usually from a large variety of the ethnicity, although one can choose a reference that are representative of the majority of samples, we are uncertain how the sample mixing can alter the LD structures, this might potentially leads to biased estimates. If the fundamental LD structure was not as expected, both SHREK and LDSC will not be able to provide an accurate estimate. For example, if a GWAS was conducted with 50% European and 50% African, population stratification may confound the results. Even if one control for the population stratification using the principle component analysis (PCA), the question remains whether if one should use the African reference panel or the European reference panel in the estimation of SNP heritability. This problem is further complicated when the information regarding population stratification was usually unavailable. Thus further researches are required to tackle the problem of population stratification before one can confidently estimate the SNP heritability on large scale meta analysis that consists of samples from a large variety of ethic background.

One important observation in our simulation study was that there was a general bias observed in all the SNP-heritability estimation algorithm. This is likely to due to the ascertainment bias introduced through case control sampling. Although the liability adjustment was performed, bias was still observed. This suggested that we will need a better liability adjustment algorithm if we would like to accurately estimate the SNP-heritability from case control studies.

As technology advances, researchers can now use the next generation sequencing (NGS) technology to sequence the genome at per base resolution. This brings great prospect in the genetic studies for now we can directly identify the causal variants and can even detect rare causal variants providing sufficient sample size. However, both SHREK and LDSC are designed to work on the test statistics of a GWAS where common SNPs are usually the focus of the studies. Because of the huge sampling error associating with rare variants, SHREK and LDSC might be unsuitable for rare variants. In fact, it was found that when all causal variants are rare (Minor Allele Frequency (maf) $< 1\%$), LDSC will often generate a negative slope, and the intercept will exceed the mean χ^2 statistic (Bulik-Sullivan et al., 2015). As a result of that, a different algorithm must be developed in order to estimates the heritability from rare variants.

4.2 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 we have finally identify 108 genetic loci that are associated with schizophrenia (Ripke et al., 2013). However, the GWAS only provides statistical association between the variants schizophrenia and does not provide direct evidence as to the functional involvement of these variants in the etiology of schizophrenia. Functional consequences of these variations in schizophrenia are therefore an important topic for the understanding of the mechanism of schizophrenia.

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. There is no doubt that by continue to increase the sample size of the GWAS, one can identify more variants associated with schizophrenia and therefore increases the SNP-heritability. However, it is also likely that they will have a very small effect size which might not be useful for clinical translations. Another possibility is that the so call "missing" heritability might be accounted for by other factors such as rare variants and epigenetic such as methylation.

There is clear evidence 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 that has a large effect size will be selected against in the population. As a result of that, causal variants with large effect size should be rare in nature (fig. 4.1). With the technological advancement in 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, we would expect to identify more rare variants that are associated with schizophrenia.

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. When considering the risk of schizophrenia, it was observed that the risk for individual born from a schizophrenic mother is larger than

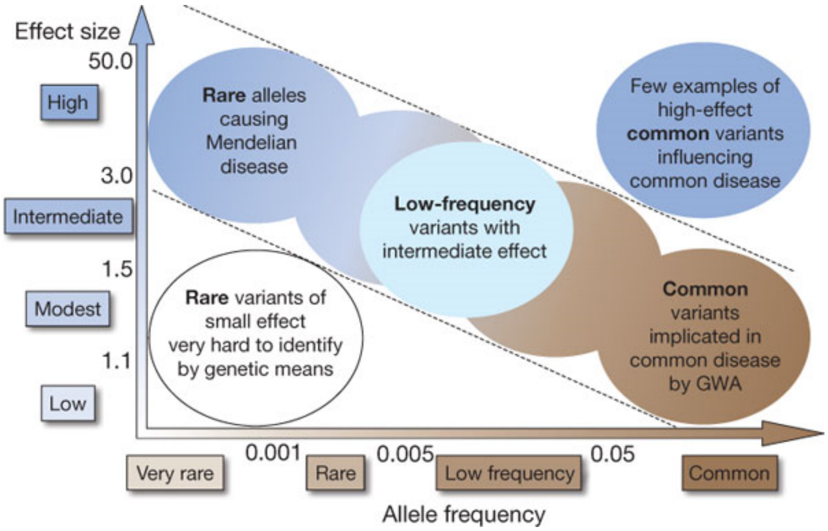

Effect size
50.0
High
3.0
Intermediate
1.5
Modest
1.1
Low
Rare alleles causing Mendelian disease
Few examples of high-effect common variants influencing common disease
Low-frequency variants with intermediate effect
Rare variants of small effect very hard to identify by genetic means
Common variants implicated in common disease by GWA
Very rare
0.001
Rare
0.005
Low frequency
0.05
Common
Allele frequency

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.

that from a schizophrenic father. It is therefore possible that the risk of schizophrenia is passed on through maternal imprinting. Epigenetic studies in schizophrenia (Wockner et al., 2014; Nishioka et al., 2012) has identified genes with differential DNA methylation patterns associated with schizophrenia, suggesting the important of epigenetic in the etiology of schizophrenia.

