Genetic and Environmental risk factors of Schizophrenia and Autism

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

August 29, 2015

Declaration

Acknowledgements

Abbreviations

ASD	Autism Spectrum Disorder
CEU	Utah residents with Northern and Western European ancestry from the CEPH collection
SCZ	Schizophrenia

Contents

Declaration i

Acknowledgments iii

Abbreviations v

Contents vii

Introduction 1

1 Literature Review I: Schizophrenia and Autism 3
- 1.1 Schizophrenia . . . 3
- 1.2 Autism Spectrum Disorder . . . 3
- 1.3 The Environmental Risk Factors of SCZ and ASD . . . 3
  - 1.3.1 Prenatal Infection . . . 3
  - 1.3.2 Parental Age . . . 4
  - 1.3.3 Prenatal Stress . . . 4
  - 1.3.4 Maternal Vitamin D Deficiency During Pregnancy . . . 4
- 1.4 The Genetic Etiology of SCZ and ASD . . . 4
- 1.5 Summary . . . 4

2 Literature Review II: Approaches to Reveal Genetic Causes 5
- 2.1 Twin Studies - Delineating Genetic and Environmental Contribution . . . 5
- 2.2 Searching for Genetic Variants . . . 5
  - 2.2.1 Role of Common Variants . . . 5
  - 2.2.2 Role of Rare Variants . . . 5
- 2.3 Searching for Gene-Environmental interaction . . . 6
  - 2.3.1 Gene Expression . . . 6
  - 2.3.2 Epigenetics . . . 6
- 2.4 Summary . . . 6

3 Environmental Risk Factor - Maternal immune activation 7
- 3.1 Study Design . . . 7
- 3.2 Materials and Method . . . 7
- 3.3 Results . . . 7
- 3.4 Discussion . . . 7
- 3.5 Conclusion . . . 7

4 Genetic Risk Factor - Heritability Estimation 9
- 4.1 Estimation of Heritability . . . 9
  - 4.1.1 Estimating the Mean . . . 9

4.1.2 Estimating the Variance . . . 9
4.2 Simulation Study . . . 11
4.2.1 Quantitative Trait . . . 11
4.2.2 Case-Control Study design . . . 11
4.3 Result . . . 11
4.4 Discussion . . . 11
4.5 Conclusion . . . 11

5 Genetic Risk Factor - Risk Prediction **13**
5.1 Risk Estimation . . . 13
5.2 Simulation Study . . . 13
5.3 Result . . . 13
5.4 Conclusion . . . 13

6 Summary and Conclusion **15**

Introduction

Chapter 1

Literature Review I: Schizophrenia and Autism

1.1 Schizophrenia

Schizophrenia (SCZ) is a Affecting roughly 1% of the human population. Detrimental to the quality of life. Limited treatment. No cure. **Detail description of the disease here**

1.2 Autism Spectrum Disorder

On the other hand, Autism Spectrum Disorder Affecting XXX of the population. Associated with mental retardation. Most are unable to take care of themselves. No cure. **Detail description of the disease here**

1.3 The Environmental Risk Factors of SCZ and ASD

Despite the difference in their phenotype, epidemiological studies suggest that they share a lot of common environment factors.

1.3.1 Prenatal Infection

Arguably one of the most important environmental risk factor for SCZ and AD. Affect $\frac{1}{3}$ of all SCZ patient. Epidemiological study of Brown. The Involvement of IL-6. No protein found in the fetus. Talk about the finding of Oskvig and Smith. Imbalance caused by trying to counter the infection

1.3.2 Parental Age

1.3.3 Prenatal Stress

1.3.4 Maternal Vitamin D Deficiency During Pregnancy

1.4 The Genetic Etiology of SCZ and ASD

Talk about the PGC studies. Previous line of evidence? What they have found with the genetic studies? (SCZ) Involvement of the PSD95. (Shaun) Most enriched area is the MHC. Other associated SNPs are also highly enriched by immune genes. (ASD) Need to read more paper on this

