2022

Examination of Common and Rare Variant Genetic Architecture of Psychiatric Disorders

Mohammad Ahangari
Virginia Commonwealth University

Examination of Common and Rare Variant Genetic Architecture of Psychiatric Disorders

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

by

Mohammad Ahangari, M.Sc.

Director: Brien P. Riley, Ph.D.
Professor, Departments of Psychiatry and Human and Molecular Genetics

Virginia Commonwealth University
Richmond, Virginia
December, 2022

Acknowledgment

I would like to start by thanking Brien for his guidance and all the insights he has shared with me as my mentor. This work was made possible by your patience, trust, and the enormous amount of time that you have dedicated to mentoring me. I couldn't have asked for a better mentor. I would also like to thank the members of my committee. Silviu, thank you for teaching me why it is important to have a solid understanding of statistics in genetic studies. I have also always enjoyed our casual conversations and will certainly miss them. Brian, I will always be grateful for your continuous support from day one. You have been instrumental in shaping my way of thinking as a scientist. I appreciate your dedication to mentoring me and teaching me how to think outside the box. Todd, I am grateful for your constant support and generosity in the past five years. Thank you for helping me become a better scientist and showing me how to conduct good science. With delays in receiving the whole genome data, I didn't think that I would get the opportunity to work with any sequence data during my Ph.D. studies. But under your guidance, I was able to work with the UK Biobank exome data to gain the necessary skills to analyze sequence data. Thank you, Mikhail, for your patience and your valuable insights and support during my training. I would also like to thank Brandon for welcoming me to the lab. You have been a great friend to me, and you made me feel right at home. Also, thank you Cathy for your kindness to me in the past five years. Most important of all, I would like to thank my loving parents and sister for always believing in me. Thank you madar va pedar. I'm only here because of your endless love and support. Thank you Nazanin for being the best sister I could have ever asked for. Doosetoon daram. I'm also thankful to you Feivel for keeping me sane in the past four years. Finally, I dedicate my dissertation work to my brave compatriots who are currently fighting for their dignity and freedom in the streets of Iran.

Table of Contents

List of Tables v

List of Figures vi

List of Abbreviations viii

Abstract xi

CHAPTERS

I. Global Introduction 1
- 1.1 Schizophrenia Relevant Background 1
- 1.2 Linkage Studies 3
- 1.3 Theoretical Candidate Genes 4
- 1.4 Genome-Wide Association Studies 5
- 1.5 Copy Number Variants 10
- 1.6 Sequencing Studies 11
- 1.7 Genetic Overlap Between Schizophrenia and Other Disorders 15
- 1.8 Multiplex Schizophrenia Families 17
- 1.9 Aims 19

II. Evaluating the role of common risk variation in the recurrence risk of schizophrenia in multiplex schizophrenia families 24
- 2.1 Abstract 24
- 2.2 Introduction 25
- 2.3 Methods 27
- 2.4 Results 36
- 2.5 Discussion 39

III. Examining the source of increased bipolar disorder and major depressive disorder common risk variation burden in multiplex schizophrenia families 44
- 3.1 Abstract 44
- 3.2 Introduction 45
- 3.3 Methods 47
- 3.4 Results 52
- 3.5 Discussion 60

IV. The relationship between polygenic risk scores and symptom dimensions of schizophrenia and schizotypy in multiplex schizophrenia families 68
- 4.1 Abstract 68
- 4.2 Introduction 69
- 4.3 Methods 71

4.4 Results 75
4.5 Discussion 82

V. Genome-wide analysis of schizophrenia and multiple sclerosis identifies shared genomic loci with mixed direction of effects 88
5.1 Abstract 88
5.2 Introduction 89
5.3 Methods 91
5.4 Results 96
5.5 Discussion 101

VI. Improving the discovery of rare variants associated with alcohol problems by leveraging machine learning phenotype prediction and functional information 109
6.1 Abstract 109
6.2 Introduction 110
6.3 Methods 114
6.4 Results 120
6.5 Discussion 125

VII. Global Discussion 131
7.1 Summary 131
7.2 Common variant analysis of multiplex schizophrenia families 132
7.3 Polygenic overlap between schizophrenia and multiple sclerosis 137
7.4 Rare variation in the genetic architecture of complex traits 138
7.5 Limitations 142
7.6 Future Directions 144

Extended Figures 148

Bibliography 150

List of Tables

1. Genetic correlation between schizophrenia and disorders used in this dissertation 16
2. Diagnoses in the Irish Study of High-Density Multiplex Schizophrenia Families.....28
3. Description of the genotyping arrays.....31
4. List of the SNPs used for polygenic risk scores construction.....32
5. Polygenic risk score comparison of schizophrenia versus population controls.....38
6. Polygenic risk score comparison of schizophrenia within the multiplex families.....39
7. Model parameters for GWAS-by-subtraction models 50
8. Heritability estimates of latent factors generated from GWAS-by-subtraction models 52
9. Polygenic risk score comparison of bipolar disorder versus population controls 55
10. Polygenic risk score comparison of major depressive disorder versus population controls....55
11. Polygenic risk score comparison of latent factors in multiplex schizophrenia families.....56
12. Concentric comparison of latent factor polygenic risk scores 66
13. Items used in the factor analysis of OPCRIT.....72
14. List of the psychiatric diagnoses with OPCRIT symptom information.....76
15. List of the psychiatric diagnoses with SIS symptom information 76
16. Model fit for factor analysis of SIS in multiplex schizophrenia families 77
17. Linear regression analysis results of polygenic risk scores and symptom dimensions 80
18. Quantile regression analysis results of polygenic risk scores and symptom dimensions 81
19. List of the significant annotation classes in predicted AUDIT-P 120
20. List of the significant annotation classes in schizophrenia.....142

List of Figures

1. Sample size of schizophrenia genome-wide association studies 7
2. Manhattan plot for the third wave of schizophrenia genome-wide association studies 8
3. Exome-wide study of schizophrenia 13
4. Imputation quality across different minor allele frequency bins 32
5. Principal component analysis of the Affymetrix array 34
6. Principal component analysis of the Illumina array 35
7. Principal component analysis of the psychArray 35
8. Fine-scale principal component analysis of the Irish sample 35
9. Mean polygenic risk score for schizophrenia across the samples 36
10. Polygenic risk score comparison of schizophrenia 38
11. Cholesky decomposition path diagrams for GWAS-by-subtraction models 50
12. Concentric diagnostic schema of multiplex schizophrenia families 53
13. Polygenic risk score comparison of bipolar disorder and major depressive disorder 54
14. Polygenic risk score comparison of latent factors in multiplex schizophrenia families 56
15. Replication of polygenic risk score of latent factors in sporadic schizophrenia cases 57
16. Polygenic transmission disequilibrium results in multiplex schizophrenia families 58
17. MiXeR analysis of the latent factors 59
18. Symptom information across diagnostic categories of multiplex schizophrenia families 75
19. Factor structure of SIS in multiplex schizophrenia families 77
20. Polygenic risk score associations with symptom dimensions 78
21. Quantile regression analysis of polygenic risk scores and symptom dimensions 79
22. MiXeR analysis of schizophrenia and multiple sclerosis 97
23. Conditional QQ plot for schizophrenia conditioned multiple sclerosis and vice versa 98
24. Conditional QQ plot for LDL conditioned on schizophrenia and multiple sclerosis 98
25. Manhattan plot for condFDR analysis of schizophrenia conditioned on multiple sclerosis 99
26. Manhattan plot for condFDR analysis of multiple sclerosis conditioned on schizophrenia 99
27. Manhattan plot for conjFDR analysis of schizophrenia and multiple sclerosis 100
28. Annotation of the SNPs jointly associated with schizophrenia and multiple sclerosis 101
29. Proportion of the variance explained by genome-wide significant SNPs 102
30. MAGIC-LASSO prediction of AUDIT in the UK Biobank 115

31. Distribution of the intervals tested in rare variant analysis of AUDIT-P 118
32. Partitioning the heritability of predicted AUDIT-P into functional categories 121
33. Distribution of empirical functional weights for AUDIT-P 122
34. SKAT-O results for observed AUDIT-P 124
35. SKAT-O results for predicted AUDIT-P 124
36. Partitioning the heritability of schizophrenia into functional categories 141

List of Abbreviations

ADH ..... alcohol dehydrogenase
ADH1B ..... alcohol dehydrogenase 1B beta polypeptide
AIC ..... Akaike information criterion
ALDH2 ..... aldehyde dehydrogenase 2
ARC ..... activity-regulated cytoskeleton-associated protein complex
AUC ..... area under the curve
AUDIT ..... Alcohol Use Disorder Identification Test
AUDIT-C .... AUDIT consumption subscale
AUDIT-P..... AUDIT problems subscale
BHR ..... Burden Heritability Regression
BMI ..... body mass index
CACNA1C ... calcium voltage-gated channel subunit alpha 1-C
CACNA1G ... calcium voltage-gated channel subunit alpha 1-G
CADD ..... Combined Annotation-Dependent Depletion
CD40 ..... CD 40 molecule
CFA ..... confirmatory factor analysis
CI ..... confidence interval
CNV ..... copy number variation
COMT ..... catechol-O-methyltransferase
condFDR ..... conditional false discovery rate
conjFDR ..... conjunctional false discovery rate
cSNV ..... common single nucleotide variation
CUL1 ..... cullin 1
DGCR2 ..... DiGeorge syndrome critical region gene 2
DISC1 ..... disrupted in schizophrenia I
DRD2 ..... dopamine receptor 2
DSM ..... Diagnostic and Statistical manual of Mental Disorders
DSM-III-R... Diagnostic and Statistical manual of Mental Disorders 3 revised
DTNBP1 ..... dystrobrevin binding protein 1
DZ ..... dizygotic
EFA ..... exploratory factor analysis
eQTL ..... expression quantitative trait loci
FANTOM.... Functional Annotation of the mouse/Mammalian Genome
FDR ..... false discovery rate
FH ..... family history
FMRP ..... fragile-X mental retardation protein
FUMA ..... Functional Mapping and Annotation
GCTA ..... genome-wide complex trait analysis
genomicSEM.... genomic structural equation modeling
GO ..... Gene Ontology
GREML ..... genome-based restricted maximum likelihood
GRIA3 ..... glutamate ionotropic receptor subunit 3
GRIN2A ..... NMDA receptor subunit 2A
GRM ..... genomic relationship matrix

GRM3 ..... metabotropic glutamate receptor 3
GRM5 ..... glutamate metabotropic receptor 5
Group-LASSO ..grouped least absolute shrinkage and selection operator LASSO
GTEx ..... genotype tissue expression
GWAS ..... genome-wide association studies
H3K27ac ..... histone 3 lysine 27 acetylation
H3K4me3 .... histone 3 lysine 4 trimethylation
HERC1 ..... HECT domain containing E3 ubiquitin protein ligase family member 1
HRC ..... haplotype reference consortium
HSD17B4 .... hydroxysteroid 17-beta dehydrogenase 4
ICD..... International Classification of Disease
IMSGC..... International Multiple Sclerosis Genetic Consortium
Indels..... insertion/deletion
ISGC ..... Irish Schizophrenia Genomics Consortium
ISHDSF..... Irish Study of High-Density Multiplex Schizophrenia Families
LASSO..... least absolute shrinkage and selection operator
LD ..... linkage disequilibrium
LDL..... low density lipoprotein
LDSC ..... LD Score Regression
MAF ..... minor allele frequency
MAGIC ..... missingness adapted group-wise informed clustered
MAPT ..... microtubule associated protein tau
MHC ..... major histocompatibility complex
MiXeR..... bivariate causal mixture modeling
ML..... machine learning
MZ..... monozygotic
NMDA ..... N-methyl-D-aspartate
NRG1..... neuregulin 1
NRGN..... neurogranin
NRX1 ..... neurexin 1
OPCRIT ..... Operational Criteria Checklist for Psychotic Disorders
OQFE ..... original quality functional equivalent
OR ..... odds ratio
PC..... principal components
PCA..... principal component analysis
PGC..... Psychiatric Genomics Consortium
PGC-SCZ .... Psychiatric Genomics Consortium Schizophrenia Group
PRS ..... polygenic risk score
pTDT..... polygenic transmission disequilibrium test
QC ..... quality control
QQ..... quantile-quantile
RB1CC1 ..... RB1 inducible coiled-coil 1
r_G ..... genetic correlation
SCHEMA .... Schizophrenia Exome Meta-Analysis
SD ..... standard deviation
SE..... standard error

SETD1A..... SET domain containing 1A, histone lysine methyltransferase
SIS..... Structured Interview for Schizotypy
SKAT ..... Sequence Kernel Association Test
SLC12A5 ..... solute carrier family 12 member 5
SNP ..... single nucleotide polymorphism
SP4 ..... Sp4 transcription factor
SV ..... structural variants
TAD ..... topologically associated domains
TCF4 ..... transcription factor 4
TF..... transcription factor
TNFAIP8 ..... TNF alpha induced protein 8
TRIO..... trio Rho guanine nucleotide exchange factor
WES ..... whole exome sequencing
WGS..... whole genome sequencing
XPO7..... exportin 7
YWHAB ..... tyrosine 3-monoxygenase activation protein beta
ZNF804A..... zinc finger protein 804A

Abstract

Psychiatric disorders are often heterogenous in their manifestation and genome-wide association studies have identified many common risk variants involved in their polygenic architectures with varying degrees of pleiotropy. In recent years, large-scale biobanks have also begun sequencing the genome of their participants to elucidate the role of rare risk variation in the genetic architecture of complex phenotypes, including psychiatric traits. This dissertation sought to better understand the role of both common and rare risk variation in the genetic architecture of psychiatric disorders with a particular focus on schizophrenia and alcohol problems. In the first three analyses, we focused on characterizing the common risk variant architecture of multiplex schizophrenia families in terms of family history, cross-disorder risk, as well as symptom severity. In the fourth analysis, we compared and contrasted the polygenic architecture of schizophrenia with multiple sclerosis, an autoimmune, neurodegenerative disorder that also shows co-occurring neuropsychiatric symptoms, beyond genetic correlation. In the fifth analysis, we used the 200k exome release of the UK Biobank to investigate the rare variant genetic architecture of alcohol problems by combining machine learning phenotype prediction and empirical information to improve rare variant discovery. Together, these studies contribute to our understanding of the genetic architecture of schizophrenia and its pleiotropic relationship with other psychiatric and neurological disorders and provides new avenues for future studies of this disease in both family and sporadic samples. Additionally, we proposed a novel framework for rare variant analysis of complex disorders that improves discovery of rare variants in biobank datasets.

CHAPTER I

Global Introduction

1.1 Schizophrenia Relevant Background

Schizophrenia is a common, highly heterogeneous psychiatric disorder with a population prevalence of ~1% (Saha et al., 2007). The age of onset for schizophrenia is in early adulthood to mid-twenties for males and late-twenties to mid-thirties for females and due to its onset in early adulthood, it requires long term care with significant cost to society (Delisi, 1992). As a psychiatric disorder, schizophrenia was first described by Emil Kraepelin as dementia praecox (Kendler, 2020), where he recognized the inherent differences between schizophrenia and bipolar disorder. Kraepelin observed that negative symptoms of schizophrenia such as anhedonia, blunted affect, and avolition were the symptoms that most clearly distinguished schizophrenia from mood disorders such as bipolar disorder. These negative symptoms reflect a loss of normal function, while positive symptoms such as delusions and hallucinations reflect a distortion in normal functioning of an individual or a gain of function. Based on the current recommendations by the Diagnostic and Statistical Manual of Mental Disorders (DSM)-V (Bhati, 2013), for a diagnosis of schizophrenia, in addition to the presence of negative symptoms and catatonia for six months, at least one positive symptom must also be present for a one-month period.

In the landmark Danish Adoption Study of Schizophrenia (Kety et al., 1975), Kety and colleagues demonstrated that although some relatives of schizophrenia probands do not satisfy

the criteria for schizophrenia, they present with signs and symptoms that closely resemble those observed in their more overtly ill relatives which clearly differentiates them from the general population. This finding provided strong evidence in support of the spectrum model of schizophrenia (Kendler, 1984; Kety et al., 1994) by showing that relatives of schizophrenia probands are at an increased risk for developing schizophrenia as well as schizophrenia related disorders. Later, these signs and symptoms were combined to create the diagnosis of schizotypal personality disorder in the DSM-III and DSM-III revised (DSM-III-R) (Squires-Wheeler et al., 1988).

While both genetic and environmental influences are thought to impact disease risk for schizophrenia, familial aggregation suggests that genetic predisposition is a major contributing factor to schizophrenia risk (Tsuang, 2000). Twin and family-based heritability of schizophrenia is estimated to be 0.7-0.8 (Hilker et al., 2018), making it one of the most highly heritable psychiatric disorders. However, despite significant progress in the past decade on the genetic basis of schizophrenia, heritability of schizophrenia using common single nucleotide variation (cSNV) from genome-wide association studies (GWAS) is currently estimated to be ~0.24 (Trubetskoy et al., 2022), suggesting that a large component of the heritability of schizophrenia remains to be identified. Schizophrenia also shows strong genetic correlation (r_G) with other major psychiatric disorders including bipolar disorder (r_G =0.67) and major depressive disorder (r_G =0.35), and a moderate genetic correlation with other psychiatric disorders such as autism spectrum disorder (r_G =0.21), anorexia nervosa (r_G =0.25), and attention deficit hyperactivity disorder (r_G =0.12) (Demontis et al., 2019; Grove et al., 2019; Howard et al., 2019; Mullins et al., 2021; Watson et al., 2019).

Despite strong evidence for genetic effects in the etiology of schizophrenia and high heritability estimates, the monozygotic (MZ) twin concordance rate for schizophrenia is around 0.48, suggesting that in addition to strong genetic components, environmental factors also play an important role in the etiology of the disorder (Tsuang, 2000). Environmental factors such as famine and war (St Clair, 2005), cannabis use (Hall & Degenhardt, 2008), living in an urban setting (J. McGrath & Scott, 2006), and season of birth (J. J. McGrath & Welham, 1999) have all been shown to contribute to the risk of schizophrenia. Additionally, infections with Herpes Simplex virus, Epstein-Barr virus and *toxpoplasma gondii* have been suggested to have some effect on the development of schizophrenia (Brown & Derkits, 2010; Dickerson et al., 2019; Khandaker et al., 2013). Taken together, these findings suggest that while genetic influences appear to have a stronger role in its etiology, schizophrenia is a complex disorder where both genetic and environmental influences act together to increase susceptibility. Therefore, genetic studies of schizophrenia should be contextualized in a frame that takes both genetic and environmental influences into account.

1.2 Linkage Studies

Due to the high familial aggregation and strong evidence for the role of genetic influences in the etiology of schizophrenia, large-scale molecular genetics efforts were designed and carried out to further understand the underlying genetic architecture of schizophrenia. Initial studies in pedigree and family samples using linkage strategies strongly suggested that there is no evidence for a single-locus Mendelian inheritance model of schizophrenia (Elston et al., 1978). In addition, the substantial drop in the concordance rate of schizophrenia from MZ twins (0.48) to dizygotic (DZ) twins (0.11), and siblings (0.086), strongly suggested that the genetic architecture of schizophrenia is unlikely to be monogenic (Mendelian), rather, a combination of

multiple genes and risk alleles are more likely to be contributing to the risk of schizophrenia (Risch, 1990).

As a result, in the mid to late 1990s and early 2000s, the single locus inheritance model of schizophrenia was scrapped in favor of a multi-locus model and a large number of linkage studies were carried out to investigate whether few loci of moderate to large effects could be responsible for the development of schizophrenia. While earlier studies identified no locus in linkage with schizophrenia (Kendler & Diehl, 1993), later studies were able to identify a number of loci with weak linkage across the genome, but with low agreement and replication across the studies (McGuffin et al., 2003; Riley, 2004). Some of the identified loci included catechol-O-methyltransferase (*COMT*), dystrobrevin binding protein 1 (*DTNBP1*), neuregulin 1 (*NRG1*), and disrupted in schizophrenia (*DISC1*) (Straub et al., 2002; Chubb et al., 2008; Egan et al., 2001; Ishizuka et al., 2006). In addition to the lack of replication across these linkage studies, the estimated effect sizes for these loci were also relatively small and no significant variants were found in these loci for follow-up analyses (Kirov et al., 2005). The lack of any replicable loci with moderate to large impact on schizophrenia risk suggest that this multi-locus, moderate effect model is also unlikely to account for schizophrenia risk. In keeping with the field-wide shift in genetic models of other complex traits, a polygenic inheritance model postulating that many variants with small effect sizes contribute to schizophrenia risk emerged and started to gain traction (Owen et al., 2005). This remains the most widely held view of the genetic architecture of most complex traits, including schizophrenia, and is strongly supported by subsequent studies.

1.3 Theoretical Candidate Genes

In addition to the genes and loci identified from linkage studies, theoretical candidate genes such as the dopamine receptor 2 (*DRD2*) or metabotropic glutamate receptor 3 (*GRM3*)

genes, involved in biological systems that are thought to be perturbed in schizophrenia) have also been studied in detail. While previous studies that focused solely on theoretical candidate genes were largely unsuccessful, in recent years, the second and third phase of the Psychiatric Genomics Consortium Schizophrenia Working Group (PGC-SCZ) found robust associations in the *DRD2* gene, encoding the target of many antipsychotic drugs, as well as genes involved in glutamatergic neurotransmission such as *GRM3* (Ripke et al., 2014; Trubetskoy et al., 2022). Additionally, large-scale exome studies of schizophrenia conducted by the Schizophrenia Exome Meta-Analysis (SCHEMA) Consortium have also identified robust associations between rare variation in the glutamate ionotropic receptor subunit 3 (*GRIA3*) and N-methyl-D-aspartate (NMDA) receptor subunit 2A (*GRIN2A*) genes involved in glutamatergic neurotransmission, providing evidence in support of some candidate genes (Singh et al., 2022).

1.4 Genome-Wide Association Studies

Lack of success in identifying replicable genomic loci in linkage and candidate gene studies suggested that schizophrenia is likely to have a polygenic architecture with many variants of small effect sizes involved in its genetic architecture. As a result, in the late 2000s, researchers started to change their approach and use case-control designs to study schizophrenia using the genome-wide association study (GWAS) framework to identify common variants with minor allele frequency (MAF) of >5%. Early studies by the Wellcome Trust Case-Control Consortium (The Wellcome Trust Case Control Consortium, 2007), The International Schizophrenia Genomics Consortium (The International Schizophrenia Genomics Consortium, 2009), and Steffanson and colleagues (Stefansson et al., 2009) among others, showed robust, replicable associations between common variants in the Major Histocompatibility Complex (MHC), transcription factor 4 (*TCF4*), nerogranin (*NRGN)*, zinc finger protein 804A (*ZNF804A*) and

other regions in the genome. These findings suggested that the genetic architecture of schizophrenia is indeed polygenic, meaning that many variants of small effect sizes are contributing to the genetic risk for this disorder. Later, it was further proposed that to capture the polygenicity of schizophrenia GWAS signals, a weighted score that sums up the log-transformed odds-ratio (OR) of the allele effect sizes can be used to aggregate the common risk variation burden of schizophrenia in cases versus controls (The International Schizophrenia Genomics Consortium, 2009). This widely used method which captures the aggregation of variants with small effect sizes is called a polygenic risk score (PRS). The PRS constructed for schizophrenia by Purcell and colleagues showed that while the combined effect sizes of these alleles can explain only a small portion of the variance for schizophrenia (~2%), the variants associated with schizophrenia risk were distributed throughout the genome in a uniform fashion. These findings suggested that if we can identify such aggregate effects in modestly sized samples (like the International Schizophrenia Consortium sample), increased sample size could lead to the discovery of many more loci.

Since then, large, collaborative efforts by the PGC-SCZ have aimed to steadily increase the available sample for analysis of schizophrenia. Figure 1 taken from the third wave of PGC-SCZ GWAS (Trubetskoy et al., 2022) demonstrates how increase in sample size across different waves of SCZ GWAS have increased statistical power for variant identification. The first wave of PGC-SCZ GWAS included 8,228 cases and 12,462 controls and identified eight loci significantly associated with schizophrenia, five of which were novel (Ripke et al., 2011). An interesting observation in the first wave of PGC-SCZ GWAS was that the vast majority of the single nucleotide polymorphisms (SNPs) identified in GWAS of schizophrenia were in intragenic regions, an observation that was further strengthened in subsequent waves of PGC-

SCZ GWAS. In 2013, a follow-up study that combined the PGC1-SCZ results with a Swedish sample of 5,001 cases and 6,243 controls, identified 13 additional loci robustly associated with schizophrenia further showing that increase in effective sample size can boost GWAS discovery power for highly polygenic complex traits such as schizophrenia (Ripke et al., 2013).

The second wave of schizophrenia GWAS (PGC2-SCZ) expanded the sample sizes to 36,989 cases and 113,075 controls (Ripke et al., 2014). At the time of its publication in 2014, this study was the largest GWAS of any neuropsychiatric disorder and uncovered 108 distinct genomic loci robustly associated with schizophrenia, 83 of which were novel. Of importance, PGC2-SCZ found strong association between schizophrenia and the *DRD2* gene for the first time. The DRD2 receptor is the target of many antipsychotic drugs used to treat schizophrenia (Zhang et al., 2015). Additionally, the PGC2-SCZ GWAS also implicated genes involved in glutamatergic neurotransmission such as *GRM3* and *GRIN2A*, providing the first evidence from large, well-powered samples for these prior candidate genes for schizophrenia.

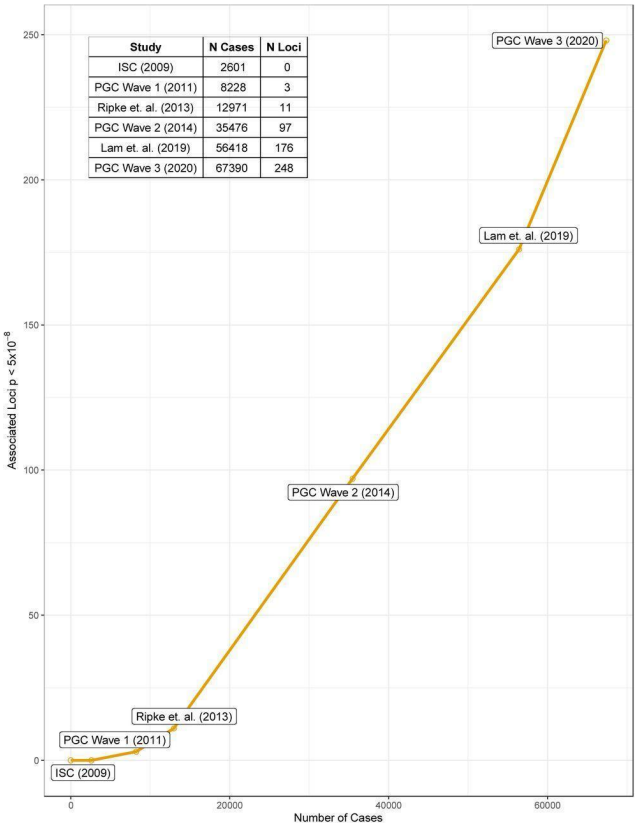


Figure 1: The relationship between sample-size and number of significant loci identified across different waves of PGC-SCZ GWAS studies. X-axis shows number of cases. Y-axis shows the number of independent loci identified in the discovery meta-analysis without replication data. Figure is reproduced from PGC3-SCZ publication.

An important recent study showed that common risk variants from PGC2-SCZ are linked to cell-specific histone modifications in

the human frontal lobe (Girdhar et al., 2018). They used psychENCODE (Wang et al., 2018) and CommonMind (Hoffman et al., 2019) resources to generate maps from neuronal, neuron-depleted, and bulk tissue chromatin for dorsolateral prefrontal cortex and anterior cingulate cortex using histone 3 lysine 4 trimethylation (H3K4me3) and histone 3 lysine 27 acetylation (H3K27ac), two histone marks associated with active promoters and enhancers. Their results indicated that common variants conferring risk to schizophrenia were significantly over-represented in neuronal H3K4me3 and H3K27ac sites, highlighting the critical role of cell-type specific signatures in uncovering the role of GWAS hits in the pathobiology of schizophrenia.

Most recently, the third wave of schizophrenia GWAS (PGC3-SCZ) increased the sample size to 76,755 cases and 243,649 controls, making it the largest GWAS of schizophrenia to date (Trubetskoy et al., 2022). The consortium reported the association between schizophrenia and 287 independent genomic loci, many of them concentrated in genes that are uniquely expressed in neurons of the central nervous system (Figure 2). Additionally, associated loci were also shown to be enriched for genes associated with rare disruptive coding variants, providing strong evidence for the convergence of common and rare variant signals in the genetic architecture of schizophrenia. Using Linkage Disequilibrium (LD) Score Regression (LDSC) (Bulik-Sullivan et al., 2015), heritability of schizophrenia from the PGC3-SCZ GWAS data is estimated to be ~0.24. Given that the twin-based heritability estimate

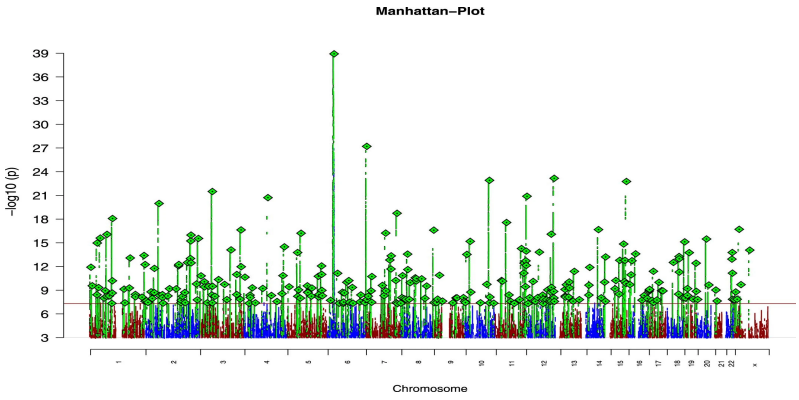


Figure 2: Current genomics landscape of schizophrenia as of 2022. The X-axis represents chromosomal position, and the Y-axis represents the significance of association. The red line represents genome-wide significance level (5 x 10^{-8}). Figure is reproduced from PGC3-SCZ publication.

of schizophrenia is estimated to be ~0.7-0.8, this suggests that at current sample sizes, only around one-third of schizophrenia heritability can be accounted for by common variant GWAS data. While it is expected that with increase in sample sizes, the portion of the heritability that is accounted for by common variant GWAS data will increase, there are other additional sources that could also contribute to schizophrenia heritability beyond common variation, such as rare variation and insertion/deletions (indels) identified through sequencing studies. For example, a recent study (Wainschtein et al., 2022) used Genome-wide Complex Trait Analysis (GCTA) Genome-based restricted maximum likelihood (GREML) method (Yang et al., 2016) to estimate the heritability of height and body mass index (BMI) by evaluating whole-genome sequencing (WGS) data. They assessed the contribution of both rare and common variants and estimated the heritability of height at 0.68 and BMI at 0.30, a number that is much closer to twin and pedigree estimates of the heritability of these two anthropometric phenotypes. These results show that most of the "missing heritability" of height and BMI that is not captured from common variant GWAS data likely resides in regions with moderate to MAF that are in low LD with each other. Using the same GREML-GCTA approach, another study (Halvorsen et al., 2020) were also able to show that in a sample of 1,162 schizophrenia cases and 936 ancestry matched controls with deep WGS data, ~0.56 of schizophrenia heritability can be accounted for when combined effects of both common and rare variants are included in the model, a number much closer to twin and pedigree estimate of the heritability of schizophrenia (~0.7-0.8). While the heritability estimates from this study had large standard errors (SE), likely due to sample size, these results nevertheless suggest that similar to other complex phenotypes such as anthropometric traits, a large portion of schizophrenia "missing heritability" may be captured through the analysis of

WGS data across different frequencies. As larger WGS sample of schizophrenia are collected, it remains to be seen whether the observation holds true with higher certainty in larger samples.

1.5 Copy Number Variants

The first evidence for the involvement of rare variants in the genetic architecture of schizophrenia was identified through copy number variation (CNV) studies (Rees et al., 2014; Walsh et al., 2008). The strongest CNV signal with impact on schizophrenia identified to date is the 22q11.2 deletion, identified as the cause of DiGeorge Syndrome (Bassett et al., 2008). This mutation is primarily *de novo* (in 90% of cases), meaning that in the majority of cases, it does not contribute to the heritability of schizophrenia. Other CNVs such as 2q15.3 deletion impacting *NRXN1*, 16p11.2, and 16p13.1 have also been shown to have significant impact on modulating the risk for schizophrenia (Ingason et al., 2011; Kirov et al., 2007; S. E. McCarthy et al., 2009).

In the most comprehensive CNV study of schizophrenia to date, the CNV Working group of the PGC-SCZ used a sample of 21,094 schizophrenia cases and 20,227 controls and showed that after accounting for all the known CNVs impacting schizophrenia risk, a global enrichment of CNV burden is still observed in cases compared to controls (Marshall et al., 2017a). Further downstream analyses suggested that CNV burden is mostly enriched in genes associated with synaptic function and neurobehavioral phenotypes. In total, there are currently eight replicable genome-wide significant CNVs with strong evidence for schizophrenia risk. These eight loci are 22q11.2, 15p11.2 proximal, 15p11.2 distal, neurexin 1 (*NRXN1*), 3q29, 7q11.2, 15q13.3, 2p16.3, and 1q21.1. An important distinction between CNVs and other genomic mutations implicated in schizophrenia is that loci identified from these studies are generally nonspecific to schizophrenia. In particular, they show pleiotropic association with epilepsy, mental retardation, and autism spectrum disorder, and it is now widely accepted that while CNVs are potent alleles with high

ORs, they are non-specific, with pleiotropic effects on multiple phenotypes (Rees & Kirov, 2021).

1.6 Sequencing Studies

Most of the sequencing studies of schizophrenia have been conducted using exome capturing strategies, with the initial studies focused on identifying *de novo* variants in trios. For example, analysis of 14 schizophrenia trios and identified an overall increased burden of *de novo* mutations in schizophrenia cases (Girard et al., 2011). In a similar study, analysis of a sample of 53 schizophrenia trios and 22 control trios identified 40 *de novo* mutations with a strong association in the DiGeorge syndrome critical region gene 2 (*DGCR2*) gene located in the 22q11.2 (DiGeorge) region (Xu et al., 2011). In a follow-up study, the same group increased the sample size to 231 schizophrenia trios and 34 control trios and were able to replicate previous findings and further show that there is an enrichment of *de novo* nonsynonymous mutations in genes with higher prenatal expression (Xu et al., 2012). In another trio design study, it was shown that there is also an increased burden of rare protein altering variation in genes involved in glutamatergic neurotransmission including glutamate metabotropic receptor 5 (*GRM5*) (Timms et al., 2013). In 2014, another study expanded the trio design to also include sporadic schizophrenia cases and were able to show that there is a 3.5-fold increase in *de novo* mutations in sporadic cases compared to familial cases of schizophrenia with excess rates of *de novo* mutations in genes with high estimated probability of haploinsufficiency, many of which showed pleiotropic association with autism and intellectual disability (McCarthy et al., 2014). A more comprehensive study expanded previous trio design studies and analyzed the exome data of 623 schizophrenia trios and 731 controls and were able to show that there is an enrichment of nonsynonymous de novo mutations in gene-sets with purported dysregulation in schizophrenia

(Fromer et al., 2014). In agreement with previous studies, they also found evidence for enrichment of loss-of-function mutations in genes with pleiotropic association with autism and intellectual disability, further underpinning the non-specificity of these associations with schizophrenia. These results also highlighted that there is an increased burden of nonsynonymous and loss-of-function mutations in genes encoding components of the activity-regulated cytoskeleton associated protein (ARC) complex, N-methyl-D-aspartate (NMDA) receptors, as well as the fragile X mental retardation protein (*FMRP*) gene, in agreement with a previous study (Kirov et al., 2012).

In contrast to the studies with a focus on trio design, Purcell and colleagues opted to use a case-control design to analyze the exome data from 2,536 schizophrenia cases and 2,543 controls (Purcell et al., 2014). While this study failed to replicate the findings implicating the genes encoding NMDA receptors and FMRP, they were able to confirm the disruptive variant signals in genes encoding components of the ARC complex from the trio study conducted by Fromer and colleagues (Fromer et al., 2014). They were also able to show that singleton disruptive variants in schizophrenia cases are significantly enriched in voltage-gated calcium ion channel genes such as calcium voltage-gated channel subunit alpha1-C (*CACNA1C*), which also shows association with schizophrenia and pleiotropic association with bipolar disorder in common variant data (Ripke et al., 2011).

An important limitation of early sequencing studies of schizophrenia is the relatively small size of the samples studied. To address this limitation, the SCHEMA Consortium was established in 2017 to aggregate existing exome data and generate the largest exome sequencing sample of schizophrenia to date. Singh and colleagues have recently published the report on the SCHEMA Consortium phase 1 project (Singh et al., 2022) which includes the exome data from

24,248 schizophrenia cases, 97,322 controls, and 3,402 trios. The sample includes individuals from diverse ancestries and shows that ultra-rare coding variants in 10 genes confer substantial risk to schizophrenia with ORs ranging from 3-50. The strongest signals were for protein truncating variants and missense variants in the SET domain containing 1A, histone lysine methyltransferase (*SETD1A*) gene. Previous studies have shown that deleterious variants are enriched in *SETD1A* in both schizophrenia and other neurodevelopmental disorders (Singh et al., 2016). *SETD1A* encodes a methyltransferase that catalyzes the methylation of lysine residues in histone H3 and other loss of function variants in this gene are shown to result in dominant Mendelian disorders as well as neurodevelopmental disorders with severe intellectual disability such as Wiedemann-Steiner syndrome, Kleefstra syndrome, and Kabuki syndrome (Fahrner & Bjornsson, 2014).

As shown in Figure 3, additional signals are identified by the SCHEMA Consortium in the trio Rho guanine nucleotide exchange factor (*TRIO)*, Sp4 transcription factor (*SP4)*, cullin 1 (*CUL1*), exportin 7 (*XPO7*), RB1 inducible coiled-coil 1 (*RB1CC1*), HECT domain containing E3 ubiquitin protein ligase family member 1 (*HERC1*), *GRIN2A*, calcium voltage-gated channel subunit alpha1-G (*CACNA1G*), and *GRIA3*, while an additional 22 genes were also identified at a false discovery rate (FDR) of <0.05. The genes identified in this study show enriched

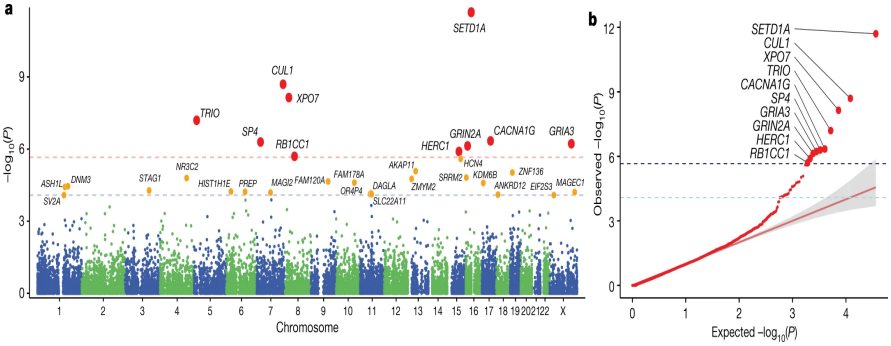


Figure 1: Manhattan plot (a) and quantile-quantile plot (b) with -log$_{10}$ transformed p-values from SCHEMA Consortium. Light blue dashed lines indicate FDR <5%. Dark blue dashed lines indicate Bonferroni correction at *P*=2.14 x 10^{-6}. Figure is reproduced from the SCHEMA Consortium publication.

expression in central nervous system neurons with diverse molecular function. Of interest, the association of *GRIN2A* and *GRIA3*, two receptor subunits involved in glutamatergic neurotransmission in the brain provide evidence for the involvement of glutamatergic signaling dysfunction in the pathogenesis of schizophrenia (McCutcheon et al., 2020). The analysis conducted by the SCHEMA Consortium, in combination with the PGC3-SCZ GWAS results described in the previous section above, strongly suggest that there is a convergence between common and rare genetic risk factors involved in pathogenesis of schizophrenia. Furthermore, SCHEMA showed that when the top genes identified were excluded from the analysis, schizophrenia cases still appeared to have a significantly increased burden of ultra-rare variants compared to controls, further suggesting that with increase in sample size over the coming years, more genes with increased rate of rare variation that confer risk to schizophrenia will be identified.

While the SCHMEA Consortium provided a framework to conduct large-scale exome analysis of schizophrenia, these studies have focused on the coding regions of the genome. With the decrease in cost of sequencing, we can expect to see more WGS studies of schizophrenia in the coming years. In the largest WGS study of schizophrenia to date, Halvorsen and colleagues analyzed high-coverage WGS data of 1,162 schizophrenia cases and 936 ancestry-matched controls from the population of Sweden and showed that ultra-rare structural variants (SV) that are near topologically associated domains (TAD) significantly increase the risk for schizophrenia (Halvorsen et al., 2020). Previous studies show that deletion of elements in TADs are associated with clear phenotypic consequences in various human diseases and are depleted of deletions in human populations compared to the rest of the noncoding regions of the genome (Redin et al., 2017). While this modestly sized WGS analysis of schizophrenia suggests that there is an excess

rate of ultra-rare SVs in TAD boundaries, the authors caution that pinpointing the exact impact of these SVs on gene expression and regulation is currently unclear and further work is needed to pinpoint their precise impact on schizophrenia risk.

Together, rare variant studies of schizophrenia using CNV, exome and WGS designs suggest that genes associated with schizophrenia are involved in brain function, with specific impact on synaptic networks. These results further suggest that many rare variants associated with schizophrenia are pleiotropic in nature with overlap with bipolar disorder (Palmer et al., 2022) and neurodevelopmental disorders. An important take-away from these studies is that while exome studies have significantly contributed to our understanding of the rare variant architecture of schizophrenia, WGS can provide us with a more comprehensive assessment of the impact of rare variation on schizophrenia risk. With continuing drop in sequencing cost over the coming years, it is expected that more study designs will focus on WGS analysis of schizophrenia instead of exome study designs. Therefore, with the increase in sample sizes for WGS studies of schizophrenia and the subsequent increase in statistical power for rare variant identification, we can expect to identify more variation in both coding and non-coding regions of the genome that confer risk to schizophrenia.

1.7 Genetic overlap Between Schizophrenia and Other Disorders

Genetic correlation describes the relationship between two traits quantified as the correlation coefficient of additive genetic effects for them (van Rheenen et al., 2019). While methods such as GCTA can be used to estimate genetic correlation using raw genotypes (Yang et al., 2011), in recent years, newly developed methods such as LDSC (Bulik-Sullivan et al., 2015) and LDAK (Speed et al., 2012) can estimate genetic correlation using only GWAS summary statistics data. Although methods that utilize GWAS summary statistics data generally produce

larger confidence intervals (CI) compared to GCTA, they have provided a much faster and more efficient method for estimating genetic correlation across many complex traits (Zheng et al., 2017). Using the GWAS summary statistics data, the Brainstorm Consortium analyzed the data from 265,218 patients with psychiatric, neurological, and behavioral disorders and 784,643 controls to quantify the genetic correlation among 25 common brain disorders and 17 phenotypic traits related to the brain (The Brainstorm Consortium., 2018). They were able to demonstrate that there is a strong genetic correlation among many psychiatric disorders at common variation level. In contrast, with the exception of migraine which showed significant genetic correlation with major depressive disorder and attention deficit hyperactivity disorder, neurological disorders appear to be distinct from one another with no significant genetic correlation observed among them or with psychiatric disorders. Furthermore, the Brainstorm Consortium was able to demonstrate that neuroticism is also significantly correlated with almost every psychiatric disorder. The high degree of genetic correlation among psychiatric disorders (Table 1) suggests that the genetic basis of many psychiatric disorders is interconnected and categorization of these disorders into distinct entities may not adequately reflect the pathogenic processes among psychiatric disorders at the genetic level (McGrath et al., 2020).

Table 1: Genetic correlation estimation between schizophrenia and other disorders analyzed in this dissertation.

Trait 1	Trait 2	Genetic Correlation
Schizophrenia	Bipolar Disorder	0.68
	Major Depressive Disorder	0.35

While genetic correlation estimates the correlation coefficient of additive genetic effects between two disorders with the same direction of effect, they fail to capture mixed direction of effects across genetic variants. For example, although there is no significant genetic correlation between schizophrenia and brain morphology, many genome-wide significant loci are shown to

be jointly associated between them (Cheng et al., 2021). Recently developed methods such as bivariate causal mixture modeling (MiXeR), or conditional/conjunctional FDR (condFDR) can quantify polygenic overlap between phenotypes with mixed direction of effects (Andreassen et al., 2013; Frei et al., 2019). For example, using the MiXeR framework, it was shown that despite no significant genetic correlation between schizophrenia and educational attainment, there is a substantial polygenic overlap between these two traits that is not captured by standard genetic correlation estimates (Frei et al., 2019). Taken together, these results suggest that while genetic correlation estimates can identify easily interpretable relationships between schizophrenia and other major psychiatric disorders, they fail to capture more complex pleiotropic relationships that exist among complex traits. However, newly developed methods such as MiXeR and condFDR can capture polygenic overlaps between complex traits beyond genetic correlation.

1.8 Multiplex Schizophrenia Families

Since the advent of GWAS, a tremendous amount of effort has gone into collecting and ascertaining case-control samples for GWAS analyses of schizophrenia, while interest in the use of multiplex families has waned. Although different waves of schizophrenia GWAS by the PGC have clearly demonstrated the utility of GWAS in identifying risk alleles associated with increased risk for schizophrenia (Ripke et al., 2014; Trubetskoy et al., 2022), family and trio samples have also shown their utility in identifying *de novo* mutations such as CNVs with larger effect sizes (Marshall et al., 2017b). In addition, decades of research using multiplex family samples suggests that schizophrenia and related disorders significantly aggregate in multiplex families (Kety et al., 1975, 1994), and epidemiological studies also suggest that multiplex families show a substantially higher recurrence risk of schizophrenia compared to sporadic cases (Käkelä et al., 2014). Furthermore, the high concordance rate of schizophrenia in MZ twins, in

combination with adoption studies showing that only biological relatives of schizophrenia are at an increased risk for developing schizophrenia (Kety et al., 1975), demonstrate that multiplex families are an important resource for understanding the underlying pathobiology of schizophrenia in parallel with case-control designs. Families also offer the chance to detect larger numbers of rare alleles than can be detected in unrelated case/control samples.