As a genetic 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 are also important factors to consider. It was estimated that prenatal infection accounts for roughly 33% of all schizophrenia cases (A S 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, 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. The problem now then isn't the cost of generating the data, but the difficulties in making sense of the data. The first problem is the alignment of sequence read to low complexity sequence or low-degeneracy repeats (Sims et al., 2014). One possible solution is to use system such as the Oxford Nanopore which can provide extra long-reads, thus allowing for better alignment. However, the Oxford Nanopore is still underdevelopment 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 we can accurately align all the reads to the genome, we still face another challenge. When it comes to complex disease such as schizophrenia, there can be a lot of causal variants observed throughout the genome. However, we are only capable to estimate the functional impact of variants on the exomic regions. The development of ENCODE project (ENCODE Project Consortium, 2012) and Genotype-Tissue Expression (GTEx) project (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 to us. 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 we have a deeper understanding of the complex disease mechanism of schizophrenia. Hopefully, in the near future, we can gain enough understanding to start translating the research findings into clinical applications to help improving the quality of life of schizophrenia patients.

Bibliography

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Publishing, p. 991. URL: http://encore.llu.edu/iii/encore/record/C%5C_%5C_Rb1280248%5C_%5C_SDSM-V%5C_%5C_P0,2%5C_%5C_Orightresult%5C_%5C_X3;jsessionid=ABB7428ECBC4BA66625EDD0E0C5AAFA5?lang=eng%5C&suite=cobalt$%5Cbackslash$nhttp://books.google.com/books?id=EIbMlwEACAAJ%5C&pgis=1 (cit. on p. 119).

Anders, S and W Huber (2010). “Differential expression analysis for sequence count data”. eng. In: *Genome Biol* 11.10, R106. URL: http://www.ncbi.nlm.nih.gov/pubmed/20979621 (cit. on p. 125).

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. 121).

Andrews, S. *FastQC A Quality Control tool for High Throughput Sequence Data.* URL: citeulike-article-id:11583827%20http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (cit. on p. 123).

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. URL: http://www.ncbi.nlm.nih.gov/pubmed/24246964 (cit. on p. 136).

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. URL: `http://www.ncbi.nlm.nih.gov/pubmed/20123911` (cit. on pp. 118, 143).

Brown, Alan S (2012). “Epidemiologic studies of exposure to prenatal infection and risk of schizophrenia and autism.” eng. In: *Developmental neurobiology* 72.10, pp. 1272–1276 (cit. on p. 119).

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. URL: `http://www.nature.com/doifinder/10.1038/ng.3211` (cit. on p. 141).

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. URL: `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3582267/` (cit. on pp. 128, 134).

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. 120).

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 p. 144).

Consortium, The GTEx (2015). “The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans”. In: *Science* 348.6235, pp. 648–660. URL: `http://www.sciencemag.org/content/348/6235/648.abstract` (cit. on p. 144).

Dobin, A et al. (2013). “STAR: ultrafast universal RNA-seq aligner”. eng. In: *Bioinformatics* 29.1, pp. 15–21. URL: http://www.ncbi.nlm.nih.gov/pubmed/23104886 (cit. on pp. 124, 125, 128).

ENCODE Project Consortium (2012). “An integrated encyclopedia of DNA elements in the human genome”. In: *Nature* 489.7414, pp. 57–74. URL: http://dx.doi.org/10.1038/nature11247%20http://www.nature.com/nature/journal/v489/n7414/abs/nature11247.html%5C#supplementary-information (cit. on p. 144).

Engstrom, Par G et al. (2013). “Systematic evaluation of spliced alignment programs for RNA-seq data”. In: *Nat Meth* 10.12, pp. 1185–1191. URL: http://dx.doi.org/10.1038/nmeth.2722%2010.1038/nmeth.2722%20http://www.nature.com/nmeth/journal/v10/n12/abs/nmeth.2722.html%7B%5C#%7Dsupplementary-information (cit. on p. 125).

Garbett, K a et al. (2012). “Effects of maternal immune activation on gene expression patterns in the fetal brain”. In: *Translational Psychiatry* 2.4, e98 (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. 125).

Javitt, Daniel C (2010). “Glutamatergic theories of schizophrenia.” eng. In: *The Israel journal of psychiatry and related sciences* 47.1, pp. 4–16 (cit. on p. 134).

Kim, Daehwan et al. (2013). “TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions”. In: *Genome Biology* 14.4, R36. URL: http://genomebiology.com/2013/14/4/R36 (cit. on p. 124).

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. 121).

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. URL: `http://www.ncbi.nlm.nih.gov/pubmed/17151600` (cit. on p. 136).

Li, Q, C Cheung, R Wei, V Cheung, et al. (2010). “Voxel-based analysis of postnatal white matter microstructure in mice exposed to immune challenge in early or late pregnancy”. eng. In: *Neuroimage* 52.1, pp. 1–8. URL: `http://www.ncbi.nlm.nih.gov/pubmed/20399275` (cit. on p. 119).

Li, Q, C Cheung, R Wei, E S Hui, 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. URL: `http://www.ncbi.nlm.nih.gov/pubmed/19629183` (cit. on pp. 119, 122).

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 p. 120).

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 pp. 125, 128).