1.5 Summary

Chapter 2

Literature Review II: Approaches to Reveal Genetic Causes

2.1 Twin Studies - Delineating Genetic and Environmental Contribution

Should briefly talk about how Twin modeling was used for finding the GE contribution. Should also mention the ACE model. At the end, we can talk about the heritability estimates of SCZ and AD

2.2 Searching for Genetic Variants

2.2.1 Role of Common Variants

Genome Wide Association Study

Should talk about what is GWAS and how it is used. Should also talk about the current GWAS studies in SCZ and AD

2.2.2 Role of Rare Variants

Exome Sequencing

Similar to the GWAS. Talk about the Pros and Cons. Need to briefly mention the Denovo paper and Shaun's paper.

Whole Genome Sequencing

Very very brief description of WGS and the current status.

2.3 Searching for Gene-Environmental interaction

2.3.1 Gene Expression

Micro-array

RNA Sequencing

2.3.2 Epigenetics

Methylation Chip

Bisulfite Sequencing

2.4 Summary

Chapter 3

Environmental Risk Factor: Maternal immune activation

3.1 Study Design

This should serves as the place where we place the mini introduction. What have people not done? early MIA. What is the importance? Earlier the worst. What are we going to do? What is the aim and goal? Brief description of what to be done.

3.2 Materials and Method

3.3 Results

3.4 Discussion

3.5 Conclusion

Short conclusion on the Environmental Risk. Also link the result to the next chapter.

Chapter 4

Genetic Risk Factor: Heritability Estimation

4.1 Estimation of Heritability

4.1.1 Estimating the Mean

4.1.2 Estimating the Variance

To calculate the variance of the estimation, we will need to obtain the variance covariance matrix of h. Because $\boldsymbol{h} = (\boldsymbol{\rho^2})^{-1}\boldsymbol{f}$, we can obtain the variance covariance matrix of h as

$$\mathbf{Cov}(\boldsymbol{h}) = (\boldsymbol{\rho^2})^{-1}\mathbf{Cov}(\boldsymbol{f})(\boldsymbol{\rho^2})^{-1}$$

As f is a function of χ^2, we can obtain the variance covariance matrix of f by first calculating the variance covariance matrix of the χ^2 variables.

First, let that χ_i be the standardized genotype with standard normal mean z_i and non-centrality parameter μ_i, we have

$$\begin{aligned}
\mathrm{E}[\chi_i] &= \mathrm{E}[z_i + \mu_i] \\
&= \mu_i \\
\mathrm{Var}(\chi_i) &= \mathrm{E}[(z_i + \mu_i)^2] + \mathrm{E}[(z_i + \mu_i)]^2 \\
&= \mathrm{E}[z_i^2 + \mu_i^2 + 2z_i\mu_i] + \mu_i^2 \\
&= 1
\end{aligned}$$

Given the LD between two genotype χ_i and χ_j are ρ_{ij}, then

$$\begin{aligned}\mathrm{Cov}(\chi_i, \chi_j) &= \mathrm{E}[(z_i + \mu_i)(z_j + \mu_j)] - \mathrm{E}[z_i + \mu_i]\mathrm{E}[z_j + \mu_j] \\ &= \mathrm{E}[z_i z_j + z_i\mu_j + \mu_i z_j + \mu_i\mu_j] - \mu_i\mu_j \\ &= \mathrm{E}[z_i z_j] + \mathrm{E}[z_i\mu_j] + \mathrm{E}[z_j\mu_i] + \mathrm{E}[\mu_i\mu_j] - \mu_i\mu_j \\ &= \mathrm{E}[z_i z_j]\end{aligned}$$