Singleton cases are considered to be the norm for most complex disorders including schizophrenia (Yang et al., 2010). However, an interesting observation in the epidemiology of schizophrenia is that while family history (FH) appears to be the strongest risk factor for developing schizophrenia (Walder et al., 2014), most singleton cases ascertained from the general population report no FH of psychotic illnesses (Esterberg et al., 2010). To expand on this, a meta-analysis (Käkelä et al., 2014) revealed that ~2/3 of schizophrenia cases report no FH of psychotic illness while most subjects with a positive FH report only a single affected relative. In contrast, members of multiplex schizophrenia families represent the extreme tail of the distribution of schizophrenia recurrence risk with substantially higher rate of schizophrenia recurrence. This discrepancy in the rate of schizophrenia recurrence risk suggests that there may be important differences in the genetic architecture of familial and sporadic schizophrenia probands that requires further investigation.

While familial samples can increase the number of putative risk alleles observed in a sample and thus, aid in rare variant identification, multiplex families may also prove to be useful for examining the role of environmental effects and assortative mating on the risk of schizophrenia, and suitable for understanding the underlying causes of schizophrenia beyond the genetic studies. Study designs that would lead to better understanding of the nature of multiplex schizophrenia families ideally require collection of large, well-ascertained samples with detailed

phenotypic information in geographical regions where large sibships or family samples with multiple affected relatives with schizophrenia can be found. While collecting such families may be more difficult than collecting case-control samples of schizophrenia, it is apparent that multiplex family samples open up new avenues for research by allowing us to explore questions that are not readily feasible to be assessed through case-control study designs. The Irish Study of High-Density Schizophrenia Families (ISHDSF) is one of the largest, well-ascertained multiplex schizophrenia family samples with genotype data (Kendler, O'Neill, et al., 1996) that can complement case-control studies of schizophrenia. It is a collaborative effort between Virginia Commonwealth University, Queen's University Belfast, and the Irish Health Research Board. Subjects were recruited in the Republic of Ireland and Northern Ireland. The full sample consists of 1,425 individuals from 270 families ascertained on the basis of two or more members with DSM-III-R diagnosis of schizophrenia or poor-outcome schizoaffective disorder. Probands in the ISHDSF sample were ascertained through public psychiatric hospitals in the island of Ireland. Interviews for the ascertainment of the subjects were conducted between April 1987 and November 1992 by psychiatrists, psychiatric nurses, and social workers after obtaining informed consent. Diagnosis in the ISHDSF Sample was carried out using modified sections of the Structured Clinical Interview for DSM-III-R for Axis I disorders and all relevant diagnostic information for each of the individuals in the families was reviewed in detail blind to their pedigree assignment. In addition to detailed structured clinical interviews with family members in the ISHDSF sample, the Operational Criteria Checklist for Psychotic Disorders (OPCRIT) (McGuffin et al., 1991) and the Structured Interview for Schizotypy (Kendler et al., 1989) were used for detailed phenotypic assessment of psychotic and non-psychotic family members, respectively. In the subsequent chapters, we will use the ISHDSF sample in combination with an

ancestry-matched case-control sample from the Irish Schizophrenia Genomics Consortium (ISGC) to explore the genetic architecture of schizophrenia in multiplex families.

1.9 Aims

Understanding the genetic basis of schizophrenia is still under active investigation. While tremendous progress has been made in recent years in identifying specific variants, loci, or genes that confer risk to schizophrenia, multiplex families have been largely overlooked in favor of case-control study designs. We hypothesized that by analyzing the available common variant array data of the ISHDSF sample, we can conduct a series of comprehensive analyses in multiplex schizophrenia families to fill gaps in the literature. Additionally, given that recently developed methods such as MiXeR and condFDR have demonstrated the presence of a polygenic overlap between complex traits in the absence of genetic correlation, we further sought to quantify the relationship between schizophrenia multiple sclerosis to further understand the shared genetic architecture of psychiatric and neurological disorders at common variation level.

In the second chapter, we investigate the role of common risk variation in the recurrence risk of schizophrenia in the ISHDSF sample. As reported in previous epidemiological studies referenced above, multiple schizophrenia families represent the upper bounds of the recurrence risk of schizophrenia in the population. This is particularly important, because sporadic cases are generally considered to be the norm for complex disorders such as schizophrenia (Yang et al., 2010) and understanding the source of this increased recurrence risk could unravel potentially important differences between familial and sporadic schizophrenia cases, and also determine the relative focus on environmental exposures as well as common and rare genetic variation in family studies of schizophrenia. To test this hypothesis, we compared the PRS loading of

schizophrenia in members of the ISHDSF samples versus ancestry-matched cases from the ISGC sample.

In the third chapter, we investigated whether members of multiplex schizophrenia families also have an increased burden of common risk variation conferring risk to other major psychiatric disorders. In addition to high aggregation of psychotic disorders in relatives of schizophrenia probands, we also observe an increased aggregation of bipolar disorder and major depressive disorder in multiplex schizophrenia families compared to the rates expected in the general population. Furthermore, as demonstrated by various cross disorder studies of psychiatric disorders (Lee et al., 2021), schizophrenia shows strong genetic correlation with bipolar disorder (r_G=0.67) and major depressive disorder (r_G=0.35). We therefore hypothesized that due to the strong genetic correlation among these three major psychiatric disorders, members of multiplex schizophrenia families may also have an increased burden of common risk variation for other correlated psychiatric disorders, and we further sought to quantify the source of this increased risk.

In the fourth chapter, we utilized the available detailed phenotypic information in the ISHDSF sample to quantify the association between polygenic risks for major psychiatric disorders and symptom dimensions in members of multiplex schizophrenia families. Epidemiological studies in schizophrenia and their relatives suggest that there is a single continuum of liability for schizophrenia and schizotypy at the phenotypic level (Fanous et al., 2001), but this relationship has not been fully established at the genetic level. While current symptom level analysis of schizophrenia has largely focused on case-control or non-clinical cohorts, we hypothesized that by using a well-ascertained sample of multiplex schizophrenia families with detailed phenotypic information, we may be able to provide empirical genetic

evidence in support of a continuum model of schizophrenia at symptom level. The concentric diagnostic approach in the ISHDSF sample reflecting narrow, intermediate and broad case definitions based on the idea of a spectrum of psychotic illness, coupled with detailed interview-based symptom level information on all subjects in the families regardless of their diagnostic status, provided us with a unique opportunity to address this important gap in the literature of multiplex schizophrenia families.

In the fifth chapter, we moved beyond the analysis of multiplex families and disorders with significant correlation with schizophrenia. Instead, we investigated the relationship and shared genetic architecture of schizophrenia and multiple sclerosis. We hypothesized that prior genetic evidence implicating immune dysregulation in patients with schizophrenia, coupled with co-occurring neuropsychiatric symptoms observed in patients with progressive multiple sclerosis, suggest that there may be a polygenic overlap between these two disorders not captured by standard genetic correlation estimation that can be exploited for further downstream analyses. We used state-of-the-art statistical methods such as MiXeR and condFDR frameworks to quantify this polygenic overlap and identify shared genomic loci between these two disorders beyond genetic correlation.

The analyses described in the sixth chapter are slightly different from the preceding chapters. Initially, these analyses were planned to be carried out on the WGS data of schizophrenia from the sequencing data of the ISHDSF and ISGC samples. However, due to various delays from our collaborators, we did not receive the WGS data in time for inclusion. As a result, we attempted to perform these analyses using the whole-exome sequencing (WES) data from the UK Biobank to test our analysis pipeline using another available dataset and to also provide training in the handling and analysis of sequence data. We note that we view the

analyses presented in this chapter as a stand-alone project, but we also emphasize that this rare variant framework can be extended to other phenotypes and sequencing datasets. Upon receiving the WGS data for the ISHDSF and ISGC samples, future lab members in the Riley Lab will use this pipeline for WGS analysis of schizophrenia.

To expand on this, in this chapter we describe a pipeline to perform rare variant analysis of problematic alcohol use by incorporating empirical functional information as a priori weights for interval-based testing on sequence data. Alcohol use disorder is a moderately heritable psychiatric condition and GWAS have identified many common variants associated with alcohol use disorder. However, rare variant investigations of alcohol use disorder and related disorders have yet to achieve large enough sample sizes to have adequate power. Here, we sought to address this gap in the literature by conducting an interval-based rare variant (MAF < 0.01) analysis of the Alcohol Use Disorder Identification Test Problems subscale (AUDIT-P) using both machine learning (ML) phenotype prediction and empirical functional weights by utilizing the 200K release of the UK Biobank WES dataset. We hypothesized that through this analysis, we would be able to show that increase in effective sample size and inclusion of functional information through ML phenotype prediction and empirical functional weights can refine and increase statistical power for rare variant association testing.

CHAPTER II

Evaluating the role of common risk variation in the recurrence risk of schizophrenia in multiplex schizophrenia families

2.1 Abstract

Multiplex families have higher recurrence risk of schizophrenia compared to the families of sporadic cases, but the source of this increased recurrence risk is unknown. We used schizophrenia genome-wide association study data (N=156,509) to construct PRS in 1,005 individuals from 257 multiplex schizophrenia families, 2,114 ancestry-matched sporadic cases, and 2,205 population controls, to evaluate whether increased PRS can explain the higher recurrence risk of schizophrenia in multiplex families compared to ancestry-matched sporadic cases. Our results show that schizophrenia PRS in familial cases does not differ significantly from sporadic cases either with, or without FH of psychotic disorders (All sporadic cases p = 0.90, FH+ cases p = 0.88, FH- cases p = 0.82), suggesting that increased schizophrenia PRS is unlikely to account for the higher recurrence risk of schizophrenia in multiplex families. In the absence of elevated PRS, segregation of rare risk variation or environmental influences unique to the families may explain the increased familial recurrence risk. These findings also further validate a genetically influenced psychosis spectrum, as shown by a continuous increase of common schizophrenia risk variation burden from unaffected relatives to schizophrenia cases in multiplex families.

2.2 Introduction

Schizophrenia is a severe, clinically heterogeneous psychiatric disorder with a population prevalence of ~1% (Saha et al., 2007). Twin, family, and adoption studies consistently show a strong genetic component, with heritability estimates of around 0.75-0.80 (Cannon et al., 1998; Cardno & Gottesman, 2000; Heston, 1966; Kendler et al., 1985; Tienari et al., 2000), and FH remains the strongest risk factor for developing schizophrenia (Walder et al., 2014). Despite high heritability, ~2/3 of schizophrenia cases report no FH of psychotic illness, and most subjects with a positive FH (FH+) report only a single affected relative (Esterberg et al., 2010; Käkelä et al., 2014), concordant with the rates of 31% FH+ and 69% family history negative (FH-) observed in the sample of sporadic schizophrenia cases analyzed in this study (Riley et al., 2010).
The ISHDSF sample (Kendler, O'Neill, et al., 1996) consists of 257 multiplex schizophrenia families with genotype data, ascertained to have two or more first-degree relatives meeting the DSM-III-R criteria for schizophrenia or poor-outcome schizoaffective disorder. Such multiplex families display substantially higher recurrence risk of schizophrenia than reported in sporadic cases, and this discrepancy in recurrence risk suggests that there may be important differences in the genetic architecture between familial and sporadic schizophrenia cases that warrant further investigation.

One explanation of this difference is that familial schizophrenia cases may carry a higher burden of common schizophrenia risk variation as measured by a higher schizophrenia PRS, than ancestry matched sporadic cases. Another explanation is that the increased recurrence risk in multiplex families may be attributable to segregation of rarer, higher risk variation, identifiable through WES or WGS likely in combination with common risk variation. Sequencing studies suggest that rare, deleterious variation in the genome is involved in the genetic etiology of

schizophrenia and other psychiatric disorders (Ament et al., 2015; Cruceanu et al., 2018; Goes et al., 2016; Homann et al., 2016; Okayama et al., 2018; Palmer et al., 2022; Singh et al., 2022; Toma et al., 2014), but the extent to which rare variation contributes to schizophrenia risk in multiplex families is currently unknown. A third hypothesis, not addressed here, is that familial cases may have increased exposure to environmental risks unique to the families that may explain the higher recurrence risk in multiplex families.

Mega-analyses of schizophrenia GWAS data by the PGC-SCZ working group have identified 287 loci associated with schizophrenia (Ripke et al., 2011, 2014; Trubetskoy et al., 2022). GWAS data from such studies are frequently used to construct PRS to index an individual's common genetic variant risk for a disorder. Although current PRS currently lack power to predict schizophrenia in the general population, they have been shown to index meaningful differences in schizophrenia liability among individuals. For example, in the European PGC3-SCZ sample, the highest PRS centile has an OR of 44 (95% CI=31-63) for schizophrenia compared to the lowest centile of PRS, and OR of 7 (95% CI=5.8-8.3) when the top centile is compared with the remaining 99% of the individuals in the sample (Trubetskoy et al., 2022).

Common risk variation analyses in multiplex family samples smaller than ISHDSF have been performed (Andlauer et al., 2021; de Jong et al., 2018; Szatkiewicz et al., 2019), and we have previously used the summary statistics from the first wave of PGC-SCZ mega-analysis (Ripke et al., 2011) to investigate whether the concept of the genetically influenced psychosis spectrum is supported by empirical data in multiplex schizophrenia families (Bigdeli et al., 2014). Here, we extend our previous work by using PRS profiling in multiplex schizophrenia families, sporadic schizophrenia cases and population controls, all from the population of the

island of Ireland, to directly test whether common schizophrenia risk variation in the genome may explain the increased recurrence risk of schizophrenia in multiplex families. Identifying the source of the increased familial recurrence risk of schizophrenia is important for future research into the genetic etiology of familial schizophrenia, and potentially for both diagnosis and treatment of schizophrenia with different familial backgrounds, as it will determine the relative focus on environmental exposures, as well as common and rare genetic variation in case-control and family studies of schizophrenia.

2.3 Methods

Sample Description

Irish Study of High-Density Schizophrenia Families (ISHDSF)

Fieldwork for the ISHDSF sample was carried out between 1987 and 1992, with probands ascertained from public psychiatric hospitals in the Republic of Ireland and Northern Ireland, with approval from local ethics committees (Kendler, O'Neill, et al., 1996). Inclusion criteria were two or more first-degree relatives meeting DSM-III-R criteria for schizophrenia or poor-outcome schizoaffective disorder, with all four grandparents being born in Ireland or the United Kingdom. Relatives of probands suspected of having psychotic illness were interviewed by trained psychiatrists, and trained social workers interviewed other relatives. Hospital and out-patient records were obtained and abstracted in > 98% of cases with schizophrenia or poor-outcome schizoaffective disorder diagnoses. To avoid bias and detect possible mistakes in diagnosis, independent review of all diagnostic information such as interview, FH reports, and hospital information was made blind to family assignments by two trained psychiatrists, with each psychiatrist making up to 3 best estimate DSM-III-R diagnoses, with high agreement between the two psychiatrists (weighted k= 0.94 +- 0.05).

The concentric diagnostic schema of the ISHDSF shown in Table 2 includes 4 case definitions: *narrow* case definition (schizophrenia, poor-outcome schizoaffective disorder and simple schizophrenia), *intermediate* case definition (adding schizotypal personality disorder, schizophreniform disorder, and delusional disorder, psychosis not otherwise specified, and good-outcome schizoaffective disorder), *broad* case definition (adding psychotic affective illness, paranoid, avoidant and schizoid personality disorders, and other disorders that significantly aggregate in relatives of probands based on previous epidemiological work in Ireland (Kendler et al., 1993)) and *very broad* case definition (adding any other psychiatric illness in the families). The ISHDSF sample also includes *unaffected* family members with no diagnosis of any psychiatric illness. The ISHDSF diagnostic schema is described extensively elsewhere (Levinson et al., 2012).

Table 2: List of the diagnoses present in the ISHDSF sample.

Irish study of high-density schizophrenia families (ISHDSF)

			Narrow	Intermediate					Broad						Very Broad					
Category	Description	Schizoph-renia	Schizoaff-ective disorder[1]	Psychosis NOS	Schizoph-reniform disorder	Delusional disorder	Schizotypal personality disorder	Avoidant personality disorder	Schizoid personality disorder	Paranoid personality disorder	Bipolar disorder	Major depressive disorder	Alcohol dependence/ abuse	Generalized anxiety disorder	Panic disorder	Other[3]	Unaffected relatives[2]
Narrow (N = 469)	Narrow Schizophrenia spectrum	409	60														
Intermediate (N = 112)	Intermediate Schizophrenia spectrum		30	34	10	7	31										
Broad (N = 52)	Disorders significantly aggregating in the relatives of probands							9	2	2	17	22					
Very Broad (N = 140)	Any other psychiatric disorders in the relatives of probands										4	80	27	21	6	2	
Unaffected (N = 232)	Unaffected relatives of probands																232
		409	90	34	10	7	31	9	2	2	21	102	27	21	6	2	232

The concentric diagnostic hierarchy of the ISHDSF contains four case definitions: narrow, Intermediate, Broad and Very Broad spectrums. These case definitions in the ISHDSF reflect core and periphery of the psychosis spectrum based on previous genetic epidemiology work referenced in the methods section. A visual representation is provided in Supplementary Fig. 4.
[1]Poor-outcome and good-outcome schizoaffective cases are represented in narrow and intermediate diagnostic category respectively.
[2]There are four individuals with intellectual disability in the unaffected relatives category.
[3]Other diagnoses include Anorexia Nervosa (1) and Cyclothymia (1) in the very-broad diagnostic category.

<u>*Irish Schizophrenia Genomics Consortium Case/Control Sample (ISGC)*</u>

The ISGC sample was assembled for a GWAS of schizophrenia in Ireland. Details of recruitment, screening, and quality control (QC) methods used for the ISGC sample have been previously described in detail elsewhere (Irish Schizophrenia Genomics Consortium,

2012). Briefly, the case sample was recruited through community mental health service and inpatient units in the Republic of Ireland and Northern Ireland following protocols with local ethics approval. All participants were interviewed using a structured clinical interview for DSM-III-R or DSM-IV, were over 18 years of age and reported all four grandparents born either in Ireland or the United Kingdom. Cases were screened to exclude substance-induced psychotic disorder or psychosis due to a general medical condition. A subset of sporadic cases sampled by Virginia Commonwealth University (N=745) have genotypic data and FH information available (Riley et al., 2010) from completion of the family history research diagnostic criteria interview (Andreasen et al., 1977). This includes 233 (~31%) FH+ cases and 512 (~69%) FH- cases, in close concordance with the other large meta-analyses (Esterberg et al., 2010; Käkelä et al., 2014). Controls from the Irish Biobank used in ISGC were blood donors from the Irish Blood Transfusion Service recruited in the Republic of Ireland. Inclusion criteria were all four grandparents born in Ireland or the United Kingdom and no reported history of psychotic illness. Due to the relatively low lifetime prevalence of schizophrenia, misclassification of controls should have minimal impact on power (Colhoun et al., 2003).

All subjects from the ISHDSF and ISGC provided informed consent to participate in the study procedures. All procedures contributing to the sample collection comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. Procedures were approved by St. James Hospital / Adelaide and Meath Hospital – National Children's Hospital Research Ethics Committee with approval number 2009/09/04, Scotland A Research Ethics Committee with approval number 11/SS/0041, and Virginia Commonwealth University Institutional Review Board with approval number HM12497.

Genotyping and QC

Samples were genotyped using 3 different arrays (Table 3). 830 individuals representing 237 families from the ISHDSF sample were genotyped on the Illumina 610-Quad Array. An additional 175 ISHDSF individuals from 52 families were later genotyped on the Infinium PsychArray V.1.13 Array. For the ISGC sample, 1,627 sporadic cases and 1,730 controls were successfully genotyped using the Affymetrix V.6.0 Array, either at the Broad Institute or by Affymetrix. An additional 487 sporadic cases and 475 controls were later genotyped on the PsychArray along with the additional ISHDSF individuals described above. The same QC protocols were applied to all three datasets and full details are described elsewhere for the ISHDSF (Levinson et al., 2012) and the case-control sample (Irish Schizophrenia Genomics Consortium, 2012). Exclusion criteria for samples were a call rate of <95%, more than one Mendelian error in the ISHDF sample, and difference between reported and genotypic sex. Exclusion criteria for SNPs were MAF <1%, call rate <98%, and $p<0.0001$ for deviation from Hardy-Weinberg expectation. The final ISHDSF sample included 1,005 individuals from 257 pedigrees, and the final case-control sample included 4,319 individuals (2,114 sporadic cases and 2,205 controls), whose SNP data passed all QC filters.

Table 3: Description of the genotyping arrays used in this dissertation. Number of individuals in each diagnostic category and pre/post imputation SNPs on each array are provided.

Array	ISGC	ISHDSF	N SNP's pre-imputation	N SNPs post-imputation
Illumina 610-Quad - **Narrow** - **Intermediate** - **Broad** - **Very Broad** - **Unaffected**	**NA**	**830** - **430** - **102** - **50** - **36** - **211**	**557,373**	**9,298,012**
Affymetrix V.60 - **Case** - **Control**	**1,730** - **1,509** - **1,731**	**NA**	**686,646**	**11,080,279**
Infinium psychArray v.1.13 - **Narrow** - **Intermediate** - **Broad** - **Very Broad** - **Unaffected** - **Case** - **Control**	**1,296** - **716** - **580**	**176** - **39** - **10** - **1** - **105** - **21**	**384,389**	**11,081,999**

Imputation

Genotypes passing QC were phased using Eagle V.2.4 (Loh et al., 2016) and phased genotypes were then imputed to the Haplotype Reference Consortium (HRC) reference panel (McCarthy et al., 2016) on the Michigan Imputation Server using Minimac4 (Das et al., 2016). The HRC reference panel includes 64,975 samples from 20 different studies that are predominantly of European ancestry, making it suitable for imputation of the samples studied here. Each of the genotype sets were imputed and the imputed genotype probabilities were extracted and used for PRS construction and downstream analyses. As part of the post-imputation QC, variants with MAF <1% and imputation quality score of <0.3 (Auton et al., 2015) were excluded for the initial merging. After imputation and all QC steps, 9,298,012 SNPs in the Illumina Array, 11,080,279 SNPs in the Affymetrix Array, and 11,081,999 SNPs in the

PsychArray remained for analysis. In total, 9,008,825 SNPs were shared across all three arrays and were used for PRS construction and all downstream analyses. As shown in Figure 4 and Table 4, the mean imputation quality for the SNPs used for PRS construction and downstream analyses on each array was high (mean for all ≥ 0.96).

Table 4: HapMap3 SNPs across the 3 arrays used for PRS construction. Mean imputation quality scores (SD) are provided on the 4th column for the imputed SNPs.

Array	Genotyped	Imputed	Mean Imputation r^2 (SD)
Affymetrix V.60	443,872	499,148	0.98 (0.041)
Illumina 610-Quad	414,052	528,968	0.98 (0.035)
Infinium PsychArray	338,265	604,755	0.96 (0.056)

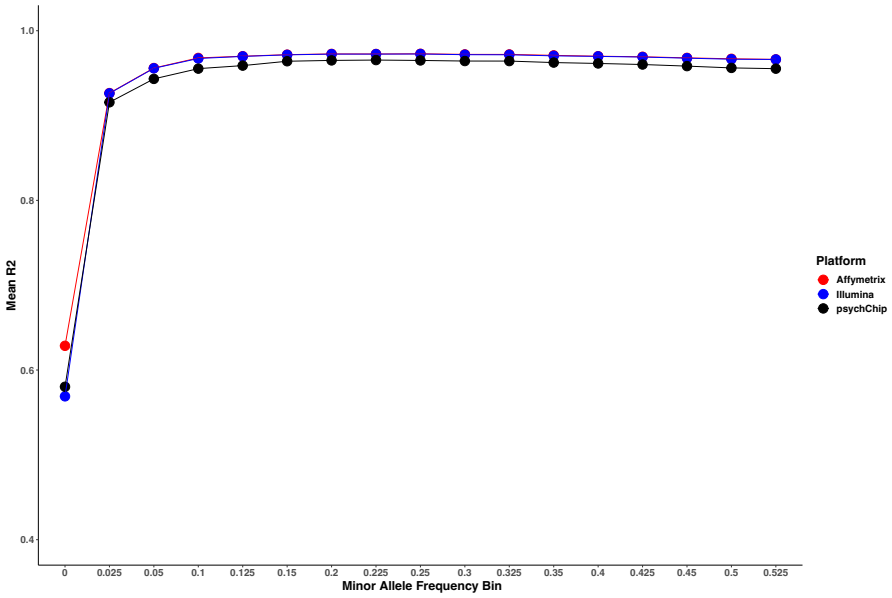


Figure 4: Comparison of imputation quality across the three arrays. Variants were binned according to their MAF and imputation R2 averaged across variants in each bin. Arrays were imputed to the HRC reference panel. X-axis shows MAF. Y-axis shows mean R2 based on the MAF bin.

Construction of Polygenic Risk Scores

The ISGC and the ISHDSF cohorts are part of the PGC3-SCZ GWAS. To avoid upward bias in PRS estimations, we acquired leave-N-out schizophrenia summary statistics from the PGC by excluding all cohorts containing any Irish subjects included in the current study. The leave-N-out GWAS summary statistics for PGC3-SCZ (N= 156,509) were first QC'd by

excluding variants with MAF < 1% and imputation quality score of < 0.9, as well as removing strand ambiguous variants and insertion deletion polymorphisms. We then constructed PRS for all subjects using a Bayesian regression framework by placing a continuous shrinkage prior on SNP effect sizes using PRS-CS with phi value of 1e-2 (Ge et al., 2019). PRS-CS uses LD information from 1000 Genomes European Phase 3 European sample (Clarke et al., 2017) to estimate the posterior effect sizes for each SNP. Although *p*-value thresholding methods have been previously used frequently, PRS-CS has shown substantial improvement in predictive power compared to those methods (Ni et al., 2021). Similar to LD Score regression (Bulik-Sullivan et al., 2015), PRS-CS limits the SNPs for PRS construction to approximately 1.2 million variants from HapMap3. By restricting the variants to HapMap3, the partitioning provides ~ 500 SNPs per LD block which substantially reduces memory and computational costs. The constructed PRS using the PRS-CS method were normalized against the score distribution in the population control for subsequent analyses.

To show the specificity of the PRS constructed from PGC3-SCZ, an additional PRS for low density lipoprotein (LDL, N=87,048) from the ENGAGE Consortium (Surakka et al., 2015) was also constructed using the same protocols described above. Genetic correlation and Mendelian Randomization studies show that there is no genetic correlation or causal relationship between schizophrenia and LDL, making LDL an appropriate comparison phenotype in which no inflation of schizophrenia PRS would be expected (Bulik-Sullivan et al., 2015; Zheng et al., 2017).

Genomic Relationship Matrix, Principal Component and Statistical Analyses

Statistical analyses were carried out using the mixed effects logistic regressions GMMAT package in R (Chen et al., 2016; R Core Team, 2020). To account for the high degree of

relatedness among individuals, we used the *glmm.wald()* function, fitted by maximum likelihood using Nelder-Mead optimization. Family structure was modeled as a random effect with genomic relationship matrix (GRM) calculated using LDAK (Speed et al., 2012) in all family members as well as sporadic cases and population controls. Principal component analysis (PCA) of the full sample is consistent with all individuals in the sample having European ancestry (Figures 5-7). However, to account for fine-scale structure within the Irish population (Figure 8), the top 10 principal components (PC) were also included as covariates in the analyses. The decision to include the top 10 PCs as covariates was made because the variance explained by PCs While none of the PCs showed association with genotype arrays or sites, to account for other possible batch effects due to genotyping carried out on different arrays or at different sites, we included platform and site as covariates in the model. The final regression models included GRM as a random effect covariate, with the top 10 PCs, genotyping platform, site, and sex as fixed effect covariates. The results were adjusted for multiple testing correction using the Holm method. While less stringent than Bonferroni, the family-wise error rate for the Holm method is similar to Bonferroni, making it suitable for multiple testing correction in modestly sized cohorts.

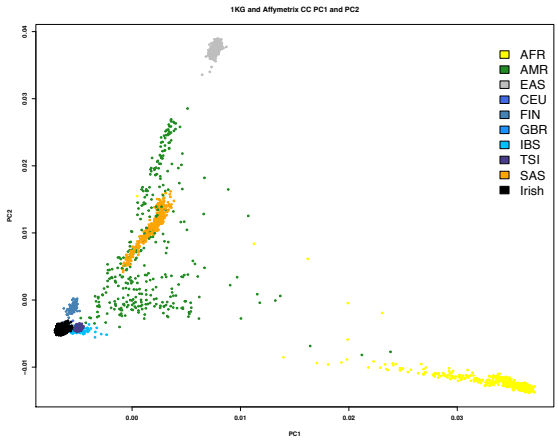


Figure 5: Continental PCA plot for the samples genotyped on the Affymetrix array projected on the 1000 Genomes Phase 3 data. X-axis PC1. Y-axis PC2.

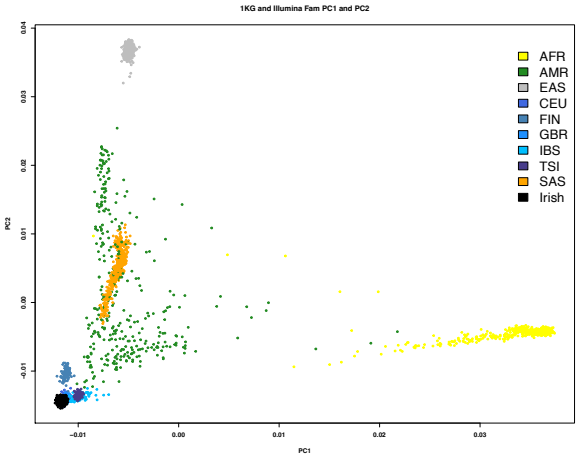


Figure 6: Continental PCA plot for the samples genotyped on the Illumina array projected on the 1000 Genomes Phase 3 data. X-axis PC1. Y-axis PC2.

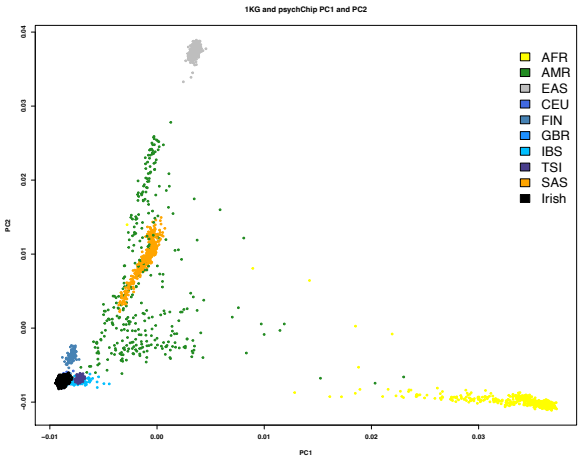


Figure 7: Continental PCA plot for the samples genotyped on the psychChip array projected on the 1000 Genomes Phase 3 data. X-axis PC1. Y-axis PC2.

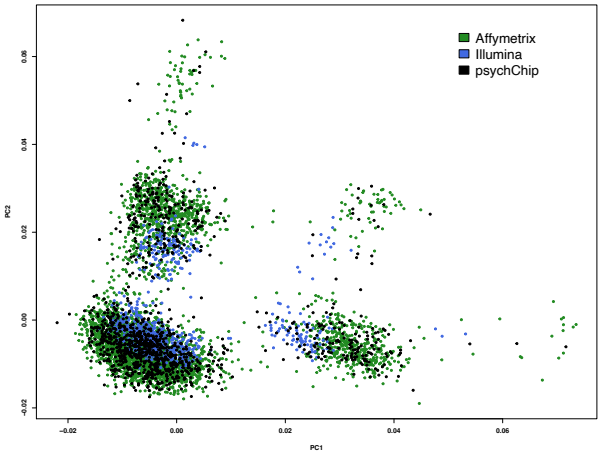


Figure 8: Fine-scale PCA analysis of the Irish population. X-axis PC1. Y-axis PC2.

2.3 Results

The mean schizophrenia PRS across the diagnostic categories for schizophrenia are displayed in Figure 8. No significant differences in LDL PRS were observed between any of the diagnostic categories compared to population controls, indicating the specificity of PGC3-SCZ PRS in this study. PGC3-SCZ PRS results (Figure 9) show that the *Narrow* case definition in the families, which includes familial cases of schizophrenia, had the highest mean PRS (Z=1.13, SE=0.09) followed by sporadic cases (Z=1.06, SE=0.09), *intermediate* case definition (Z=0.81, SE=0.10), *broad* case definition (Z=0.67, SE=0.11), *very-broad* case definition (Z=0.53, SE=0.098), *unaffected* family members (Z=0.36, SE=0.10) and population controls (Z=0.004, SE=0.07).

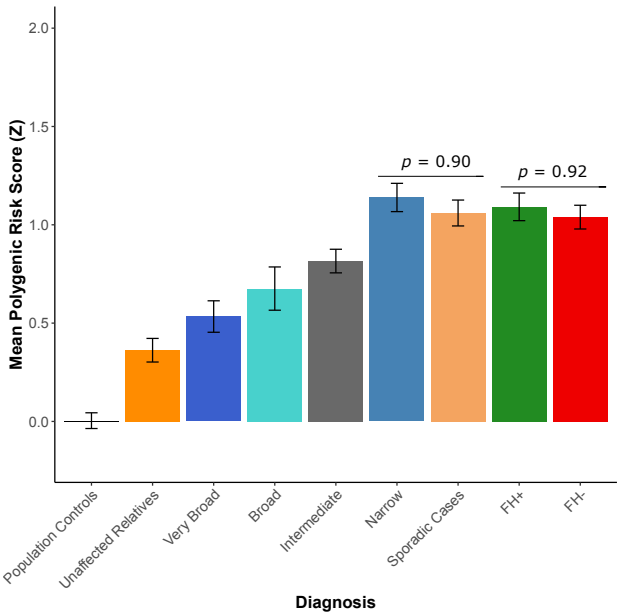


Figure 9: Mean Leave-N-Out PGC3-SCZ PRS for each of the diagnostic categories in the ISHDSF, sporadic cases, and ancestry-matched population controls. Bars represent standard errors.

No significant difference between familial and sporadic cases of schizophrenia

We observe no significant difference in PRS between familial schizophrenia cases and all sporadic schizophrenia cases, (p = 0.90), nor between familial schizophrenia cases and either FH+ (p = 0.88) or FH- (p = 0.82) sporadic schizophrenia cases. These results suggest that an increased burden of common schizophrenia risk variation is unlikely to account for the higher recurrence risk of schizophrenia in multiplex families (Figure 9). Additionally, we show that there is no significant difference in schizophrenia PRS between FH+ and FH- sporadic schizophrenia cases (p = 0.92), suggesting that the inclusion of all sporadic cases in the

comparison is unlikely to cause an upward bias in the mean PRS for the full cohort of sporadic cases, and further supporting the hypothesis that increased PRS is unlikely to account for FH of schizophrenia in the cohort studied here.

All family members carry a high burden of common schizophrenia risk variants

Familial and sporadic schizophrenia cases show a significantly higher mean schizophrenia PRS compared to all other diagnostic categories in the ISHDSF sample and ancestry-matched population controls (Figure 10, Table 5), underlining the important role of common risk variation in the genetic architecture of both familial and sporadic schizophrenia cases. All other ISHDSF diagnostic categories also show a significantly higher schizophrenia PRS compared to the population controls. PRS comparison within the ISHDSF sample (Table 6) shows no significant difference between mean PRS for *intermediate* and *broad* categories, indicating that individuals in both categories have a similar burden of common schizophrenia risk variants despite the presence of a range of diagnoses on the psychosis spectrum such as atypical psychosis and delusional disorder in the *intermediate* category, and disorders such as major depressive disorder with psychotic features, and bipolar disorder in the *broad* category. We observed no significant difference in schizophrenia PRS loading between the *broad* category and the *very-broad* category, which includes any other psychiatric disorder in the ISHDSF sample. The mean schizophrenia PRS in the *very broad* category is not significantly different from the *unaffected* members of the families, indicating a similar burden of common schizophrenia risk variation in these two distinct diagnostic categories. Finally, we observe a significantly higher schizophrenia PRS in *unaffected* family members compared to the population controls ($P = 4.13 \times 10^{-3}$), indicating a high baseline risk for schizophrenia in all members of multiplex families compared to population controls, regardless of their diagnostic

status. This observation is consistent with schizophrenia transmission through some unaffected family members observed in the ISHDSF and other family samples.

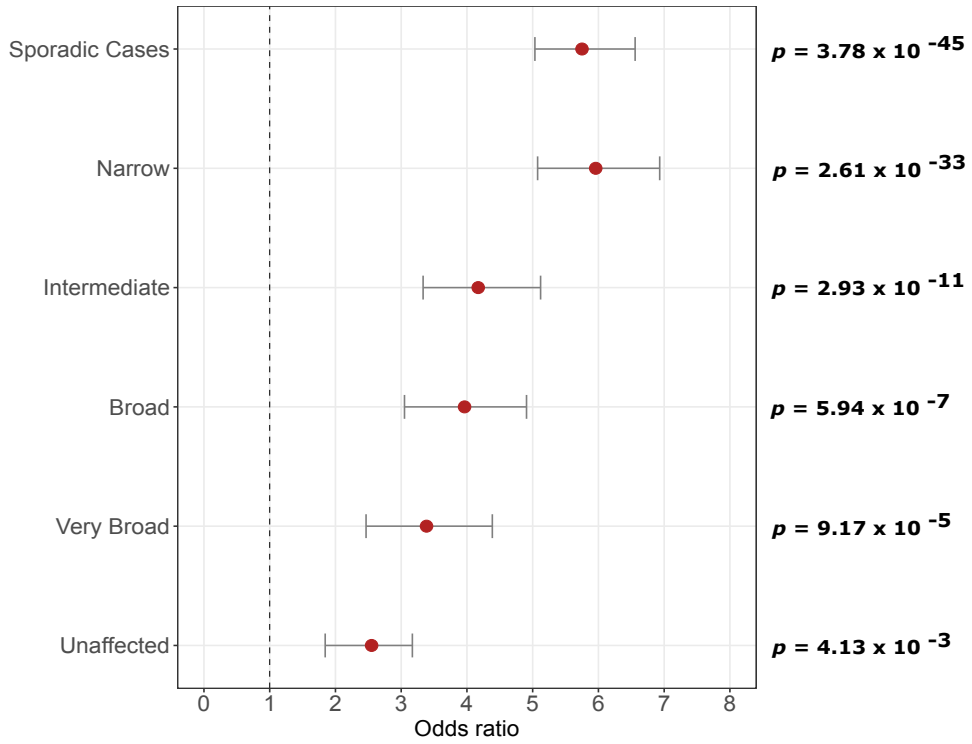


Figure 10: Comparison of PRS between ISHDSF diagnostic categories and sporadic cases, versus population controls. Odds ratios (OR) and confidence intervals (CI) are shown for each comparison. *P*-values are corrected for multiple testing comparison.

Table 5: PRS comparison between ISHDSF and sporadic cases versus population controls.

Comparison Groups	**OR**	**CI (95%)**	***P*-value**	**Adjusted *P*-value**
Sporadic Cases vs Control	5.75	5.03-6.55	1.64E-48	3.78E-45
Narrow vs Control	5.95	5.07-6.93	1.18E-39	2.61E-33
Intermediate vs Control	4.17	3.33-5.11	1.54E-13	2.93E-11
Broad vs Control	3.96	3.05-4.90	3.72E-08	5.94E-7
Very Broad vs Control	3.38	2.46-4.38	6.12E-06	9.17E-5
Unaffected vs Control	2.55	1.84-3.16	2.75E-4	4.13E-3

Table 6: Full PRS comparison results within the ISHDSF sample.

Comparison Groups	***P*-value**	**Adjusted *P*-value**
Narrow vs Sporadic Cases	0.29	0.90
Narrow vs FH+ Sporadic Cases	0.18	0.82
Narrow vs FH- Sporadic Cases	0.23	0.88
FH+ vs FH- Sporadic Cases	0.32	0.92
Sporadic Cases vs Intermediate	6E-4	9.1E-3
Sporadic Cases vs Broad	1.4E-5	5.9E-4
Sporadic Cases vs Very Broad	2.71E-5	6.1E-4
Sporadic Cases vs Unaffected	4.81E-15	7.61E-15
Narrow vs Intermediate	4E-4	8E-3
Narrow vs Broad	3.5E-5	4.1E-4
Narrow vs Very Broad	1.87E-05	3.17E-4
Narrow vs Unaffected	5.73E-16	1.20E-14
Intermediate vs Broad	0.41	1
Intermediate vs Very Broad	1.1E-4	8.3E-3
Intermediate vs Unaffected	3.5E-5	4.6E-4
Broad vs Unaffected	9.1E-4	8.3E-3
Broad vs Very Broad	0.45	1
Very Broad vs Unaffected	0.46	1

2.4 Discussion

Multiplex schizophrenia families represent the upper bounds of the distribution of recurrence risk for schizophrenia, and this study aimed to investigate the source of this increased recurrence risk. Since sporadic cases are considered to be the norm for most complex diseases including schizophrenia (Yang et al., 2010), this makes sporadic schizophrenia cases a good comparison group to assess whether elevated schizophrenia PRS can account for the increase in recurrence risk in familial cases. We observed that familial schizophrenia cases do not have a significantly increased PRS compared to sporadic schizophrenia cases in our modestly sized sample. We further show that this observation holds true regardless of the FH status of sporadic cases. Therefore, our finding provides empirical evidence that increased recurrence risk of

schizophrenia in the ISHDSF sample is unlikely to be attributable to an increased burden of common schizophrenia risk variation as identified from genome-wide association studies. Therefore, the hypothesis that high familial recurrence risk of schizophrenia in multiplex families may be attributable to excess rare variation in the genome specific to schizophrenia, warrants further investigation. Furthermore, these results validate the concept of a genetically influenced psychosis spectrum in multiplex schizophrenia families as shown by a continuous increase of common schizophrenia risk variation burden across all members of the ISHDSF, from *unaffected* family members to *narrow* category in the ISHDSF sample.

This analysis reveals potentially important differences in the genetic architecture of familial schizophrenia cases compared to familial bipolar disorder cases. An analysis conducted in bipolar multiplex families (Andlauer et al., 2021) has shown that unlike the familial schizophrenia cases studied here, familial bipolar cases have a significantly higher bipolar PRS compared to ancestry matched sporadic cases. There is also currently limited evidence for the involvement of rare risk variation in bipolar disorder (Palmer et al., 2022), and taken together, these results suggest greater importance of common risk variation in bipolar disorder, whereas studies of schizophrenia have demonstrated the importance of both common and rare risk variation (Fromer et al., 2014; Howrigan et al., 2020; Singh et al., 2022).

Although sequencing studies are only now reaching sample sizes sufficiently powered to detect individual rare variants and rare variant enriched genes associated with schizophrenia (Singh et al., 2022), early sequencing and rare variation studies observe consistent enrichment of rare variation in certain gene-sets and functional categories related to schizophrenia (Fromer et al., 2014). In addition, SNP signals from PGC3-SCZ GWAS are shown to be highly enriched in non coding functional sequences in the genome (Trubetskoy et al., 2022), further underscoring

the importance of conducting large scale whole-genome sequencing to identify rare variation in non-coding regions of the genome linked to schizophrenia. Results from the 1000 Genomes Project demonstrates that rare functional variation is frequent in the genome (Auton et al., 2015) and shows strong population specificity (MacArthur et al., 2012). For example, using GWAS probe intensity data in the ISGC sample used in this study, we have previously detected a rare, novel 149kb duplication overlapping the protein activated kinase 7 (*PAK7*) gene only found in the Irish population (Morris et al., 2014). This duplication is associated with schizophrenia in the ISGC ($p = 0.007$), and a replication sample of Irish and UK case-controls with 22 carriers in 11,707 cases and 10 carriers in 21,204 controls ($p = 0.0004$, OR=11.3). This duplication in the *PAK7* gene is in strong LD with local haplotypes ($p = 2.5 \times 10^{-21}$), indicating a single ancestral event and inheritance identical by descent in carriers.

We note that the liability that is captured by PRS constructed from PGC3-SCZ is currently insufficient for predicting a diagnosis of schizophrenia with area under the curve (AUC) = 0.71 (Trubetskoy et al., 2022), meaning that PRS alone cannot be used as a diagnostic tool. The results of our study further suggest that current PRS alone is unlikely to be predictive of schizophrenia recurrence risk in the families of index probands. To address both of these predictive limitations of schizophrenia PRS, additional components of genetic risk must be identified and included in order to improve both identification of future cases and recurrence risk prediction in the relatives of probands.

The results presented in this chapter should be interpreted in the context of some limitations. First, current PGC3-SCZ PRS accounts for ~2.6% of the total variance in schizophrenia liability, and genetic risks from rare and structural variation are not represented in the PRS. As a result, some known genetic risk factors for schizophrenia such as the 22-q11

deletion (Marshall et al., 2017b) are not included in PRS construction, and such genetic risk factors are best measured through direct assessment of structural variation or whole genome sequencing studies. Despite these limitations, PRS provides the most reliable measurement of common risk variation in the genome and is suitable for indexing an individual's risk for schizophrenia in this study. Second, the various diagnostic categories in the ISHDSF sample contain different numbers of subjects. For example, the lower number of individuals satisfying *broad* and *very broad* diagnostic schema in the families, means that the power of analysis in those subgroups is lower. However, the *narrow* category which includes familial schizophrenia cases in the ISHDSF sample, has the highest number of individuals across all the diagnostic categories in the ISHDSF, making the sample suitable for the main hypothesis being tested in this study. Third, FH information is only available for a subset of sporadic cases as described in the methods. However, the ratio of FH+ (~31%) and FH- (~69%) sporadic cases studied here is in close agreement with FH data from large meta-analyses samples referenced above, suggesting the subset of sporadic FH+ and FH- cases available are representative. Fourth, this analysis did not assess the common risk variant burden of each family separately, and the degree to which common risk variation may impact each family could vary between different families. Fifth, since the environmental factors unique to the families have also not been systematically assessed here, integrating rare genetic variation from whole sequencing studies with environmental influences in future analyses could further elucidate the role of rare variation and environmental influences on the recurrence risk of schizophrenia in multiplex families. Finally, as more samples from under-represented populations are collected, it is essential to replicate and show the generalizability of these findings in more diverse populations.