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. URL: `http://dx.doi.org/10.1016/S0140-6736(09)60072-6` (cit. on p. 117).

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. 125).

Marioni, J C et al. (2008). “RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays”. eng. In: *Genome Res* 18.9, pp. 1509–1517. URL: http://www.ncbi.nlm.nih.gov/pubmed/18550803 (cit. on p. 125).

Martin, Marcel (2011). “Cutadapt removes adapter sequences from high-throughput sequencing reads”. In: *EMBnet.journal; Vol 17, No 1: Next Generation Sequencing Data Analysis*. URL: http://journal.embnet.org/index.php/embnetjournal/article/view/200 (cit. on p. 123).

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. URL: http://www.ncbi.nlm.nih.gov/pubmed/17519367 (cit. on pp. 119, 143).

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. 144).

Nakazawa, Kazu et al. (2012). “GABAergic interneuron origin of schizophrenia pathophysiology”. In: *Neuropharmacology* 62.3, pp. 1574–1583. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3090452/ (cit. on p. 134).

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. 143).

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. 121).

Ojeda, Luis et al. (2011). “Critical Role of PI3K/Akt/GSK3β in Motoneuron Specification from Human Neural Stem Cells in Response to FGF2 and EGF”. In: *PLoS ONE* 6.8. Ed. by Mai Har Sham, e23414. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3160859/ (cit. on p. 134).

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. 123).

Oskviga, Devon B. et al. (2012). "Maternal immune activation by LPS selectively alters specific gene expression profiles of interneuron migration and oxidative stress in the fetus without triggering a fetal immune response". In: *Brain, Behavior, and Immunity* 26.4, pp. 623–634. URL: http://www.sciencedirect.com/science/article/pii/S0889159112000177 (cit. on p. 119).

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. 121).

Purcell, S M et al. (2014). "A polygenic burden of rare disruptive mutations in schizophrenia". eng. In: *Nature* 506.7487, pp. 185–190. URL: http://www.ncbi.nlm.nih.gov/pubmed/24463508 (cit. on pp. 133, 142).

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. 122).

Ripke, S et al. (2013). "Genome-wide association analysis identifies 13 new risk loci for schizophrenia". eng. In: *Nat Genet* 45.10, pp. 1150–1159. URL: http://www.ncbi.nlm.nih.gov/pubmed/23974872 (cit. on pp. 117, 141).

Robinson, M D, D J McCarthy, and G K Smyth (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data". eng. In: *Bioinformatics* 26.1, pp. 139–140. URL: http://www.ncbi.nlm.nih.gov/pubmed/19910308 (cit. on p. 125).

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. 142).

Schlessinger, Joseph (2004). "Common and distinct elements in cellular signaling via EGF and FGF receptors." eng. In: *Science (New York, N.Y.)* 306.5701, pp. 1506–1507 (cit. on p. 134).

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. URL: `http://bib.oxfordjournals.org/content/16/1/59.abstract` (cit. on p. 125).

Sims, David et al. (2014). "Sequencing depth and coverage: key considerations in genomic analyses". In: *Nat Rev Genet* 15.2, pp. 121–132. URL: `http://dx.doi.org/10.1038/nrg3642%2010.1038/nrg3642` (cit. on p. 144).

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. URL: `http://www.ncbi.nlm.nih.gov/pubmed/17913903` (cit. on pp. 120, 143).

Subramanian, Aravind et al. (2005). "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles". In: *Proceedings of the National Academy of Sciences* 102.43, pp. 15545–15550. URL: `http://www.pnas.org/content/102/43/15545.abstract` (cit. on p. 127).

Sullivan, Patrick F (2005). "The Genetics of Schizophrenia". In: *PLoS Med* 2.7, e212. URL: `http://dx.doi.org/10.1371/journal.pmed.0020212` (cit. on p. 118).

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 p. 117).

Trapnell, Cole et al. (2012). "Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks". In: *Nat. Protocols* 7.3, pp. 562–578. URL: http://dx.doi.org/10.1038/nprot.2012.016 (cit. on p. 125).

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. 120).

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. 121).

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 pp. 136, 137).

Wang, K et al. (2010). "MapSplice: accurate mapping of RNA-seq reads for splice junction discovery". eng. In: *Nucleic Acids Res* 38.18, e178. URL: http://www.ncbi.nlm.nih.gov/pubmed/20802226 (cit. on p. 124).

Wang, Zhong, Mark Gerstein, and Michael Snyder (2009). "RNA-Seq: a revolutionary tool for transcriptomics". In: *Nat Rev Genet* 10.1, pp. 57–63. URL: http://dx.doi.org/10.1038/nrg2484 (cit. on p. 120).

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. URL: http://www.ncbi.nlm.nih.gov/pubmed/14624191 (cit. on p. 134).

Wockner, L F et al. (2014). "Genome-wide DNA methylation analysis of human brain tissue from schizophrenia patients". In: *Transl Psychiatry* 4, e339. URL: http://dx.doi.org/10.1038/tp.2013.111%2010.1038/tp.2013.111 (cit. on p. 143).

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. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3118440/ (cit. on p. 121).

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 pp. 125, 126).