As the genotypes are standardized, therefore $\mathrm{Cov}(\chi_i, \chi_j) == \mathrm{Cor}(\chi_i, \chi_j)$ and we can obtain

$$\mathrm{Cov}(\chi_i, \chi_j) = \mathrm{E}[z_i z_j] = \rho_{ij}$$

Given these information, we can then calculate $\mathrm{Cov}(\chi_i^2, \chi_j^2)$ as:

$$\begin{aligned}\mathrm{Cov}(\chi_i^2, \chi_j^2) &= \mathrm{E}[(z_i + \mu_i)^2(z_j + \mu_j)^2] - \mathrm{E}[z_i + \mu_i]\mathrm{E}[z_j + \mu_j] \\ &= \mathrm{E}[(z_i^2 + \mu_i^2 + 2z_i\mu_i)(z_j^2 + \mu_j^2 + 2z_j\mu_j)] - \mathrm{E}[z_i^2 + \mu_i^2 + 2z_i\mu_i]\mathrm{E}[z_j^2 + \mu_j^2 + 2z_j\mu_j] \\ &= \mathrm{E}[(z_i^2 + \mu_i^2 + 2z_i\mu_i)(z_j^2 + \mu_j^2 + 2z_j\mu_j)] - (\mathrm{E}[z_i^2] + \mathrm{E}[\mu_i^2] + 2\mathrm{E}[z_i\mu_i])(\mathrm{E}[z_j^2] + \mathrm{E}[\mu_j^2] + 2\mathrm{E}[z_j\mu_j]) \\ &= \mathrm{E}[z_i^2(z_j^2 + \mu_j^2 + 2z_j\mu_j) + \mu_i^2(z_j^2 + \mu_j^2 + 2z_j\mu_j) + 2z_i\mu_i(z_j^2 + \mu_j^2 + 2z_j\mu_j)] - (1 + \mu_i^2)(1 + \mu_j^2) \\ &= \mathrm{E}[z_i^2(z_j^2 + \mu_j^2 + 2z_j\mu_j)] + \mu_i^2\mathrm{E}[z_j^2 + \mu_j^2 + 2z_j\mu_j] + 2\mu_i\mathrm{E}[z_i(z_j^2 + \mu_j^2 + 2z_j\mu_j)] - (1 + \mu_i^2)(1 + \mu_j^2) \\ &= \mathrm{E}[z_i^2 z_j^2 + z_i^2\mu_j^2 + 2z_i^2 z_j\mu_j] + \mu_i^2 + \mu_i^2\mu_j^2 + 2\mu_i\mathrm{E}[z_i z_j^2 + z_i\mu_j^2 + 2z_i z_j\mu_j] - (1 + \mu_i^2)(1 + \mu_j^2) \\ &= \mathrm{E}[z_i^2 z_j^2] + \mu_j^2 + \mu_i^2 + \mu_i^2\mu_j^2 + 4\mu_i\mu_j\mathrm{E}[z_i z_j] - (1 + \mu_i^2 + \mu_j^2 + \mu_i\mu_j) \\ &= \mathrm{E}[z_i^2 z_j^2] + 4\mu_i\mu_j\mathrm{E}[z_i z_j] - 1\end{aligned}$$

Remember that $\mathrm{E}[z_i z_j] = \rho_{ij}$, we then have

$$\mathrm{Cov}(\chi_i^2, \chi_j^2) = \mathrm{E}[z_i^2 z_j^2] + 4\mu_i\mu_j\rho_{ij} - 1$$

By definition,

$$z_i | z_j \sim N(\mu_i + \rho_{ij}(z_j - \mu_j), 1 - \rho_{ij}^2)$$

We can then calculate $\mathrm{E}[z_i^2 z_j^2]$ as

$$\begin{aligned}\mathrm{E}[z_i^2 z_j^2] &= \mathrm{Var}[z_i z_j] + \mathrm{E}[z_i z_j]^2 \\ &= \mathrm{E}[\mathrm{Var}(z_i z_j | z_i)] + \mathrm{Var}[\mathrm{E}[z_i z_j | z_i]] + \rho_{ij}^2 \\ &= \mathrm{E}[z_j^2\mathrm{Var}(z_i | z_j)] + \mathrm{Var}[z_j\mathrm{E}[z_i | z_j]] + \rho_{ij}^2 \\ &= (1 - \rho_{ij}^2)\mathrm{E}[z_j^2] + \mathrm{Var}(z_j(\mu_i + \rho_{ij}(z_j - \mu_j))) + \rho_{ij}^2 \\ &= (1 - \rho_{ij}^2) + \mathrm{Var}(z_j\mu_i + \rho_{ij}z_j^2 - \mu_j z_j\rho_{ij}) + \rho_{ij}^2 \\ &= 1 + \mu_i^2\mathrm{Var}(z_j) + \rho_{ij}^2\mathrm{Var}(z_j^2) - \mu_j^2\rho_{ij}^2\mathrm{Var}(z_j) \\ &= 1 + 2\rho_{ij}^2\end{aligned}$$