In conclusion, in this chapter, we showed that differences in common risk variation as indexed by current PRS, is unlikely to account for the increased recurrence risk of schizophrenia in our cohort of multiplex schizophrenia families and ancestry matched sporadic cases. Therefore, our results suggest that both common and rare schizophrenia risk variation needs to be indexed to potentially improve diagnostic and familial recurrence prediction of schizophrenia.

CHAPTER III

Examining the source of increased bipolar disorder and major depressive disorder common risk variation burden in multiplex schizophrenia families

3.1 Abstract

Psychotic and affective disorders often aggregate in the relatives of probands with schizophrenia and genetic studies show substantial genetic correlation among schizophrenia, bipolar disorder, and major depressive disorder. In this chapter, we examined the polygenic risk burden of bipolar disorder and major depressive disorder in 257 multiplex schizophrenia families (N=1,005) from the ISHDSF sample versus 2,205 ancestry-matched controls. Our results indicate that members of multiplex schizophrenia families have an increased polygenic risk for bipolar disorder and major depressive disorder compared to population controls. However, this observation is largely attributable to part of the genetic risk that bipolar disorder or major depressive disorder share with schizophrenia due to genetic correlation, rather than the affective portion of the genetic risk unique to them. These findings suggest that a complete interpretation of cross-disorder polygenic risks in multiplex families requires assessment of the relative contribution of shared versus unique genetic factors to account for genetic correlations across psychiatric disorders.

3.2 Introduction

Psychotic and affective disorders have long been viewed as two separate axes of mental illness, and early practitioners of psychiatry like Emil Kraepelin and Eugen Bleuler observed that relatives of patients with schizophrenia have an increased rate of psychiatric disorders ranging from atypical psychosis to schizophrenia spectrum personality disorders, many of which appeared to be milder versions of the symptoms observed in patients with schizophrenia (Kendler, 1985). Some of the first family studies of schizophrenia conducted in the early 20th century, confirmed that in addition to schizophrenia, a range of other psychiatric disorders on the psychosis spectrum also aggregate in the relatives of probands with schizophrenia. These findings were later solidified by the Danish Adoption Study of Schizophrenia, which showed that biological relatives of patients with schizophrenia were at an increased risk for schizophrenia as well as milder syndromes on the psychosis spectrum (Kety et al., 1975, 1994).

Large-scale GWAS conducted by the PGC have shown that common risk variation in the genome (MAF $\geq$1%) can explain a modest portion of the heritability of major psychiatric disorders (Mullins et al., 2021; Trubetskoy et al., 2022; Wray et al., 2018). Additionally, the Cross-Disorder Group of the PGC have provided robust, replicable evidence for strong genetic correlation (r_G) between schizophrenia and both bipolar disorder (r_G = 0.68) and major depressive disorder (r_G = 0.35) (Lee et al., 2019). Bipolar disorder and major depressive disorder also have a significant positive genetic correlation (r_G = 0.44) (Mullins et al., 2021). Together, these results indicate that there is substantial genetic overlap among these three disorders with varying degrees of psychotic and affective features, suggesting widespread genetic pleiotropy in psychiatric disorders at common variation level (Lee et al., 2021).

As described in the previous chapter, the ISHDSF sample (Kendler, O'Neill, et al., 1996) consists of 257 multiplex schizophrenia families with genotype data, ascertained to have two or more first-degree relatives meeting the DSM-III-R criteria for schizophrenia or poor-outcome schizoaffective disorder. In line with previous epidemiological observations in the relatives of probands with schizophrenia (Asarnow et al., 2001; Baron et al., 1983; Kendler et al., 1995), in addition to a significant aggregation of psychotic disorders, other psychiatric diagnoses including affective, personality, and substance use disorders, are also present in the relatives of the ISHDSF probands. Furthermore, our previous PRS profiling in the ISHDSF sample described in the previous chapter showed that all ISHDSF family members, including the unaffected relatives, have an increased burden of common schizophrenia risk variation compared to population controls, consistent with the polygenic architecture of psychiatric disorders, and the observation of schizophrenia transmission through some non-psychotic, or unaffected family members in this sample (Ahangari, Gentry, et al., 2022; Bigdeli et al., 2014).

The high baseline risk for schizophrenia observed across all case definitions of the ISHDSF sample, coupled with the evidence for strong genetic correlation among schizophrenia, bipolar disorder, and major depressive disorder, suggests that members of the ISHDSF may have an increased polygenic risk for bipolar disorder and major depressive disorder. In this chapter, we sought to test this hypothesis by constructing univariate bipolar disorder and major depressive disorder PRS in 1,005 subjects from 257 multiplex schizophrenia families and 2,205 ancestry-matched population controls all from the population of the island of Ireland. Given that the strong genetic correlation among schizophrenia, bipolar disorder, and major depressive disorder makes standard univariate cross-disorder PRS profiling in members of multiplex families less informative, we also used GWAS-by-subtraction (Demange et al., 2021) as

implemented in the genomic structural equation modeling (genomicSEM) framework (Grotzinger et al., 2019), to disentangle bipolar disorder and major depressive disorder polygenic signals into underlying genetic factors. By doing so, we investigated whether the increased polygenic risk for bipolar disorder or major depressive disorder in multiplex schizophrenia families is attributable to the portion of the genetic risk that bipolar disorder or major depressive disorder share with schizophrenia due to their genetic correlation, or the affective portion of the genetic risk that is unique to them. To further investigate whether polygenic risks for bipolar disorder and major depressive disorder, and their unique and shared genetic factors are over-transmitted from parents to probands in the families, we performed polygenic transmission disequilibrium tests (Weiner et al., 2016) in a subset of the ISHDSF sample with full parent-offspring information. Based on epidemiological findings in multiplex schizophrenia families and the substantial genetic correlation among schizophrenia, bipolar disorder, and major depressive disorder, we hypothesized that increased burden of common risk variation burden for bipolar disorder or major depressive disorder in multiplex schizophrenia families is likely to be due to the portion of the genetic risk that these disorders share with schizophrenia due to genetic correlation, rather than the portion of the genetic risk unique to one or both affective disorders. Therefore, by addressing these questions in this chapter, we attempted to clarify the complexity of cross-disorder PRS analyses in multiplex families.

3.3 Methods

Sample Description

Details about the sample collection and ascertainment for the ISHDSF sample and ISGC are provided in chapter 2.

Genotyping and Imputation

Details about genotyping and imputation are provided in chapter 2.

GWAS-by-subtraction

We performed GWAS-by-subtraction using the genomicSEM framework by analyzing summary statistics data for schizophrenia, bipolar disorder, and major depressive disorder. Briefly, summary statistics from leave-N-out schizophrenia GWAS excluding all Irish participants, (N=156,509) (Trubetskoy et al., 2022), bipolar disorder GWAS (N=413,466) (Mullins et al., 2021), and major depressive disorder GWAS (N=500,199) (Howard et al., 2019) were acquired and the genetic covariances between them were estimated using LDSC. SNPs were filtered for MAF < 0.01 and imputation quality < 0.8, and only SNPs that are present in both the schizophrenia and bipolar disorder datasets, or the schizophrenia and major depressive disorder datasets were used to generate GWAS-by-subtraction models. This left us with 6,361,243 SNPs for bipolar disorder, and 6,599,052 SNPs for major depressive disorder. We then used the QC'd summary statistics by first regressing them on two latent factors, a *SCZ* factor and *Affective* factor underlying bipolar disorder or major depressive disorder. Therefore, *SCZ* factors in bipolar disorder or major depressive disorder capture part of the genetic risk that each of these two disorders share with schizophrenia due to their genetic correlation, whereas *Affective* factors in bipolar disorder or major depressive disorder capture the affective portion of the genetic risk that is unique to these two disorders and not shared with schizophrenia. We then regressed *SCZ* factor and *Affective* factor on each SNP from the summary statistics that passed QC measurements as described above, which allowed for two separate paths of association with bipolar disorder or major depressive disorder for each SNP: 1) a path that is fully mediated by *SCZ* factor, and 2) a path that is fully independent of *SCZ* factor, called *Affective* factor. The models assume that genetic effects on schizophrenia are also impacting bipolar disorder and

major depressive disorder to some degree given that both these disorders have a strong genetic correlation with schizophrenia. The path diagrams for the Cholesky decomposition used to disentangle the polygenic signals is provided in Figure 11. In brief, the regression equations composing GWAS-by-subtraction model in this study are:

Bipolar disorder:

$SCZ = \lambda_1 SCZ\ factor$ $BIP = \lambda_2\ SCZ\ factor + \lambda_3 Affective\ factor$

$SCZ\ factor = \boldsymbol{\beta}_1 SNP + \mu_{SCZ\ factor}$

$Affective\ factor = \boldsymbol{\beta}_2 SNP + \mu_{Affective\ factor}$

Major depressive disorder:

$SCZ = \lambda_1 SCZ\ factor$ $MDD = \lambda_2 SCZ\ factor + \lambda_3 Affective\ factor$

$SCZ\ factor = \boldsymbol{\beta}_1 SNP + \mu_{SCZ\ factor}$

$Affective\ factor = \boldsymbol{\beta}_2 SNP + \mu_{Affective\ factor}$

We used a method suggested by Mallard and colleagues (Mallard et al., 2019) to calculate the effective sample size. First, we restricted the study to SNPs with MAF between 10% and 40% using the following script in R. This formula is prone to error for SNPs with low MAF. Therefore, it is suggested by Mallard and colleagues that we set a lower and upper MAF limit of approximately 10% and 40% when estimating the effective sample size:

df <- subset(df, df$MAF <= 0.4 & df$MAF >= 0.1)

Where df is the GWAS-by-subtraction file. Since we have performed a Cholesky model in the GWAS-by-subtraction models, we also needed to adjust the estimates '*est'* by multiplying them by the residual heritability for each GWAS-by-subtraction model shown in Table 7. We then calculated the effective sample size using the following script in R:

*effective_n <- (mean((df$Z_Estimate/df$est*λ)^2/(2*df$MAF*(1-df$MAF))))*

Where Z_Estimate is the Z statistic from the GWAS and est is the path estimates and λ is the residual for each GWAS.

Table 7: Model parameters for GWAS-by-subtraction analyses of bipolar disorder and major depressive disorder.

			Unstandadrized_Estimate	Unstandardized_SE	STD_Genotype	STD_Genotype_SE	STD_All
Affective-BIP	=~	BIP	0.2502653	0.007488661	0.7305109	0.021858999	0.7305109
Affective-BIP	=~	Affective-BIP	1		1		1
SCZ-BIP	=~	BIP	0.2339547	0.008028231	0.6829011	0.023433975	0.6829011
SCZ-BIP	=~	SCZ-BIP	1		1		1
Affective-MDD	=~	MDD	0.27151584	0.0060926	0.9455374	0.021217108	0.9455374
Affective-MDD	=~	Affective-MDD	1		1		1
SCZ-MDD	=~	MDD	0.09347285	0.006708394	0.3255135	0.023361575	0.3255135
SCZ-MDD	=~	SCZ-MDD	1		1		1

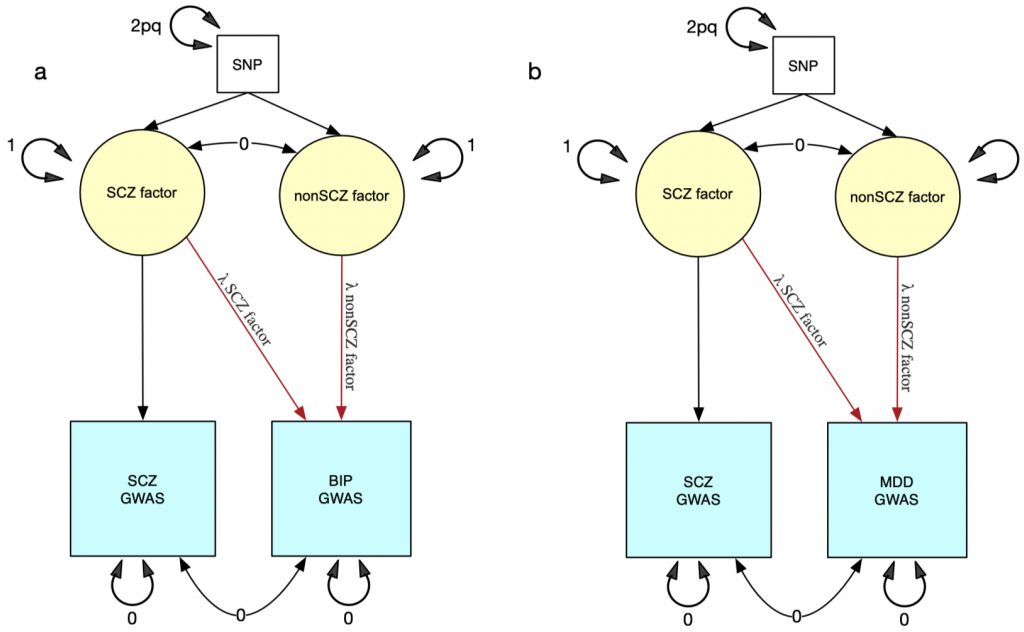


Figure 11: The Cholesky model fitted using genomicSEM. Circles represent latent variables and squares represent observed variables. a. path diagram for bipolar disorder genomicSEM model. b. Path diagram for major depressive disorder genomicSEM model.

Construction of Polygenic Risk Scores

Summary statistics for bipolar disorder (N=413,466), major depressive disorder (N=500,199), *SCZ* factor in bipolar disorder (N_{eff}=146,420), *Affective* factor in bipolar disorder (N_{eff}= 310,018), *SCZ* factor in major depressive disorder (N_{eff} = 147,014), and *Affective* factor in major depressive disorder (N_{eff} = 458,356) were first QC'd by excluding variants with MAF < 1% or imputation quality score of < 0.9 and removing strand ambiguous and indel polymorphisms. We then constructed PRS using the same method described in the previous

chapter by employing PRS-CS. Based on current recommendations (Ge et al., 2019) we used the phi value of 1e-2 for bipolar disorder and major depressive disorder due to their high polygenicity, whereas the "auto" function in PRS-CS was used to automatically learn the phi value for SNP weights for *SCZ* and *Affective* factors underlying bipolar disorder and major depressive disorder.

Polygenic transmission disequilibrium test (pTDT)

We used pTDT in a subset of the ISHDSF sample (41 families) with full parent-offspring data to test for over-transmission of polygenic risks for bipolar disorder, major depressive disorder, and their unique and shared (with schizophrenia) underlying genetic factors from parents to probands in the families. Additionally, to detect possible bias or systematic issues in the analyses, we also assessed the over-transmission of polygenic risks generated from schizophrenia and LDL as positive and negative controls, respectively with details of the PRS construction for schizophrenia and LDL described in more detail in the previous chapter. The pTDT deviation scores were generated for each multiplex family by subtracting the mean parental polygenic risks from the proband polygenic risks. The pTDT deviation scores were then standardized by dividing them by the cohort-specific mean parental polygenic risks standard deviations. To test whether the mean pTDT deviation was significantly greater than zero, representing an over-transmission of polygenic risks from parents to probands, a one-sided, one-sample *t*-test was employed.

Statistical Analyses

Statistical analyses were carried out using a mixed-effects logistic regression and as described in chapter two.

Polygenic overlap analysis

We used the MiXeR framework (Frei et al., 2019) to quantify the polygenicity and the polygenic overlap of schizophrenia with bipolar disorder and major depressive disorder. MiXeR uses GWAS summary statistics to estimate the polygenicity of each phenotype and constructs a bivariate Gaussian mixture model to estimate the number of shared and unique variants that explains 90% of SNP heritability for each GWAS. We also estimated the heritability of the derived factors underlying bipolar disorder and major depressive disorder using LDSC as shown in Table 8.

Table 8: Heritability estimates and LD score intercepts for GWAS-by-subtraction results. Note that heritability estimates are on the observed scale.

	SNP h2	SNP h2 SE	LDScore intercept	Ratio
Affective factor in BIP	0.0478	0.0026	0.9967	< 0
SCZ factor in BIP	0.3261	0.0105	0.9901	< 0
Affective factor in MDD	0.0548	0.0022	0.997	< 0
SCZ factor MDD	0.317	0.0105	0.985	< 0

3.4 Results

The diagnostic schema in the ISHDSF shown Table 1 and described in detail in the previous chapter follows a concentric pattern ranked by the degree to which they reflect narrow versus broad case definition within the psychosis spectrum. Below we also provide a visual representation of these case definitions in Figure 12.

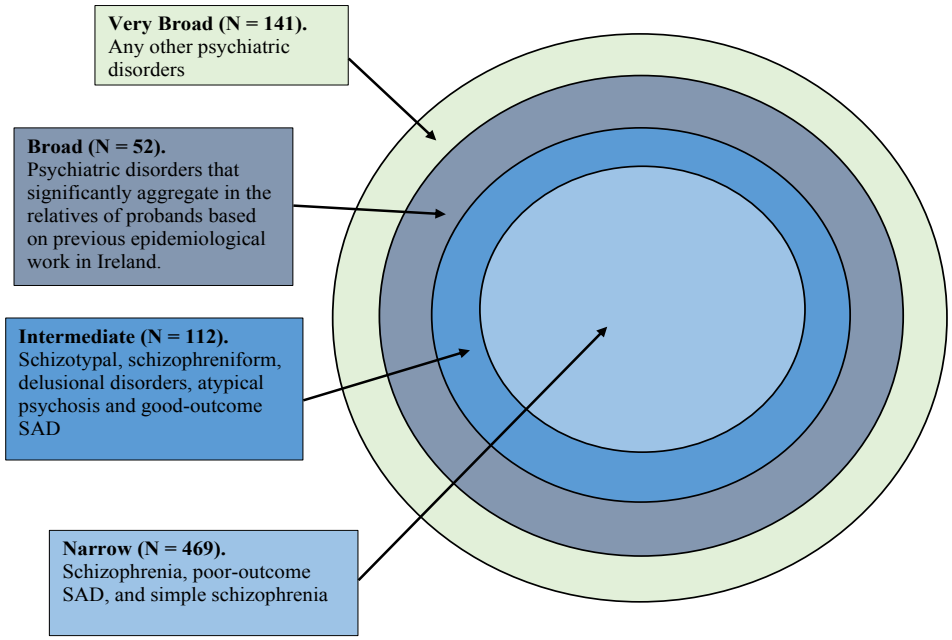


Figure 12: Concentric diagnostic schema of the ISHDSF sample representing the narrow versus broad case definitions of the psychosis spectrum.

Increased univariate bipolar disorder and major depressive disorder PRS in multiplex families.

Figure 13 shows the results for the univariate bipolar disorder and major depressive disorder PRS analysis in multiplex schizophrenia families versus population controls. All case definitions in multiplex schizophrenia families show a significantly increased bipolar disorder PRS compared to population controls (Figure 13a). The highest OR was observed in the *broad* case definition (OR = 2.21, 95% CI = 1.62-3.03) which includes 17 of the 21 bipolar disorder cases in the ISHDSF sample with psychotic features, excluding bipolar disorder cases without psychotic features which are represented in the *very broad* case definition. Except for the *unaffected* relatives, all diagnostic case definitions in multiplex schizophrenia families also show a significantly increased major depressive disorder PRS compared to population controls (Figure 13b). The highest OR was observed in the *very broad* case definition (OR = 1.45, 95% CI =

1.20-1.76), which includes 80 of 102 major depressive disorder cases in the ISHDSF sample, excluding major depressive disorder cases with psychotic features that are represented in the *broad* case definition. Full results for univariate bipolar disorder and major depressive disorder PRS are also provided in Tables 8 and 9.

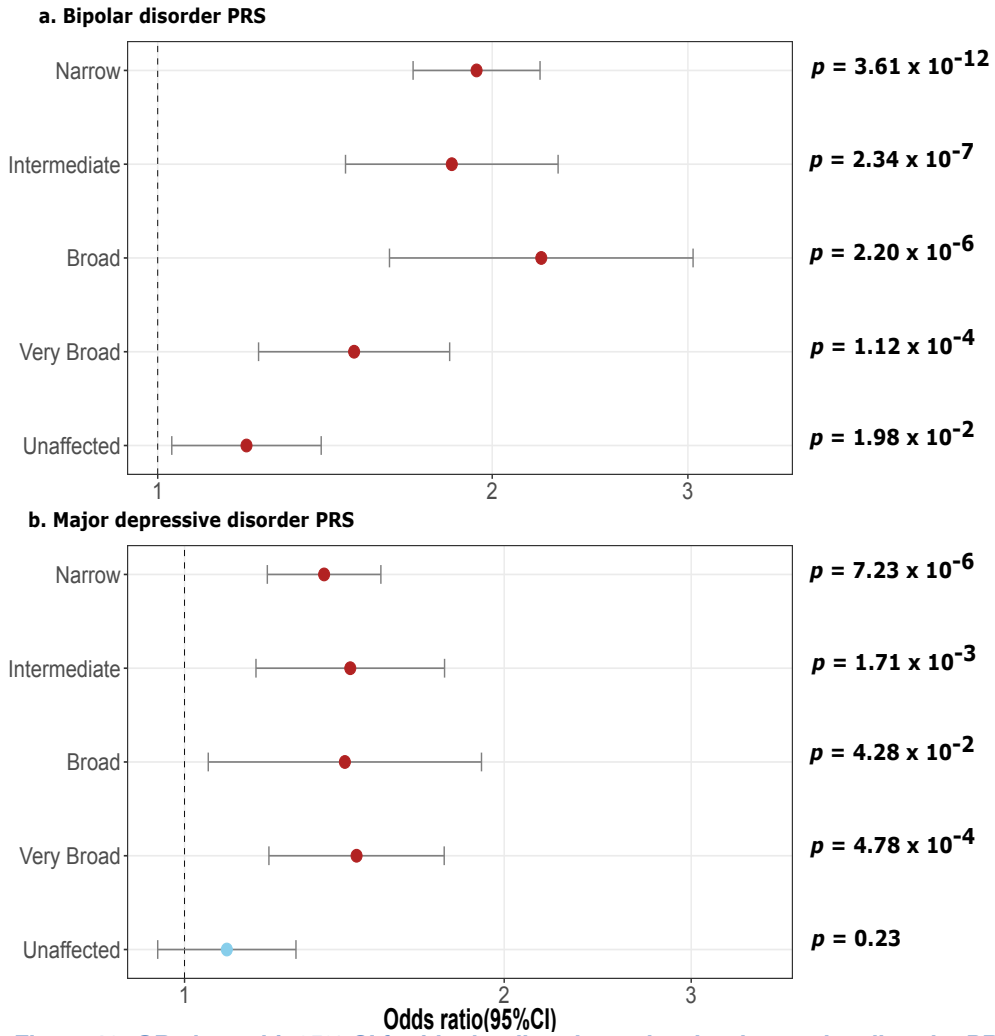


Figure 13: OR plots with 95% CI for bipolar disorder and major depressive disorder PRS comapred to population controls. OR and CI are shown for each comparison. *P*-values are corrected for multiple testing comparison. Red dot represents significant and blue dot represent non-significant results.

Table 9: Full bipolar disorder PRS results for ISHDSF diagnostic categories versus population controls. *p*-values were adjusted using the holm method in R.

Comparison Groups	**OR**	**CI (95%)**	***P*-value**	**Adjusted *P*-value**
Narrow vs Control	1.93	1.70-2.21	7.21E-23	3.61E-22
Intermediate vs Control	1.83	1.48-2.29	5.85E-8	2.34E-7
Broad vs Control	2.21	1.62-3.03	7.32E-7	2.2E-6
Very Broad vs Control	1.50	1.23-1.83	6.60E-5	2.11E-5
Unaffected vs Control	1.20	1.03-1.40	1.98-2	1.98E-2

Table 10: Full major depressive disorder PRS results for ISHDSF diagnostic categories versus population controls. *p*-values were adjusted using the holm method in R.

Comparison Groups	**OR**	**CI (95%)**	***P*-value**	**Adjusted *P*-value**
Narrow vs Control	1.35	1.19-1.53	1.45E-6	7.25E-6
Intermediate vs Control	1.43	1.16-1.76	5.72E-4	1.71E-3
Broad vs Control	1.41	1.05-1.90	2.14E-2	4.29E-2
Very Broad vs Control	1.45	1.20-1.76	1.20E-4	4.79E-4
Unaffected vs Control	1.09	0.94-1.27	0.23	0.38

No increased *Affective* factor PRS in multiplex families

Figure 14 shows the results for *SCZ* factor and *Affective* factor components derived from bipolar disorder and major depressive disorder polygenic risks in multiplex schizophrenia families versus population controls. The PRS results constructed for *SCZ* factor in bipolar disorder and *SCZ* factor in major depressive disorder (Figures 14a and 14c), representing the part of the polygenic risk that these two disorders share with schizophrenia due to genetic correlation, are significantly increased in all diagnostic case definitions of multiplex schizophrenia families compared to population controls. In contrast, the PRS constructed from the *Affective* factor in bipolar disorder, and the *Affective* factor in major depressive disorder (Figures 14b and 14d), representing the affective portion of the polygenic risk unique to these two disorders, show no significant increase in members of multiplex schizophrenia families compared to population controls. Full results are provided in the Table 11 below.

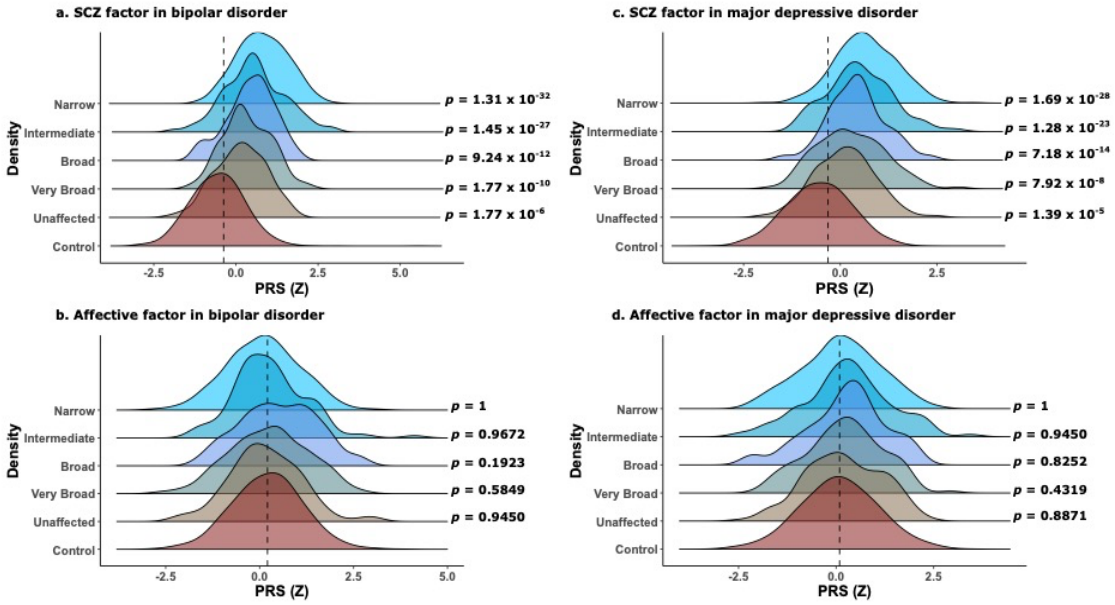


Figure 14: Density plots showing the distribution of *SCZ* factor and *Affective* factor PRS results. Each color represents one of the case definitions in the ISHDSF sample shown on the left side of the panel. The dashed line shows the mean PRS for population controls. *P*-values are corrected for multiple testing comparison.

Table 11: GWAS-by-subtraction PRS results for each underlying factor. *p*-values were adjusted using the holm method in R.

	Comparison Groups	SE	*P*-value	Adjusted *P*-value
SCZ Factor in BIP	Narrow versus Control	0.0879127	6.53E-39	1.31E-32
	Intermediate versus Control	0.0911982	7.63E-32	1.45E-27
	Broad versus Control	0.0991201	6.16E-15	9.24E-12
	VeryBroad versus Control	0.0812791	3.64E-11	5.10E-10
	Unaffected versus Control	0.070181	1.36E-10	1.77E-06
Affective Factor in BIP	Narrow versus Control	0.0712679	0.31210127	1
	Intermediate versus Control	0.0728196	0.16120981	9.67E-01
	Broad versus Control	0.0812717	0.08216128	0.19235886
	VeryBroad versus Control	0.08768171	0.13182091	0.5849128
	Unaffected versus Control	0.08168101	0.33091218	0.94502232
SCZ Factor in MDD	Narrow versus Control	0.0918278	9.40E-31	1.69E-28
	Intermediate versus Control	0.0728918	7.50E-25	1.28E-23
	Broad versus Control	0.0980198	4.49E-15	7.18E-14
	VeryBroad versus Control	0.0849812	7.20E-09	7.92E-08
	Unaffected versus Control	0.08812791	1.16E-09	1.39E-05
Affective Factor in MDD	Narrow versus Control	0.0821791	0.37129817	1
	Intermediate versus Control	0.0799821	0.11812779	0.94502232
	Broad versus Control	0.0979102	0.091810281	0.826292529
	VeryBroad versus Control	0.0817812	0.21012881	0.431928319
	Unaffected versus Control	0.0819717	0.35210909	0.88712087

To further assess the generalizability of these observations beyond multiplex schizophrenia families, we replicated the PRS analyses for *SCZ* and *Affective* factors underlying bipolar disorder and major depressive disorder in an independent cohort of ancestry matched sporadic schizophrenia cases from ISGC (N=2,225). As shown in Figure 15, the observed pattern of PRS enrichment in familial cases from multiplex schizophrenia families is similar to ancestry-matched sporadic cases from ISGC, demonstrating the generalizability of these observations in an independent cohort of schizophrenia cases.

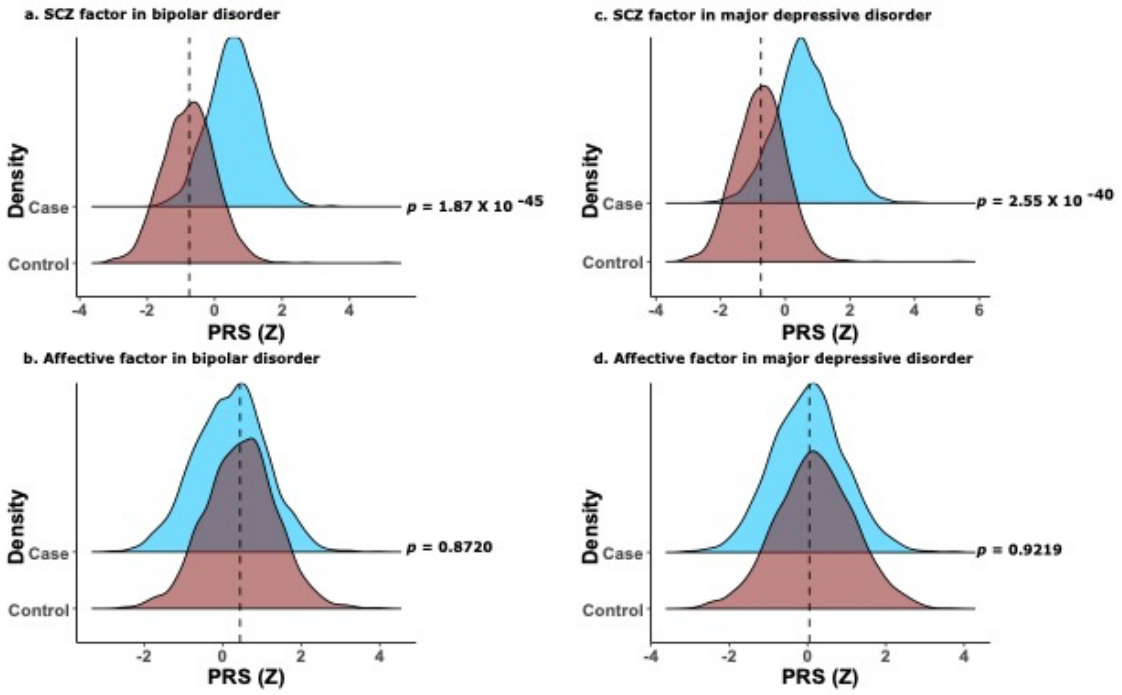


Figure 15: Density plots showing the distribution of SCZ factor and Affective factor PRS results. The colors represent sporadic cases and controls from the ISGC sample shown on the left side of the panel. Dashed line represents the mean PRS for population controls. *P*-values are corrected for multiple testing comparison.

Polygenic transmission equilibrium test (pTDT) in multiplex families

Figure 16 shows the pTDT results in multiplex schizophrenia families. In panel 15a we show that as expected, schizophrenia polygenic risks described in the previous chapter are significantly over-transmitted from parents to probands, while LDL polygenic risks show no significant over-transmission from parents to probands, suggesting the absence of systematic biases in the results. We next show that univariate bipolar disorder and major depressive disorder polygenic risks are significantly over-transmitted from parents to probands in multiplex schizophrenia families (Figure 16a). Polygenic risks for *SCZ* factors derived from bipolar disorder and major depressive disorder are also significantly over-transmitted from parents to probands in the families. In contrast, no significant over-transmission of polygenic risks derived from *Affective* factors unique to these two disorders and not shared with schizophrenia were observed in the families (Figure 16b).

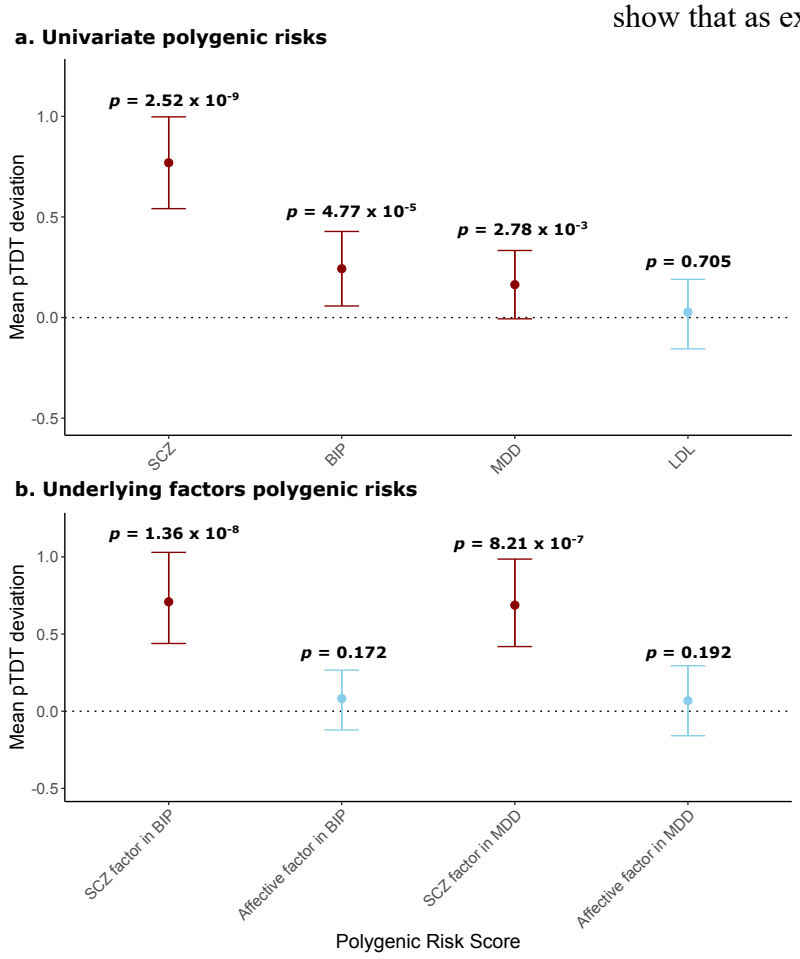


Figure 16: pTDT results presented as SD on the mid-parent distribution with 95% CI. to evaluate polygenic over-transmission in the families.

Polygenicity and polygenic overlap of underlying factors

We used MiXeR to estimate the polygenicity, polygenic overlap, and the number of shared and unique causal variants between bipolar disorder, major depressive disorder, and their underlying latent genetic factors (Figure 17). In agreement with previous findings, the polygenic signals of schizophrenia substantially overlap with bipolar disorder and major depressive disorder (17a and 17b respectively), while schizophrenia is estimated to be more polygenic than bipolar disorder, but less polygenic than major depressive disorder. Similarly, we observe that polygenic signals from *SCZ* factors and *Affective* factors derived from bipolar disorder and major depressive disorder also substantially overlap (17c and 17d respectively). The *SCZ* factor underlying bipolar disorder is estimated to be more polygenic than the *Affective* factor underlying bipolar disorder, whereas the *Affective* factor underlying major depressive disorder is estimated to be more polygenic than the *SCZ* factor underlying major depressive disorder.

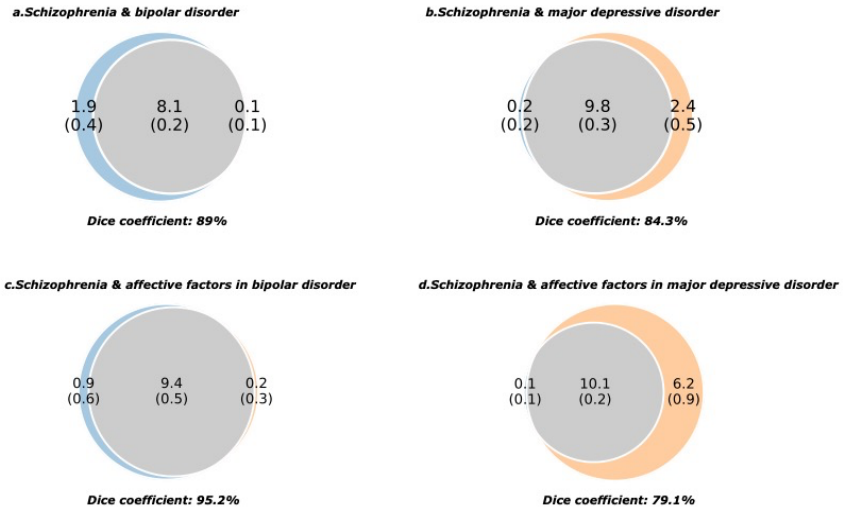


Figure 17: MiXeR results for latent factors. Venn diagrams show the estimated portion of causal variants shared and unique to each of the phenotypes. The grey part represents the overlap, and the size of the circles represent the polygenicity. The polygenic overlap is measured as the Dice coefficient on a 0-100% scale.

3.5 Discussion

Large-scale GWAS of schizophrenia, bipolar disorder, and major depressive disorder have shown that many common risk variants with small effect sizes contribute to disease risk in major psychiatric disorders. Additionally, cross-disorder analyses of psychiatric disorders have also provided consistent evidence that these three disorders share substantial genetic risk at common variation level (Lee et al., 2021). Based on these observations, we sought to investigate the source of increased common risk variation burden of bipolar disorder and major depressive disorder, two disorders with prominent affective features but also with strong genetic correlation with schizophrenia, in members of multiplex schizophrenia families.

Our results indicate that members of multiplex schizophrenia families, including the unaffected individuals, have an increased burden of common risk variation conferring risk to bipolar disorder compared to ancestry matched population controls. With the exception of unaffected relatives, we also observe that multiplex schizophrenia families have an increased burden of common risk variation conferring risk to major depressive disorder compared to ancestry matched population controls. We used GWAS-by-subtraction to disentangle bipolar disorder and major depressive disorder polygenic risks into two underlying genetic factors that they share with schizophrenia due to genetic correlation which we called *SCZ* factors, and genetic factors unique to them, which we called *Affective* factors, and our results suggest that increased polygenic risks for these two disorders in multiplex schizophrenia families is largely driven by part of the genetic liability that these two disorders share with schizophrenia due to genetic correlation. This observation is in agreement with previous epidemiological findings that show significant increase in the incidence of psychotic, but not affective disorders in relatives of patients with schizophrenia in multiplex families (Asarnow et al., 2001; Baron et al., 1983;

Kendler et al., 1985). While in addition to schizophrenia, non-schizophrenic psychotic disorders (Jablensky, 2001) also show significant familial aggregation in multiplex families (Kendler, McGuire, Gruenberg, Spellman, et al., 1993), affective or anxiety disorders are generally not considered to be on the same continuum as psychotic disorders (Kendler, McGuire, Gruenberg, O'Hare, et al., 1993b). Therefore, these results provide empirical genetic evidence in support of previous epidemiological findings in multiplex schizophrenia families by suggesting that although members of multiplex schizophrenia families have an increased polygenic risk burden for two disorders with prominent affective features, the source of this increased polygenic risk in a sample with high incidence of schizophrenia is largely due to strong genetic correlation between bipolar and/or major depressive disorder with schizophrenia. We also replicated these results in an independent sample of sporadic schizophrenia cases and showed that this observation is also generalizable to sporadic schizophrenia cases beyond multiplex families.

Using the MiXeR framework, we next quantified the polygenicity and polygenic overlap between the latent factors generated in this study. MiXeR has been used in recent years to quantify the polygenicity and the polygenic overlap between schizophrenia and other complex traits regardless of a genetic correlation (Ahangari, Everest, et al., 2022; W. Cheng et al., 2021; Smeland, Frei, Dale, et al., 2020). In agreement with previous findings (Frei et al., 2019), we first showed that schizophrenia shows substantial polygenic overlap with bipolar disorder and major depressive disorder at common variation level. Next, we showed that while *SCZ* and *Affective* factors underlying bipolar disorder and major depressive disorder have a substantial overlap, and *SCZ* factor appears to be more polygenic than *Affective* factor in bipolar disorder, while *SCZ* factor appears to be less polygenic than *Affective* factor in major depressive disorder. We note that this observation is expected, as previous findings show that while schizophrenia is

more polygenic than bipolar disorder, major depressive disorder is considered to be more polygenic than either bipolar disorder or schizophrenia (Holland et al., 2020). These results suggest that although both bipolar disorder and major depressive disorder have prominent affective features in their etiology, the polygenic signals conferring risk to affective features of major depressive disorder appear to be more polygenic than bipolar disorder.

We used an extension of standard TDT, called pTDT to investigate whether polygenic risks for bipolar disorder, major depressive disorder, and their shared and unique genetic factors are over-transmitted from parents to probands. Children are expected to inherit half of their parents' risk alleles and the transmission disequilibrium test (Spielman et al., 1993) queries whether a genetic variant (or aggregate polygenic scores in case of pTDT) is significantly over-transmitted from parents to probands. In addition to being robust against population stratification, pTDT is also robust to different sources of unmeasured biases such as socioeconomic status or environmental influences since matched, un-transmitted chromosomes in families are employed as controls. In agreement with expectation and prior family studies (Rees et al., 2020), polygenic liability to schizophrenia was over-transmitted from parents to probands in the ISHDSF sample, while we observed no over-transmission of polygenic liability to LDL which increases the risk to cardiovascular diseases with no direct correlation or causation to psychiatric disorders. Next, we showed that univariate bipolar disorder and major depressive disorder polygenic risks are also significantly over-transmitted from parents to probands in the ISHDSF sample, indicating that the proband's polygenic risks for these two disorders were on average higher than that of their parents. Disentangling the polygenic risks using genomicSEM further showed that only the portion of the polygenic risk that bipolar disorder or major depressive disorder share with schizophrenia due to genetic correlation is over-transmitted from

parents to probands, while no over-transmission of the affective portion of the risk unique to them was observed. Members of multiplex schizophrenia families show higher incidence of psychotic, but not affective disorders compared to the general population, making these observations consistent with schizophrenia transmission through some non-psychotic or unaffected family members in the ISHDSF sample (Ahangari et al., 2022; Käkelä et al., 2014; Yang et al., 2010).

Other family and pedigree studies of psychiatric disorders have also demonstrated the presence of an increased cross-disorder polygenic risk for psychiatric disorders. For example, Andlauer and colleagues (Andlauer et al., 2021) analyzed multiplex bipolar disorder families (N=395) consisting of 166 bipolar disorder and 78 major depressive disorder cases and showed that familial bipolar disorder cases and their unaffected relatives, had an increased PRS for bipolar disorder and schizophrenia compared to population controls. Another study that used a densely affected psychiatric pedigree (N=418) (Szatkiewicz et al., 2019) showed an increased schizophrenia PRS in affected members compared to unaffected members and population controls. De Jong and colleagues (de Jong et al., 2018) also used a dense pedigree (N=300) with bipolar disorder and major depressive disorder cases and showed nominally significant bipolar disorder and schizophrenia PRS in affected members compared to unaffected members and population controls. In contrast to the findings from families with multiple bipolar disorder, major depressive disorder, or schizophrenia cases, Halvorsen and colleagues (Halvorsen et al., 2021) analyzed a densely affected pedigree (N=122) with Tourette syndrome, a neurodevelopmental disorder characterized by recurrent nonrhythmic tics that shows significant genetic correlation with obsessive-compulsive disorder and attention deficit hyperactivity disorder (Arnold et al., 2018; Demontis et al., 2019). While a significantly increased PRS for

Tourette syndrome were observed in cases compared to controls, they did not show a significantly increased cross-disorder PRS for obsessive compulsive disorder or attention deficit hyperactivity disorder in cases compared to controls. This apparent lack of cross-disorder PRS loading in families with Tourette syndrome cases could reflect the lower power of obsessive-compulsive disorder and attention deficit hyperactivity disorder GWAS compared to bipolar disorder or major depressive disorder. Alternatively, there may be important differences in the genetic architecture of neurodevelopmental disorders such as Tourette syndrome that differentiates their polygenic profile from other major psychiatric disorders such as schizophrenia which warrants further investigation. Here, we show that similar to the studies noted above, members of multiplex schizophrenia families analyzed here (N=1,005) also have an increased cross-disorder polygenic risk for correlated psychiatric disorders. Furthermore, our PRS profiling in the full sample of multiplex schizophrenia families, in combination with pTDT results in a subset of families, provide empirical genetic evidence that the source of increased cross-disorder polygenic risk for bipolar disorder and major depressive disorder in multiplex schizophrenia families is largely attributable to the portion of the genetic risk that these two disorders share with schizophrenia. Despite distinct manifestations of psychotic and affective illnesses, these two separate axes of mental illness share significant genetic portions at common risk variation level and our results offer new insights into the nature of the elevated risk for affective disorders such as bipolar disorder and major depressive disorder in multiplex schizophrenia families.

We note that although GWAS-by-subtraction can be extended from bivariate to multivariate models, we opted to use two separate bivariate models in this study. This is due to our specific hypothesis about bipolar disorder and major depressive disorder as two separate

disorders with varying degrees of affective and psychotic features in their symptomatology. Additionally, caution is warranted when including more phenotypes in a Cholesky decomposition due to the sample size and power limitations of current GWAS results in psychiatric disorders. For example, if we extended models from two bivariate models to a single multivariate model and subtracted out both bipolar disorder and major depressive disorder signals from schizophrenia in a single model, we may be left with inadequate signals for the affective portions of the risks since a large portion of signals would be subtracted out due to strong genetic correlation among these three disorders. As the sample size and power of GWAS for psychiatric disorders increases, future work could extend the bivariate models to multivariate models in order to empirically test for this.

The analyses presented in this chapter should be interpreted in the context of some limitations. The predictive power of current PRS methods is mostly limited to individuals of European ancestry. As cross-ancestry PRS methods become more sophisticated and samples of multiplex families from more diverse populations become available, we could replicate these findings in ancestrally diverse multiplex families. Similarly, testing the predictive power of *SCZ* and *Affective* factors in independent schizophrenia, bipolar disorder, and major depressive disorder cohorts could provide further genetic evidence in support of the derived factors from genomicSEM models. While we did not have access to an independent bipolar disorder and major depressive disorder cohorts, we were able to show that *SCZ* factors underlying bipolar disorder and major depressive disorder can significantly distinguish between schizophrenia cases and controls from the ISGC cohort. Future work could also test the predictive power of *SCZ* and *Affective* factors in independent bipolar disorder and major depressive disorder cohorts. Some case definitions in the ISHDSF sample (e.g the *broad* case definition), have a lower number of

subjects, which may potentially bias some of the results due to lower power. We addressed this potential issue by repeating the PRS analyses by grouping the subjects in a concentric manner as described in the original ISHDSF publication (Kendler et al., 1996) The concentric comparison versus population controls shown in Table 12 suggests a similar patterns of PRS enrichment, indicating that lower numbers of subjects in some of the case definitions is unlikely to be a source of bias for the central findings of this study.

Table 12: Concentric comparison of GWAS-by-subtraction and univariate bipolar and major depressive disorder PRS in the ISHDSF sample versus population controls. *p*-values were adjusted using the holm method in R.

Narrow + Intermediate		
PRS	***P*-value**	**Adjusted *P*-value**
BIP	2.44E-24	2.68E-23
SCZ factor in BIP	1.51E-59	2.27E-38
Affective factor in BIP	0.749	1
MDD	6.43E-08	4.50E-07
SCZ factor in MDD	1.43E-55	1.86E-34
Affective factor in MDD	0.0931	0.5586
Narrow + Intermediate + Broad		
PRS	***P*-value**	**Adjusted *P*-value**
BIP	1.77E-25	2.12E-24
SCZ factor in BIP	1.44E-67	2.30E-46
Affective factor in BIP	0.475	1
MDD	3.01E-08	2.41E-07
SCZ factor in MDD	1.21E-58	1.69E-41
Affective factor in MDD	0.219	0.876
Narrow + Intermediate + Broad + Very Broad		
PRS	***P*-value**	**Adjusted *P*-value**
BIP	3.19E-24	3.19E-23
SCZ factor in BIP	4.83E-73	8.69E-52
Affective factor in BIP	0.572	1
MDD	4.60E-09	4.14E-08
SCZ factor in MDD	8.69E-69	1.48E-47
Affective factor in MDD	0.148	0.74

Given that environmental factors have not been assessed here, future analyses could integrate environmental influences unique to the families to further elucidate the role of

environmental factors on the elevated polygenic risk for bipolar disorder and major depressive disorder in multiplex schizophrenia families. Current PRS methods are limited to common risk variation in the genome and omit important genetic risk factors such as structural and copy number variation in the genome. Despite sparse evidence for the involvement of rare risk variation in the genetic architecture of major depressive disorder (Cheng et al., 2022) and bipolar disorder (Palmer et al., 2022) at current sample sizes, copy number variants in the genome often show strong pleiotropy among psychiatric disorders (Marshall et al., 2017b). Finally, due to the sample collection and genotyping strategies, not all the multiplex families have full parent-offspring genotypic or phenotypic information available. Therefore, the pTDT analyses results presented here should be interpreted with the caveat that only 41 families out of the full sample of 257 had full parent-offspring and genotype information available.

In conclusion, in this chapter we showed that in addition to increased burden of common risk variation conferring risk to schizophrenia (Ahangari, Gentry, et al., 2022), members of multiplex schizophrenia families studied here also have an increased polygenic vulnerability to bipolar disorder and major depressive disorder. However, we further show that this observation is likely to be largely attributable to part of the genetic risk that these two disorders share with schizophrenia due to their genetic correlation, rather than the affective portion of the genetic risk unique to these two disorders. Therefore, these results suggest that a complete interpretation of elevated cross-disorder PRS across correlated psychiatric disorders in multiplex families requires consideration of the relative contribution of the shared and unique genetic factors to account for the known genetic correlations across psychiatric disorders.

CHAPTER IV

The relationship between polygenic risk scores and symptom dimensions of schizophrenia and schizotypy in multiplex schizophrenia families

4.1 Abstract

Psychotic disorders and schizotypal traits aggregate in the relatives of probands with schizophrenia. It is currently unclear how variability in symptom dimensions in schizophrenia probands and their relatives is associated with polygenic liability to psychiatric disorders. In this study, we investigated whether PRS can predict symptom dimensions in members of multiplex schizophrenia families. The largest GWAS datasets for schizophrenia, bipolar disorder, and major depressive disorder were used to construct PRS in 861 subjects from the ISHDSF sample. Symptom dimensions were derived using the OPCRIT in subjects with a history of a psychotic episode, and the SIS in subjects without a history of a psychotic episode. Mixed-effects linear regression models were used to assess the relationship between PRS and symptom dimensions across the psychosis spectrum. Our results indicate that schizophrenia PRS is significantly associated with negative/disorganized symptom dimension in psychotic ($p = 2.31 \times 10^{-4}$) and negative dimension in non-psychotic ($p = 1.42 \times 10^{-3}$) subjects. Bipolar disorder PRS is significantly associated with manic symptom dimension in psychotic subjects ($p = 3.70 \times 10^{-4}$). No association with major depressive disorder PRS was observed. These findings suggest that polygenic liability to schizophrenia is associated with higher negative/disorganized symptoms in psychotic and negative symptoms in non-psychotic subjects in multiplex schizophrenia families. These results provide genetic evidence in support of the spectrum model of schizophrenia and support the view that negative and disorganized symptoms may have greater genetic basis than positive symptoms, making them better indices of familial liability to schizophrenia.

4.2 Introduction

Schizophrenia is a clinically heterogeneous psychiatric disorder with a population prevalence of ~ 1% (Saha et al., 2007). In the past decade, GWAS, copy number variation, and rare variant studies have significantly improved our understanding of the genetic basis of schizophrenia (Marshall et al., 2017; Singh et al., 2022; Trubetskoy et al., 2022). Due to the heterogenous manifestation of schizophrenia symptoms, studies have attempted to capture this clinical heterogeneity in terms of symptom dimensions derived from factor analyses. Although these derived dimensions vary across studies, they often result in positive, negative/disorganized, and affective dimensions (Wickham et al., 2001).

In recent years, the relationship between aggregate common risk variation indexed by PRS and clinical dimensions of schizophrenia has garnered much attention. Early studies using the first wave of PGC-SCZ GWAS found no association between schizophrenia PRS and symptom dimensions, likely due to the smaller sample size and lower power of the first wave of PGC-SCZ GWAS (Derks et al., 2012). Recent analyses using the second wave of PGC-SCZ GWAS have found significant associations between schizophrenia PRS and negative and disorganized dimensions, suggesting that polygenic liability to schizophrenia can explain part of the variance in negative and disorganized symptoms (Bigdeli et al., 2017; Ruderfer et al., 2018). Most recently, a study in a sample of schizophrenia and schizoaffective cases (Smigielski et al., 2021) showed that PGC3-SCZ PRS is also significantly associated with dimensions from the Positive and Negative Syndrome Scale.

In addition to the heterogeneous manifestation of schizophrenia, some relatives of schizophrenia probands, though never psychotic, exhibit clinical features that closely resemble those observed in their ill relatives (Kendler, 1995). In the Danish Adoption Study of

schizophrenia, these symptoms and signs differentiated the relatives of schizophrenia probands from controls and were later combined into the classification of schizotypal personality disorder in the DSM-III (Kety et al., 1975). Since then, considerable evidence from family, adoption, and twin studies suggests that schizotypal traits aggregate in relatives of schizophrenia probands (Fanous et al., 2001). While earlier studies have linked specific genes to schizotypal traits (Meller et al., 2019), the relationship between schizophrenia PRS and schizotypal traits has not been fully established (Nenadić et al., 2020). Recently, subclinical phenotypes such as psychotic-like experiences have been proposed to be used as proxies to capture subclinical liability to psychosis. For example, analysis of the UK Biobank cohort shows that psychotic-like experiences have pleiotropic association with polygenic liability to schizophrenia and other psychiatric and neurodevelopmental disorders (Legge et al., 2019). However, these findings indicate that unlike schizotypal traits which significantly aggregate in the relatives of schizophrenia probands, psychotic-like experiences are not specific to schizophrenia. This is further strengthened by studies showing that rates of psychotic-like experiences do not differ significantly between relatives and non-relatives of schizophrenia patients in clinically ascertained samples (Landin-Romero et al., 2016).

Subjects ascertained from multiplex schizophrenia families represent the upper bounds of schizophrenia risk in the population, and a major question is the extent to which symptom severity in schizophrenia can be attributed to genetic differences among subjects. We have previously shown that members of the ISHDSF (Kendler et al., 1996) have an increased PRS for schizophrenia, bipolar disorder, and major depressive disorder compared to population controls (Ahangari et al., 2022). In this study, we sought to examine the differential relationship between PRS for these three major psychiatric disorders, and quantitative measurement of symptom

severity in the ISHDSF sample. We hypothesized that by using a well-ascertained sample of multiplex schizophrenia families, we will be able to identify associations between schizophrenia PRS and core schizophrenia symptom dimensions in psychotic subjects, while also maximizing power to uncover specific associations between schizophrenia PRS and schizotypal dimensions in non-psychotic subjects in the families. To the best of our knowledge, this study is the first that aims to establish a relationship between schizophrenia PRS and symptom dimensions in schizophrenia cases and their relatives across the extended psychosis spectrum.

4.3 Methods

Irish Study of High-Density Multiplex Schizophrenia Families (ISHDSF)

Details about the sample collection and ascertainment for the ISHDSF sample is provided in chapter two.

Symptom dimensions in subjects with a history of a psychotic episode

For subjects with a lifetime occurrence of a psychotic episode (N=539), The OPCRIT (McGuffin et al., 1991) was completed based on the review of detailed hospital records and interviews to assess the symptom dimensions. A full description of the factor analysis of OPCRIT items in the ISHDSF sample is provided elsewhere (Fanous et al., 2005). Briefly, 55 of the 75 items of the OPCRIT were entered into the factor analysis. These items were selected because they represent signs and symptoms rather than the course of illness. Five factors were derived, and factor-derived scores were generated. These five factors were identified as 1) negative/disorganized, 2) hallucinations, 3) delusions, 4) manic symptoms, and 5) depressive symptoms. The full list of items and their loadings are provided in Table 13 below.

Table 13: 55 items used in the factor analysis of OPCRIT data. Factor loadings of OPCRIT items onto the 5 dimensions. Items highlighted with an asterisk are the disorganized items.

Factor	Item	Loading
Negative/ Disorganized Dimension	1. Bizarre behavior*	0.538
	2. Catatonia*	0.544
	3. Speech difficult to understand*	0.623
	4. Incoherent	0.842
	5. Positive formal thought disorder*	0.621
	6. Negative formal thought disorder	0.777
	7. Restricted affect	0.679
	8. Blunted affect	0.817
	9. Inappropriate affect*	0.6
	10. Rapport difficult	0.815
	11. Information not credible	0.813
	12. Deterioration from premorbid level of function	0.783
Delusions Dimension	13. Persecutory delusions	0.725
	14. Well-organized delusions	0.707
	15. Delusions of influence	0.709
	16. Bizarre delusions	0.86
	17. Widespread Delusions	0.806
	18. Delusions of passivity	0.944
	19. Thought insertion	0.898
	20. Thought withdrawal	0.707
	21. Thought broadcast	0.723
Hallucinations Dimension	22. Delusions and hallucinations > 1week	0.971
	23. Persecutory/jealous delusions with hallucinations	0.968
	24. Third person auditory hallucinations	0.734
	25. Running commentary voices	0.659
	26. Abusive/accusatory/persecutory voices	0.796
	27. Other nonaffective auditory hallucinations	0.568
	28. Nonaffective hallucination in any modality	0.986
Manic Dimension	29. Affective symptoms predominate	0.866
	30. Grandiose delusions	0.482
	31. Elevated mood	0.962
	32. Irritable mood	0.711
	33. Excessive activity	0.935
	34. Reckless activity	0.842
	35. Pressured speech	0.953
	36. Increased self-esteem	0.879
	37. Thoughts racing	0.94
	38. Distractibility	0.892
	39. Reduced need for sleep	0.886
Depressive Dimension	40. Schizophrenia symptoms at the same time as affective symptoms	0.934
	41. Dysphoria	0.968
	42. Agitated activity	0.615
	43. Slowed activity	0.759
	44. Loss of energy/tiredness	0.886
	45. Loss of pleasure	0.897
	46. Poor concentration	0.826
	47. Excessive self-reproach	0.865
	48. Suicidal ideation	0.724
	49. Initial insomnia	0.863
	50. Early morning waking	0.869
	51. Excessive sleep	0.523
	52. Poor appetite	0.904
	53. Weight loss	0.86
	54. Increased appetite	0.487
	55. Weight gain	0.489

Symptom dimensions in subjects without a history of a psychotic episode

For subjects without a lifetime occurrence of a psychotic episode (N=322), the SIS (Kendler et al., 1989) was used to assess schizotypal signs and symptoms across the psychosis spectrum. The items used included the DSM-III-R major signs and symptoms of schizotypal personality disorder. The SIS was originally developed for family studies of schizophrenia in Ireland (Kendler et al., 1993a). It includes signs and symptoms that are specific to schizotypy, with a contextual assessment of the pathological nature of symptoms that can significantly discriminate the relatives of schizophrenia probands from that of controls. Based on our hypothesis that schizotypy captures a continuous measure of liability to schizophrenia in relatives of schizophrenia probands, SIS was also administered to unaffected relatives to capture the symptom dimensions on the extended psychosis spectrum.

Factor analysis of SIS in non-psychotic subjects

Exploratory (EFA) and confirmatory (CFA) factor analyses were conducted to determine and verify the least number of factors explaining the maximum amount of variance in the SIS data, using the R packages Psych (Revelle, 2021) and OpenMx (Neale et al., 2016). The maximum likelihood polychoric correlations were estimated using the R package polycor (Fox, 2022) to obtain the eigenvalues and eigenvectors, using a minimum eigenvalue of one as the cut off. EFA with two and three factors were conducted using an oblique and orthogonal rotation. The minimum cutoff for each indicator factor loadings was set at ≥ 0.3, considering only the highest loading if one indicator loaded into more than one factor. Two independent fits of CFA with two and three factors were implemented to corroborate the factor structure, and factor scores were generated using the ML method.

Genotyping and Imputation

Details about genotyping and imputation are provided in chapter two.

Polygenic risk score construction

Details about PRS construction are provided in chapters two and three.

Statistical Analyses

As described in chapter two, to account for the family structure in the sample, GRM was constructed using LDAK (Speed et al., 2012) and included in the mixed models as a random effect. Sex, genotyping platform, genotyping sites, age at interview, and the top 10 principal components were also included as additional covariates. More information on the principal component analysis, covariates and handling of possible batch and site effects are also provided in chapter two.

Association analyses were carried out using a two-step approach. First, mixed-effects linear regression analyses were performed using lmekin in R (Therneau, 2015). Given that the floor effect in some of the symptom dimensions could violate the assumptions of a linear regression (Cook & Manning, 2013), we also conducted mixed-effects quantile-regression analyses on dimensions that showed significant association with PRS using the qrLMM package in R (Galarza et al., 2017). Quantile regression is an extension of linear regression that estimates the effects at different locations in the distribution without the need to have normality assumptions met. In situations where the assumption of normality of the data may be violated (e.g., with an abundance of zero responses in the symptom measurements being analyzed here), the quantile regression method provides a more accurate estimation of centrality of the data at different locations in the distribution of symptom scores. While the mean is used as the measure of centrality in linear regression, the means of a selected number of pre-defined quantiles are used as the measure of centrality in quantile regression. We used three tau (τ) values

(0.25,0.5,0.75) as the quantiles, representing the lowest 25% (Q1), the lowest 50% (Q2), and the lowest 75% (Q3) of symptom severity scores, respectively. The nominal significance for all analyses was set at $p < 0.05$ and the p-values were adjusted for multiple testing using the Holm method.

4.4 Results

Sample structure

Figure 18 below provides a visual representation of the diagnostic schema of the ISHDSF sample with their symptom information. Subjects in the *Narrow* case definition represent schizophrenia and poor-outcome schizoaffective disorder cases. Subjects in the *Intermediate* case definition represent cases with diagnoses of other psychotic disorders in the families. Symptom severity in these two case definitions were measured using OPCRIT. The *Broad* case definition includes subjects with a diagnosis of a psychiatric disorder that significantly aggregate in the relatives of schizophrenia probands. Symptom severity for these subjects was measured using OPCRIT or SIS, depending on whether an individual had a history of a psychotic episode.

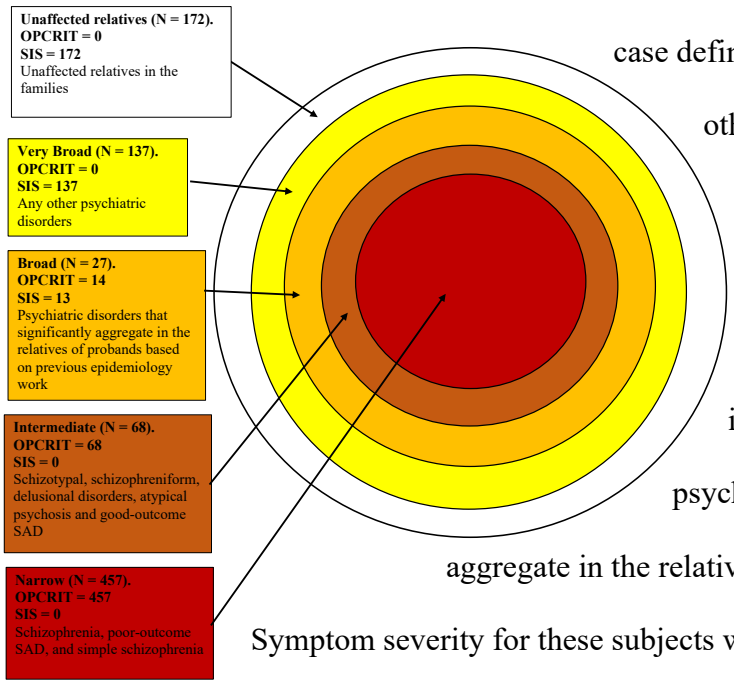


Figure 18: Concentric diagnostic schema of the ISHDSF sample with symptom information. SIS = Structured Interview for Schizotypy. OPCRIT = Operational Criteria Checklist for Psychotic and Affective Illness.

The *Very Broad* case definition includes any other psychiatric disorder present in the families. Symptom severity for these subjects, and *unaffected relatives* was measured using SIS. The full list of the psychiatric diagnoses in each case definition is provided in Tables 13-14.

Table 14: List of the diagnoses in narrow, intermediate, and broad case definitions with OPCRIT information available.

Subjects with Operational Criteria Checklist for Psychotic Disorders (OPCRIT) Assessment	
Narrow (N=457)	Schizophrenia N=261
	Schizophrenia Paranoid-type N = 60
	Schizophrenia Disorganized-type N = 62
	Schizophrenia Catatonic-type N = 11
	Schizophrenia Simple-type N = 7
	Poor-outcome schizoaffective disorder N = 56
Intermediate (N =68)	Delusional disorder N = 4
	Psychotic disorder NOS N = 29
	Good-outcome schizoaffective disorder N = 28
	Schizophreniform N = 1
	Schizotypal personality disorder N = 6
Broad (N=14)	Bipolar disorder = 8
	Major depression with psychotic features N = 5
	Paranoid personality disorder = 1

Table 15: List of the diagnoses in narrow, intermediate, and broad case definitions with OPCRIT information available.

Subjects with Structured Interview for Schizotypy (SIS) Assessment	
Broad (N=13)	Delusional disorder (N=2)
	Avoidant personality disorder (N=9)
	Schizoid personality disorder (N=2)
Very Broad (N=137)	Bipolar disorder (N =6)
	Generalized anxiety disorder (N=22)
	Major depressive disorder (N=77)
	alcohol abuse + alcohol dependence (N=25)
	Panic disorder (N=6)

Factor analysis of schizotypy symptoms in non-psychotic subjects

Factor analysis of SIS is shown in Figure 19. Two eigenvalues were above the minimum cut off value of one, suggesting the factor solution of retaining two factors (Figure 19A). The

EFA two-factor model fit under an oblique rotation explained the same cumulative variance

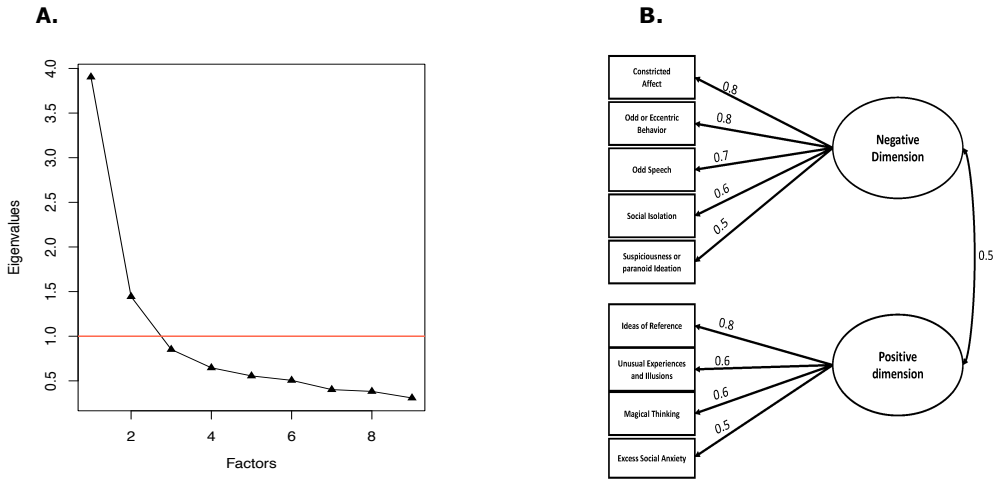


Figure 19: Factor analysis of SIS in the ISHDSF sample. A. Scree plot. B. Path diagram of the factor structure of SIS. Red line in panel A represents the eigenvalue of 1 used as the cut-off.

(0.49) as the two-factor orthogonal rotation, with similar ranges for their factor loadings (0.5-0.8; Figure 19B). An oblique rotation was selected to allow for correlation (0.51) between the two factors. The CFA models supported the two-factor solution (3-factor: $-2lnL$=4842.22, AIC=4920.22; 2-factor: $-2lnL$=4809.67, AIC=4883.67, $\Delta\chi^2$ (2) =-32.55, p=1). The two schizotypy dimensions were discernible as positive and negative dimensions of schizotypy (Figure 18B). The model fits is provided in Table 16.

Table 16: Model fit comparisons for the CFA on SIS data. CFA=Confirmatory factor analysis. Ep= estimated parameters. LL=Log-likelihood. AIC=Akaike's information criterion.

Base	Comparison	Ep	Minus2LL	Df	CFI	RMSEA	AIC	diffLL	Diffdf	P
CFA3 factors		39	4842.219	3066	0.823	0.096	4920.219			
CFA 3 factors	CFA 2 factors	37	4809.670	3068	0.903	0.068	4883.670	-32.548	2	1

Association of polygenic risks with symptom dimensions in psychotic subjects

Figure 20A shows the results for the associations between PRS and OPCRIT symptom dimensions in psychotic subjects. The schizophrenia PRS was a significant predictor of the negative/disorganized symptom dimension ($\beta = 0.198$; 95% CI 0.099 - 0.305; $P = 2.31 \times 10^{-4}$). Bipolar disorder PRS was a significant predictor of the manic symptom dimension ($\beta = 0.181$; 95% CI 0.061 - 0.241; $P = 3.70 \times 10^{-4}$). Both schizophrenia and major depressive disorder PRS

also showed suggestive associations with delusional and depressive symptoms respectively, but these did not survive multiple testing corrections. To narrow down the association between schizophrenia PRS and the negative/disorganized dimension, two additional mixed-effects linear regression analyses were carried out by separating the symptoms into two groups representing negative and disorganized symptoms separately. A significant association was observed for both negative only ($\beta = 0.188$; 95% CI 0.081 - 0.291; $p = 9.10 \times 10^{-3}$) and disorganized only ($\beta = 0.199$; 95% CI 0.091 - 0.312; $p = 1.41 \times 10^{-5}$) symptoms. More information is provided in Table 17.

Association of polygenic risks with symptom dimensions in non-psychotic subjects

Figure 20B shows the results for the association of polygenic risks with SIS dimensions in non-psychotic subjects. The schizophrenia PRS was found to be a significant predictor of negative symptom dimension ($\beta = 0.186$; 95% CI 0.080 - 0.0.289; $P = 1.42 \times 10^{-3}$), while no significant association was observed with positive symptom dimension. Additionally, bipolar disorder and major depressive disorder PRS showed no association with SIS dimensions. Full results are provided in Table 17.

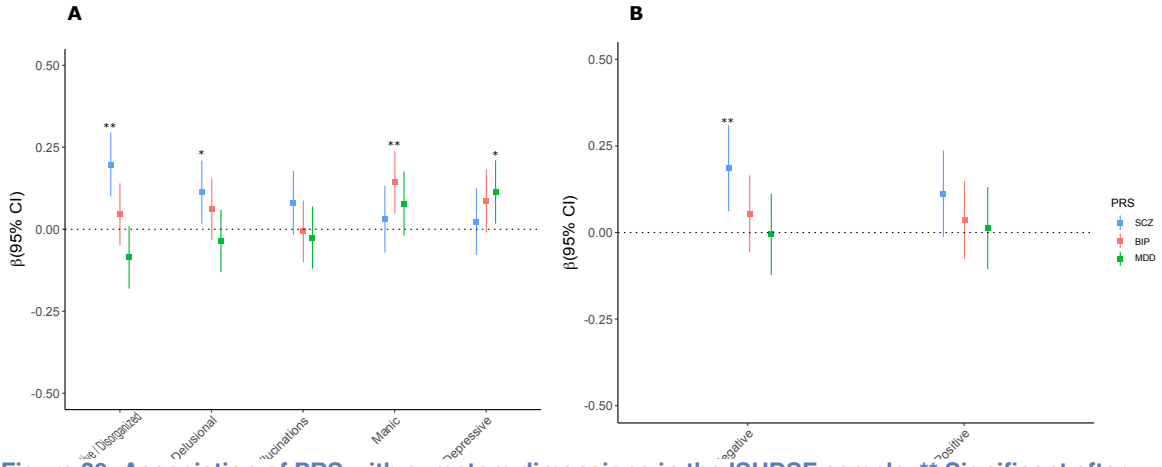


Figure 20: Association of PRS with symptom dimensions in the ISHDSF sample. ** Significant after multiple testing correction. * Nominally significant. Error bars represent 95% CI of the β value.

Quantile regression analysis of the significant associations

Figure 21 shows the follow-up mixed-effects quantile regression analyses for the three dimensions that showed significant association with the polygenic risks. The schizophrenia PRS is significantly associated with the negative/disorganized symptom dimension in psychotic subjects at first ($t = 2.47$, $p = 8.16 \times 10^{-3}$), second ($t = 2.33$ $p = 1.14 \times 10^{-2}$) and third ($t = 2.55$ p $= 6.76 \times 10^{-4}$) quantiles of symptom severity. In contrast, schizophrenia PRS is significantly associated with negative symptom dimension in non-psychotic subjects only at the third quantile ($t = 3.29$ $p = 6.3 \times 10^{-4}$), while a suggestive association was also observed at the second quantile. Similarly, bipolar disorder PRS is also significantly associated with manic symptoms in psychotic subjects only at the third quantile ($t = 3.14$ $p = 5.0 \times 10^{-3}$). Full quantile regression results are reported in Table 17.

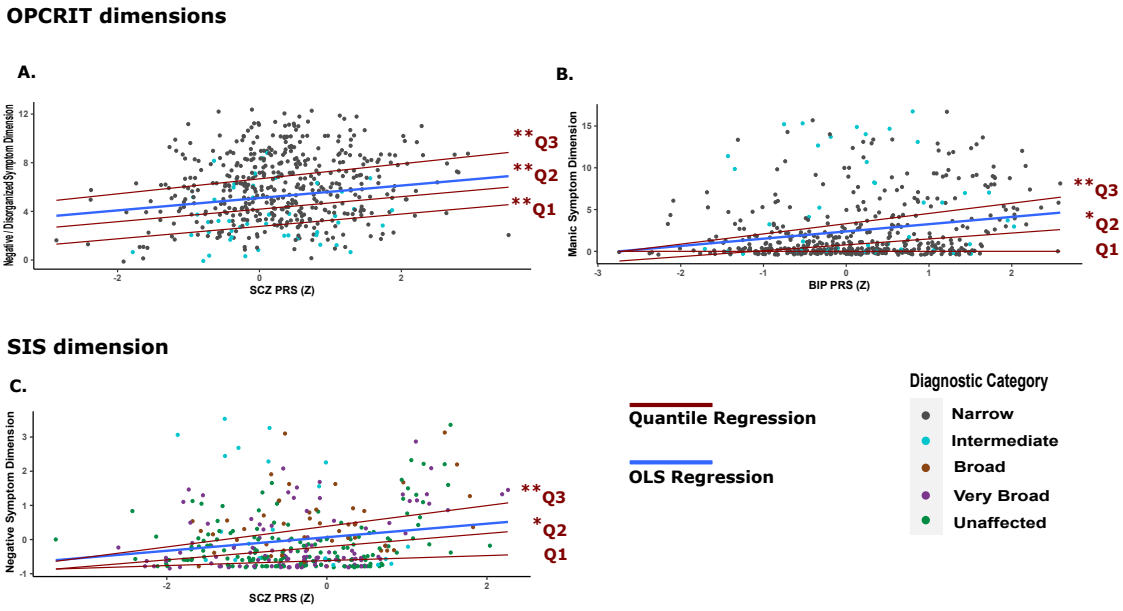


Figure 21: Quantile regression analysis of symptom dimensions that show significant association with polygenic risks. Three quantiles were tested (Q1=0.25, Q2=0.5, Q3=0.75), corresponding to the first, second and third quantile of symptom distribution. Quantiles are shown in red; OLS is shown in blue.

Table 17: Mixed-effects linear regression results for association between PRSs and symptom dimensions. Adjusted *p*-values were generated using the Holm method in R.

PRS	Symptom Dimensions	β	SE	*p*-value	Adjusted *p*-value
SCZ PRS	**Operational Criteria Checklist for Psychotic Illness (OPCRIT) Symptom dimensions**				
	Negative symptoms	0.19753552	0.04987663	0.000011	**2.31E-04**
	Delusional symptoms	0.1320749	0.04921818	**0.0023**	0.0937
	Hallucinations	0.08097746	0.04971806	0.078	0.2932
	Manic symptoms	0.03094654	0.05191875	0.53	1
	Depressive symptoms	0.02324251	0.05186413	0.96	1
	Structured Interview for Schizotypy (SIS) symptom dimensions				
	Negative symptoms	0.1857	0.0634	0.000084	**1.42E-03**
	Positive symptoms	0.1123	0.0635	0.184	1
BIP PRS	**Operational Criteria Checklist for Psychotic Illness (OPCRIT) Symptom dimensions**				
	Negative symptoms	0.04565756	0.04832537	0.13	1
	Delusional symptoms	0.0608749	0.04793015	0.17	1
	Hallucinations	-0.0066352	0.04791145	0.76	1
	Manic symptoms	0.18113907	0.0480533	0.000015	**3.70E-04**
	Depressive symptoms	0.08657643	0.04915946	0.15	1
	Structured Interview for Schizotypy (SIS) symptom dimensions				
	Negative symptoms	0.0544	0.0564	0.023	1
	Positive symptoms	0.0362	0.0574	0.064	1
MDD PRS	**Operational Criteria Checklist for Psychotic Illness (OPCRIT) Symptom dimensions**				
	Negative symptoms	-0.095136	0.04868627	0.098	1
	Delusional symptoms	-0.0357596	0.04852334	0.72	1
	Hallucinations	-0.0253366	0.04843669	0.73	1
	Manic symptoms	0.07828989	0.04985869	0.21	1
	Depressive symptoms	0.1664895	0.04929699	**0.0041**	0.1138
	Structured Interview for Schizotypy (SIS) symptom dimensions				
	Negative symptoms	0.04013	0.0601	0.091	1
	Positive symptoms	0.0352	0.0605	0.36	1

Table 18: Mixed-effects quantile regression results for association between PRSs and symptom dimensions. Adjusted *p*-values were generated using the Holm method in R.

PRS	Symptom Dimensions	Tau	Coefficient	SE	t-value	*p*-value	Adjusted *p*-value
SCZ PRS	Operational Criteria Checklist for Psychotic Illness (OPCRIT) Symptom dimensions						
	Negative	0.25	0.50898	0.20536	2.47844	**1.36E-04**	**8.16E-03**
		0.5	0.49937	0.21341	2.33995	**1.97E-03**	**1.14E-02**
		0.75	0.56876	0.2224	2.55743	**1.09E-05**	**6.76E-04**
	Delusional	0.25	0	0.2783	0	1.00E+00	1
		0.5	0.48873	0.26779	1.82502	6.87E-02	1
		0.75	0.6204	0.22939	2.70463	**7.09E-03**	0.050413
	Hallucinations	0.25	0.30245	0.15111	2.00146	0.4594	1
		0.5	0.37412	0.13256	2.82229	0.498	1
		0.75	0.30651	0.13095	0.34058	0.1968	1
	Manic	0.25	0	0.15606	0	1	1
		0.5	0	0.19144	0	1	1
		0.75	-0.261	0.35465	-0.73593	0.46215	1
	Depressive	0.25	0	0.35992	0	1	1
		0.5	0.27213	0.43388	0.6272	0.53084	1
		0.75	0.38704	0.7228	0.53547	0.59258	1
	Structured Interview for Schizotypy (SIS) symptom dimensions						
	Negative	0.25	0.07318	0.06327	1.15668	0.24822	1
		0.5	0.14287	0.07212	1.98096	**4.84E-02**	1
		0.75	0.31668	0.0715	3.29	**1.00E-05**	**6.30E-04**
	Positive	0.25	0.05863	0.05922	0.99008	0.32284	1
		0.5	0.1523	0.06831	2.37602	**1.81E-02**	0.95718
		0.75	0.34591	0.07194	4.8048	**8.24E-03**	0.4532
BIP PRS	Operational Criteria Checklist for Psychotic Illness (OPCRIT) Symptom dimensions						
	Negative	0.25	0	0.19147	0	1	**1**
		0.5	0	0.22117	0	1	**1**
		0.75	0	0.22589	0	1	**1**
	Delusional	0.25	0	0.06725	0	1	**1**
		0.5	0.15145	0.06985	2.16814	0.3066	**1**
		0.75	0	0.0763	0	1	**1**
	Hallucinations	0.25	0	0.16538	0	1	1
		0.5	0	0.13158	0	1	1
		0.75	-0.19611	0.12863	-1.52465	0.12804	1
	Manic	0.25	0	0.1678	0	1	1
		0.5	0.47935	0.17955	2.6697	**7.86E-03**	0.44016
		0.75	1.37394	0.22371	3.14172	**8.20E-05**	**5.00E-03**
	Depressive	0.25	0	0.32312	0	1	1
		0.5	0.5109	0.3997	1.27823	0.20182	1
		0.75	0.55366	0.66223	1.11996	0.26332	1
	Structured Interview for Schizotypy (SIS) symptom dimensions						
	Negative	0.25	0.0252	0.04791	0.52605	0.5992	1
		0.5	0.1686	0.05802	0.90572	0.391	1
		0.75	0.35392	0.06083	0.81778	0.17827	1
	Positive	0.25	0.05863	0.05922	0.99008	0.32284	**1**
		0.5	0.1623	0.06831	0.37602	0.10806	**1**
		0.75	0.34591	0.07194	0.80848	0.12871	**1**
MDD PRS	Operational Criteria Checklist for Psychotic Illness (OPCRIT) Symptom dimensions						
	Negative	0.25	0	0.19482	0	1	**1**
		0.5	-0.39424	0.1994	-1.97712	0.4863	**1**
		0.75	0	0.21209	0	1	**1**
	Delusional	0.25	-0.08083	0.07735	-1.04502	0.29657	**1**
		0.5	0	0.08239	0	1	**1**
		0.75	0	0.07445	0	1	**1**
	Hallucinations	0.25	0	0.17495	0	1	1
		0.5	0	0.139	0	1	1
		0.75	-0.26326	0.12139	-2.16866	0.3062	1
	Manic	0.25	0	0.14832	0	1	1
		0.5	0	0.17884	0	1	1
		0.75	0.63826	0.25476	2.5039	**1.26E-02**	0.67932
	Depressive	0.25	0	0.36156	0	1	1
		0.5	0.67795	0.43051	1.57476	0.11601	1
		0.75	1.5665	0.64002	2.44759	**1.48E-03**	0.087084
	Structured Interview for Schizotypy (SIS) symptom dimensions						
	Negative	0.25	0.03858	0.04753	0.81166	0.41756	1
		0.5	0.09275	0.05877	0.57802	0.11549	1
		0.75	0.14698	0.08375	0.75504	0.08016	1
	Positive	0.25	0.05863	0.05922	0.99008	0.32284	**1**
		0.5	0.1623	0.06831	0.37602	0.11806	**1**
		0.75	0.34591	0.07194	0.80848	0.30824	**1**

4.5 Discussion

In this study, we investigated the relationship between PRS for three major psychiatric disorders and symptom dimensions in multiplex schizophrenia families. Our results indicate that polygenic liability to schizophrenia is significantly associated with increased negative/disorganized symptoms in psychotic subjects, and negative symptoms in non-psychotic subjects. These findings suggest that polygenic liability to negative and disorganized symptoms appear to be specific to schizophrenia, as no significant association between these core schizophrenia symptoms and bipolar disorder or major depressive disorder PRS were observed.

Examination of the scree plot of schizotypy factor structure suggested that a 2-factor solution fit the data best. The symptoms and signs included in each dimension were consistent with the observation that schizotypal traits are generally divided into positive and negative dimensions (Siever & Gunderson, 1983). While the use of self-report questionnaires in combination with interview-based measures is likely to provide a more comprehensive assessment of schizotypy, previous work in another family sample from Ireland suggests that interview-based scales have significantly greater predictive power than self-report measures, in particular for negative symptoms (Kendler et al., 1996). Self-report measures of schizotypy may be inherently limited in their ability to assess certain signs and symptoms. For example, if an individual has little insight into their guardedness in answering the questions, asking them to describe this characteristic in a self-report questionnaire may be ineffective. We attribute our ability to detect a significant association between schizophrenia PRS and the negative dimension of schizotypy to the increased power of PGC3-SCZ derived PRS, and the use of interview-based measurement of schizotypy. It is also possible that our use of a family-based sample (with a high incidence of psychotic disorders, worse premorbid functioning (Peralta et al., 1991) and

potentially more severe symptoms (Fanous et al., 2001)), instead of a population-based cohort also contributed to our ability to detect a significant association. We further note that the association of schizophrenia PRS with only the negative dimension of schizotypy is in agreement with previous epidemiological findings that show familial predisposition to schizophrenia in non-psychotic relatives of probands is likely to be better indexed by negative symptoms (Kendler, 1995).

Our findings on the association between PRS and symptom dimensions in multiplex schizophrenia families provide new insights into the relationship between polygenic liability to schizophrenia and symptom dimensions across the psychosis spectrum. Previous studies have addressed the existence of a single continuum of liability for schizophrenia and schizotypy at the phenotypic level by showing that negative symptoms in psychotic probands were correlated with negative schizotypy symptoms in their non-psychotic relatives (Fanous et al., 2001). Our results further show that polygenic liability to schizophrenia is also significantly associated with negative/disorganized symptom dimension in psychotic subjects, and negative symptom dimension in their non-psychotic relatives in multiplex families. Familial aggregation of negative symptoms has been reported in several studies including the Danish Adoption Study of Schizophrenia (Kety et al., 1975), the Roscommon Family Study of Schizophrenia (Kendler, McGuire, Gruenberg, O'Hare, et al., 1993a), and Maudsley Twin Studies of Schizophrenia (Cardno et al., 2001). These findings are further reinforced by PRS examinations that show strong polygenic associations with negative symptoms (Bigdeli et al., 2017; Ruderfer et al., 2018), while other studies have also reported polygenic associations with disorganized symptoms in schizophrenia (Fanous et al., 2012). Thus, our findings provide genetic evidence in support of previous epidemiological findings that negative and disorganized symptoms are likely to have a

greater familial basis than positive symptoms (Vassos et al., 2008), making them better indices of the familial liability to schizophrenia across the psychosis spectrum. This is further supported by studies that suggest the correlation between negative schizophrenia and schizotypal symptoms appears higher than the correlation between positive schizophrenia and schizotypal symptoms (Kendler, 1995).

We note that the factor structure of OPCRIT in our sample differs slightly from other factor analyses of schizophrenia (Wickham et al., 2001). While hallucinations and delusions often load on a single factor called positive symptoms, they loaded on two distinct factors in our study. However, these results are supported by neurological studies that show etiological discontinuities between hallucinations and delusions (Shergill et al., 2000). Perhaps more importantly, negative and disorganized symptoms loaded on the same factor in our study instead of forming two distinct factors. While we acknowledge this as a potential limitation in our study, we note that this factor structure is consistent with our previous factor analysis in this sample using the Major Symptoms of Schizophrenia Scale (Kendler et al., 1997), as well as factor structures in other studies (Bigdeli et al., 2017; Fanous et al., 2012). To address this further, we narrowed down the loadings from the negative/disorganized factor into "negative only" and "disorganized only" symptoms and showed that while both negative and disorganized symptoms were still independently associated with schizophrenia PRS, this association appears stronger with the disorganized symptoms. This result is in agreement with the observation in another study that suggests while both negative and disorganized symptoms show a strong familial basis, disorganized symptoms are likely to have a more direct association with polygenic liability to schizophrenia (Fanous et al., 2012).

Our results also provide genetic evidence in support of the spectrum model of schizophrenia at the symptom level. Previous studies on symptom dimensions of schizophrenia have focused on sporadic schizophrenia cases (Derks et al., 2012; Nenadić et al., 2020; Smigielski et al., 2021). In contrast, we utilized a well-ascertained schizophrenia family sample with detailed interview-based symptom information to provide a full assessment of the relationship between polygenic liability to major psychiatric disorders and symptom dimensions across the psychosis spectrum. While we observed a significant association between schizophrenia PRS and negative/disorganized dimension in psychotic subjects, and negative dimension in non-psychotic subjects, no significant association between bipolar disorder or major depressive disorder polygenic risks and these core schizophrenia symptom dimensions was observed. This result suggests that, unlike schizophrenia PRS which shows specific associations with negative symptoms in multiplex schizophrenia families, polygenic liability to bipolar disorder or major depressive disorder lack the specificity for core schizophrenia symptom dimensions.

The long-term prognosis of schizophrenia depends on the severity of negative symptoms, and a major question about the clinical heterogeneity of schizophrenia is the extent to which these clinical differences are attributable to genetic differences. Our findings suggest that polygenic liability to schizophrenia is associated with increased negative/disorganized symptoms in psychotic and negative symptoms in non-psychotic subjects from multiplex schizophrenia families. We further showed that in agreement with previous work, polygenic liability to schizophrenia appears to be more strongly associated with disorganized symptoms, while the quantile regression analyses suggest that the association between schizophrenia PRS and negative schizotypy in non-psychotic subjects appears to be strongest at the highest level of

symptom severity. Together, these findings across the extended psychosis spectrum provide genetic evidence for the spectrum model of schizophrenia at symptom level and corroborates previous epidemiological findings that show negative and disorganized symptoms are likely to have a greater genetic basis than positive symptoms, resulting in better indices of familial liability to schizophrenia.

The analyses presented here should be interpreted in the context of some limitations. First, the number of subjects in this sample is modest. Therefore, future studies should replicate these findings in larger family samples. However, to the best of our knowledge, this is the largest family study to date that establishes a link between schizophrenia PRS and negative/disorganized symptoms across the psychosis spectrum. Second, the factor structure of schizophrenia symptoms in our sample differs slightly from other studies. These differences could be attributable to instruments, sample ascertainment, phase of illness, or the rotation used for determining the factors. Third, given that no follow-up assessment on the ISHDSF sample was conducted, we cannot conclusively rule out the possibility that the association between schizophrenia PRS and negative schizotypy dimension could be driven by some unaffected relatives who may have developed schizophrenia later in life. However, we note that of the 172 unaffected relatives, only 11 are in the risk age group for developing schizophrenia (3 males between 18 to 25 and 8 females between 25 to 35), suggesting that this is an unlikely source of bias. Fourth, negative and disorganized symptoms are associated with cognitive deficits in schizophrenia. Given that no cognitive measurements were available, we also cannot rule out the possibility that the association between schizophrenia PRS and negative/disorganized symptoms in psychotic subjects might be driven by cognitive deficits. Fifth, PRS predictions are currently constrained to individuals of European ancestry. Thus, as sophisticated cross-ancestry PRS

methods become available, these findings should be replicated in ethnically and geographically diverse backgrounds. Furthermore, as current PRS methods exclude rare and structural variants, some potentially relevant rare and structural variants were omitted. Finally, we did not consider the role of neuroleptics in ameliorating the symptom severity in the subjects.