As a result, the variance covariance matrix of the χ^2 variances can be calculated as

$$\text{Cov}(\chi_i^2, \chi_j^2) = 2\rho_{ij}^2 + 4\rho_{ij}\mu_i\mu_j$$

Now that we have calculated the variance covariance matrix of χ^2, we can get the variance covariance matrix of f as

$$\begin{aligned} Cov(f_i, f_j) &= \frac{d}{d(\chi_i^2)}\frac{\chi_i^2 - 1}{n - 2 + \chi_i^2}\frac{d}{d(\chi_j^2)}\frac{\chi_j^2 - 1}{n - 2 + \chi_j^2}Cov(\chi_i^2, \chi_j^2) \\ &= \frac{(n-1)^2}{(n - 2 + \chi_i^2)^2(n - 2 + \chi_j^2)^2}Cov(\chi_i^2, \chi_j^2) \end{aligned}$$

4.2 Simulation Study

4.2.1 Quantitative Trait

4.2.2 Case-Control Study design

4.3 Result

4.4 Discussion

4.5 Conclusion

Chapter 5

Genetic Risk Factor: Risk Prediction

5.1 Risk Estimation

5.2 Simulation Study

5.3 Result

5.4 Conclusion

Chapter 6

Summary and Conclusion

Supplementary Materials

Table S1: Primer Sequences used in real time PCR

Gene Name	Primer Sequence
Actb	ACTGAGCTGCGTTTTACACCCTTTC
Akt3	CTTCTCAGTGGCAAAATGTCAGTTA
Eomes	AATAACATGCAGGGCAATAAGATGT
Lama5	ACACGAGCGAGACCAGTGAGAAGAT
Robo3	AAGGGAGTCAAGTCCTGCTTTTCCC

Table S2: Gene set enrichment results based on the RNA Seq data. All p-values were bonferroni corrected. Details of the gene sets can be found on http://www.inside-r.org/packages/cran/WGCNA/docs/userListEnrichment

Gene Set	RNA Seq	Denovo				GWAS	
		Fromer et al. [2]	Neale et al. [3]	Sanders et al. [4]	O'Roak et al. [5]	Anney et al. [6]	Ripke et al. [7]
		Scz	ASD	ASD	ASD	ASD	PGC Scz
Post-Synaptic Density proteins (Bayes)	3.35×10^{-20}	9.14×10^{-9}	1	0.0784	9.99×10^{-3}	0.588	0.965
Neuron probable (Cahoy)	6.46×10^{-19}	2.13×10^{-7}	1	1	7.47×10^{-6}	0.607	0.11
Up CD40 stimulation in MG (AitGhezala)	4.43×10^{-10}	5.73×10^{-3}	1	1	1	0.132	0.0208
Down With Alzheimers (Blalock)	2.24×10^{-9}	0.212	1	0.0142	9.89×10^{-3}	0.145	0.887
Neuron definite (Cahoy)	6.05×10^{-6}	1	1	1	0.114	0.555	0.122
Ribosome (HumanMeta)	3.01×10^{-5}	1	1	1	1	0.476	0.418
Autism associated module (Voineagu)	3.86×10^{-5}	1	1	1	1	0.847	0.61
Cytoplasm (Foster)	5.44×10^{-5}	1	1	1	1	0.396	0.34
Down With Alzheimers (Liang)	1.18×10^{-4}	0.298	1	1	1	0.739	0.215
Up With ABeta MGactivation (GSE772)	1.53×10^{-4}	0.381	1	1	1	0.274	0.0949