CHAPTER V

Genome-wide analysis of schizophrenia and multiple sclerosis identifies shared genomic loci with mixed direction of effects

5.1 Abstract

Common genetic variants identified in GWAS show varying degrees of genetic pleiotropy across complex human disorders. Clinical studies of schizophrenia suggest that in addition to neuropsychiatric symptoms, patients with schizophrenia also show variable immune dysregulation. Epidemiological studies of multiple sclerosis, an autoimmune, neurodegenerative disorder of the central nervous system, suggest that in addition to the manifestation of neuroinflammatory complications, patients with multiple sclerosis may also show co-occurring neuropsychiatric symptoms with disease progression. In this study, we analyzed the largest available GWAS datasets for schizophrenia (N=161,405) and multiple sclerosis (N=41,505) using MiXeR and condFDR frameworks to explore and quantify the shared genetic architecture of these two complex disorders at common variant level. Despite detecting only a negligible genetic correlation (r_G=0.057), we observe polygenic overlap between schizophrenia and multiple sclerosis, and a substantial genetic enrichment in schizophrenia conditional on associations with multiple sclerosis, and vice versa. By leveraging this cross-disorder enrichment, we identified 36 loci jointly associated with schizophrenia and multiple sclerosis at conjunctional FDR <0.05 with mixed direction of effects. Follow-up functional analysis of the shared loci implicates candidate genes and biological processes involved in immune response and B-cell receptor signaling pathways. In conclusion, this study demonstrates the presence of polygenic overlap between schizophrenia and multiple sclerosis in the absence of a genetic correlation and provides new insights into the shared genetic architecture of these two disorders at the common variant level.

5.2 Introduction

In the last decade, GWAS have identified a large number of common genetic risk variants associated with complex human phenotypes (Visscher et al., 2017). Many genetic variants identified by GWAS exhibit varying degrees of genetic pleiotropy (Solovieff et al., 2013), and investigating the nature of these shared genetic risks is important for improving our understanding of the etiology and underlying genetic architecture of complex human disorders. A widely used method for assessing the genetic relationship between two disorders is to estimate genetic correlation defined as the correlation coefficient of additive genetic effects between two disorders (Bulik-Sullivan et al., 2015a; Bulik-Sullivan et al., 2015b). However, genetic correlation estimates do not capture mixtures of effect directions across shared genetic variants, limiting application and interpretation. For example, only a negligible genetic correlation is observed between schizophrenia and brain morphology, however, many genome-wide significant loci are shown to be jointly associated with schizophrenia and brain morphology with mixed direction of effects not captured by genetic correlation estimates (Cheng et al., 2021). Recently developed methods can quantify polygenic overlap between two phenotypes in the absence of a genetic correlation and detect mixed effect directions (Andreassen et al., 2013; Frei et al., 2019). Utilizing such methods to jointly analyze complex human disorders with no direct genetic correlation, but purported dysregulation in similar biological systems, can improve statistical power for locus discovery, reveal biological connections, and provide better understanding of their etiology and potential treatment strategies.

Schizophrenia is a severe psychiatric disorder with a population prevalence of ~ 1% (Saha et al., 2007). Twin, family, and adoption studies consistently estimate the heritability of schizophrenia to be ~0.7-0.8 (Cannon et al., 1998; Cardno & Gottesman, 2000; Heston, 1966;

Tienari et al., 2000), and the largest GWAS of schizophrenia to date have identified 287 loci robustly associated with schizophrenia and account for common variant schizophrenia heritability of ~0.24 (Trubetskoy et al., 2022). Across different waves of schizophrenia GWAS mega-analyses, the MHC region on chromosome 6 involved in the adaptive immune system is shown to be strongly associated with schizophrenia. In addition, schizophrenia GWAS loci are enriched for enhancers that are active in tissues related to immune functions (Ripke et al., 2014). Although many of these immune-related signals can be traced to the MHC region, some of these enrichments remain significant even after exclusion of the MHC region, demonstrating that these findings are not solely driven by the MHC region in the genome (Corvin & Morris, 2014). Furthermore, clinical studies of schizophrenia also suggest cytokine abnormalities in patients with schizophrenia (Arias et al., 2012; Hope et al., 2009).

Multiple sclerosis is a chronic, neurodegenerative disease of the central nervous system with evidence for both genetic and environmental risk factors (Goodin et al., 2021). Multiple sclerosis has a strong genetic component (de Jager et al., 2009; Patsopoulos et al., 2019; Roger Bobowick et al., 1978; Sadovnick et al., 1993), and the largest GWAS of multiple sclerosis to date has identified 32 MHC and 200 non-MHC loci associated with multiple sclerosis and accounts for common variant heritability of ~0.48 (Patsopoulos et al., 2019). In addition to the manifestation of neurological complications, patients with multiple sclerosis sometimes present co-occurring neuropsychiatric symptoms with disease progression (Murphy et al., 2017). For example, the rate of psychosis-like symptoms in multiple sclerosis patients is reported to be ~2-4%, approximately three times higher than in the general population (Patten et al., 2005). However, these symptoms often follow the multiple sclerosis diagnosis and thus, the pathogenesis of these co-occurring neuropsychiatric symptoms remains elusive.

Despite only a negligible genetic correlation between schizophrenia and multiple sclerosis (The Brainstorm Consortium. 2018), both schizophrenia and multiple sclerosis patients show varying degrees of immune and brain dysfunction, suggesting possible shared mechanisms. Additionally, a joint analysis of schizophrenia and multiple sclerosis common risk variation in the MHC region shows that the same alleles are found to be associated with schizophrenia and multiple sclerosis with opposite direction of effects (Andreassen et al., 2015). In the current study, we used the largest available schizophrenia (N=161,405) and multiple sclerosis (N=41,505) GWAS datasets to explore the shared genetic architecture of these two complex disorders by employing two different approaches. First, we applied MiXeR (Frei et al., 2019) to explore and contrast the genetic architecture of these two complex disorders, and to estimate the polygenic overlap between them in the absence of a genetic correlation. Second, we applied condFDR (Andreassen et al., 2013) to leverage the pleiotropic enrichment between these two disorders to increase statistical power for genomic loci discovery and to also identify overlapping loci. By using these statistical approaches, we aim to characterize and contrast the polygenic architecture of schizophrenia and multiple sclerosis, and to enhance our understanding of the underlying genetic mechanisms shared between these two disorders at common variant level.

5.3 Methods

Participants and GWAS data acquisition

The schizophrenia GWAS summary statistics used in this study consists of 67,390 cases and 94,015 controls from the PGC-SCZ. The multiple sclerosis GWAS summary statistics from the International Multiple Sclerosis Genetics Consortium (IMSGC), consists of 14,802 cases and 26,703 controls. Individuals in both studies were predominantly of European ancestry, and detailed descriptions on sample recruitment and subsequent GWAS analyses are available in the

original publications for schizophrenia (Trubetskoy et al., 2022), and multiple sclerosis (Patsopoulos et al., 2019).

Statistical analyses

LDSC was used to estimate the SNP heritability (h^2_{SNP}) and genetic correlation between schizophrenia and multiple sclerosis using the default settings. We applied Gaussian mixture modeling using univariate MiXeR (Holland et al., 2020) to 1) estimate the polygenicity of schizophrenia and multiple sclerosis (defined as π, representing the proportion of causal variants in each disorder), which is expressed as the fraction of the SNPs in the reference panel that have a non-null, true effect on the disorder, 2) estimate the discoverability of these causal variants which is defined as σ, representing the average effect size of the causal variants that shows mean strength of association, representing the mean strength of association, and 3) estimate the distribution of non-null causal variants associated with schizophrenia and multiple sclerosis. Next, we extended the model to bivariate MiXeR (Frei et al., 2019) to estimate the number of trait-specific and shared causal variants associated with schizophrenia and multiple sclerosis. We used SNPs from the European subset of the 1000 Genomes Phase 3 data (Auton et al., 2015) to compute MiXeR parameters. Due to intricate LD patterns and based on current recommendations, variants in the MHC region [GRCh37 6:26000000-34000000] and chromosome 8 inversion [GRCh37 8:7200000-12500000], were excluded from MiXeR analyses (Frei et al., 2019).

To provide a visual representation and assess cross-disorder polygenic enrichment between schizophrenia and multiple sclerosis, we constructed conditional quantile-quantile (QQ) plots. A standard QQ plot visualizes the statistical association of a polygenic disorder relative to the expectation under the null hypothesis. In contrast, a conditional QQ plot provides a visual

representation of the differential enrichment and overlap in associations between two disorders with successively smaller *p*-value thresholds ($p \leq 0.1$, $p \leq 0.01$, $p \leq 0.001$) for association with the secondary disorder. The presence of a leftward deflection observed across more stringent *p*-value thresholds for the secondary disorder in conditional QQ plots represents cross-disorder enrichment and polygenic overlap between the two disorders at the common variant level.

The condFDR framework was then applied to schizophrenia and multiple sclerosis GWAS data to leverage the cross-disorder enrichment shown in conditional QQ plots, to improve statistical power for locus discovery in schizophrenia and multiple sclerosis, and to identify shared genomic loci between them (Andreassen et al., 2013). The condFDR framework has been successfully used in recent years to increase discovery power for loci identification and to reveal shared genomic loci among various polygenic traits or disorders, by leveraging the combined power of two GWAS regardless of the presence of a genetic correlation (Bahrami et al., 2021; Hindley et al., 2021; Rødevand et al., 2021; Smeland, Bahrami, et al., 2020; Smeland, Frei, Shadrin, et al., 2020). This framework is an extension of the standard FDR method that builds on an empirical Bayesian framework by employing all *p*-values from GWAS results of the two disorders, by re-ranking the test statistics for a primary disorder conditioned on a secondary disorder, while controlling for type 1 error by using Benjamini-Hochberg-like FDR correction. An extension of condFDR method called the conjunctional FDR method (conjFDR), was then employed to identify shared genomic loci associated with both disorders. conjFDR is defined as the maximum of the two condFDR values, providing a conservative estimate of the FDR for association. Based on the current recommendations (Smeland, Frei, Shadrin, et al., 2020), the significance threshold was set at condFDR < 0.01 and conjFDR < 0.05. Similar to the preparation for the MiXeR analysis described above, variants in the MHC region [GRCh37

6:26000000-34000000] and chromosome 8 inversion [GRCh37 8:7200000-12500000] were also excluded from the analyses because of the possible impact of intricate regional LD in these two intervals. To correct for variance inflation, all *p*-values were corrected using a genomic inflation control procedure, and random pruning of all SNPs across 100 iterations in QQ-plot constructions and condFDR analyses were also carried out. To achieve this, for each pruning iteration, only one random SNP was retained to represent each LD-independent block ($r^2 > 0.1$), with the final result averaged across all iterations.

Definition of genomic loci

We defined independent genomic loci according to the Functional Mapping and Annotation (FUMA) protocol recommendations (Watanabe et al., 2017). First, independent significant SNPs were defined as SNPs that are independent from each other at $r^2 < 0.6$ with condFDR < 0.01 or conjFDR < 0.05. A subset of these independent SNPs with $r^2 < 0.1$ were then selected as lead SNPs. A distinct genomic locus was then defined by merging all loci that were less than 250kb apart and selecting the SNP with the most significant *p*-value as the lead SNP for the merged locus. In regions with complex LD and multiple overlapping signals, we used 1 independent lead SNP to represent the signals. All the LD information was calculated using the European subset of the 1000 Genomes Project Phase 3 reference panel (Auton et al., 2015) and the direction of effects of the loci shared between schizophrenia and multiple sclerosis were evaluated by comparing their respective Z-scores. A locus was defined as novel using current recommendations (Smeland, Frei, Shadrin, et al., 2020) if: 1) it was not reported in the original GWAS and 2) the association was not previously reported in the NHGRI-EBI GWAS catalog (Buniello et al., 2019) (GWAS catalog last accessed on March 3, 2022).

Functional annotation, gene mapping, and gene-set analyses

FUMA V.1.3.6a (Watanabe et al., 2017) was used to functionally annotate all the candidate SNPs with $r^2 > 0.6$ with one of the independent significant SNPs in the genomic loci with condFDR <0.01 or conjFDR <0.05. These SNPs were annotated with ANNOVAR (Wang et al., 2010) and combined annotation-dependent depletion (CADD) scores (Rentzsch et al., 2019) to predict the deleteriousness of SNP effects on the structure and function of protein products, with the CADD scores of ≥ 12.37, signifying the most deleterious variants. In addition, we used RegulomeDB scores to predict the likelihood of regulatory function of the SNPs based on the overlap of existing functional data including annotation to cis-expression quantitative trait loci and transcription factor binding (Boyle et al., 2012). Furthermore, SNPs were annotated with their chromatin states using 15 categorical states as predicted by ChromHMM based on 5 chromatin marks and 127 epigenomes to predict their transcription and regulatory effects (Enrst & Kellis, 2017).

FUMA V.1.3.6a (Watanabe et al., 2017) was used to link SNPs from condFDR and conjFDR analyses to candidate genes using three gene mapping strategies according to FUMA recommendations: First, we mapped SNPs based on their physical proximity. Second, we used the expression quantitative trait loci (eQTL) mapping method to match cis-eQTL SNPs to genes whose expression is likely to be associated with variation at the SNP level. Third, we used chromatin interaction mapping to link SNPs to genes based on three-dimensional chromatin interactions between each SNP region and specific genes.

Finally, we evaluated gene ontology (GO) (Carbon et al., 2021) gene-set enrichment for the candidate genes mapped to the shared loci using the hypergeometric gene-set analysis implemented in FUMA V 1.3.6a to test whether genes of interest mapped using our 3 gene-mapping strategies were over-represented in any of the pre-defined gene-sets. Moreover, we used

genotype tissue expression (GTEx) data resource (Aguet et al., 2020) to assess gene expression and look at the eQTL functionality of likely regulatory SNPs in the shared loci. All reported *p*-values were corrected for multiple testing using the Bonferroni method.

5.4 Results

Univariate MiXeR suggests that schizophrenia and multiple sclerosis are both polygenic disorders but with different common variant genetic architectures

Our LDSC genetic correlation analysis of current GWAS data shows a negligible, non-significant genetic correlation between schizophrenia and multiple sclerosis (r_G=0.057, *p*=0.41), in agreement with results from the Brainstorm Consortium. Using univariate MiXeR, we found that schizophrenia ($\pi = 3.14 \times 10^{-3}$) is more polygenic than multiple sclerosis ($\pi = 1.77 \times 10^{-4}$). Furthermore, multiple sclerosis has an estimated 566 causal variants with discoverability estimate (causal effect size variance) of $\sigma = 8.92 \times 10^{-4}$, whereas schizophrenia has an estimated 10,002 causal variants with discoverability estimate of $\sigma = 3.87 \times 10^{-5}$. Together, these results indicate that schizophrenia is ~20 times more polygenic, and its genetic determinants are approximately 23 times less discoverable, than multiple sclerosis with the current GWAS sample sizes.

Bivariate MiXeR reveals polygenic overlap between schizophrenia and multiple sclerosis in the absence of a genetic correlation

Bivariate MiXeR (Figure 22) indicates that of 10,002 schizophrenia and 566 multiple sclerosis causal variants, 327 (SD=77) causal variants are shared between these two disorders (Figure 22a), with the proportion of variants estimated to have concordant effect direction of 0.63 (SD=0.04). Furthermore, as shown in the bivariate density plot (Figure 22b), the distribution of Z-scores for schizophrenia and multiple sclerosis show a mixed direction of

effects for variants, with the net result considered to be a negligible positive genetic correlation (r_G=0.057).

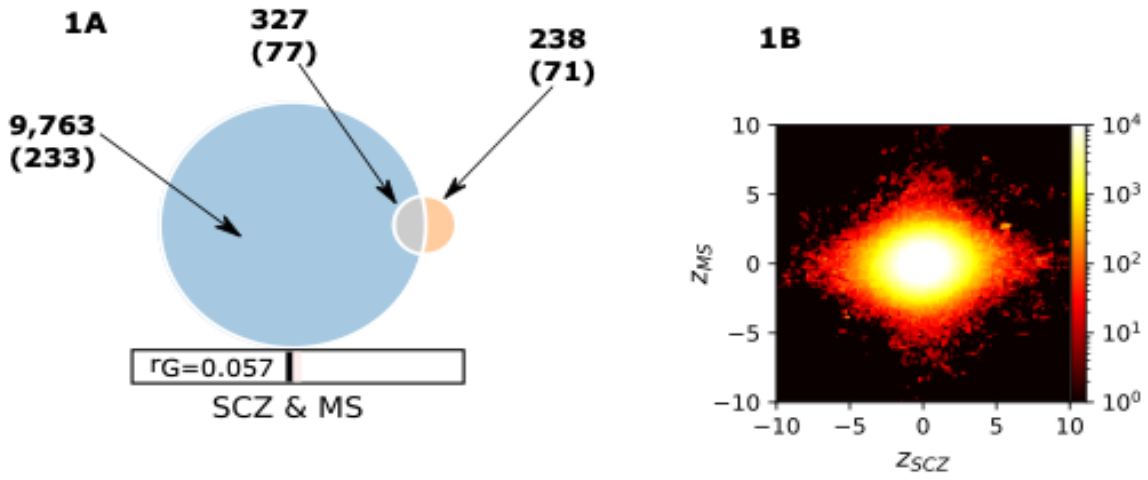


Figure 22: Polygenic overlap between schizophrenia and multiple sclerosis. Venn diagram showing the estimated number of causal variants shared between the disorders and unique to each one. Size of the circle represents polygenicity and the grey portion represents the overlap. Density plot shows the relationship between Z-scores for schizophrenia and multiple sclerosis GWAS visualized using 100 x 100 grid bins with the color indicating log10(N) where N is the number of SNPs projected into each bin.

Conditional/conjunctional FDR analyses increase statistical power and identify shared genomic loci between schizophrenia and multiple sclerosis

Figure 23 shows the conditional QQ plots for schizophrenia and multiple sclerosis. We observe strong enrichment for schizophrenia conditional on associations with multiple sclerosis and vice versa. As a comparison, we also provide conditional QQ plots for schizophrenia and multiple sclerosis conditioned on LDL, which shows no cross-disorder genetic enrichment, indicating the specificity of these enrichments between schizophrenia and multiple sclerosis (Figure 24).

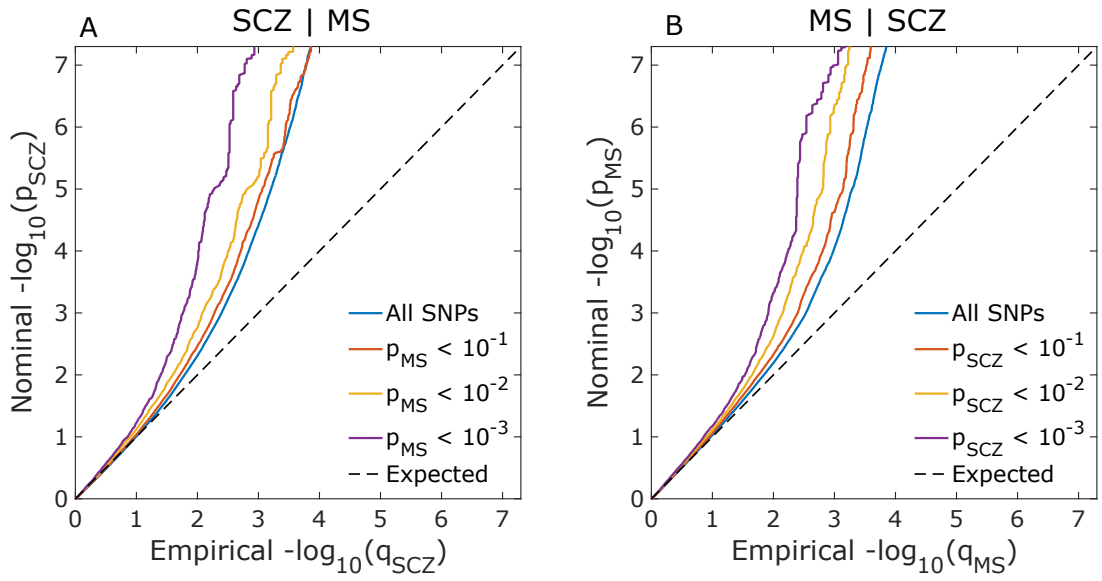


Figure 23: QQ plots for nominal versus empirical schizophrenia -log10 p-values as a function of significant association with multiple sclerosis (A) and vice versa (B)

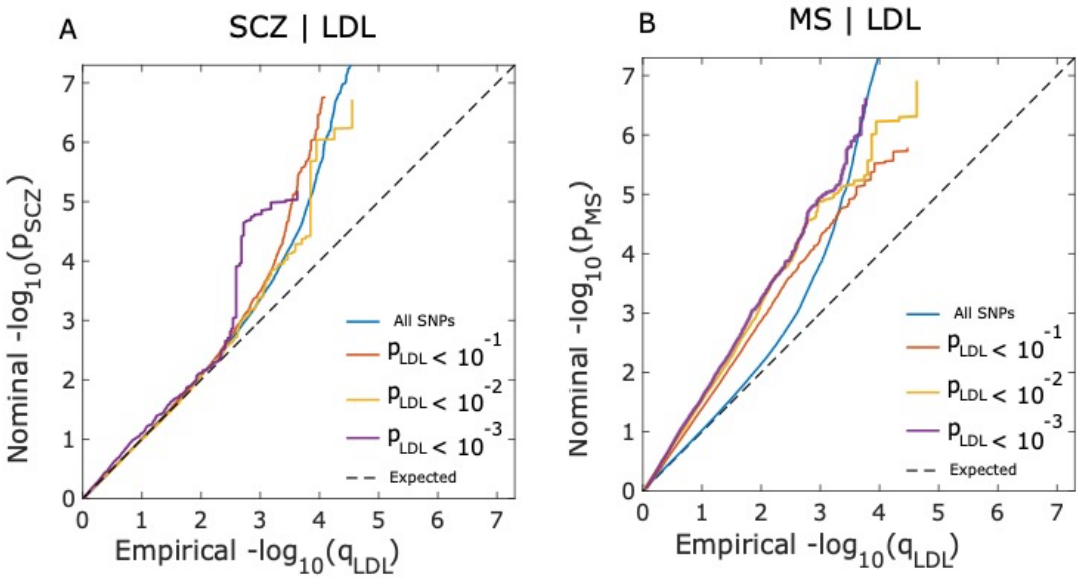


Figure 24: QQ plots for nominal versus empirical schizophrenia -log10 p-values as a function of significant association with multiple sclerosis (A) and vice versa (B)

We leveraged this pleiotropic genetic enrichment between schizophrenia and multiple sclerosis to improve the statistical power for locus discovery through the condFDR framework.

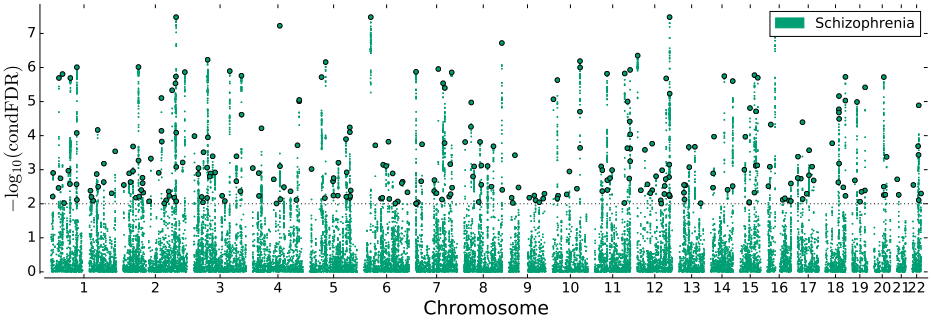


Figure 25: Manhattan plot depicting the -log$_{10}$ condFDR values for SNPs associated with schizophrenia conditional on multiple sclerosis. The dotted line represents the threshold for significant association set at condFDR < 0.01.

At a condFDR <0.01 and identified 247 loci associated with schizophrenia conditional on multiple sclerosis (Figure 25).

Of these 247 loci, 232 loci had been identified in previous schizophrenia GWAS studies and 15 are novel (see Ahangari et al., 2022for the list of novel loci). Furthermore, by conditioning multiple sclerosis GWAS signals on schizophrenia, we identified 139 loci associated with multiple sclerosis at condFDR < 0.01 (Figure 26). Of these 139 loci, 119 loci had been identified in previous multiple sclerosis GWAS studies and 20 are novel (see Ahangari et al., 2022 for the full list of SNPs in the shared loci). We note that some of the significant results from the condFDR analyses, are driven by marginal signals in the second trait, while others show more stronger enrichments.

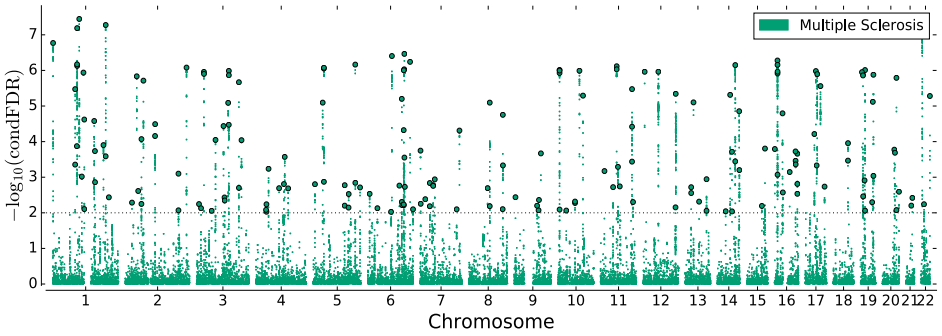


Figure 26: Manhattan plot depicting the -log$_{10}$ condFDR values for SNPs associated with multiple sclerosis conditional on schizophrenia. The dotted line represents the threshold for significant association set at condFDR < 0.01.

Figure 27 (and see Table 1 in Ahangari et al., 2022) show the results for the joint conjFDR analysis of schizophrenia and multiple sclerosis. We identified 36 genomic loci jointly associated with schizophrenia and multiple sclerosis at conjFDR < 0.05. Overall, only ~58% of the lead SNPs in shared loci show the same direction of effect in schizophrenia and multiple sclerosis.

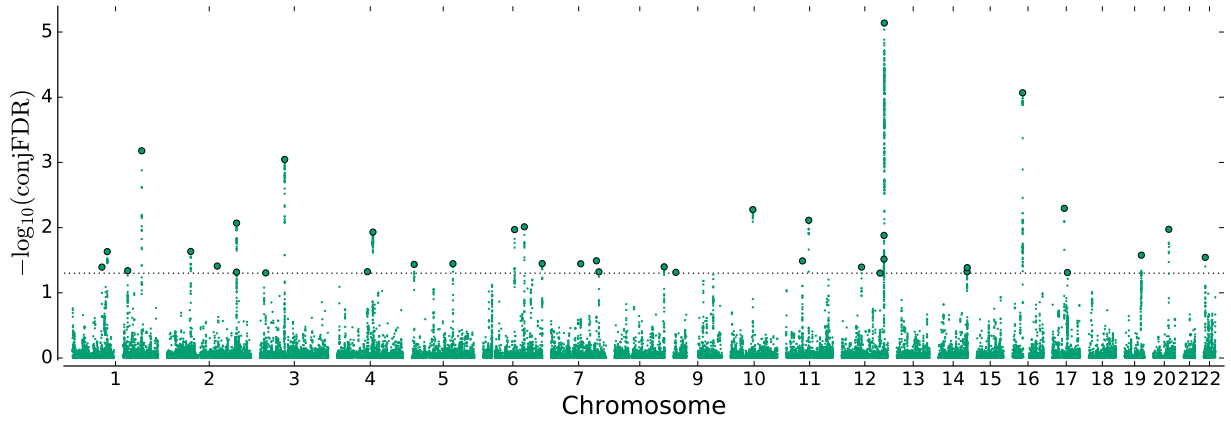


Figure 27: Manhattan plot depicting the -log$_{10}$ conjFDR values for SNPs associated with multiple sclerosis conditional on schizophrenia. The dotted line represents the threshold for significant association set at condFDR < 0.05.

Functional and gene-set analyses of the shared loci between schizophrenia and multiple sclerosis

Functional annotation of all the SNPs in the jointly associated loci are shown in Figure 28. Annotation of SNPs using ANNOVAR in the jointly associated loci revealed that the majority of the SNPs are in intronic regions of the genome with only 1.7% in exons. The distribution of minimum chromatin state shows that 84.3% of the SNPs in the jointly associated loci are in open chromatin state regions, making them more accessible to DNA regulatory elements. Additionally, we observe that 3 of the lead SNPs in the 36 shared loci (rs6065926, rs35866622, and rs16917546) have CADD scores above 12.37, the threshold suggested to signify high deleteriousness.

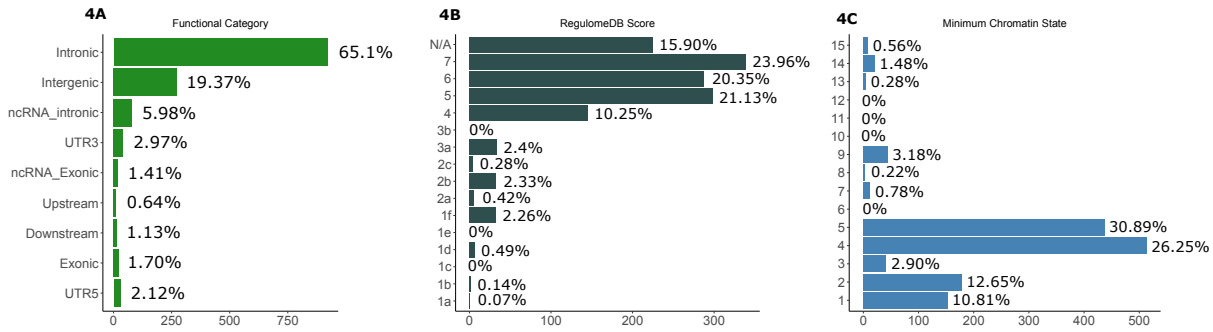


Figure 28: Annotation of the SNPs jointly associated with schizophrenia and multiple sclerosis at conjFDR < 0.05.

GO gene-set analysis for the candidate genes mapped to the jointly associated loci reveals 9 gene-sets significantly associated with schizophrenia and multiple sclerosis (all p <0.001 after multiple testing correction). Of the 9 identified gene-sets, 4 are involved in immune regulation processes (B-cell receptor signaling pathway, immune response regulating cell surface receptor signaling pathway, antigen receptor mediated signaling pathway, and immune response regulating signaling pathway). The p-values for all gene-set analyses were corrected for multiple testing using the Bonferroni method (see supplementary tables in https://doi.org/10.1016/j.bbi.2022.06.007).

5.5 Discussion

In this chapter, we used the largest available GWAS datasets for schizophrenia and multiple sclerosis to characterize the polygenic architecture and overlap between these two complex disorders at common variation level. We first showed that despite the complex, polygenic nature of these two disorders, schizophrenia and multiple sclerosis has different genetic architectures at the common variant level as evident from varying levels of polygenicity, and discoverability of the causal variants associated with each disorder. Estimation of polygenicity and discoverability of the causal variants associated with each disorder, in combination with the estimation of the narrow-sense SNP heritability captured by common

SNPs, can provide us with empirical information to explain why disorders such as multiple sclerosis show higher power for SNP discovery (as explained by higher narrow-sense SNP heritability from current GWAS), than disorders such as schizophrenia. Our results suggest that lower polygenicity, and higher discoverability of causal variants associated with multiple sclerosis are likely to be contributing factors to the differences in the narrow-sense SNP-based heritability multiple sclerosis and schizophrenia at current sample sizes. Furthermore, these results indicate that at current sample sizes, only a fraction of the heritability for schizophrenia or multiple sclerosis can be accounted for by common, genome-wide significant SNPs (Figure 29).

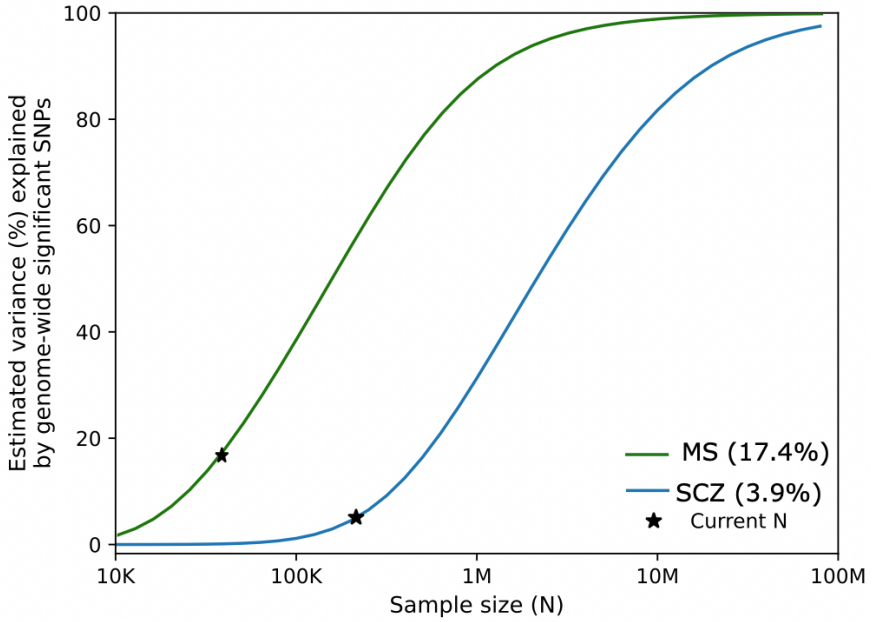


Figure 29: Estimation of the genetic variance explained by SNPs at genome-wide significance level on the Y-axis as a function of sample size on the X-axis (log_{10} scale) for schizophrenia and multiple sclerosis.

Therefore, a substantial increase in current sample sizes is required for both disorders, particularly for schizophrenia due to its higher polygenicity and lower causal variant discoverability, to detect the majority of the common risk variation at the genome-wide significance level.

Genetic correlation estimates rely on globally consistent directions of effects between the two disorders. In contrast, by using MiXeR, we show that despite the negligible genetic correlation between schizophrenia and multiple sclerosis, we can observe a polygenic overlap between these two disorders that is not captured by genetic correlation estimates due to a mixed

direction of effects. Other recent joint analyses of psychiatric and neurological disorders (Monereo-Sánchez et al., 2021; Smeland et al., 2021) also suggest that despite no genetic correlation between psychiatric and neurological disorders, we can observe a substantial polygenic overlap among them with mixed direction of effects. The Brainstorm Consortium have investigated the shared genetic architecture among various psychiatric and neurological disorders and found that neurological disorders do not have strong genetic correlations with neuropsychiatric disorders. In particular, they showed that multiple sclerosis shows only a negligible genetic correlation with other psychiatric and neurological disorders. Based on our univariate MiXeR analysis, we suggest that the very weak genetic correlation between multiple sclerosis and schizophrenia may be partly attribute to the relatively small number of causal variants associated with multiple sclerosis compared to schizophrenia, or the mixed direction of effects among the variants shared across these two disorders, or a combination of both. This is particularly true for genetic correlation estimations that are calculated using the LDSC method, as benchmarking shows that small sample sizes and low polygenicity (specially as evident with multiple sclerosis), can cause bias in LDSC genetic correlation estimation and make the results difficult to interpret (Bulik-Sullivan et al., 2015).

Based on the observed polygenic overlap between schizophrenia and multiple sclerosis, we hypothesize that increased incidence of psychotic-like features in some multiple sclerosis patients, may be partially attributable to the genetic risk from loci shared between schizophrenia and multiple sclerosis with concordant direction of effects that is otherwise not captured through standard genetic correlation estimations. However, we caution that since these psychotic-like features tend to occur after the diagnosis of patients with multiple sclerosis, pinpointing their true cause is challenging. Thus, we should be careful not to conclude that genetic effects, or aberrant

neurological dysfunctions individually can be the cause of these psychotic-like features in patients with multiple sclerosis, rather, a combination of these concordant genetic effects and brain dysfunctions is likely to be a more complete explanation.

By leveraging the cross-disorder enrichment between schizophrenia and multiple sclerosis through condFDR analyses, we identified 15 and 20 novel loci associated with schizophrenia and multiple sclerosis respectively. We note that some of the significant signals in the conditional analyses are driven by quite marginal signals in the second trait, while others are not. In general, we see that multiple sclerosis signals tend to be stronger in the condFDR analysis of schizophrenia, than schizophrenia signals are in the condFDR analysis of multiple sclerosis. This observation may reflect the higher discoverability and larger effect sizes of multiple sclerosis variants compared to schizophrenia. Here we discuss one locus of potential interest for each analysis. One of the identified novel multiple sclerosis loci is in the upstream region of a protein-coding gene hydroxysteroid 17-beta dehydrogenase 4 (*HSD17B4*) which encodes peroxisomal multifunctional enzyme type 2, and mutations in this gene are associated with autosomal-recessive Perrault syndrome and D-bifunctional protein deficiency, both of which show evidence of the involvement of the nervous system (Pierce et al., 2010; Suzuki et al., 1997). A homozygous variant in *HSD17B4* is proposed to be causative of middle-age onset, autosomal recessive spinocerebellar ataxia in two consanguineous families (Matsuda et al., 2020). *HSD17B4* is shown to be ubiquitously expressed in a wide variety of tissues including the brain and immune cells (Uhlén et al., 2015), and *Hsd17b4* knock-out mice also show neurodegeneration and demyelination (Bult et al., 2019), both of which are characteristics of multiple sclerosis pathology, which further supports the notion that variation in the *HSD17B4* gene may modulate the risk of developing multiple sclerosis with low to intermediate effects.

Additionally, the TNF alpha induced protein 8 (*TNFAIP8*) gene that is preferentially expressed in human immune cell types also resides in the same locus (Sun et al., 2021), providing further evidence for the involvement of this locus in multiple sclerosis pathobiology. In contrast, one of the identified novel schizophrenia loci is in the upstream of the *YWHAB* gene that is part of the YWHA gene cluster, encoding 14-3-3 proteins, a family of conserved genes highly expressed in the brain (Aitken, 2006) that play an important role in synapse development and plasticity (Li et al., 2006; Toyo-Oka et al., 2003). Recently, dysregulation of peripheral expression of the *YWHAB* gene was shown to be associated with the onset of psychosis (Demars et al., 2020), while other members of the YWHA cluster of genes have also been implicated in schizophrenia (Navarette et al., 2022). This further supports the involvement of this locus in schizophrenia pathobiology. Together, these results indicate that condFDR analyses of schizophrenia and multiple sclerosis can boost statistical power for genetic discovery and identify novel loci with meaningful associations.

We note that the majority of the SNPs identified in the shared loci between schizophrenia and multiple sclerosis through the conjFDR are in intronic regions that are likely to impact expression/regulation and are not predicted to be highly deleterious based on CADD scores. This observation is expected, as highly deleterious variants are typically predicted to be removed by selection rapidly in the population, and therefore are more likely to contribute to disorders that are rare and severe, rather than more common, polygenic disorders such as schizophrenia or multiple sclerosis (Dudley et al., 2012; Stover et al., 2022). Considering this expectation, we identified 3 SNPs (rs6065926, rs35866622, and rs16917546) with CADD scores above 12.37 in the conjFDR analysis, implying significant deleteriousness, yet with no fatal effects. One of these SNPs, rs6065926, also had RegulomeDB score of 1f, which indicates that it is likely to

affect binding sites and possibly linked to expression of a gene target with QTL + transcription factor (TF) binding/DNase peak evidence. We followed up this finding in the GTEx database and found that this SNP is a significant eQTL in different brain regions for the CD40 molecule (*CD40*) and solute carrier family 12 member 5 (*SLC12A5)* genes. The *CD40* gene codes for CD40 protein found in antigen-presenting cells and is required for their activation, and its deficiency is linked to autoimmune diseases (Karnell et al., 2019). The product of *SLC12A5* is a potassium-chloride cotransporter that maintains chloride homeostasis in neurons and is also involved in GABAergic signaling; variants within it are associated with neurodevelopmental and epileptic disorders (Saitsu et al., 2016). Additionally, one of the jointly associated loci between schizophrenia and multiple sclerosis is linked to the microtubule associated protein tau (*MAPT*) gene that codes for tau protein, which is differentially expressed in the nervous system (Caillet-Boudin et al., 2015); mutations in this gene are linked to neurodegenerative diseases such as Alzheimer's Disease and other forms of dementia. A homolog of the *MAPT* gene is found to be differentially phosphorylated in patients with schizophrenia (Grubisha et al., 2021), while abnormally phosphorylated tau protein is also shown to be associated with multiple sclerosis (Anderson et al., 2008). Furthermore, patients with early onset frontotemporal dementia caused by mutations in *MAPT,* sometimes show psychotic-like symptoms that resemble those seen in patients with schizophrenia (Momeni et al., 2010). This finding provides further preliminary genetic evidence in support of a possible role of tau protein in the pathobiology of both schizophrenia and multiple sclerosis at common variant level.

Although genetic components play a substantial role in the etiology of schizophrenia and multiple sclerosis, we note that the role of environmental influences, in particular with multiple sclerosis, should not be omitted. A recently published longitudinal analysis conducted on a large

sample of United States Army recruits showed that a prior infection with Epstein-Barr virus increases the risk of multiple sclerosis by 32-fold, making it by far the most important risk factor for developing multiple sclerosis (Bjornevik et al., 2022). Additionally, patients with schizophrenia are also shown to have marked elevation in the levels of antibodies to Epstein-Barr virus compared to controls, suggesting that Epstein-Barr virus may also have a role in the etiology of schizophrenia (Dickerson et al., 2019).

The results presented in this study should be interpreted in the context of some limitations. Current large-scale GWAS studies of complex disorders such schizophrenia and multiple sclerosis are largely conducted in individuals of European ancestry, which limits the generalizability of these findings to other ancestral groups. Additionally, current methods employed in this study rely on common variant data from published GWAS studies, therefore, only the common variant data for schizophrenia and multiple sclerosis were analyzed in this study. As a result, some rare and structural variations important to the genetic architecture of these two disorders were not taken into consideration. Furthermore, due to intricate LD patterns in the MHC region, and based on the current recommendations for both MiXeR and condFDR frameworks, variants in the MHC that play an important role in the genetic architecture of both schizophrenia and multiple sclerosis were excluded from the current study. However, we note that another study (Andreassen et al., 2015) which specifically focused on common MHC genetic variation, also observed a balance of mixed direction of effects for the variants shared between schizophrenia and multiple sclerosis in the MHC region, making the current study an extension of the prior findings in the MHC region at genome-wide level. Finally, with a recent study showing the strong effect of Epstein-Barr virus infection on the onset of multiple sclerosis, as well as some previous studies proposing a link between viral infections prior to the onset of

schizophrenia, future studies could also look at the effects of shared environmental factors between schizophrenia and multiple sclerosis.

In conclusion, this chapter provides new insights into the genetic relationship between schizophrenia and multiple sclerosis at the common variant level. To our knowledge, prior joint analyses of schizophrenia and multiple sclerosis have largely focused on either global genetic correlation measurements, or specifically on variants inside the MHC region. We hypothesized that because of prior evidence for immune dysregulation in schizophrenia patients, and co-occurring neuropsychiatric symptoms in multiple sclerosis patients, we may observe polygenic overlap between these two disorders that can be exploited for further downstream analyses. In agreement with prior work, we detected no significant genetic correlation between schizophrenia and multiple sclerosis by LDSC analysis. This observation appears to be attributable to the balance of shared and opposite effect directions detected in both bivariate MiXeR and conjFDR analyses. Our results further suggest that the increased rate of psychotic-like features in some multiple sclerosis patients may be partially attributable to the genetic loci shared between schizophrenia and multiple sclerosis that have concordant direction of effects. Finally, our results suggest that despite substantial increase in GWAS sample sizes in recent years, a considerable growth in sample sizes are still required to increase our understanding of the shared genetic etiology among complex, polygenic disorders such as schizophrenia and multiple sclerosis at common variant level.

CHAPTER VI

Improving the discovery of rare variants associated with alcohol problems by leveraging machine learning phenotype prediction and functional information

6.1 Abstract

Alcohol use disorder is moderately heritable with significant social and economic impact. GWAS have identified common variants associated with alcohol use disorder, however, rare variant investigations have yet to achieve well-powered sample sizes. In this study, we conducted an interval-based exome-wide analysis of the AUDIT-P using ML predicted risk and empirical functional weights. Filtering the UK Biobank 200k exome release to unrelated individuals of European ancestry resulted in 147,386 individuals with 51,357 observed and 96,029 unmeasured but predicted AUDIT-P. Sequence Kernel Association Test (SKAT) was used for rare variant (MAF < 0.01) interval analyses using default and empirical weights. Empirical weights were constructed using annotations found significant by stratified LDSC analysis of predicted AUDIT-P GWAS. Samples with observed AUDIT-P yielded no significantly associated intervals, but alcohol dehydrogenase 1C (*ADH1C*) and thyroid hormone receptor alpha (*THRA*) genes were significant (FDR q<0.05) using default and empirical weights in the predicted AUDIT-P sample. These findings provide evidence for rare variant association of *ADH1C* and *THRA* with AUDIT-P and highlight the successful leveraging of ML to increase effective sample size and prior empirical functional weights based on common variant GWAS data to refine and increase the statistical significance in underpowered phenotypes.

6.2 Introduction

Alcohol use disorder and problematic alcohol use are moderately heritable conditions with significant social and economic impact (Griswold et al., 2018; Verhulst et al., 2015). Large-scale GWAS have been successful in identifying many common variants associated with alcohol consumption and alcohol use disorders. Together, common risk variants can account for up to 7-12% of the variance in alcohol use disorder and related disorders (Sanchez-Roige et al., 2019; Walters et al., 2018; Zhou et al., 2020). Due to the heterogeneous manifestation of alcohol use disorder, assembling large, meticulously diagnosed cohorts is challenging. When clinical diagnoses using DSM or International Classification of Disease (ICD) are not feasible or available, screening instruments such as the AUDIT can be used (Saunders et al., 1993). The AUDIT is a screening questionnaire designed to identify hazardous alcohol use that consists of 10 items that produce a total quantitative measurement from 0-40 (AUDIT-T) composed of two subscales of alcohol consumption (AUDIT-C) and problematic use (AUDIT-P). Previous work has demonstrated that AUDIT-C shows strong genetic correlation with other consumption phenotypes such as drinks per week (Kranzler et al., 2019; Liu et al., 2019), but only a moderate genetic correlation with DSM diagnosis of alcohol use disorder, suggesting that alcohol consumption measurements such as AUDIT-C are not on the same phenotypic continuum as alcohol use disorder. In contrast, there is a strong genetic correlation between AUDIT-P and DSM diagnosis of alcohol use disorder, as well as other major psychiatric disorders such as schizophrenia and major depressive disorder, suggesting that AUDIT-P is useful for studies of alcohol related psychopathology when alcohol use disorder diagnosis is unavailable because its use can increase both sample size and statistical power for discovery of common variants with meaningful biological functions (Sanchez-Roige et al., 2019).