Mitochondria (HumanMeta)	1.89×10^{-4}	1	1	1	1	3.29×10^{-3}	9.91×10^{-3}
GABAergic Neurons In Mouse Cortex (Sugino)	1.89×10^{-4}	1	1	1	1	0.221	0.0673
Schizophrenia possible (DiseaseGenes)	3.11×10^{-4}	0.197	1	1	0.588	0.507	0.0215
Cortex (HumanChimp)	4.10×10^{-4}	1	0.698	1	1	0.946	0.824
Down Aging mitochondria synapse (Lu)	8.09×10^{-4}	1	1	1	1	0.652	0.601
Neuron (CTX)	1.08×10^{-3}	2.74×10^{-3}	1	0.481	1	0.0528	0.115
noChangeAD heat Shock Protein Activity (Blalock)	1.48×10^{-3}	1	1	1	1	0.0315	0.118
Autism differential expression across at least one comparison (Voineagu)	1.90×10^{-3}	3.86×10^{-4}	1	1	1	0.813	0.929
Microglia(Type1) (HumanMeta)	3.20×10^{-3}	1	1	1	1	0.906	0.0187
Astrocyte (CTX)	3.58×10^{-3}	1	1	1	1	0.513	0.0308
Pr10-synaptic Compartment Proteins (Morciano)	8.13×10^{-3}	1	1	1	1	0.127	0.599
Oligodendrocyte (CTX)	0.0134	5.71×10^{-3}	0.208	1	0.0822	0.383	0.0315
Mitochondria (MouseMeta)	0.0285	1	1	1	1	0.487	0.36

downAD synaptic Transmission (Blalock)	0.0302	1	1	1	1	0.437	0.275
Up In Frontal Cortex (EarlyAD)	0.0319	1	1	1	1	0.223	7.52×10^{-3}
Glutamatergic Synaptic Function (CTX)	0.0422	1	1	1	0.187	0.7	0.0252
Glutatmatergic Synapse (MouseMeta)	0.0498	1	1	1	1	0.0312	0.969

Bibliography

[1] Simon Anders. *HTSeq: Analysing high-throughput sequencing data with Python*. 2011. URL: http:%20//www-huber.embl.de/users/anders/HTSeq/.

[2] M Fromer et al. "De novo mutations in schizophrenia implicate synaptic networks". eng. In: *Nature* 506.7487 (2014), pp. 179–184. DOI: 10.1038/nature12929nature12929[pii]. URL: http://www.ncbi.nlm.nih.gov/pubmed/24463507.

[3] B M Neale et al. "Patterns and rates of exonic de novo mutations in autism spectrum disorders". eng. In: *Nature* 485.7397 (2012), pp. 242–245. DOI: 10.1038/nature11011nature11011[pii]. URL: http://www.ncbi.nlm.nih.gov/pubmed/22495311.

[4] S J Sanders et al. "De novo mutations revealed by whole-exome sequencing are strongly associated with autism". eng. In: *Nature* 485.7397 (2012), pp. 237–241. DOI: 10.1038/nature10945nature10945[pii]. URL: http://www.ncbi.nlm.nih.gov/pubmed/22495306.

[5] B J O'Roak et al. "Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations". eng. In: *Nature* 485.7397 (2012), pp. 246–250. DOI: 10.1038/nature10989nature10989[pii]. URL: http://www.ncbi.nlm.nih.gov/pubmed/22495309.

[6] R Anney et al. "A genome-wide scan for common alleles affecting risk for autism". eng. In: *Hum Mol Genet* 19.20 (2010), pp. 4072–4082. DOI: 10.1093/hmg/ddq307ddq307[pii]. URL: http://www.ncbi.nlm.nih.gov/pubmed/20663923.

[7] S Ripke et al. "Genome-wide association analysis identifies 13 new risk loci for schizophrenia". eng. In: *Nat Genet* 45.10 (2013), pp. 1150–1159. DOI: 10.1038/ng.2742ng.2742[pii]. URL: http://www.ncbi.nlm.nih.gov/pubmed/23974872.

Appendix