Although pinpointing the biological effects of variants identified from GWAS of complex disorders are challenging (Visscher et al., 2017), some of the results from studies of alcohol-related disorders represent an exception to this general pattern. Across different studies, variants in genes in the alcohol dehydrogenase (ADH) cluster, which are central to ethanol metabolism in the body, are shown to be significantly associated with alcohol use disorders (Walters et al., 2018; Zhou et al., 2020) or AUDIT-P (Sanchez-Roige et al., 2019).

While the use of screening questionnaires such as the AUDIT has allowed researchers to increase the effective sample size and the discovery power for identifying common variants associated with AUDIT, rare variant investigations have yet to achieve sample sizes similar to well-powered GWAS. With the decrease in sequencing costs, large-scale biobank samples such as the UK Biobank cohort have begun performing large-scale WES and WGS on the participants to facilitate the study of the rare variant architecture of complex traits and disorders in parallel with common variant GWAS studies (Szustakowski et al., 2021). Despite these large-scale sequencing efforts, biobanks do not necessarily have consistent measurements for all variables across all subjects and may show extensive non-random block-wise missingness in their phenotypic and survey data which limits the effective sample size for genetic association studies (Gentry et al., 2022). This problem becomes more apparent for rare variant studies as large effective sample sizes are required to identify rare variants with low MAF in the population. For example, a recent exome-wide analysis of the AUDIT using the 200k exome release of the UK Biobank failed to identify any significant genes or variants associated with heavy drinking or problem drinking (Curtis, 2022). This observation could be attributed to inadequate power for

rare variant association testing in this study (N=~50k), because of the full 500k UK Biobank participants, only ~30% have completed the AUDIT questionnaire, and of those, only ~50k have exome data in the 200k release of the UK Biobank exome data.

One way to address the issue of inadequate power and effective sample sizes for rare variant studies is to utilize the available information through ML to predict the phenotype of interest in subjects for whom it is missing. We have previously developed an application of the Group Least Absolute Shrinkage and Selection Operator (LASSO) called the Missingness Adapted Group-wise Informed Clustered LASSO (MAGIC-LASSO) which predicts unmeasured quantitative outcomes such as AUDIT with high phenotypic and genotypic accuracy in the full UK Biobank sample (Gentry et al., 2022). Another way to potentially increase the statistical power for rare variant identification is to incorporate functional genomics information as *a priori* weights.

Previous work also shows that incorporating functional information can increase predictive power in variant identification in GWAS studies (Gusev et al., 2014; Pickrell, 2014; Weissbrod et al., 2020). There is strong evidence across complex traits that functional annotations show enrichment, and the specific annotations and magnitude of enrichment vary across disorders. Genomic regions conserved in mammals show strong enrichment across many complex traits ranging from psychiatric and immunological disorders to anthropometric traits. In contrast, some annotation classes show enrichment patterns that are unique to specific traits or diseases. For example, Functional Annotation of the Mouse/Mammalian Genome (FANTOM)-5 enhancers (Andersson et al., 2014) show strong enrichment for immunological disorders such as Crohn's Disease or Ulcerative Colitis, while H3K4me3 annotations (marking active promoters)

from neuronal cells are significantly enriched for psychiatric disorders such as schizophrenia (Tansley & Hill. 2018). Furthermore, complex disorders show convergence between common and rare variant signals, suggesting that incorporating information from GWAS and functional genomics could refine and increase discovery power for rare variant testing (Johansen et al., 2010; Rivas et al., 2011; Trubetskoy et al., 2022).

In this chapter, we analyzed unrelated European subjects within the 200k exome release of the UK Biobank to conduct an interval-based rare variant analysis on AUDIT-P. SKAT-O analyses were performed using subjects for whom AUDIT-P is directly measured (N=51,357) and all predicted subjects (N=147,376) irrespective of whether AUDIT was measured or unmeasured. The two analyses were compared to evaluate whether the increase in effective sample size can improve discovery power for rare variant association testing in AUDIT-P. Additionally, we evaluated the impact of including disease-specific functional information as *a priori* weights to investigate whether inclusion of functional information can improve statistical power for rare variant association testing compared to default SKAT weights.

To our knowledge, this is the largest available rare variant study of AUDIT-P to date. We hypothesize that by increasing the effective sample size and including disease-specific functional information as *a priori* weights for the interval testing, we may be able to uncover rare variant signals associated with AUDIT-P that would be missed using only directly measured subjects and default SKAT variants weights that only consider allele frequency. These results are expected to complement the growing literature on common variant studies on alcohol related-phenotypes and the methods described in this chapter can be extended to rare variant analysis of other traits including schizophrenia.

6.3 Methods

UK Biobank Cohort

The analysis described in this chapter has been conducted using the UK Biobank dataset (application 30782). The UK Biobank dataset is a large, population-based sample that includes more than 500,000 participants aged between 40 and 69 years (Sudlow et al., 2015). A wide range of phenotypic measurements and biological samples have been assessed and collected for these participants at different centers located in the UK. The UK Biobank has obtained ethics approval from the Northwest Multi-Center Research Ethics Committee, as well as informed consent from all the participants. In addition to the exome sequencing described in more detail below, individuals in the UK Biobank were also genotyped using the Affymetrix UK BiLEVE or Affymetrix UK Biobank Axiom arrays and the genotypes were imputed to the Haplotype Reference Consortium reference panel using IMPUTE2 (Loh et al., 2016) with full details provided elsewhere (Bycroft et al., 2018).

Phenotypic description and imputation of AUDIT using MAGIC-LASSO:

The AUDIT is a ten-item, screening questionnaire instrument for alcohol consumption and problems containing three questions that survey consumption (AUDIT-C), and seven that survey problematic alcohol use (AUDIT-P). The survey was completed as part of the Mental Health Questionnaire battery of questionnaires in a subset of 157,162 out of the full 500k UK Biobank participants.

The MAGIC-LASSO procedure, described previously (Gentry et al., 2022), was applied to predict AUDIT scores in participants for whom the AUDIT questionnaire was not directly administered in the UK Biobank. In brief, the MAGIC-LASSO procedure is an adaptation of the

Group-LASSO ML method for penalized regression that can predict variables in the presence of non-random, block-wise missingness. The procedure represents a new implementation of established ML algorithms and employs a regression-based solution that is suited for the penalization of categorical predictors which are prevalent in the UK Biobank. MAGIC-LASSO procedure involves 1) characterizing missingness, 2) filtering variables for general missingness and balance across training and target sets, 3) variable clustering based on missingness, 4) iterative Group-LASSO and variable selection within clusters, and 5) cross-cluster model building with variables prioritized by informativeness. The phenotypic correlation between measured and predicted scores was 0.69, while genetic correlations between observed AUDIT-P in measured subjects and predicted AUDIT-P in unmeasured subjects (who are completely independent) was 0.91, demonstrating the method has significant accuracy and utility (Figure 30).

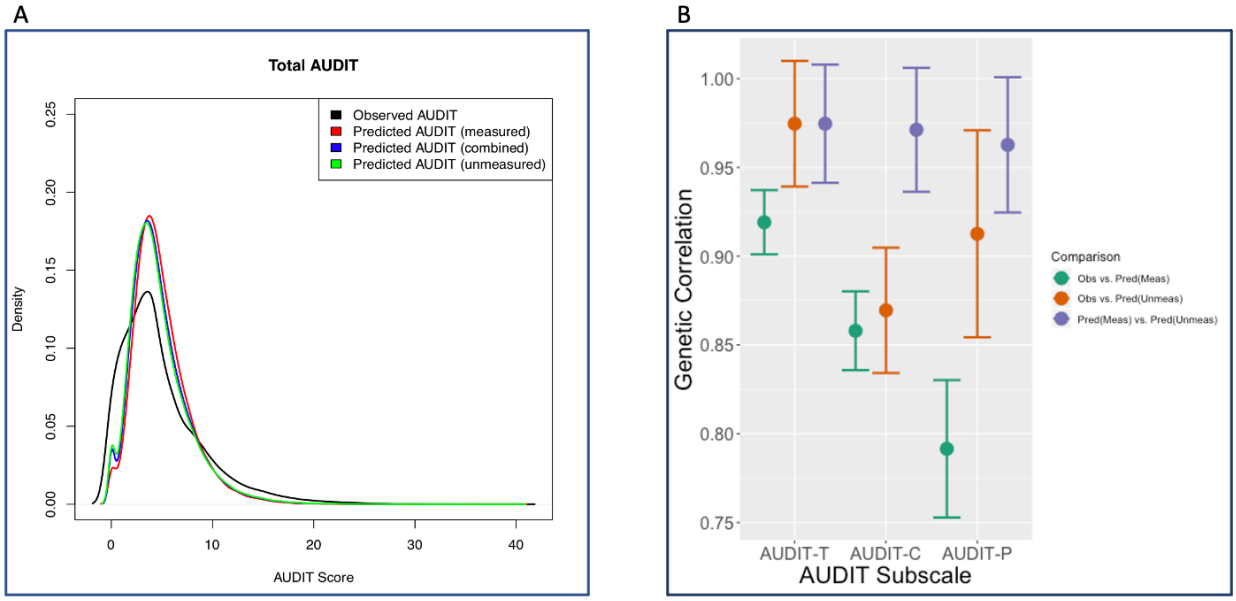


Figure 30: Prediction of AUDIT in the UK Biobank using the MAGIC-LASSO framework. the figure is reproduced from Gentry et al 2022. Panel A shows the distribution of AUDIT scores. Panel B shows the genetic correlation estimates of predicted and observed AUDIT scores.

Importantly, these correlations indicate that the MAGIC-LASSO predictions effectively recapture unmeasured phenotypic information and likely lie closely along the same genetic continuum as observed AUDIT-P. More information is provided elsewhere in the original publication (Gentry et al., 2022).

Whole exome-sequencing in the UK Biobank:

The 200k exome release of the UK Biobank dataset was used in this analysis. The exomes were downloaded for 200,643 subjects who had undergone exome sequencing. Exomes were captured using the IDT xGen Exome Research Panel v.1.0 with an average coverage of 20X at 95.6% of sites. The Original Quality Functional Equivalent (OQFE) Pipeline was used to map raw FASTQ files with BWA-MEM to the GRCh38 reference genome while retaining all other alignments. The OQFE CRAMs were then called using DeepVariant to generate per-sample gVCFs that were jointly called using GLnexus. The OQFE version of the Plink formatted exome files was then downloaded and utilized for all the analyses described in this study (field 23155). Samples were initially filtered to retain only unrelated subjects of British ancestry (N=359,980) as in previous analyses (Gillespie et al., 2022). This yielded 147,376 participants with exome data from the 200k exome release of whom 51,357 had AUDIT directly measured.

Empirical functional weights from predicted AUDIT-P GWAS:

GWAS of common variants associated with predicted AUDIT-P was used as the basis for empirical functional weights for downstream rare variant testing. As described previously (Gillespie et al., 2022), GWAS was conducted using the BGENIE software (Bycroft et al., 2018, version 1.3). Briefly, pre-GWAS filtering excluded markers with MAF <0.5%, imputation quality < 0.8, and p-value < 10^{-6} for deviations from the Hardy Weinberg expectation.

Association analyses included age, sex, and the first 20 PCs as covariates. After filtering, the sample of independent European subjects used for GWAS was 359,980 with 117,559 and 242,421 subjects in the measured and unmeasured AUDIT sets, respectively.

Stratified LDSC (Finucane et al., 2015) was used to partition the heritability of predicted AUDIT-P GWAS described above (unpublished work) into functional annotation classes while accounting for the overlap between different functional classes using the *overlap-annot* flag. In addition to the baseline functional annotation classes from LDSC version 2.2., four custom brain-specific functional annotation classes acquired from the psychENCODE Consortium (Wang et al., 2018) were also used for stratified LDSC analysis which resulted in 87 annotation classes in total. The four custom annotations classes from the psychENCODE Consortium included 1) psychENCODE Enhancers, 2) h3k27ac markers in the prefrontal cortex, 3) h3k27ac markers in the temporal cortex, and 4) h3k27ac markers in the cerebellum. These datasets are publicly available and can be downloaded as part of the derived datasets from www.resource.psychencode.org

Empirical weights based on functional annotations were constructed as follows: 1) each observed exome variant (in the vcf) was annotated for the presence or absence for each functional annotation class, 2) annotations were retained if significant by Akaike information criterion (AIC) in the stratified LDSC analysis, 3) the enrichment scores for significant annotation classes, which represents the fold enrichment of that annotation class derived from stratified LDSC, were added up for each position in the exome. In cases where a variant did not fall within any of the significantly enriched annotation classes, the variant would not get up-weighted for the analysis and receive a weight of zero.

Interval definitions:

GENCODE V.39 was used as the basis for defining intervals to include the most comprehensive list of possible intervals in this analysis (Frankish et al., 2021). In addition to protein-coding genes, GENCODE includes gene models from multiple classes of noncoding RNA genes and other locus types, allowing for a more comprehensive examination of variation in expressed sequences. A comprehensive set of intervals across the genome including those between transcripts (e.g. intergenic intervals) was constructed. Due to overlapping protein coding transcripts, nested small RNAs, and long non-coding RNAs, constructed intervals were not necessarily independent. Although the current study used exome data, some intergenic intervals were observed to contain variants possibly due to off-target capture or annotation errors. All intervals regardless of class (protein coding, intergenic, or others) were included in the analysis if they contained observed rare variation. Figure 31 shows the distribution of interval classes tested in our analyses. Of the 53,171 defined intervals, association testing was limited to intervals with at least 2 rare (MAF<0.01) exome variants. In total, 21,105 intervals were tested with most (N=17,968) being protein-coding genes, as expected. An additional 490

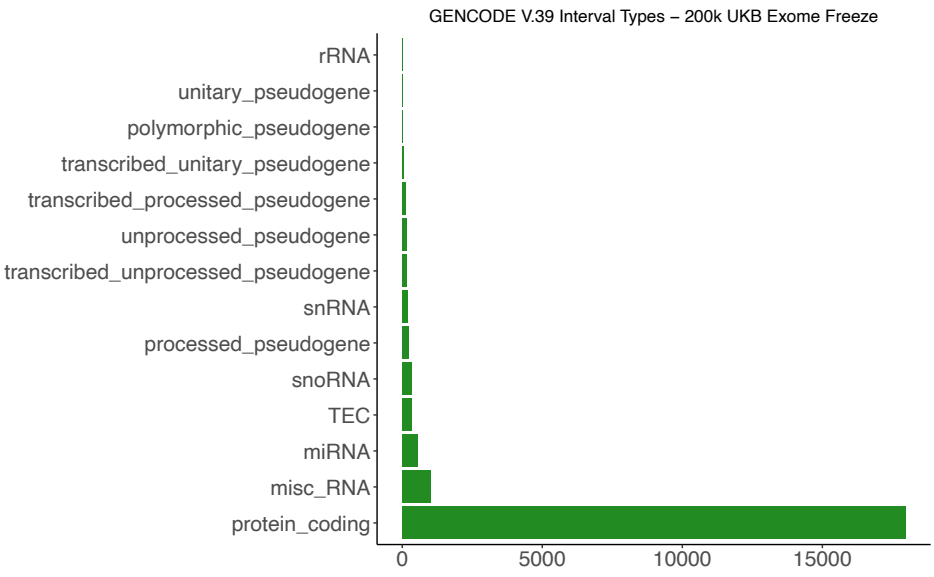


Figure 31: Distribution of the intervals tested in AUDIT-P rare variant analysis based on GENCODE V.39 definition of genic and non-genic intervals. In total 21,105 intervals were tested across the exomes.

intervals were not tested due to insufficient variants. While not informative for the current exome-based study, these may be important and informative for future WGS based studies.

Rare-variant interval-based association testing:

Interval-based rare variant aggregate testing was performed using the SKAT package in R (Ionita-Laza et al., 2013). False discovery rate (FDR) analysis was performed, and Q-values were generated for each test using the qvalue package in R (Storey J, 2015). In addition to FDR, we also report corrected *p*-values using the conservative Bonferroni correction method.

Four sets of SKAT-O interval tests were carried out varying two conditions. First, SKAT-O tests were performed using default weights, which are based on allele frequency, versus empirical functional weights constructed based on significant annotation classes from stratified LDSC analysis in subjects for whom AUDIT-P scores were directly measured (N=51,357). The goal was to evaluate whether *a priori* information from the enrichment of common variant GWAS data in combination with functional annotation information can improve statistical power for rare variant testing. Second, SKAT-O tests using the default versus empirical weights were performed in the full sample of 147,376 unrelated subjects with predicted AUDIT-P. The goal here was to investigate whether increase in effective sample size can increase our statistical power for rare variant testing. We hypothesized that using both a) empirical functional weights based on *a priori* functional annotation enrichments and b) increased effective sample size using predicted AUDIT-P would improve detection of intervals containing rare variants influencing the trait.

We opted to use SKAT-O for interval testing in this study. Burden tests are powerful when most variants in the interval are causal, with the effects in the same direction. Conversely,

SKAT is more powerful when a fraction of the variants are noncausal, or have causal variant effects with mixed direction of effects. The SKAT-O approach maximizes power by adaptively using the data to combine burden and SKAT tests while maintaining the power for rare variant testing, making it an appropriate method when the direction of effects for causal variants are not known *a priori*.

6.4 Results

Empirical functional weights for AUDIT-P

Using LDSC, we estimated the heritability of the predicted AUDIT-P (N=359,980) to be h^2_{SNP} = 0.0647 (SE=0.0034), in close agreement with measured AUDIT-P (N=121,604) in the UK Biobank (Sanchez-Roige et al., 2019). Partitioning the heritability of predicted AUDIT-P GWAS into functional categories resulted in 12 significant annotations after multiple testing corrections (Figure 32). Of these annotation classes, three were specific to the brain from the psychENCODE consortium (Table 19), demonstrating the utility of using annotations beyond the available baseline set.

Table 19: Significant annotation classes enriched in the heritability of predicted AUDIT-P after multiple testing correction. Grey categories are custom classes.

Annotation Category	S-LDSC-Enrichment	LDSC-Enrichment *p*-value
Nucleotide Diversity	0.8288	3.22E-12
Background Selection	1.1871	1.23E-06
Conserved Primate phastCons46Way	1.856	3.99E-05
CpG Content	1.0937	2.43E-05
psychENCODE Enhancers	6.4294	1.72E-04
Conserved Lindblad-Toh	7.6886	2.43E-04
psychENCODE Cerebellum H3K27ac	8.6998	3.09E-04
Bivalent Chromatin State	6.28	9.03E-03
Human Promoter Villar	5.0106	3.92E-03
Human Enhancer Villar	3.8983	4.44E-03
Introns UCSC	6.2072	1.77E-02
psychENCODE Temporal Cortex H3K27ac	5.39	2.85E-02

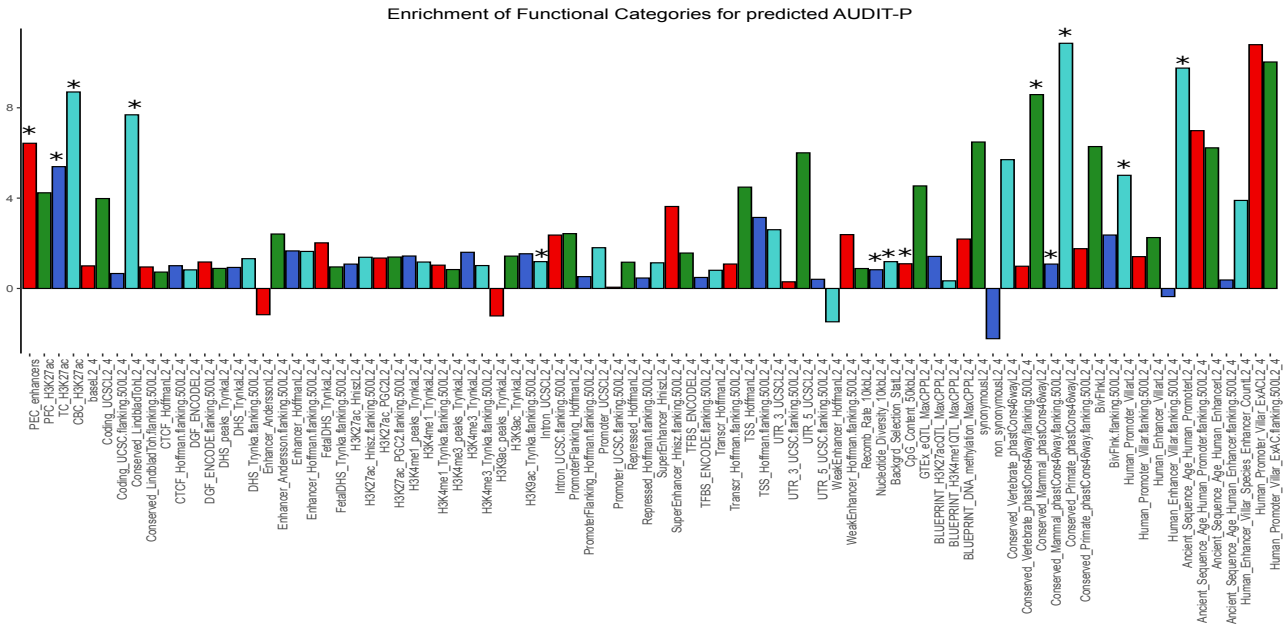


Figure 32: Partitioning the heritability of predicted AUDIT-P into functional categories using stratified LDSC. X-axis represents the functional categories. Y-axis represents the enrichment value for each category. Asterisks represent significant intervals after multiple testing correction.

Comparing the partitioned heritability results of predicted AUDIT-P with other complex traits revealed potentially important insights. For example, heritability of height GWAS (Yengo et al., 2018) is significantly enriched in 36 functional classes. However, none of the brain-specific annotation classes from the psychENCODE consortium showed significant enrichment for heritability of height. In contrast, heritability of schizophrenia GWAS (Trubetskoy et al., 2022) is significantly enriched in 21 annotation classes, which includes all four brain-specific annotation classes from the psychENCODE consortium. Finally, heritability of drinks per week GWAS (Liu et al., 2019) is significantly enriched in 10 functional classes, none of which are specific to the brain. While both schizophrenia and drinks per week show genetic correlation with predicted AUDIT-P, the pattern and magnitude of enriched annotations are different. Together, these results suggest that while there are important differences in heritability enrichment of these complex traits, inclusion of brain-specific annotation classes for psychiatric

traits such as schizophrenia or AUDIT-P can improve partitioning heritability analyses and provide better disorder-specific empirical functional weights for rare variant testing.

Empirical functional weights for AUDIT-P

The majority of the variants (63.4% of 17,549,750 variants) did not fall into any significant annotation. Therefore, these variants were not informative for empirical weights construction and received a weight of zero regardless of frequency. This is in contrast to the default weighting of SKAT which assigns weights based on MAF regardless of functional information. However, 36.6% of variants fell into at least one of the annotation classes. The resulting quantitative score based on combining the 12 significantly enriched annotations ranged from 0 to 71.23. Figure 33 shows the distribution of the empirical functional weights generated for all observed variants by allele frequency bin. Default SKAT weights are based on MAF, and therefore rare variants have higher weights by design. While MAF information is not incorporated in empirical functional weights, we still observe an enrichment of low frequency variants with higher weights. This observation is in agreement with the expectation that coding variation in the genome is rare, recent, and deleterious (Zuk et al., 2014).

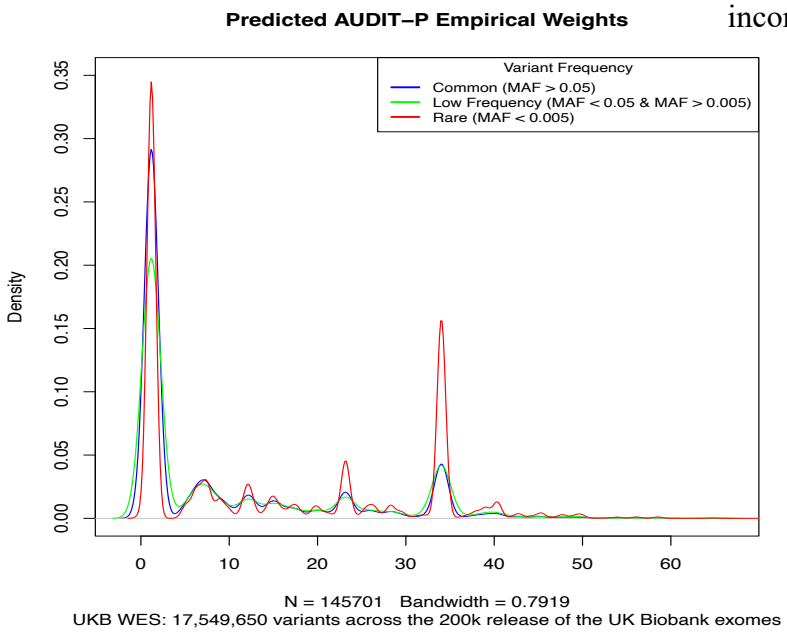


Figure 33: Distribution of empirical functional weights across the UK Biobank exome dataset. The majority of the variants (63.4%) received the weight of zero. the quantitative weights ranged from 0 to 71.23.

Interval-based tests

Post filtering, interval-based analyses were performed using 17,549,650 qualifying variants that mapped onto 21,105 intervals. Figures 34 and 35 show SKAT-O interval-based results using subjects with directly measured (N=51,357) and predicted (measured and unmeasured, N=147,376) AUDIT-P, respectively. No interval was significantly associated with measured AUDIT-P using the default (Figure 34A) or the empirical functional weights (Figure 34B). However, as shown in the QQ-plot (34C), empirical functional weights provide an improvement in statistical significance compared to default weights with no evidence of inflation when using empirical versus default weights. SKAT results for measured AUDIT-P are provided in the Extended Figure 1.

In contrast to using directly measured AUDIT-P, we observed significant associations with *ADH1C* and *THRA* genes when using both default (Figure 35A) and empirical functional weights (Figure 35B). A two-fold increase in statistical significance was observed when using empirical functional weights compared to default weights (Figure 35C), with the most significant association found with predicted AUDIT-P and empirical weights in the *ADH1C* gene after a conservative Bonferroni correction (SKAT-O $P_{\text{Default}} = 1.06 \times 10^{-9}$ and $P_{\text{Empirical weight}} = 6.25 \times 10^{-11}$). SKAT results for predicted AUDIT-P are provided in the Extended Figure 2. The *ADH1C* gene interval which encodes class I alcohol dehydrogenase gamma subunit has been previously implicated in alcohol-related phenotypes. Of the 670 markers in the *ADH1C*, 43 had MAF < 0.01 and were tested in the analysis. In contrast, the *THRA* gene interval encodes for one of the receptors for thyroid hormone and mutations in this gene are associated with intellectual disability and reduction in brain size (Krieger et al., 2019). Of the 1,245 markers in *THRA*, 54 had MAF <0.01 and were tested in the analysis.

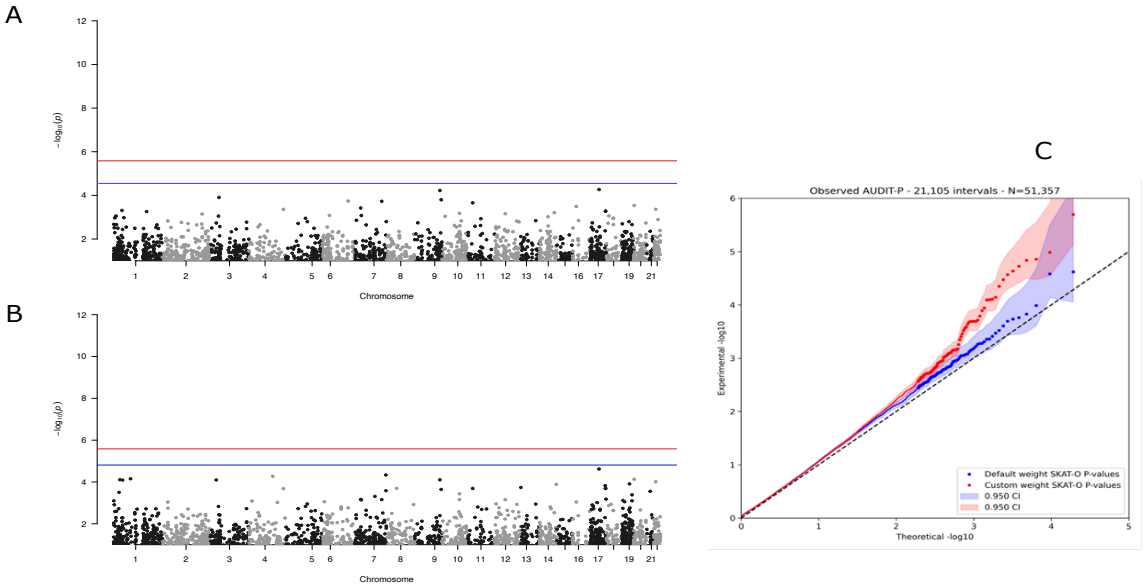


Figure 34: SKAT-O analysis on observed AUDIT-P using default (A) and empirical (B) weights. QQ plots for default and empirical weights are shown in panel (C). Red line represents Bonferroni and blue line represents FDR threshold of 5%.

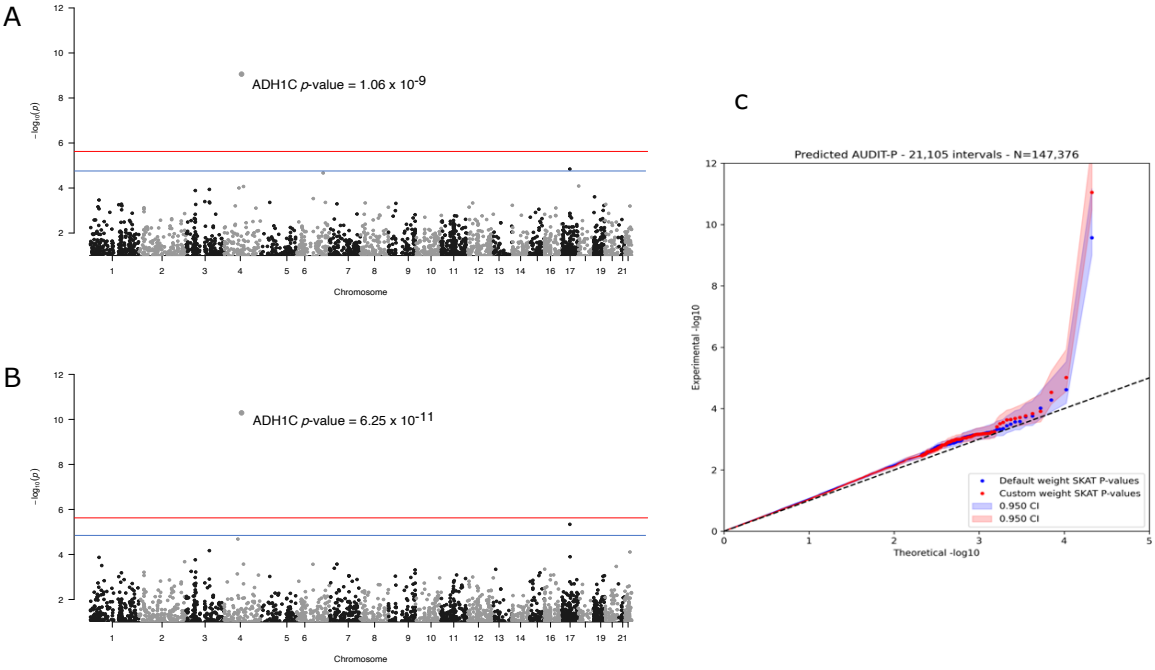


Figure 35: SKAT-O analysis on predicted AUDIT-P using default (A) and empirical (B) weights. QQ plots for default and empirical weights are shown in panel C. Red line represents Bonferroni and blue line represents FDR threshold of 5%.

6.5 Discussion

In this chapter, we sought to perform an exome-wide rare variant analysis on potential alcohol problems using the AUDIT-P subscale in the UK Biobank sample. While common variant GWAS of alcohol consumption have successfully identified many common variant loci, studies of problematic alcohol use and alcohol use disorder have been less successful (Sanchez-Roige et al., 2019; Zhou et al., 2020), and large-scale rare variant analyses have yet to identify robustly associated variants, genes, or intervals (Curtis, 2022). To address this, we used ML phenotype prediction in the UK Biobank to increase the effective sample size for AUDIT-P for rare variant association testing to perform the largest exome-wide analysis of AUDIT-P to date. Additionally, while default SKAT weights are based on the frequency of the variants and thus not disorder specific, we incorporated evidence from common variant GWAS data and functional information as *a priori* weights to provide disorder specific weights for rare variant interval testing. Our findings show that in addition to common risk variation (Zhou et al., 2020), rare risk variation in the *ADH1C* gene is also associated with AUDIT-P in the UK Biobank exome dataset, thus providing evidence for the involvement of rare risk variation in the genetic architecture of AUDIT-P at the population level.

In our first-pass analysis, we used the subset of the UK Biobank cohort with measured AUDIT-P (N=51,357) and found no significant association with rare variation using default or empirical functional weights. While we were able to demonstrate that including disorder-specific empirical functional weights appears to increase the statistical power, we note that these *a priori* weights were insufficient to rescue an underpowered phenotype for rare variant association testing. Thus, larger sample sizes are required to adequately increase statistical power for rare variant identification in AUDIT-P. Conversely, by leveraging ML phenotype prediction to

increase the effective sample size for rare variant association testing, we show that *ADH1C* and *THRA* genes are robustly associated with AUDIT-P in the UK Biobank at rare variation level. Furthermore, we were able to demonstrate that inclusion of disorder-specific empirical functional weights generated from common variant GWAS data and functional genomics information, can also refine, and increase the statistical significance for rare variation testing. These findings highlight the successful leveraging of ML to substantially increase effective sample size to improve statistical power, and disorder-specific empirical functional weights to refine and increase the statistical significance in interval-based rare variant testing.

The *ADH1C* gene plays an important role in the hepatic and gastrointestinal catabolism of ethanol (Edenberg & McClintick, 2018) and variants in this gene are shown to be robustly associated with both alcohol use disorder diagnosis and AUDIT at common variation level (Gelernter et al., 2014; Shen et al., 1997; Walters et al., 2018, Biernacka et al., 2013; Kranzler et al., 2019; Sanchez-Roige et alk., 2019). Therefore, our finding builds on previous common variant evidence for the involvement of the *ADH1C* gene in the genetic architecture of alcohol-related phenotypes by demonstrating that rare variation in *ADH1C* is also associated with alcohol problems at rare variation level. Similar to other psychiatric disorders such as schizophrenia (Singh et al., 2022), these results also suggest that there is a convergence between common and rare variant signals in the genetic architecture of AUDIT-P, and as sample sizes continue to increase, we can expect to identify more rare variant signals in previously implicated genes from common variant studies of alcohol-related phenotypes. In contrast to *ADH1C*, the *THRA* gene, which encodes the thyroid hormone receptor alpha, has not been previously implicated in alcohol-related phenotypes. Thyroid hormone deficiency during pregnancy is a common cause of intellectual disability (Bath et al., 2013) as well as neurocognitive deficits and reduction in

cerebellar volume and decreased white matter density in adults (Krieger et al., 2019). A recent study conducted in the UK Biobank has also demonstrated that increase in alcohol consumption is associated with decrease in global brain volume measurements as well as white matter microstructure (Daviet et al., 2022). Therefore, our finding suggests that rare variation in thyroid hormone receptors encoded by the *THRA* gene could be involved in the development of potential alcohol problems.

Applying the MAGIC-LASSO to predict unmeasured AUDIT-P outcomes increased the sample size from 51,357 to 147,376, representing a 129% increase in effective sample size, after accounting for the phenotypic correlation between the observed and predicted AUDIT-P scores. Importantly, these represent AUDIT-P scores for 96,019 independent subjects with exome variant data who could otherwise not have been included in this analysis without ML predicted scores. While ML-predicted outcomes are not without error, we have demonstrated an approach for efficiently and reliably predicting missing outcomes to maximally leverage measured exome data for rare variant study of a phenotype such as AUDIT-P in the UK Biobank.

Predicting the likely functional impact of variation in the genome is challenging, as it requires taking appropriate account of the different kinds and levels of prior evidence of the likely function for every position. Previous analyses (Gusev et al., 2014; Pickrell, 2014) have shown that joint modeling of available functional information can improve power to detect putative causal variants in common variant studies across complex traits. In this analysis, we hypothesized that based on the evidence for the convergence of common and rare variant signals in the genetic architecture of complex psychiatric traits, similar approaches can be utilized to also improve signal detection in rare variant studies of complex psychiatric traits.

To achieve this, we first analyzed predicted AUDIT-P GWAS data by partitioning its heritability into functional sequence classes using stratified LDSC which accounts for the overlap among different functional classes. Enrichment values from statistically significant classes were then used to determine the overall variant weight at each position in the exome, by simultaneously accounting for the SNP-based heritability of common variant signals conferring risk to AUDIT-P, as well variant membership in each of the significantly enriched functional classes. We hypothesized that these disorder-specific weights would outperform default SKAT weights which assigns the weights to the variants based on the MAF without taking *a priori* functional information into consideration.

Our findings show that by leveraging common variant GWAS results specific to the phenotype to be used in a rare variant investigation as well as functional information, we can refine and increase the statistical significance. In this case, the application resulted in discovering two independent gene intervals containing rare variants influencing alcohol problems as measured by AUDIT-P. Additionally, of the four custom annotation sources acquired from the psychENCODE Consortium, three were significantly enriched for the heritability of AUDIT-P. This finding demonstrates that by going beyond default annotation classes from LDSC and including custom annotation classes such as enhancer and acetylation marks from the psychENCODE Consortium, we can provide more specific weights that can further increase detection power. While these findings can be seen as a proof of principle for the utility of including functional information as *a priori* weights in rare variant testing, given that most of the functional variation in the genome lies outside of the coding regions of the genome (Auton et al., 2015), future work could use whole-genome sequencing data to further explore the utility and

feasibility of using empirical functional weights in non-coding regions of the genome. Furthermore, while AUDIT-P is moderately heritable, empirical functional weights derived from more heritable disorders such as schizophrenia (Trubetskoy et al., 2022) may show better predictive power for rare variant identification than AUDIT-P. Therefore, we could also explore the utility of empirical functional weights in more heritable disorders.

The analyses presented in this chapter should be interpreted in the context of some limitations. First, there are currently no large-scale samples for rare variant studies of alcohol-related phenotypes which motivated us to use ML to predict AUDIT-P in the full 200k exome release of the UK Biobank for rare variant analysis. While we note that our predicted AUDIT-P shows strong genetic correlation with measured AUDIT-P in the UK Biobank (r_G=0.91), replication of the results presented in this study in other adequately powered cohorts is important. Second, due to sample size and power limitations, the current analysis utilized interval-based rare variant testing and thus, we did not conduct single marker tests. As larger sample sizes become available, it is important to extend these analyses to also perform single marker tests on AUDIT-P. Third, while our empirical functional weighting scheme shows improvement over the default SKAT weights, we note that future work could focus on refining the weights and applying it to WGS data and other phenotypes with higher heritability. Fourth, the current analysis was limited to the European subset of the UK Biobank exome data release. As larger, more ancestrally diverse samples with adequate power for rare variant testing become available, future studies should conduct rare variant testing on AUDIT-P in under-represented and diverse populations.

In conclusion, in this study, we show that rare variation in the *ADH1C* and *THRA* genes are significantly associated with AUDIT-P in the UK Biobank. These results suggest that

leveraging ML phenotype prediction and empirical functional weights can help increase effective sample size and subsequent discovery power for rare variant association testing in underpowered phenotypes such as AUDIT-P in large-scale Biobank samples. As sample sizes continue to increase, future directions of this work include improvement of the empirical functional weights and conducting these analyses in the non-coding regions of the genome using WGS data.

CHAPTER VII

Global Discussion

7.1 Summary

The preceding chapters presented in this dissertation describe a series of analyses at common variant level on schizophrenia, while the final chapter focused on a stand-alone rare variant analysis of AUDIT-P in the UK Biobank. Although not comprehensive, we highlighted several important gaps in the current literature and described our attempts to address them. In the first chapter, we provided an overview of the current genomics landscape of schizophrenia as of 2022. In second, third, and fourth chapters, we described three different analyses conducted in members of multiplex schizophrenia families. First, we showed that common risk variation as indexed by current PRS from PGC3-SCZ cannot account for the higher recurrence risk of schizophrenia in members of multiplex schizophrenia families. Next, we showed that while members of multiplex schizophrenia families appear to have a significantly increased PRS for bipolar disorder and major depressive disorder, two psychiatric disorders displaying high genetic correlation with schizophrenia, the source of this increased polygenic risk appears to be due to part of the polygenic risk that bipolar disorder and major depressive disorder share with schizophrenia due to their strong genetic correlation. Finally, we took this observation further and explored the shared genetic architecture of these three highly correlated psychiatric disorders at symptom level in multiplex schizophrenia families and showed that negative and disorganized symptoms in psychotic, and negative symptoms in non-psychotic members of multiplex schizophrenia are specifically associated with schizophrenia PRS, while no association was observed with PRS constructed using bipolar disorder or major depressive disorder GWAS.

In the fifth chapter, we went beyond multiplex families and used the summary statistics data from PGC3-SCZ and IMSGC-MS to compare and contrast the common variant genetic

architecture of schizophrenia and multiple sclerosis. We showed that while there is no significant genetic correlation between these two disorders, there is a substantial polygenic overlap between them that can be used for downstream analyses. By leveraging this overlap, we identified 16 novel schizophrenia and 20 novel multiple sclerosis loci, while an additional 36 genomic loci were also identified to be jointly associated with both.

Finally, in the sixth chapter, we performed a stand-alone rare variant analysis on AUDIT-P using the whole-exome data from the 200k release of the UK Biobank dataset and showed that by using ML phenotype prediction and empirical functional weights, we can increase the effective sample size and refine the statistical significance of rare variant association testing in AUDIT-P and showed that the *ADH1C* gene is significantly associated with AUDIT-P in the UK Biobank at rare variation level.

In the subsequent sections below, we frame these analyses in a unified context by first describing their implications, followed by the limitation and future directions on how to further explore these questions in more detail.

7.2 Common variant analysis of multiplex schizophrenia families

In the first three chapters, we performed extensive common variant analysis of multiplex schizophrenia families to answer three key questions. While sporadic cases are generally considered to be the norm for complex traits, including schizophrenia (Yang et al., 2010), familial cases represent an interesting exception. Individuals ascertained from multiplex families such as the ISHDSF sample are distinct because they represent the upper bounds of the recurrence risk of schizophrenia in the population. Although FH remains the strongest risk factor for developing schizophrenia (Walder et al., 2014), only ~1/3 of sporadic schizophrenia cases have a positive FH of a psychotic disorder (Käkelä et al., 2014). We showed that unlike familial

bipolar disorder cases (Andlauer et al., 2021), familial schizophrenia cases do not appear to have a significantly increased PRS for schizophrenia compared to ancestry-matched sporadic cases from the population. While this finding did not tell us what the source of increased recurrence risk of schizophrenia in multiplex families is, it suggests that the higher recurrence risk of schizophrenia in multiplex families is unlikely to be due to an increased burden of common risk variation as indexed by current PRS (PGC3-SCZ in 2022). Additionally, we were able to show that all members of our multiplex schizophrenia families, including the unaffected relatives in the families, have an increased PRS for schizophrenia compared to ancestry-matched population controls. This finding also further validated the hypothesis of a genetically influenced psychosis spectrum by showing the continuous increase of schizophrenia PRS from unaffected relatives in the families to the familial cases which includes our core schizophrenia and poor-outcome schizoaffective cases.

Next, we extended our PRS profiling in the ISHDSF sample to investigate whether members of our multiplex schizophrenia families also have an increased PRS for bipolar disorder and major depressive disorder. The main motivation behind this work was two important observations. First, epidemiological studies consistently suggest that there is a high aggregation of disorders on the psychosis spectrum (besides schizophrenia) in multiplex schizophrenia families (Kendler et al., 1993). Second, cross-disorder analyses of psychiatric disorders such as the ones conducted by the PGC (Lee et al., 2021), show that there is a strong genetic correlation between schizophrenia and bipolar disorder (r_G=0.67), and schizophrenia and major depressive disorder (r_G=0.35). While cross-disorder PRS profiling across correlated psychiatric disorders have been performed in other samples, most of these analyses have been conducted in case-control samples and the source of these increased cross-disorder PRS across psychiatric disorders

remains unclear. To address this gap in the literature, we used our multiplex schizophrenia family sample (N=1,005 with genotype data) in combination with state-of-the-art statistical methods such as genomicSEM (Demange et al., 2021; Grotzinger et al., 2019) and pTDT (Weiner et al., 2016) to identify the source of increased cross-disorder PRS in the ISHDSF sample. First, we showed that similar to schizophrenia PRS, members of multiplex schizophrenia families also have an increased PRS for bipolar disorder and major depressive disorder, with the highest PRS loadings observed in the broad case definition, and very-broad case definition, respectively. The bulk of our bipolar disorder cases in the ISHDSF are in the broad case definition, while all of our major depressive disorder cases are in the very-broad case definition. Therefore, these results show that as expected, the highest loadings of bipolar disorder and major depressive disorder polygenic risks in our sample are observed in case definitions that have the highest number of bipolar disorder and major depressive disorder cases, respectively. Our univariate PRS profiling the ISHDSF sample was further supported through our pTDT analyses where we showed that PRS for schizophrenia, bipolar disorder, and major depressive disorder are significantly over-transmitted from parents to probands in the families, suggesting that offspring in multiplex schizophrenia families have, on average, higher polygenic risks for these three correlated psychiatric disorders compared to their parents which signifies over-transmission of risk for these disorders in the ISHDSF, while no over-transmission of LDL PRS was observed.

However, while interpretation of increased schizophrenia PRS in a sample with high incidence of schizophrenia is straightforward, interpretation of cross-disorder PRS, particularly in multiplex families, is challenging. Due to the pervasive genetic pleiotropy that exists in the genome, it is difficult to determine whether the increased polygenic risk we observe for these

two correlated disorders is due to the portion of the polygenic risk that these two disorders share with schizophrenia due to genetic correlation, or the "affective" portion of the polygenic risk that is unique to them and not shared with schizophrenia. To answer this question, we used genomicSEM to parse out the signals from univariate bipolar disorder and major depressive disorder into underlying latent genetic factors that 1) capture the portion of the polygenic risk that these two disorders share with schizophrenia due to genetic correlation, and 2) the affective portion of the genetic risk that is unique to them. After validating the newly generated polygenic signals from genomicSEM using downstream analyses, we were able to show that the source of this increased cross-disorder polygenic risk in members of our multiplex schizophrenia families is due to part of the polygenic risk that bipolar disorder or major depressive disorder share with schizophrenia due to genetic correlation. While this observation agreed with our initial hypothesis, we note that this is the first empirical evidence supporting this observation. These findings suggest that given the pervasive pleiotropy across the genome, it is important to pay close attention to the relative contribution of the shared and unique genetic components of polygenic risks across correlated psychiatric disorders for a complete interpretation of cross-disorder PRS profiling. We further replicated these findings in a sample of independent sporadic schizophrenia from the ISGC sample and showed that these findings can be extended to samples beyond multiplex families.

The ISHDSF sample has detailed symptom level information available across all case definitions on the extended psychosis spectrum, including the unaffected relatives. Upon establishing that members of multiplex schizophrenia families have an increased polygenic risk for schizophrenia, bipolar disorder, and major depressive disorder, we sought to examine the association of the polygenic liability to these three disorders and symptom dimensions across the

ISHDSF sample. While previous symptom level analyses of schizophrenia have mostly focused on case-control designs (Derks et al., 2012; Smigielski et al., 2021), we used our multiplex schizophrenia family sample to investigate these associations in a more detailed manner. What made our analysis stand out from previous case-control designs, is the availability of symptom measurements not only in psychotic subjects in the families, but even in family members without a diagnosis of a psychotic disorder. While subjects with a diagnosis of a psychotic disorder (narrow and intermediate case definitions in the ISHDSF sample) were given the OPCRIT, a strength of our analysis was that subjects without a diagnosis of a psychotic disorder (broad, very broad, and unaffected relatives in the ISHDSF sample) were given the SIS to assess their symptoms on the extended psychosis spectrum.

A major question about the clinical heterogeneity that we observe in schizophrenia is the extent to which clinical features and differences observed in cases are attributable to genetic differences. Through this analysis, we showed that similar to previous reports (Bigdeli et al., 2017; Fanous et al., 2012), polygenic liability to schizophrenia is associated with increased negative/disorganized symptoms in subjects with a diagnosis of a psychotic disorder. Additionally, we showed that polygenic liability to schizophrenia is also associated with negative symptoms in non-psychotic subjects in the families. Together, these findings across the extended psychosis spectrum provided genetic evidence for the spectrum model of schizophrenia at symptom level and corroborated previous epidemiological findings that show negative and disorganized symptoms are likely to have a greater genetic basis than positive symptoms, making them better indices of familial liability to schizophrenia.

7.3 Polygenic overlap between schizophrenia and multiple sclerosis

While in the third and fourth chapters we investigated the shared genetic architecture of schizophrenia with two correlated psychiatric disorders (bipolar disorder and major depressive disorder), we took a different approach in the fifth chapter. Most of the cross-disorder analyses of complex disorders have largely focused on disorders with observed genetic correlation. As discussed in the first chapter, newly developed methods such as MiXeR and condFDR (Andreassen et al., 2013; Frei et al., 2019) methods provide a framework to extend cross-disorder analysis of complex disorders beyond those with significant genetic correlation. In this chapter, we extended our cross-disorder analyses of schizophrenia by examining the polygenic overlap between schizophrenia and multiple sclerosis beyond genetic correlation. Prior joint analyses of schizophrenia and multiple sclerosis have largely focused on either global genetic correlation measurements, or specifically on variants inside the MHC region. Here, we hypothesized that because of prior evidence for immune dysregulation in schizophrenia patients and co-occurring neuropsychiatric symptoms in multiple sclerosis patients, we may observe polygenic overlap between these two disorders that can be exploited for further downstream analyses. In agreement with prior work, we detected no significant genetic correlation between schizophrenia and multiple sclerosis by LDSC analysis and showed that this observation appears to be attributable to the balance of shared and opposite effect directions detected in both bivariate MiXeR and conjFDR analyses. Our results further suggested that the increased rate of psychotic-like features in some multiple sclerosis patients may be partially attributable to the genetic loci shared between schizophrenia and multiple sclerosis that have concordant direction of effects. Finally, our results suggest that despite substantial increase in GWAS sample sizes in recent years, a considerable growth in sample sizes are still required to increase our understanding of the shared

genetic etiology among complex, polygenic disorders such as schizophrenia and multiple sclerosis at common variant level. As shown in the preceding chapters, although genetic correlation estimates and PRS profiling are the dominant methods for cross-disorder analysis of complex disorders, state-of-the-art methods used in this chapter showed that complex disorders with no direct genetic correlation but purported biological dysfunction provides new avenues for analyzing the shared genetic architecture of schizophrenia with other disorders. Therefore, we can expect to see an increase in cross-disorder analyses of complex psychiatric disorders beyond PRS profiling or genetic correlation studies.

7.4 Rare variation in the genetic architecture of complex traits

The analyses described in the second to fifth chapters focused on common variation in the genome while rare variants were omitted. In the sixth chapter, we expanded our work to also investigate the rare variant architecture of complex disorders. The analysis described in the fifth chapter were initially planned to be carried out on schizophrenia using the WGS data of ISHDSF and ISGC samples. However, due to various delays with our collaborators, we have not yet received the WGS data to conduct these analyses. As a result, instead we opted to use the available UK Biobank data to investigate the rare variant architecture of AUDIT-P. While the analysis described in the sixth chapter may seem disjointed from the preceding chapters, we note that the pipeline we described there is applicable to schizophrenia as well. First, we will discuss the implications of the results in this chapter, and we will follow that by a discussion on how these results could be applied to WGS data for schizophrenia.

We described the largest rare variant study of AUDIT-P in the UK Biobank as of 2022. We used ML phenotype prediction and empirical functional weights to increase the sample size and statistical power for rare variant interval testing and showed that while current sample sizes

are insufficient for rare variant study of AUDIT-P in the UK Biobank, we can use ML phenotype prediction to improve effective sample size for rare variant discovery with meaningful associations. Additionally, we showed that by including disorder specific functional information as a priori weights in interval testing, we can refine and increase the statistical significance of rare variant testing for AUDIT-P. While genes in the ADH cluster (such as *ADH1C*) have been robustly associated with alcohol phenotypes at common variation level across different studies, this study provide evidence for the first time that rare variation in the *ADH1C* gene is associated with AUDIT-P in the UK Biobank. While most of the improvement in discovery power came from increased effective sample size, here we will focus on the impact of including empirical functional weights in rare variant analysis of AUDIT-P. This is because the portion of the work related to ML phenotype prediction was carried out by Dr. Gentry (Assistant Professor of Psychiatry at VCU), while the construction of empirical functional weights was specifically developed and implemented into the pipeline by me as part of this dissertation.

Important studies have shown that inclusion of functional information can improve discovery power for common variants associated with complex traits (Gusev et al., 2014; Pickrell, 2014). Here, we hypothesized that the same framework can be applied to rare variant discovery for complex traits. This hypothesis relied on two important observations: Recent studies suggest that there is a convergence between common and rare variant signals in the genetic architecture of complex psychiatric disorders. Recent sequencing study of schizophrenia from the SCHEMA Consortium (Singh et al., 2022) showed that genes identified to be robustly associated with schizophrenia using WES data were enriched for common variant signals from PGC3-SCZ (Trubetskoy et al., 2022). An interesting pair of examples from schizophrenia studies are the *GRIN2A* and *SP4* genes. *GRIN2A* encodes one of the secondary subunits of the

glutamatergic NDMA receptor, while the *SP4* gene is a transcription factor that is highly expressed, and regulated by NMDA transmission and NMDA receptor abundance, in the brain. Common variant signals in these two genes have been previously associated with schizophrenia. The SCHEMA Consortium provided evidence for the first time that rare variation in this gene is also involved in the genetic architecture of schizophrenia. Another important observation from common variant studies of complex traits is that many identified signals lie in functional regions of the genome that impact gene regulation and expression. This observation has led to the development of fine-mapping methods such as PolyFun (Weissbrod et al., 2020) which performs functionally informed fine-mapping that uses functional information as a-priori evidence for detecting the likely causal variants through fine-mapping with better accuracy. We hypothesized that by combining information from common variant GWAS data and functional information from various sources such as ENCDOE (Encode Consortium., 2012) and psychENCODE (Akbarian et al., 2015), we may be able to increase our statistical power for rare variant identification. While this framework is likely to work more robustly for identification of rare variant in non-coding regions of the genome, using empirical functional weights as a-priori evidence in our rare variant analysis of AUDIT-P in the UK Biobank, we were able to increase the statistical significance for our interval-based testing by around 2-fold.

The initial plan was to apply this framework to rare variant study of schizophrenia. While we did not get access to our schizophrenia sequence data in time to conduct that analysis, here we provide our preliminary stratified LDSC analysis results applied to PGC3-SCZ GWAS data (Figure 35). Our analysis indicated that 27 functional categories were significantly enriched for the heritability of schizophrenia.

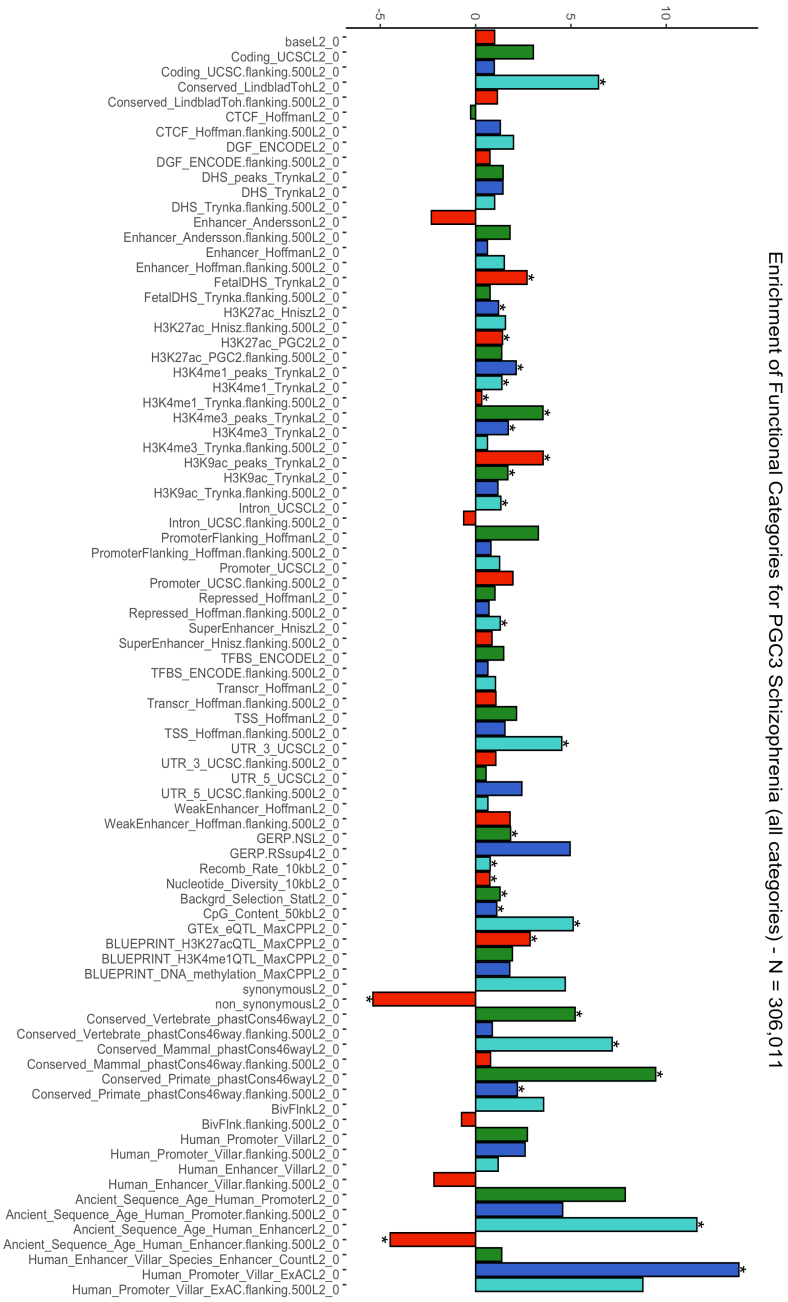


Figure 36: Partitioning the heritability of PGC3 schizophrenia GWAS into functional categories. Asterisk represents significant categories after multiple testing correction.

As shown in Table 20, many cell-specific and tissue-specific functional classes acquired from psychENCODE were shown to be significantly enriched for schizophrenia heritability. This suggests that our framework for functional annotation and empirical weighting of variants using common variant GWAS results is likely to improve the power of rare variant studies of schizophrenia. Furthermore, functional studies of schizophrenia are much more comprehensive compared to AUDIT-P, which will allow us to use a more comprehensive set of functional information to provide more accurate weights for a well-studied disorder like schizophrenia. We anticipate that upon receiving the WGS data for ISHDSF and ISGC samples, our lab will test the utility of this framework using our schizophrenia WGS data.

Table 20: Significant annotation classes enriched in the heritability of PGC3 schizophrenia after multiple testing correction. Grey categories are custom classes.

Annotation Category	S-LDSC-Enrichment	LDSC-Enrichment *p*-value
Girdhar H3K27ac NeuN+	2.87	3.11E-24
Girdhar H3K27ac NeuN-	2.13	1.81E-22
Girdhar H3K4me3 NeuN+	11.07	3.37E-14
Conserved Primate phastCons46Way	11.71	3.67E-09
Conserved LindBlad-Toh	8.99	2.33E-08
CpG Content	1.11	1.17E-07
Conserved Vertebrate PhastCons46Way	7.29	2.64E-07
H3K27ac Hnisz	1.27	1.45E-05
Human promoter Villar	16.21	5.28E-05
Ancient Sequence Age Human Enhancer	12.4	5.51E-05
H3K4me3 Trynka	2.11	1.20E-04
GTEx eQTL MaxCPP	4.87	1.40E-04
Conserved Primate phastCons46Way flanking	2.21	2.40E-04
UTR 5 UCSC flanking	3.95	1.90E-03
Fetal DHS Trynka	3.04	2.40E-03
H3K27ac PGC2	1.41	2.70E-03
UTR 3 UCSC	4.55	3.40E-03
DGF ENCODE	2.24	7.30E-03
Coding UCSC	3.91	8.40E-03
psychENCODE Temporal Cortex H3K27ac	4.15	9.00E-03
H3K4me1 peaks Trynka	1.82	9.30E-03
psychENCODE Enhancers	3.57	1.10E-02
psychENCODE Cerebellum H3K27ac	4.6	1.40E-01
Blueprint h3k27acQTL MaxCPP	2.65	1.70E-02
TFBS ENCODE	1.81	2.60E-02
Girdhar H3K4me3 NeuN-	3.65	3.30E-02
H3K4me3 peaks Trynka	2.76	3.80E-02

7.5 Limitations

The results presented in each of the preceding chapters must be interpreted in the context of several limitations. While the analyses in each chapter can be seen as stand-alone work and a more comprehensive description of the limitation of each analysis is provided separately at the end of each chapter, here we attempt to describe some of the major limitations of complex traits

genetics studies that require careful attention. We used PRS profiling method to index common variation burden of complex disorders in the ISHDSF and ISGC samples. While PRS is widely used for this purpose, there are several limitations that are inherent in current PRS construction methods, and these results in this dissertation should be interpreted in that context. For example, while schizophrenia GWAS is currently the strongest GWAS of any psychiatric disorder, current PRS of schizophrenia using PGC3-SCZ GWAS (as of 2022) can account for only ~2.6% of the variance in schizophrenia, suggesting that PRS predictive power is still very limited. Additionally, some known genetic risk factors for schizophrenia, such as indels or rare variation, are not captured by current PRS methods. While it is reasonable to predict that PRS methods will become more sophisticated in the coming years, most of the complex psychiatric disorders also have strong environmental influences that are not going to be captured by polygenic indexing. Furthermore, although more sophisticated methods such as PRS-CSx (Ruan et al., 2022) have been developed that allows for application of PRS profiling in diverse populations, most of the well-powered GWAS of complex traits are conducted in individuals of predominantly European ancestry. With recent efforts to increase both the continental diversity of GWAS samples and, as a result, the generalizability of association signals and PRS to other ancestral backgrounds, it is also reasonable to expect that well-powered GWAS in diverse ancestries will become available. But as of 2022, this is still a major limitation of current common variant studies of complex traits.

Another important limitation in the current complex trait genetics landscape is the lack of cross-disorder analysis frameworks at rare variation level. In one of the analyses described in this dissertation, we used state-of-the-art methods to conduct a cross-disorder analysis of schizophrenia and multiple sclerosis, but that analysis was limited to common variation in the

genome. While study of rare variation in the genome is inherently more difficult than common variation due to weaker LD among rare variants, we can expect that as the cost of sequencing continues to drop, sequencing the whole genome for rare variant analysis is going to become more feasible. Therefore, another reasonable expectation is that in the coming years, it is likely that we will see more integration of common and rare variant studies that will pave the way for the development of more sophisticated cross-disorder methods that allow for lower frequency variants that are captured through sequencing studies to be used in cross-disorder analyses. For example, the recently developed Burden Heritability Regression (BHR) method extends heritability and genetic correlation estimates from common variant GWAS data to rare variation in the genome (Weiner et al., 2022). Applying the BHR method to 400,000 UK Biobank exomes shows that rare coding variation in the exome can only explain ~1.3% of the phenotypic variance of complex traits on average, which is a much less estimate than the contribution of common variants from GWAS data. It was further shown that burden heritability (MAF < 0.01) of complex traits is also strongly concentrated in constrained genes. It was further shown that burden genetic correlations computed using BHR generally conform to common variant genetic correlation estimates from LDSC. Taken together, these results suggest that while both common and rare variation in the genome are convergent and have correlated effects across complex traits, rare coding variants in the genome contribute modestly to the missing heritability of complex traits.

7.6 Future Directions

While each of the preceding chapters can be seen as stand-alone analyses and their future directions are specific to them, here we attempt to provide an overview of the general themes discussed through a unified lens and describe possible future directions for complex traits

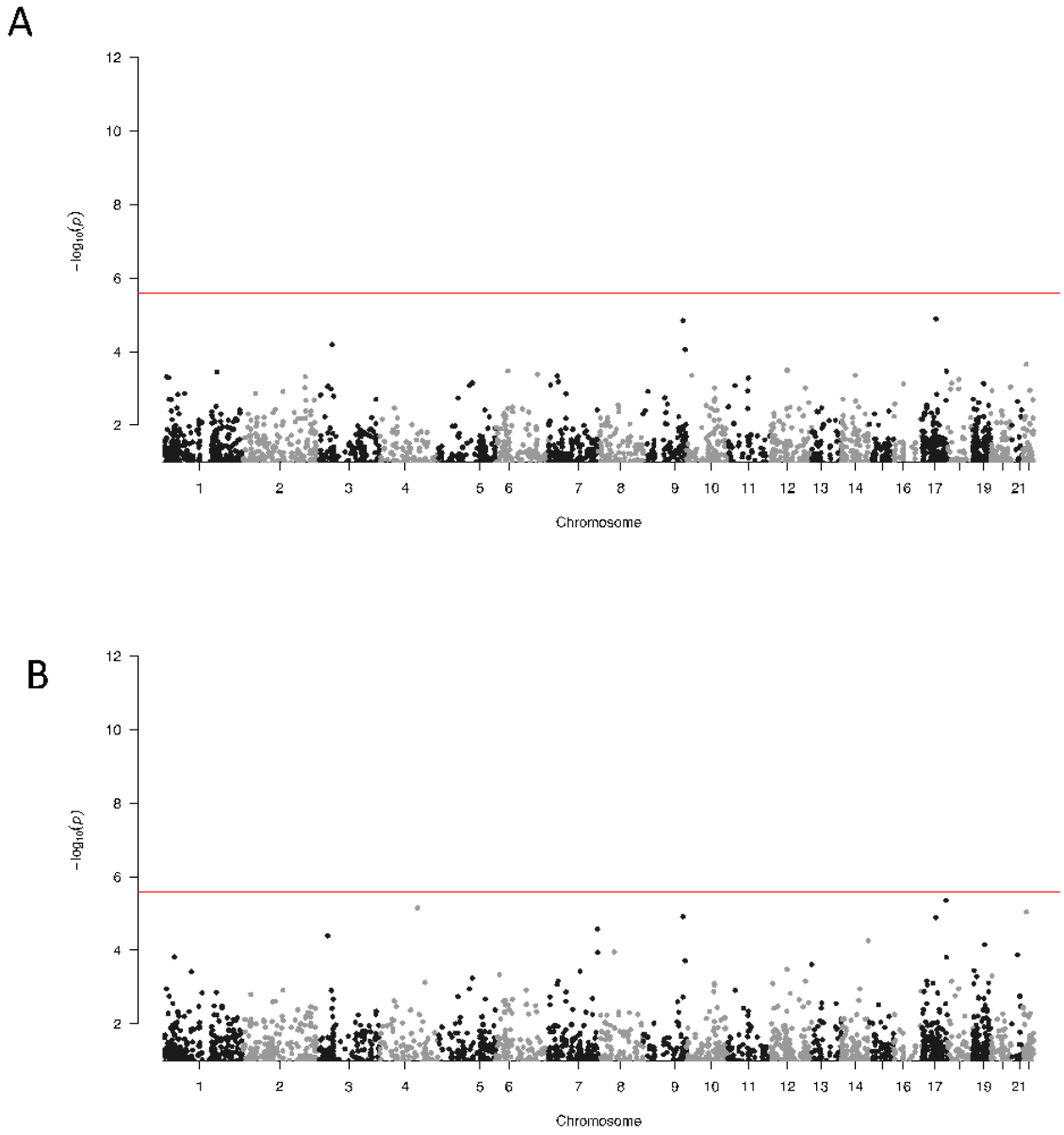
genetics. In the past decade, Genome-wide association and sequencing studies have shown that many variants with a range of effect sizes are involved in the genetic architecture of complex traits. An important takeaway from these studies is that in order to increase the efficacy and statistical power for identification of both common and rare variants, a large boost in sample sizes needs to be achieved. As a result, major genotyping, and sequencing efforts in large-scale, nationwide biobanks such as UK Biobank (Bycroft et al., 2018; Szustakowski et al., 2021), Biobank Japan (Nagai et al., 2017), and All of Us ("The 'All of Us' Research Program," 2019) are actively working to address the need to increase effective sample sizes for genetic studies. While increase in sample sizes will undoubtedly improve our statistical power for variant identification, a limitation of biobanks is that many phenotypes are lightly screened, which limits their interpretation. Another major open question moving forward is the missing heritability problem in complex traits. Recent studies show that many complex traits in humans involve a large number of variants which suggest that the heritability of complex traits may be explained by a joint consideration of both common and rare signals. While thousands of SNPs have been shown to be robustly associated with complex traits through GWAS studies, a consistent theme across all these studies is that individual effects of these SNPs are quite small and common variant GWAS analyses is unlikely to be able to capture the heritability of complex traits with current sample sizes. Different hypotheses such as epigenetics, epistatic, and gene-environment interactions have been proposed as possible sources of the missing heritability problem and recent findings suggest that rare variants with moderate to low frequency that are in weak LD with common variant signals may be a major contributor to the heritability of complex traits. For example, Wainschtein and colleagues were able to capture most of the twin-based heritability estimate of height and BMI by jointly modeling both common and rare variant signals using

WGS data (Wainschtein et al., 2022). In another study, a similar approach was used to capture the heritability of schizophrenia across different MAF bins, and it was shown that a large portion of twin-based heritability of schizophrenia can be captured when variants with low and intermediate frequencies are also included in heritability estimation (Halvorsen et al., 2020). Of note, a recently published paper (Yengo et al., 2022) used 5.4 million individuals from diverse ancestries and showed that they were able to account for the majority of the phenotypic variance (heritability) of height in this highly powered GWAS. This work suggests that it remains to be seen if large-scale sequencing efforts to identify rare variation in the genome, or extremely well-powered GWAS studies with very large sample sizes, will provide better solutions to the missing heritability across all complex traits.

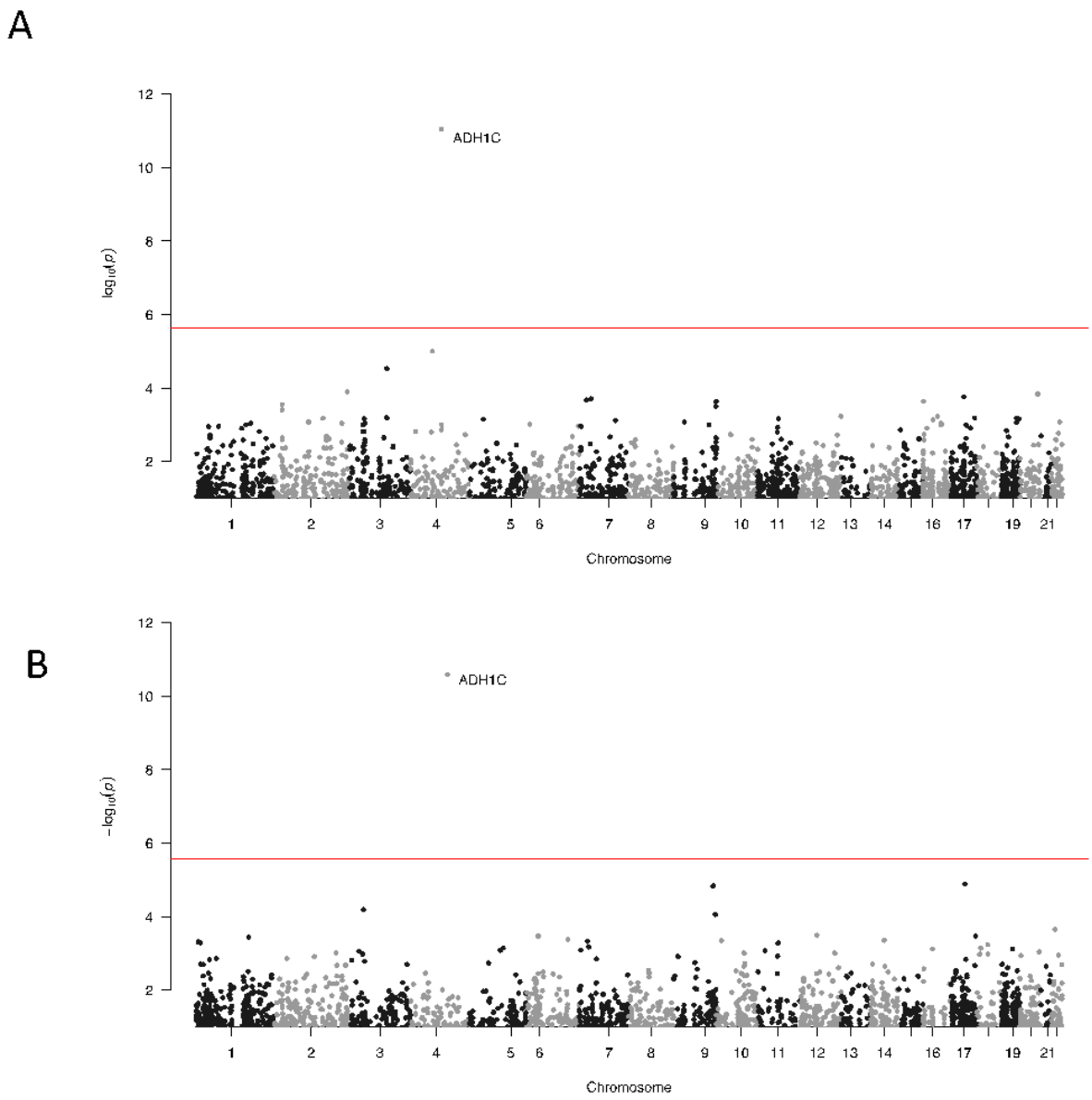
An important point to consider is that conducting large-scale GWAS results using array data may not be a feasible strategy in the long term. Given that genotyping arrays have limited utility for future studies, with the decrease in sequencing cost, it may be reasonable to assume that future studies will likely opt for sequencing strategies instead of array genotyping. There are two reasons for this. First, sequencing provides an agnostic overview of the genetic makeup of an individual while genotyping arrays only tag a set of pre-defined SNPs which are usually population specific. Thus, genetic data generated from sequencing studies will have longevity and future use beyond genotyping arrays. Second, sequencing platforms will probe the whole genome which in turn allows researchers to conduct rare variant studies of complex traits to complement the existing common variant literature. However, although sequencing can identify rare variation in the genome, an important bottleneck in rare variant studies are the challenges in parsing out putative causal signals from normal variation in the genome. We proposed a framework to combine evidence from common variant GWAS data with functional information

to improve rare variant discovery in complex traits and tested our proposed framework using the UK Biobank exome data on AUDIT-P which showed that inclusion of common variant GWAS data signals and functional information as a-priori evidence can increase and refine statistical significance for rare variant identification. We can also expect to see more sophisticated methods that combine common and rare variant data to improve rare variant discovery from sequence data. We also note that a major setback in our rare variant analysis of AUDIT-P was the use of exome data instead of whole-genome data. As described extensively in chapter seven, most of the GWAS signals and functional elements in the genome lie outside the coding regions of the genome and one of our main directions moving forward is to test the effectiveness of this framework for rare variant identification in non-coding regions of the genome.

Extended Figures



Extended Figure 1: SKAT analysis of observed AUDIT-P using the UK Biobank exome data with default (A) and empirical (B) weights.



Extended Figure 2: SKAT analysis of predicted AUDIT-P using the UK Biobank exome data with default (A) and empirical (B) weights.

List of References

Aguet, F., Barbeira, A. N., Bonazzola, R., Brown, A., Castel, S. E., Jo, B., Kasela, S., Kim-Hellmuth, S., Liang, Y., Oliva, M., Flynn, E. D., Parsana, P., Fresard, L., Gamazon, E. R., Hamel, A. R., He, Y., Hormozdiari, F., Mohammadi, P., Muñoz-Aguirre, M., … Volpi, S. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, *369*(6509), 1318–1330. https://doi.org/10.1126/SCIENCE.AAZ1776

Ahangari, M., Everest, E., Nguyen, T. H., Verrelli, B. C., Webb, B. T., Bacanu, S. A., Tahir Turanli, E., & Riley, B. P. (2022). Genome-wide analysis of schizophrenia and multiple sclerosis identifies shared genomic loci with mixed direction of effects. *Brain, Behavior, and Immunity*, *104*, 183–190. https://doi.org/10.1016/j.bbi.2022.06.007

Ahangari, M., Gentry, A. E., Riley, B. P., Morris, D. W., O'Dushlaine, C. T., Cormican, P., Kenny, E. M., Wormley, B., Donohoe, G., Quinn, E., Judge, R., Coleman, K., Tropea, D., Roche, S., Cummings, L., Kelleher, E., McKeon, P., Dinan, T., McDonald, C., … Riley, B. P. (2022). Evaluating the role of common risk variation in the recurrence risk of schizophrenia in multiplex schizophrenia families. *Translational Psychiatry*, *12*(1), 291. https://doi.org/10.1038/s41398-022-02060-3

Aitken, A. (2006). 14-3-3 proteins: A historic overview. *Seminars in Cancer Biology*, *16*(3), 162–172. https://doi.org/10.1016/j.semcancer.2006.03.005

Akbarian, S., Liu, C., Knowles, J. A., Vaccarino, F. M., Farnham, P. J., Crawford, G. E., Jaffe, A. E., Pinto, D., Dracheva, S., Geschwind, D. H., Mill, J., Nairn, A. C., Abyzov, A., Pochareddy, S., Prabhakar, S., Weissman, S., Sullivan, P. F., State, M. W., Weng, Z., … Sestan, N. (2015). The PsychENCODE project. *Nature Neuroscience*, *18*(12), 1707–1712. https://doi.org/10.1038/nn.4156

Ament, S. A., Szelinger, S., Glusman, G., Ashworth, J., Hou, L., Akula, N., Shekhtman, T., Badner, J. A., Brunkow, M. E., Mauldin, D. E., Stittrich, A. B., Rouleau, K., Detera-Wadleigh, S. D., Nurnberger, J. I., Edenberg, H. J., Gershon, E. S., Schork, N., Price, N. D., Gelinas, R., … Roach, J. C. (2015). Rare variants in neuronal excitability genes influence risk for bipolar disorder. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(11), 3576–3581. https://doi.org/10.1073/pnas.1424958112

Anderson, J. M., Hampton, D. W., Patani, R., Pryce, G., Crowther, R. A., Reynolds, R., Franklin, R. J. M., Giovannoni, G., Compston, D. A. S., Baker, D., Spillantini, M. G., & Chandran, S. (2008). Abnormally phosphorylated tau is associated with neuronal and axonal loss in experimental autoimmune encephalomyelitis and multiple sclerosis. *Brain*, *131*(7), 1736–1748. https://doi.org/10.1093/brain/awn119

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., Ntini, E., Arner, E., Valen, E., Li, K., Schwarzfischer, L., Glatz, D., Raithel, J., Lilje, B., Rapin, N., … Sandelin, A. (2014). An atlas of active enhancers across human cell types and tissues. *Nature*, *507*(7493), 455–461. https://doi.org/10.1038/nature12787

Andlauer, T. F. M., Guzman-Parra, J., Streit, F., Strohmaier, J., González, M. J., Gil Flores, S., Cabaleiro Fabeiro, F. J., del Río Noriega, F., Perez, F. P., Haro González, J., Orozco Diaz, G., de Diego-Otero, Y., Moreno-Küstner, B., Auburger, G., Degenhardt, F., Heilmann-Heimbach, S., Herms, S., Hoffmann, P., Frank, J., … Rietschel, M. (2021). Bipolar multiplex families have an increased burden of common risk variants for psychiatric disorders. *Molecular Psychiatry*, *26*(4), 1286–1298. https://doi.org/10.1038/s41380-019-0558-2

Andreasen, N. C., Endicott, J., Spitzer, R. L., & Winokur, G. (1977). The Family History Method Using Diagnostic Criteria: Reliability and Validity. *Archives of General Psychiatry*, *34*(10), 1229–1235. https://doi.org/10.1001/archpsyc.1977.01770220111013

Andreassen, O. A., Harbo, H. F., Wang, Y., Thompson, W. K., Schork, A. J., Mattingsdal, M., Zuber, V., Bettella, F., Ripke, S., Kelsoe, J. R., Kendler, K. S., O'Donovan, M. C., Sklar, P., McEvoy, L. K., Desikan, R. S., Lie, B. A., Djurovic, S., & Dale, A. M. (2015). Genetic pleiotropy between multiple sclerosis and schizophrenia but not bipolar disorder: Differential involvement of immune-related gene loci. *Molecular Psychiatry*, *20*(2), 207–214. https://doi.org/10.1038/mp.2013.195

Andreassen, O. A., Thompson, W. K., Schork, A. J., Ripke, S., Mattingsdal, M., Kelsoe, J. R., Kendler, K. S., O'Donovan, M. C., Rujescu, D., Werge, T., Sklar, P., Roddey, J. C., Chen, C. H., McEvoy, L., Desikan, R. S., Djurovic, S., & Dale, A. M. (2013). Improved Detection of Common Variants Associated with Schizophrenia and Bipolar Disorder Using Pleiotropy-Informed Conditional False Discovery Rate. *PLoS Genetics*, *9*(4). https://doi.org/10.1371/journal.pgen.1003455

Arias, I., Sorlozano, A., Villegas, E., Luna, J. de D., McKenney, K., Cervilla, J., Gutierrez, B., & Gutierrez, J. (2012). Infectious agents associated with schizophrenia: A meta-analysis. *Schizophrenia Research*, *136*(1–3), 128–136. https://doi.org/10.1016/j.schres.2011.10.026

Arnold, P. D., Askland, K. D., Barlassina, C., Bellodi, L., Bienvenu, O. J., Black, D., Bloch, M., Brentani, H., Burton, C. L., Camarena, B., Cappi, C., Cath, D., Cavallini, M., Conti, D., Cook, E., Coric, V., Cullen, B. A., Cusi, D., Davis, L. K., … Zai, G. (2018). Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Molecular Psychiatry*, *23*(5), 1181–1188. https://doi.org/10.1038/mp.2017.154

Asarnow, R. F., Nuechterlein, K. H., Fogelson, D., Subotnik, K. L., Payne, D. A., Russell, A. T., Asamen, J., Kuppinger, H., & Kendler, K. S. (2001). Schizophrenia and schizophrenia-spectrum personality disorders in the first-degree relatives of children with schizophrenia: The UCLA Family Study. *Archives of General Psychiatry*, *58*(6), 581–588. https://doi.org/10.1001/archpsyc.58.6.581

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., Lehrach, H., … Schloss, J. A. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. https://doi.org/10.1038/nature15393

Bahrami, S., Shadrin, A., Frei, O., O'Connell, K. S., Bettella, F., Krull, F., Fan, C. C., Røssberg, J. I., Hindley, G., Ueland, T., Dale, A. M., Djurovic, S., Steen, N. E., Smeland, O. B., & Andreassen, O. A. (2021). Genetic loci shared between major depression and intelligence with mixed directions of effect. *Nature Human Behaviour*, *5*(6), 795–801. https://doi.org/10.1038/s41562-020-01031-2

Baron, M., Gruen, R., Asnis, L., & Kane, J. (1983). Familial relatedness of schizophrenia and schizotypal states. *American Journal of Psychiatry*, *140*(11), 1437–1442. https://doi.org/10.1176/ajp.140.11.1437

Bassett, A. S., Marshall, C. R., Lionel, A. C., Chow, E. W. C., & Scherer, S. W. (2008). Copy number variations and risk for schizophrenia in 22q11.2 deletion syndrome. *Human Molecular Genetics*, *17*(24), 4045–4053. https://doi.org/10.1093/hmg/ddn307

Bath, S. C., Steer, C. D., Golding, J., Emmett, P., & Rayman, M. P. (2013). Effect of inadequate iodine status in UK pregnant women on cognitive outcomes in their children: results from the

Avon Longitudinal Study of Parents and Children (ALSPAC). *The Lancet*, *382*(9889), 331–337. https://doi.org/10.1016/S0140-6736(13)60436-5

Bhati, M. T. (2013). Defining Psychosis: The Evolution of DSM-5 Schizophrenia Spectrum Disorders. *Current Psychiatry Reports*, *15*(11), 409. https://doi.org/10.1007/s11920-013-0409-9

Biernacka, J. M., Geske, J. R., Schneekloth, T. D., Frye, M. A., Cunningham, J. M., Choi, D.-S., Tapp, C. L., Lewis, B. R., Drews, M. S., L.Pietrzak, T., Colby, C. L., Hall-Flavin, D. K., Loukianova, L. L., Heit, J. A., Mrazek, D. A., & Karpyak, V. M. (2013). Replication of Genome Wide Association Studies of Alcohol Dependence: Support for Association with Variation in ADH1C. *PLoS ONE*, *8*(3), e58798. https://doi.org/10.1371/journal.pone.0058798

Bigdeli, T. B., Bacanu, S. A., Webb, B. T., Walsh, D., O'Neill, F. A., Fanous, A. H., Riley, B. P., & Kendler, K. S. (2014). Molecular validation of the schizophrenia spectrum. *Schizophrenia Bulletin*, *40*(1), 60–65. https://doi.org/10.1093/schbul/sbt122

Bigdeli, T., Peterson, R., Ripke, S., Bacanu, S.-A., Amdur, R., Gejman, P., Levinson, D., Riley, B., Clair, D. st., Rietschel, M., Walters, J., Ophoff, R., McQuillin, A., Gurling, H., Rujescu, D., Sullivan, P., Kirov, G., Pato, M., Pato, C., … Fanous, A. (2017). Genome-wide Association Study of Clinical Features in the Schizophrenia Psychiatric Genomics Consortium: Confirmation of Polygenic Effect on Negative Symptoms. *BioRxiv*, 161349. https://www.biorxiv.org/content/early/2017/07/09/161349

Bjornevik, K., Cortese, M., Healy, B. C., Kuhle, J., Mina, M. J., Leng, Y., Elledge, S. J., Niebuhr, D. W., Scher, A. I., Munger, K. L., & Ascherio, A. (2022). Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science*, *375*(6578), 296–301. https://doi.org/10.1126/science.abj8222

Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M., Karczewski, K. J., Park, J., Hitz, B. C., Weng, S., Cherry, J. M., & Snyder, M. (2012). Annotation of functional variation in personal genomes using RegulomeDB. *Genome Research*, *22*(9), 1790–1797. https://doi.org/10.1101/gr.137323.112

Brainstorm Consortium, T. B., Anttila, V., Bulik-Sullivan, B., Finucane, H. K., Walters, R. K., Bras, J., Duncan, L., Escott-Price, V., Falcone, G. J., Gormley, P., Malik, R., Patsopoulos, N. A., Ripke, S., Wei, Z., Yu, D., Lee, P. H., Turley, P., Grenier-Boley, B., Chouraki, V., … Neale, B. M. (2018). Analysis of shared heritability in common disorders of the brain. *Science*, *360*(6395), https://doi.org/10.1126/science.aap8757

Brown, A. S., & Derkits, E. J. (2010). Prenatal Infection and Schizophrenia: A Review of Epidemiologic and Translational Studies. *American Journal of Psychiatry*, *167*(3), 261–280. https://doi.org/10.1176/appi.ajp.2009.09030361

Bulik-Sullivan, B. K., & Neale, B. M. (2015). LD Score Regression Distinguishes Confounding from Polygenicity in GWAS. *Nature Genetics*, *47*(3), 291–295. https://doi.org/10.1038/ng.3211.LD

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P. R., Duncan, L., Perry, J. R. B., Patterson, N., Robinson, E. B., Daly, M. J., Price, A. L., & Neale, B. M. (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, *47*(11), 1236–1241. https://doi.org/10.1038/ng.3406

Bult, C. J., Blake, J. A., Smith, C. L., Kadin, J. A., Richardson, J. E., Anagnostopoulos, A., Asabor, R., Baldarelli, R. M., Beal, J. S., Bello, S. M., Blodgett, O., Butler, N. E., Christie, K. R., Corbani, L. E., Creelman, J., Dolan, M. E., Drabkin, H. J., Giannatto, S. L., Hale, P., … Zhu, Y. (2019). Mouse Genome Database (MGD) 2019. *Nucleic Acids Research*, *47*(D1), D801–D806. https://doi.org/10.1093/nar/gky1056

Buniello, A., Macarthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., Suveges, D., Vrousgou, O., Whetzel, P. L., Amode, R., Guillen, J. A., Riat, H. S., Trevanion, S. J., Hall, P., Junkins, H., … Parkinson, H. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*, *47*(D1), D1005–D1012. https://doi.org/10.1093/nar/gky1120

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., & Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203–209. https://doi.org/10.1038/s41586-018-0579-z

Caillet-Boudin, M. L., Buée, L., Sergeant, N., & Lefebvre, B. (2015). Regulation of human MAPT gene expression. *Molecular Neurodegeneration*, *10*(1). https://doi.org/10.1186/s13024-015-0025-8

Cannon, T. D., Kaprio, J., Lönnqvist, J., Huttunen, M., & Koskenvuo, M. (1998). The genetic epidemiology of schizophrenia in a Finnish twin cohort: A population-based modeling study. *Archives of General Psychiatry*, *55*(1), 67–74. https://doi.org/10.1001/archpsyc.55.1.67

Carbon, S., Douglass, E., Good, B. M., Unni, D. R., Harris, N. L., Mungall, C. J., Basu, S., Chisholm, R. L., Dodson, R. J., Hartline, E., Fey, P., Thomas, P. D., Albou, L. P., Ebert, D., Kesling, M. J., Mi, H., Muruganujan, A., Huang, X., Mushayahama, T., … Elser, J. (2021). The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Research*, *49*(D1), D325–D334. https://doi.org/10.1093/nar/gkaa1113

Cardno, A. G., & Gottesman, I. I. (2000). Twin studies of schizophrenia: From bow-and-arrow concordances to Star Wars Mx and functional genomics. *American Journal of Medical Genetics*, *97*(1), 12–17. https://doi.org/10.1002/(sici)1096-8628(200021)97:1<12::aid-ajmg3>3.3.co;2-l

Cardno, A. G., Sham, P. C., Murray, R. M., & McGuffin, P. (2001). Twin study of symptom dimensions in psychoses. *British Journal of Psychiatry*, *178*(JULY), 39–45. https://doi.org/10.1192/bjp.179.1.39

Chen, H., Wang, C., Conomos, M. P., Stilp, A. M., Li, Z., Sofer, T., Szpiro, A. A., Chen, W., Brehm, J. M., Celedón, J. C., Redline, S., Papanicolaou, G. J., Thornton, T. A., Laurie, C. C., Rice, K., & Lin, X. (2016). Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models. *American Journal of Human Genetics*, *98*(4), 653–666. https://doi.org/10.1016/j.ajhg.2016.02.012

Cheng, S., Cheng, B., Liu, L., Yang, X., Meng, P., Yao, Y., Pan, C., Zhang, J., Li, C., Zhang, H., Chen, Y., Zhang, Z., Wen, Y., Jia, Y., & Zhang, F. (2022). Exome-wide screening identifies novel rare risk variants for major depression disorder. *Molecular Psychiatry*, *27*(7), 3069–3074. https://doi.org/10.1038/s41380-022-01536-4

Cheng, W., Frei, O., van der Meer, D., Wang, Y., O'Connell, K. S., Chu, Y., Bahrami, S., Shadrin, A. A., Alnæs, D., Hindley, G. F. L., Lin, A., Karadag, N., Fan, C. C., Westlye, L. T., Kaufmann, T., Molden, E., Dale, A. M., Djurovic, S., Smeland, O. B., & Andreassen, O. A. (2021). Genetic Association between Schizophrenia and Cortical Brain Surface Area and Thickness. *JAMA Psychiatry*, *78*(9), 1020–1030. https://doi.org/10.1001/jamapsychiatry.2021.1435

Chubb, J. E., Bradshaw, N. J., Soares, D. C., Porteous, D. J., & Millar, J. K. (2008). The DISC locus in psychiatric illness. *Molecular Psychiatry*, *13*(1), 36–64. https://doi.org/10.1038/sj.mp.4002106

Clarke, L., Fairley, S., Zheng-Bradley, X., Streeter, I., Perry, E., Lowy, E., Tassé, A. M., & Flicek, P. (2017). The international Genome sample resource (IGSR): A worldwide collection of genome

variation incorporating the 1000 Genomes Project data. *Nucleic Acids Research*, *45*(D1), D854–D859. https://doi.org/10.1093/nar/gkw829

Colhoun, H. M., McKeigue, P. M., & Smith, G. D. (2003). Problems of reporting genetic associations with complex outcomes. In *Lancet* (Vol. 361, Issue 9360, pp. 865–872). https://doi.org/10.1016/S0140-6736(03)12715-8

Cook, B. L., & Manning, W. G. (2013). Thinking beyond the mean: a practical guide for using quantile regression methods for health services research. *Shanghai Archives of Psychiatry*, *25*(1), 55–59. https://doi.org/10.3969/j.issn.1002-0829.2013.01.011

Corvin, A., & Morris, D. W. (2014). Genome-wide association studies: Findings at the major histocompatibility complex locus in psychosis. *Biological Psychiatry*, *75*(4), 276–283. https://doi.org/10.1016/j.biopsych.2013.09.018

Cruceanu, C., Schmouth, J. F., Torres-Platas, S. G., Lopez, J. P., Ambalavanan, A., Darcq, E., Gross, F., Breton, B., Spiegelman, D., Rochefort, D., Hince, P., Petite, J. M., Gauthier, J., Lafrenière, R. G., Dion, P. A., Greenwood, C. M., Kieffer, B. L., Alda, M., Turecki, G., & Rouleau, G. A. (2018). Rare susceptibility variants for bipolar disorder suggest a role for G protein-coupled receptors. *Molecular Psychiatry*, *23*(10), 2050–2056. https://doi.org/10.1038/mp.2017.223

Curtis, D. (2022). Investigation of Association of Rare, Functional Genetic Variants With Heavy Drinking and Problem Drinking in Exome Sequenced UK Biobank Participants. *Alcohol and Alcoholism*, *57*(4), 421–428. https://doi.org/10.1093/alcalc/agab031

Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., Vrieze, S. I., Chew, E. Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P. R., Iacono, W. G., Swaroop, A., Scott, L. J., Cucca, F., Kronenberg, F., Boehnke, M., … Fuchsberger, C. (2016). Next-generation genotype imputation service and methods. *Nature Genetics*, *48*(10), 1284–1287. https://doi.org/10.1038/ng.3656

Daviet, R., Aydogan, G., Jagannathan, K., Spilka, N., Koellinger, P. D., Kranzler, H. R., Nave, G., & Wetherill, R. R. (2022). Associations between alcohol consumption and gray and white matter volumes in the UK Biobank. *Nature Communications*, *13*(1), 1175. https://doi.org/10.1038/s41467-022-28735-5

de Jager, P. L., Jia, X., Wang, J., de Bakker, P. I. W., Ottoboni, L., Aggarwal, N. T., Piccio, L., Raychaudhuri, S., Tran, D., Aubin, C., Briskin, R., Romano, S., Baranzini, S. E., McCauley, J. L., Pericak-Vance, M. A., Haines, J. L., Gibson, R. A., Naeglin, Y., Uitdehaag, B., … Oksenberg, J. R. (2009). Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. *Nature Genetics*, *41*(7), 776–782. https://doi.org/10.1038/ng.401

de Jong, S., Diniz, M. J. A., Saloma, A., Gadelha, A., Santoro, M. L., Ota, V. K., Noto, C., Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., Adams, M. J., Agerbo, E., Air, T. M., Andlauer, T. F. M., Bacanu, S. A., Bækvad-Hansen, M., … Breen, G. (2018). Applying polygenic risk scoring for psychiatric disorders to a large family with bipolar disorder and major depressive disorder. *Communications Biology*, *1*(1). https://doi.org/10.1038/s42003-018-0155-y

Delisi, L. E. (1992). The significance of age of onset for schizophrenia. *Schizophrenia Bulletin*, *18*(2), 209–215. https://doi.org/10.1093/schbul/18.2.209

Demange, P. A., Malanchini, M., Mallard, T. T., Biroli, P., Cox, S. R., Grotzinger, A. D., Tucker-Drob, E. M., Abdellaoui, A., Arseneault, L., van Bergen, E., Boomsma, D. I., Caspi, A., Corcoran, D. L., Domingue, B. W., Harris, K. M., Ip, H. F., Mitchell, C., Moffitt, T. E., Poulton, R., … Nivard, M. G. (2021). Investigating the genetic architecture of noncognitive skills using

GWAS-by-subtraction. *Nature Genetics*, *53*(1), 35–44. https://doi.org/10.1038/s41588-020-00754-2

Demars, F., Kebir, O., Marzo, A., Iftimovici, A., Schramm, C., Amado, I., Bourgin, J., Daban Huard, C., Jantac Mam-Lam-Fook, C., Plaze, M., Rivollier, F., Krebs, M. O., & Chaumette, B. (2020). Dysregulation of peripheral expression of the YWHA genes during conversion to psychosis. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-66901-1

Demontis, D., Walters, R. K., Martin, J., Mattheisen, M., Als, T. D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Churchhouse, C., Dumont, A., Eriksson, N., Gandal, M., Goldstein, J. I., Grasby, K. L., Grove, J., … Neale, B. M. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature Genetics*, *51*(1), 63–75. https://doi.org/10.1038/s41588-018-0269-7

Derks, E. M., Vorstman, J. A. S., Ripke, S., Kahn, R. S., Ophoff, R. A., O'Donovan, M. C., Craddock, N., Holmans, P. A., Hamshere, M., Williams, H. J., Moskvina, V., Dwyer, S., Georgieva, L., Zammit, S., Owen, M. J., Sullivan, P. F., Lin, D. Y., van den Oord, E., Kim, Y., … Kendler, K. S. (2012). Investigation of the genetic association between quantitative measures of psychosis and schizophrenia: A polygenic risk score analysis. *PLoS ONE*, *7*(6). https://doi.org/10.1371/journal.pone.0037852

Dickerson, F., Jones-Brando, L., Ford, G., Genovese, G., Stallings, C., Origoni, A., O'Dushlaine, C., Katsafanas, E., Sweeney, K., Khushalani, S., & Yolken, R. (2019). Schizophrenia is Associated With an Aberrant Immune Response to Epstein–Barr Virus. *Schizophrenia Bulletin*, *45*(5), 1112–1119. https://doi.org/10.1093/schbul/sby164

Dudley, J. T., Kim, Y., Liu, L., Markov, G. J., Gerold, K., Chen, R., Butte, A. J., & Kumar, S. (2012). Human genomic disease variants: A neutral evolutionary explanation. *Genome Research*, *22*(8), 1383–1394. https://doi.org/10.1101/gr.133702.111

Edenberg, H. J., & McClintick, J. N. (2018). Alcohol Dehydrogenases, Aldehyde Dehydrogenases, and Alcohol Use Disorders: A Critical Review. *Alcoholism: Clinical and Experimental Research*, *42*(12), 2281–2297. https://doi.org/10.1111/acer.13904

Egan, M. F., Goldberg, T. E., Kolachana, B. S., Callicott, J. H., Mazzanti, C. M., Straub, R. E., Goldman, D., & Weinberger, D. R. (2001). Effect of COMT Val $^{108/158}$ Met genotype on frontal lobe function and risk for schizophrenia. *Proceedings of the National Academy of Sciences*, *98*(12), 6917–6922. https://doi.org/10.1073/pnas.111134598

Elston, R. C., Namboodiri, K. K., Spence, M. A., & Rainer, J. D. (1978). A Genetic Study of Schizophrenia Pedigrees. *Neuropsychobiology*, *4*(4), 193–206. https://doi.org/10.1159/000117633

ENCODE Consortium, E., Bernstein, B. E., Birney, E., Dunham, I., Green, E. D., Gunter, C., & Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74. https://doi.org/10.1038/nature11247

Ernst, J., & Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. *Nature Protocols*, *12*(12), 2478–2492. https://doi.org/10.1038/nprot.2017.124

Esterberg, M. L., Trotman, H. D., Holtzman, C., Compton, M. T., & Walker, E. F. (2010). The impact of a family history of psychosis on age-at-onset and positive and negative symptoms of schizophrenia: A meta-analysis. *Schizophrenia Research*, *120*(1–3), 121–130. https://doi.org/10.1016/j.schres.2010.01.011

Fahrner, J. A., & Bjornsson, H. T. (2014). Mendelian Disorders of the Epigenetic Machinery: Tipping the Balance of Chromatin States. *Annual Review of Genomics and Human Genetics*, *15*(1), 269–293. https://doi.org/10.1146/annurev-genom-090613-094245

Fanous, A. H., van den Oord, E. J., Riley, B. P., Aggen, S. H., Neale, M. C., O'Neill, F. A., Walsh, D., & Kendler, K. S. (2005). Relationship between a high-risk haplotype in the DTNBP1 (dysbindin) gene and clinical features of schizophrenia. *American Journal of Psychiatry*, *162*(10), 1824–1832. https://doi.org/10.1176/appi.ajp.162.10.1824

Fanous, A. H., Zhou, B., Aggen, S. H., Bergen, S. E., Amdur, R. L., Duan, J., Sanders, A. R., Shi, J., Mowry, B. J., Olincy, A., Amin, F., Cloninger, C. R., Silverman, J. M., Buccola, N. G., Byerley, W. F., Black, D. W., Freedman, R., Dudbridge, F., Holmans, P. A., … Levinson, D. F. (2012). Genome-wide association study of clinical dimensions of schizophrenia: Polygenic effect on disorganized symptoms. *American Journal of Psychiatry*, *169*(12), 1309–1317. https://doi.org/10.1176/appi.ajp.2012.12020218

Fanous, A., Gardner, C., Walsh, D., & Kendler, K. S. (2001). Relationship between positive and negative symptoms of schizophrenia and schizotypal symptoms in nonpsychotic relatives. *Archives of General Psychiatry*, *58*(7), 669–673. https://doi.org/10.1001/archpsyc.58.7.669

Finucane, H. K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P. R., Anttila, V., Xu, H., Zang, C., Farh, K., Ripke, S., Day, F. R., Purcell, S., Stahl, E., Lindstrom, S., Perry, J. R. B., Okada, Y., Raychaudhuri, S., Daly, M. J., … Price, A. L. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics*, *47*(11), 1228–1235. https://doi.org/10.1038/ng.3404

Fox, J. (2022). polycor: Polychoric and Polyserial Correlations. *R Package Version 0.8-1 Https://CRAN.R-Project.Org/Package=polycor*. http://cran.r-project.org/package=polycor

Frankish, A., Diekhans, M., Jungreis, I., Lagarde, J., Loveland, J. E., Mudge, J. M., Sisu, C., Wright, J. C., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Boix, C., Carbonell Sala, S., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I. T., García Girón, C., … Flicek, P. (2021). GENCODE 2021. *Nucleic Acids Research*, *49*(D1), D916–D923. https://doi.org/10.1093/nar/gkaa1087

Frei, O., Holland, D., Smeland, O. B., Shadrin, A. A., Fan, C. C., Maeland, S., O'Connell, K. S., Wang, Y., Djurovic, S., Thompson, W. K., Andreassen, O. A., & Dale, A. M. (2019). Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *Nature Communications*, *10*(1), 2417. https://doi.org/10.1038/s41467-019-10310-0

Fromer, M., Pocklington, A. J., Kavanagh, D. H., Williams, H. J., Dwyer, S., Gormley, P., Georgieva, L., Rees, E., Palta, P., Ruderfer, D. M., Carrera, N., Humphreys, I., Johnson, J. S., Roussos, P., Barker, D. D., Banks, E., Milanova, V., Grant, S. G., Hannon, E., … O'Donovan, M. C. (2014). De novo mutations in schizophrenia implicate synaptic networks. *Nature*, *506*(7487), 179–184. https://doi.org/10.1038/nature12929

Galarza, C. E., Lachos, V. H., & Bandyopadhyay, D. (2017). Quantile regression in linear mixed models: A stochastic approximation EM approach. *Statistics and Its Interface*, *10*(3), 471–482. https://doi.org/10.4310/SII.2017.v10.n3.a10

Ge, T., Chen, C. Y., Ni, Y., Feng, Y. C. A., & Smoller, J. W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-09718-5

Gelernter, J., Kranzler, H. R., Sherva, R., Almasy, L., Koesterer, R., Smith, A. H., Anton, R., Preuss, U. W., Ridinger, M., Rujescu, D., Wodarz, N., Zill, P., Zhao, H., & Farrer, L. A. (2014). Genome-wide association study of alcohol dependence:significant findings in African- and

European-Americans including novel risk loci. *Molecular Psychiatry*, *19*(1), 41–49. https://doi.org/10.1038/mp.2013.145

Gentry AE, Kirkpatrick RM, & Webb BT. (2022). Missingness Adapted Group Informed Clustered (MAGIC)-LASSO: A novel paradigm for prediction in data with widespread non-random missingness. *Biorxiv*.

Gillespie, N. A., Gentry, A. E., Kirkpatrick, R. M., Reynolds, C. A., Mathur, R., Kendler, K. S., Maes, H. H., Webb, B. T., & Peterson, R. E. (2022). Determining the stability of genome-wide factors in BMI between ages 40 to 69 years. *PLOS Genetics*, *18*(8), e1010303. https://doi.org/10.1371/journal.pgen.1010303

Girard, S. L., Gauthier, J., Noreau, A., Xiong, L., Zhou, S., Jouan, L., Dionne-Laporte, A., Spiegelman, D., Henrion, E., Diallo, O., Thibodeau, P., Bachand, I., Bao, J. Y. J., Tong, A. H. Y., Lin, C.-H., Millet, B., Jaafari, N., Joober, R., Dion, P. A., … Rouleau, G. A. (2011). Increased exonic de novo mutation rate in individuals with schizophrenia. *Nature Genetics*, *43*(9), 860–863. https://doi.org/10.1038/ng.886

Girdhar, K., Hoffman, G. E., Jiang, Y., Brown, L., Kundakovic, M., Hauberg, M. E., Francoeur, N. J., Wang, Y. chih, Shah, H., Kavanagh, D. H., Zharovsky, E., Jacobov, R., Wiseman, J. R., Park, R., Johnson, J. S., Kassim, B. S., Sloofman, L., Mattei, E., Weng, Z., … Akbarian, S. (2018). Cell-specific histone modification maps in the human frontal lobe link schizophrenia risk to the neuronal epigenome. *Nature Neuroscience*, *21*(8), 1126–1136. https://doi.org/10.1038/s41593-018-0187-0

Goedde, H. W., Agarwal, D. P., Fritze, G., Meier-Tackmann, D., Singh, S., Beckmann, G., Bhatia, K., Chen, L. Z., Fang, B., Lisker, R., Paik, Y. K., Rothhammer, F., Saha, N., Segal, B., Srivastava, L. M., & Czeizel, A. (1992). Distribution of ADH2 and ALDH2 genotypes in different populations. *Human Genetics*, *88*(3). https://doi.org/10.1007/BF00197271

Goes, F. S., Pirooznia, M., Parla, J. S., Kramer, M., Ghiban, E., Mavruk, S., Chen, Y. C., Monson, E. T., Willour, V. L., Karchin, R., Flickinger, M., Locke, A. E., Levy, S. E., Scott, L. J., Boehnke, M., Stahl, E., Moran, J. L., Hultman, C. M., Landén, M., … Potash, J. B. (2016). Exome sequencing of familial bipolar disorder. *JAMA Psychiatry*, *73*(6), 590–597. https://doi.org/10.1001/jamapsychiatry.2016.0251

Goodin, D. S., Khankhanian, P., Gourraud, P. A., & Vince, N. (2021). The nature of genetic and environmental susceptibility to multiple sclerosis. *PLoS ONE*, *16*(3 March). https://doi.org/10.1371/journal.pone.0246157

Griswold, M. G., Fullman, N., Hawley, C., Arian, N., Zimsen, S. R. M., Tymeson, H. D., Venkateswaran, V., Tapp, A. D., Forouzanfar, M. H., Salama, J. S., Abate, K. H., Abate, D., Abay, S. M., Abbafati, C., Abdulkader, R. S., Abebe, Z., Aboyans, V., Abrar, M. M., Acharya, P., … Gakidou, E. (2018). Alcohol use and burden for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet*, *392*(10152), 1015–1035. https://doi.org/10.1016/S0140-6736(18)31310-2

Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S. J., Mallard, T. T., Hill, W. D., Ip, H. F., Marioni, R. E., McIntosh, A. M., Deary, I. J., Koellinger, P. D., Harden, K. P., Nivard, M. G., & Tucker-Drob, E. M. (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature Human Behaviour*, *3*(5), 513–525. https://doi.org/10.1038/s41562-019-0566-x

Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R. K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O. A., Anney, R., Awashti, S., Belliveau, R., Bettella, F., Buxbaum, J. D., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Christensen, J. H., … Børglum,

A. D. (2019). Identification of common genetic risk variants for autism spectrum disorder. *Nature Genetics*, *51*(3), 431–444. https://doi.org/10.1038/s41588-019-0344-8

Grubisha, M. J., Sun, X., MacDonald, M. L., Garver, M., Sun, Z., Paris, K. A., Patel, D. S., DeGiosio, R. A., Lewis, D. A., Yates, N. A., Camacho, C., Homanics, G. E., Ding, Y., & Sweet, R. A. (2021). MAP2 is differentially phosphorylated in schizophrenia, altering its function. *Molecular Psychiatry*, *26*(9), 5371–5388. https://doi.org/10.1038/s41380-021-01034-z

Gusev, A., Lee, S. H., Trynka, G., Finucane, H., Vilhjálmsson, B. J., Xu, H., Zang, C., Ripke, S., Bulik-Sullivan, B., Stahl, E., Kähler, A. K., Hultman, C. M., Purcell, S. M., McCarroll, S. A., Daly, M., Pasaniuc, B., Sullivan, P. F., Neale, B. M., Wray, N. R., … Sullivan, P. F. (2014). Partitioning Heritability of Regulatory and Cell-Type-Specific Variants across 11 Common Diseases. *The American Journal of Human Genetics*, *95*(5), 535–552. https://doi.org/10.1016/j.ajhg.2014.10.004

Hall, W., & Degenhardt, L. (2008). Cannabis use and the risk of developing a psychotic disorder. *World Psychiatry*, *7*(2), 68–71. https://doi.org/10.1002/j.2051-5545.2008.tb00158.x

Halvorsen, M., Huh, R., Oskolkov, N., Wen, J., Netotea, S., Giusti-Rodriguez, P., Karlsson, R., Bryois, J., Nystedt, B., Ameur, A., Kähler, A. K., Ancalade, N., Farrell, M., Crowley, J. J., Li, Y., Magnusson, P. K. E., Gyllensten, U., Hultman, C. M., Sullivan, P. F., & Szatkiewicz, J. P. (2020). Increased burden of ultra-rare structural variants localizing to boundaries of topologically associated domains in schizophrenia. *Nature Communications*, *11*(1), 1842. https://doi.org/10.1038/s41467-020-15707-w

Halvorsen, M., Szatkiewicz, J., Mudgal, P., Yu, D., Nordsletten, A. E., Mataix-Cols, D., Mathews, C. A., Scharf, J. M., Mattheisen, M., Robertson, M. M., McQuillin, A., & Crowley, J. J. (2021). Elevated common variant genetic risk for tourette syndrome in a densely-affected pedigree. *Molecular Psychiatry*, *26*(12), 7522–7529. https://doi.org/10.1038/s41380-021-01277-w

Heston, L. L. (1966). Psychiatric disorders in foster home reared children of schizophrenic mothers. *The British Journal of Psychiatry : The Journal of Mental Science*, *112*(489), 819–825. https://doi.org/10.1192/bjp.112.489.819

Hilker, R., Helenius, D., Fagerlund, B., Skytthe, A., Christensen, K., Werge, T. M., Nordentoft, M., & Glenthøj, B. (2018). Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide Danish Twin Register. *Biological Psychiatry*, *83*(6), 492–498. https://doi.org/10.1016/j.biopsych.2017.08.017

Hindley, G., Bahrami, S., Steen, N. E., O'Connell, K. S., Frei, O., Shadrin, A., Bettella, F., Rødevand, L., Fan, C. C., Dale, A. M., Djurovic, S., Smeland, O. B., & Andreassen, O. A. (2021). Characterising the shared genetic determinants of bipolar disorder, schizophrenia and risk-taking. *Translational Psychiatry*, *11*(1). https://doi.org/10.1038/s41398-021-01576-4

Hoffman, G. E., Bendl, J., Voloudakis, G., Montgomery, K. S., Sloofman, L., Wang, Y.-C., Shah, H. R., Hauberg, M. E., Johnson, J. S., Girdhar, K., Song, L., Fullard, J. F., Kramer, R., Hahn, C.-G., Gur, R., Marenco, S., Lipska, B. K., Lewis, D. A., Haroutunian, V., … Roussos, P. (2019). CommonMind Consortium provides transcriptomic and epigenomic data for Schizophrenia and Bipolar Disorder. *Scientific Data*, *6*(1), 180. https://doi.org/10.1038/s41597-019-0183-6

Holland, D., Frei, O., Desikan, R., Fan, C. C., Shadrin, A. A., Smeland, O. B., Sundar, V. S., Thompson, P., Andreassen, O. A., & Dale, A. M. (2020). Beyond SNP heritability: Polygenicity and discoverability of phenotypes estimated with a univariate Gaussian mixture model. *PLoS Genetics*, *16*(5). https://doi.org/10.1371/journal.pgen.1008612

Homann, O. R., Misura, K., Lamas, E., Sandrock, R. W., Nelson, P., Mcdonough, S. I., & de Lisi, L. E. (2016). Whole-genome sequencing in multiplex families with psychoses reveals mutations in

the SHANK2 and SMARCA1 genes segregating with illness. *Molecular Psychiatry*, *21*(12), 1690–1695. https://doi.org/10.1038/mp.2016.24

Hope, S., Melle, I., Aukrust, P., Steen, N. E., Birkenaes, A. B., Lorentzen, S., Agartz, I., Ueland, T., & Andreassen, O. A. (2009). Similar immune profile in bipolar disorder and schizophrenia: Selective increase in soluble tumor necrosis factor receptor I and von Willebrand factor. *Bipolar Disorders*, *11*(7), 726–734. https://doi.org/10.1111/j.1399-5618.2009.00757.x

Howard, D. M., Adams, M. J., Clarke, T. K., Hafferty, J. D., Gibson, J., Shirali, M., Coleman, J. R. I., Hagenaars, S. P., Ward, J., Wigmore, E. M., Alloza, C., Shen, X., Barbu, M. C., Xu, E. Y., Whalley, H. C., Marioni, R. E., Porteous, D. J., Davies, G., Deary, I. J., … McIntosh, A. M. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience*, *22*(3), 343–352. https://doi.org/10.1038/s41593-018-0326-7

Howrigan, D. P., Rose, S. A., Samocha, K. E., Fromer, M., Cerrato, F., Chen, W. J., Churchhouse, C., Chambert, K., Chandler, S. D., Daly, M. J., Dumont, A., Genovese, G., Hwu, H. G., Laird, N., Kosmicki, J. A., Moran, J. L., Roe, C., Singh, T., Wang, S. H., … Neale, B. M. (2020). Exome sequencing in schizophrenia-affected parent–offspring trios reveals risk conferred by protein-coding de novo mutations. *Nature Neuroscience*, *23*(2), 185–193. https://doi.org/10.1038/s41593-019-0564-3

Ingason, A., Rujescu, D., Cichon, S., Sigurdsson, E., Sigmundsson, T., Pietiläinen, O. P. H., Buizer-Voskamp, J. E., Strengman, E., Francks, C., Muglia, P., Gylfason, A., Gustafsson, O., Olason, P. I., Steinberg, S., Hansen, T., Jakobsen, K. D., Rasmussen, H. B., Giegling, I., Möller, H.-J., … Clair, D. M. S. (2011). Copy number variations of chromosome 16p13.1 region associated with schizophrenia. *Molecular Psychiatry*, *16*(1), 17–25. https://doi.org/10.1038/mp.2009.101

Ionita-Laza, I., Lee, S., Makarov, V., Buxbaum, J. D., & Lin, X. (2013). Sequence Kernel Association Tests for the Combined Effect of Rare and Common Variants. *The American Journal of Human Genetics*, *92*(6), 841–853. https://doi.org/10.1016/j.ajhg.2013.04.015

Irish Schizophrenia Genomics Consortium (2012). Genome-wide association study implicates HLA-C*01:02 as a risk factor at the major histocompatibility complex locus in schizophrenia. *Biological Psychiatry*, *72*(8), 620–628. https://doi.org/10.1016/j.biopsych.2012.05.035

Ishizuka, K., Paek, M., Kamiya, A., & Sawa, A. (2006). A Review of Disrupted-in-Schizophrenia-1 (disc1): Neurodevelopment, Cognition, and Mental Conditions. *Biological Psychiatry*, *59*(12), 1189–1197. https://doi.org/10.1016/j.biopsych.2006.03.065

Jablensky, A. (2001). Classification of nonschizophrenic psychotic disorders: a historical perspective. *Current Psychiatry Reports*, *3*(4), 326–331. https://doi.org/10.1007/s11920-001-0029-7

Johansen, C. T., Wang, J., Lanktree, M. B., Cao, H., McIntyre, A. D., Ban, M. R., Martins, R. A., Kennedy, B. A., Hassell, R. G., Visser, M. E., Schwartz, S. M., Voight, B. F., Elosua, R., Salomaa, V., O'Donnell, C. J., Dallinga-Thie, G. M., Anand, S. S., Yusuf, S., Huff, M. W., … Hegele, R. A. (2010). Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nature Genetics*, *42*(8), 684–687. https://doi.org/10.1038/ng.628

Käkelä, J., Panula, J., Oinas, E., Hirvonen, N., Jääskeläinen, E., & Miettunen, J. (2014). Family history of psychosis and social, occupational and global outcome in schizophrenia: A meta-analysis. *Acta Psychiatrica Scandinavica*, *130*(4), 269–278. https://doi.org/10.1111/acps.12317

Karnell, J. L., Rieder, S. A., Ettinger, R., & Kolbeck, R. (2019). Targeting the CD40-CD40L pathway in autoimmune diseases: Humoral immunity and beyond. *Advanced Drug Delivery Reviews*, *141*, 92–103. https://doi.org/10.1016/j.addr.2018.12.005

Kendler, K. S. (1984). An Independent Analysis of the Danish Adoption Study of Schizophrenia. *Archives of General Psychiatry*, *41*(6), 555. https://doi.org/10.1001/archpsyc.1984.01790170029004

Kendler, K. S. (1985). Diagnostic approaches to schizotypal personality disorder: a historical perspective. *Schizophrenia Bulletin*, *11*(4), 538–553. https://doi.org/10.1093/schbul/11.4.538

Kendler, K. S. (1995). Schizotypal Symptoms and Signs in the Roscommon Family Study. *Archives of General Psychiatry*, *52*(4), 296. https://doi.org/10.1001/archpsyc.1995.03950160046009

Kendler, K. S. (2020). The Development of Kraepelin's Concept of Dementia Praecox: A Close Reading of Relevant Texts. *JAMA Psychiatry*, *77*(11), 1181–1187. https://doi.org/10.1001/jamapsychiatry.2020.1266

Kendler, K. S., & Diehl, S. R. (1993). The Genetics of Schizophrenia: A Current, Genetic-epidemiologic Perspective. *Schizophrenia Bulletin*, *19*(2), 261–285. https://doi.org/10.1093/schbul/19.2.261

Kendler, K. S., Gruenberg, A. M., & Tsuang, M. T. (1985). Psychiatric Illness in First-degree Relatives of Schizophrenic and Surgical Control Patients: A Family Study Using DSM-III Criteria. *Archives of General Psychiatry*, *42*(8), 770–779. https://doi.org/10.1001/archpsyc.1985.01790310032004

Kendler, K. S., Karkowski-Shuman, L., Anthony O'Neill, F., Straub, R. E., MacLean, C. J., & Walsh, D. (1997). Resemblance of psychotic symptoms and syndromes in affected sibling pairs from the irish study of high-density schizophrenia families: Evidence for possible etiologic heterogeneity. *American Journal of Psychiatry*, *154*(2), 191–198. https://doi.org/10.1176/ajp.154.2.191

Kendler, K. S., Lieberman, J. A., & Walsh, D. (1989). The structured interview for schizotypy (SIS): A preliminary report. *Schizophrenia Bulletin*, *15*(4), 559–571. https://doi.org/10.1093/schbul/15.4.559

Kendler, K. S., McGuire, M., Gruenberg, A. M., O'Hare, A., Spellman, M., & Walsh, D. (1993a). The Roscommon Family Study: III. Schizophrenia-Related Personality Disorders in Relatives. *Archives of General Psychiatry*, *50*(10), 781–788. https://doi.org/10.1001/archpsyc.1993.01820220033004

Kendler, K. S., McGuire, M., Gruenberg, A. M., O'Hare, A., Spellman, M., & Walsh, D. (1993b). The Roscommon Family Study: IV. Affective Illness, Anxiety Disorders, and Alcoholism in Relatives. *Archives of General Psychiatry*, *50*(12), 952–960. https://doi.org/10.1001/archpsyc.1993.01820240036005

Kendler, K. S., McGuire, M., Gruenberg, A. M., Spellman, M., O'Hare, A., & Walsh, D. (1993). The Roscommon Family Study: II. The Risk of Nonschizophrenic Nonaffective Psychoses in Relatives. *Archives of General Psychiatry*, *50*(8), 645–652. https://doi.org/10.1001/archpsyc.1993.01820200059006

Kendler, K. S., Neale, M. C., & Walsh, D. (1995). Evaluating the spectrum concept of schizophrenia in the Roscommon Family Study. *American Journal of Psychiatry*, *152*(5), 749–754. https://doi.org/10.1176/ajp.152.5.749

Kendler, K. S., O'Neill, F. A., Burke, J., Murphy, B., Duke, F., Straub, R. E., Shinkwin, R., Nuallain, M. N., MacLean, C. J., & Walsh, D. (1996). Irish study of high-density schizophrenia families: Field methods and power to detect linkage. *American Journal of Medical Genetics - Seminars in Medical Genetics*, *67*(2), 179–190. https://doi.org/10.1002/(SICI)1096-8628(19960409)67:2<179::AID-AJMG8>3.0.CO;2-N

Kendler, K. S., Thacker, L., & Walsh, D. (1996). Self-report measures of schizotypy as indices of familial vulnerability to schizophrenia. *Schizophrenia Bulletin*, *22*(3), 511–520. https://doi.org/10.1093/schbul/22.3.511

Kety, S. S., Rosenthal, D., Wender, P. H., Schulsinger, F., & Jacobsen, B. (1975). Mental illness in the biological and adoptive families of adopted individuals who have become schizophrenic: a preliminary report based on psychiatric interviews. *Proceedings of the Annual Meeting of the American Psychopathological Association*, *63*, 147–165

Kety, S. S., Wender, P. H., Jacobsen, B., Ingraham, L. J., Janson, L., Faber, B., & Kinney, D. K. (1994). Mental Illness in the Biological and Adoptive Relatives of Schizophrenic Adoptees: Replication of the Copenhagen Study in the Rest of Denmark. *Archives of General Psychiatry*, *51*(6), 442–455. https://doi.org/10.1001/archpsyc.1994.03950060006001

Khandaker, G. M., Zimbron, J., Lewis, G., & Jones, P. B. (2013). Prenatal maternal infection, neurodevelopment and adult schizophrenia: a systematic review of population-based studies. *Psychological Medicine*, *43*(2), 239–257. https://doi.org/10.1017/S0033291712000736

Kirov, G., Gumus, D., Chen, W., Norton, N., Georgieva, L., Sari, M., O'Donovan, M. C., Erdogan, F., Owen, M. J., Ropers, H.-H., & Ullmann, R. (2007). Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Human Molecular Genetics*, *17*(3), 458–465. https://doi.org/10.1093/hmg/ddm323

Kirov, G., O'Donovan, M. C., & Owen, M. J. (2005). Finding schizophrenia genes. *The Journal of Clinical Investigation*, *115*(6), 1440–1448. https://doi.org/10.1172/JCI24759

Kirov, G., Pocklington, A. J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., Moran, J., Chambert, K., Toncheva, D., Georgieva, L., Grozeva, D., Fjodorova, M., Wollerton, R., Rees, E., Nikolov, I., van de Lagemaat, L. N., Bayés, À., Fernandez, E., Olason, P. I., … Owen, M. J. (2012). De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Molecular Psychiatry*, *17*(2), 142–153. https://doi.org/10.1038/mp.2011.154

Kranzler, H. R., Zhou, H., Kember, R. L., Vickers Smith, R., Justice, A. C., Damrauer, S., Tsao, P. S., Klarin, D., Baras, A., Reid, J., Overton, J., Rader, D. J., Cheng, Z., Tate, J. P., Becker, W. C., Concato, J., Xu, K., Polimanti, R., Zhao, H., & Gelernter, J. (2019). Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nature Communications*, *10*(1), 1499. https://doi.org/10.1038/s41467-019-09480-8

Krieger, T. G., Moran, C. M., Frangini, A., Visser, W. E., Schoenmakers, E., Muntoni, F., Clark, C. A., Gadian, D., Chong, W. K., Kuczynski, A., Dattani, M., Lyons, G., Efthymiadou, A., Vargha-Khadem, F., Simons, B. D., Chatterjee, K., & Livesey, F. J. (2019). Mutations in thyroid hormone receptor α1 cause premature neurogenesis and progenitor cell depletion in human cortical development. *Proceedings of the National Academy of Sciences*, *116*(45), 22754–22763. https://doi.org/10.1073/pnas.1908762116

Landin-Romero, R., McKenna, P. J., Romaguera, A., Álvarez-Moya, E., Sarró, S., Aguirre, C., Sarri, C., Compte, A., Bosque, C., Salvador, R., & Pomarol-Clotet, E. (2016). Examining the continuum of psychosis: Frequency and characteristics of psychotic-like symptoms in relatives and non-relatives of patients with schizophrenia. *Schizophrenia Research*, *178*(1–3), 6–11. https://doi.org/10.1016/j.schres.2016.07.015

Lee, P. H., Anttila, V., Won, H., Feng, Y. C. A., Rosenthal, J., Zhu, Z., Tucker-Drob, E. M., Nivard, M. G., Grotzinger, A. D., Posthuma, D., Wang, M. M. J., Yu, D., Stahl, E. A., Walters, R. K., Anney, R. J. L., Duncan, L. E., Ge, T., Adolfsson, R., Banaschewski, T., … Smoller, J. W.

(2019). Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell*, *179*(7), 1469-1482.e11. https://doi.org/10.1016/j.cell.2019.11.020

Lee, P. H., Feng, Y. C. A., & Smoller, J. W. (2021). Pleiotropy and Cross-Disorder Genetics Among Psychiatric Disorders. *Biological Psychiatry*, *89*(1), 20–31. https://doi.org/10.1016/j.biopsych.2020.09.026

Legge, S. E., Jones, H. J., Kendall, K. M., Pardiñas, A. F., Menzies, G., Bracher-Smith, M., Escott-Price, V., Rees, E., Davis, K. A. S., Hotopf, M., Savage, J. E., Posthuma, D., Holmans, P., Kirov, G., Owen, M. J., O'Donovan, M. C., Zammit, S., & Walters, J. T. R. (2019). Association of Genetic Liability to Psychotic Experiences with Neuropsychotic Disorders and Traits. *JAMA Psychiatry*, *76*(12), 1256–1265. https://doi.org/10.1001/jamapsychiatry.2019.2508

Levinson, D. F., Shi, J., Wang, K., Oh, S., Riley, B., Pulver, A. E., Wildenauer, D. B., Laurent, C., Mowry, B. J., Gejman, P. v., Owen, M. J., Kendler, K. S., Nestadt, G., Schwab, S. G., Mallet, J., Nertney, D., Sanders, A. R., Williams, N. M., Wormley, B., … Holmans, P. A. (2012). Genome-wide association study of multiplex schizophrenia pedigrees. *American Journal of Psychiatry*, *169*(9), 963–973. https://doi.org/10.1176/appi.ajp.2012.11091423

Li, Y., Wu, Y., & Zhou, Y. (2006). Modulation of Inactivation Properties of CaV2.2 Channels by 14-3-3 Proteins. *Neuron*, *51*(6), 755–771. https://doi.org/10.1016/j.neuron.2006.08.014

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., Datta, G., Davila-Velderrain, J., McGuire, D., Tian, C., Zhan, X., Choquet, H., Docherty, A. R., Faul, J. D., Foerster, J. R., Fritsche, L. G., Gabrielsen, M. E., Gordon, S. D., Haessler, J., … Vrieze, S. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics*, *51*(2), 237–244. https://doi.org/10.1038/s41588-018-0307-5

Loh, P. R., Danecek, P., Palamara, P. F., Fuchsberger, C., Reshef, Y. A., Finucane, H. K., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G. R., Durbin, R., & Price, A. L. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. *Nature Genetics*, *48*(11), 1443–1448. https://doi.org/10.1038/ng.3679

MacArthur, D. G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J. K., Montgomery, S. B., Albers, C. A., Zhang, Z. D., Conrad, D. F., Lunter, G., Zheng, H., Ayub, Q., DePristo, M. A., Banks, E., Hu, M., … Tyler-Smith, C. (2012). A systematic survey of loss-of-function variants in human protein-coding genes. *Science*, *335*(6070), 823–828. https://doi.org/10.1126/science.1215040

Mallard, T., Linnér, R., Grotzinger, A., Sanchez-Roige, S., Seidlitz, J., Okbay, A., de Vlaming, R., Meddens, S. F. W., Palmer, A., Davis, L., Tucker-Drob, E., Kendler, K., Keller, M., Koellinger, P., & Harden, K. P. (2019). Multivariate GWAS of psychiatric disorders and their cardinal symptoms reveal two dimensions of cross-cutting genetic liabilities. *BioRxiv*, 603134. https://doi.org/10.1101/603134

Marshall, C. R., Howrigan, D. P., Merico, D., Thiruvahindrapuram, B., Wu, W., Greer, D. S., Antaki, D., Shetty, A., Holmans, P. A., Pinto, D., Gujral, M., Brandler, W. M., Malhotra, D., Wang, Z., Fajarado, K. V. F., Maile, M. S., Ripke, S., Agartz, I., Albus, M., … Sebat, J. (2017a). Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nature Genetics*, *49*(1), 27–35. https://doi.org/10.1038/ng.3725

Marshall, C. R., Howrigan, D. P., Merico, D., Thiruvahindrapuram, B., Wu, W., Greer, D. S., Antaki, D., Shetty, A., Holmans, P. A., Pinto, D., Gujral, M., Brandler, W. M., Malhotra, D., Wang, Z., Fuentes Fajarado, K. v., Maile, M. S., Ripke, S., Agartz, I., Albus, M., … Sebat, J. (2017b). Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nature Genetics*, *49*(1), 27–35. https://doi.org/10.1038/ng.3725

Matsuda, Y., Morino, H., Miyamoto, R., Kurashige, T., Kume, K., Mizuno, N., Kanaya, Y., Tada, Y., Ohsawa, R., Yokota, K., Shimozawa, N., Maruyama, H., & Kawakami, H. (2020). Biallelic mutation of HSD17B4 induces middle age-onset spinocerebellar ataxia. *Neurology: Genetics*, *6*(1). https://doi.org/10.1212/NXG.0000000000000396

McCarthy, S. E., Gillis, J., Kramer, M., Lihm, J., Yoon, S., Berstein, Y., Mistry, M., Pavlidis, P., Solomon, R., Ghiban, E., Antoniou, E., Kelleher, E., O'Brien, C., Donohoe, G., Gill, M., Morris, D. W., McCombie, W. R., & Corvin, A. (2014). De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. *Molecular Psychiatry*, *19*(6), 652–658. https://doi.org/10.1038/mp.2014.29

McCarthy, S. E., Makarov, V., Kirov, G., Addington, A. M., McClellan, J., Yoon, S., Perkins, D. O., Dickel, D. E., Kusenda, M., Krastoshevsky, O., Krause, V., Kumar, R. A., Grozeva, D., Malhotra, D., Walsh, T., Zackai, E. H., Kaplan, P., Ganesh, J., Krantz, I. D., … Sebat, J. (2009). Microduplications of 16p11.2 are associated with schizophrenia. *Nature Genetics*, *41*(11), 1223–1227. https://doi.org/10.1038/ng.474

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., Kang, H. M., Fuchsberger, C., Danecek, P., Sharp, K., Luo, Y., Sidore, C., Kwong, A., Timpson, N., Koskinen, S., Vrieze, S., Scott, L. J., Zhang, H., Mahajan, A., … Marchini, J. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*, *48*(10), 1279–1283. https://doi.org/10.1038/ng.3643

McCutcheon, R. A., Krystal, J. H., & Howes, O. D. (2020). Dopamine and glutamate in schizophrenia: biology, symptoms and treatment. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, *19*(1), 15–33. https://doi.org/10.1002/wps.20693

McGrath, J. J., & Welham, J. L. (1999). Season of birth and schizophrenia: a systematic review and meta-analysis of data from the Southern Hemisphere1Details of this paper were presented at the Winter Workshop on Schizophrenia Research, Davos, Switzerland, February 1998.1. *Schizophrenia Research*, *35*(3), 237–242. https://doi.org/10.1016/S0920-9964(98)00139-X

McGrath, J. J., Lim, C. C. W., Plana-Ripoll, O., Holtz, Y., Agerbo, E., Momen, N. C., Mortensen, P. B., Pedersen, C. B., Abdulmalik, J., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., Bromet, E. J., Bruffaerts, R., Bunting, B., de Almeida, J. M. C., de Girolamo, G., de Vries, Y. A., Florescu, S., … de Jonge, P. (2020). Comorbidity within mental disorders: a comprehensive analysis based on 145 990 survey respondents from 27 countries. *Epidemiology and Psychiatric Sciences*, *29*, e153. https://doi.org/10.1017/S2045796020000633

McGrath, J., & Scott, J. (2006). Urban birth and risk of schizophrenia: a worrying example of epidemiology where the data are stronger than the hypotheses. *Epidemiology and Psychiatric Sciences*, *15*(4), 243–246. https://doi.org/10.1017/S1121189X00002104

McGuffin, P., Farmer, A., & Harvey, I. (1991). A polydiagnostic application of operational criteria in studies of psychotic illness: Development and reliability of the OPCRIT system. *Archives of General Psychiatry*, *48*(8), 764–770. https://doi.org/10.1001/archpsyc.1991.01810320088015

McGuffin, P., Tandon, K., & Corsico, A. (2003). Linkage and association studies of schizophrenia. *Current Psychiatry Reports*, *5*(2), 121–127. https://doi.org/10.1007/s11920-003-0028-y

Meller, T., Schmitt, S., Stein, F., Brosch, K., Mosebach, J., Yüksel, D., Zaremba, D., Grotegerd, D., Dohm, K., Meinert, S., Förster, K., Redlich, R., Opel, N., Repple, J., Hahn, T., Jansen, A., Andlauer, T. F. M., Forstner, A. J., Heilmann-Heimbach, S., … Nenadić, I. (2019). Associations of schizophrenia risk genes ZNF804A and CACNA1C with schizotypy and modulation of attention in healthy subjects. *Schizophrenia Research*, *208*, 67–75. https://doi.org/10.1016/j.schres.2019.04.018

Momeni, P., Wickremaratchi, M. M., Bell, J., Arnold, R., Beer, R., Hardy, J., Revesz, T., Neal, J. W., & Morris, H. R. (2010). Familial early onset frontotemporal dementia caused by a novel S356T MAPT mutation, initially diagnosed as schizophrenia. *Clinical Neurology and Neurosurgery*, *112*(10), 917–920. https://doi.org/10.1016/j.clineuro.2010.07.015

Monereo-Sánchez, J., Schram, M. T., Frei, O., O'Connell, K., Shadrin, A. A., Smeland, O. B., Westlye, L. T., Andreassen, O. A., Kaufmann, T., Linden, D. E. J., & van der Meer, D. (2021). Genetic Overlap Between Alzheimer's Disease and Depression Mapped Onto the Brain. *Frontiers in Neuroscience*, *15*. https://doi.org/10.3389/fnins.2021.653130

Morris, D. W., Pearson, R. D., Cormican, P., Kenny, E. M., O'Dushlaine, C. T., Perreault, L. P. L., Giannoulatou, E., Tropea, D., Maher, B. S., Wormley, B., Kelleher, E., Fahey, C., Molinos, I., Bellini, S., Pirinen, M., Strange, A., Freeman, C., Thiselton, D. L., Elves, R. L., … Corvin, A. (2014). An inherited duplication at the gene p21 protein-activated Kinase 7 (PAK7) is a risk factor for psychosis. *Human Molecular Genetics*, *23*(12), 3316–3326. https://doi.org/10.1093/hmg/ddu025

Mullins, N., Forstner, A. J., O'Connell, K. S., Coombes, B., Coleman, J. R. I., Qiao, Z., Als, T. D., Bigdeli, T. B., Børte, S., Bryois, J., Charney, A. W., Drange, O. K., Gandal, M. J., Hagenaars, S. P., Ikeda, M., Kamitaki, N., Kim, M., Krebs, K., Panagiotaropoulou, G., … Andreassen, O. A. (2021). Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nature Genetics*, *53*(6), 817–829. https://doi.org/10.1038/s41588-021-00857-4

Murphy, R., O'Donoghue, S., Counihan, T., McDonald, C., Calabresi, P. A., Ahmed, M. A. S., Kaplin, A., & Hallahan, B. (2017). Neuropsychiatric syndromes of multiple sclerosis. *Journal of Neurology, Neurosurgery and Psychiatry*, *88*(8), 697–708. https://doi.org/10.1136/jnnp-2016-315367

Nagai, A., Hirata, M., Kamatani, Y., Muto, K., Matsuda, K., Kiyohara, Y., Ninomiya, T., Tamakoshi, A., Yamagata, Z., Mushiroda, T., Murakami, Y., Yuji, K., Furukawa, Y., Zembutsu, H., Tanaka, T., Ohnishi, Y., Nakamura, Y., Kubo, M., Shiono, M., … Yanai, H. (2017). Overview of the BioBank Japan Project: Study design and profile. *Journal of Epidemiology*, *27*(3), S2–S8. https://doi.org/10.1016/j.je.2016.12.005

Navarrete, M., & Zhou, Y. (2022). The 14-3-3 Protein Family and Schizophrenia. Frontiers in Molecular Neurosciences, 15. https://doi.org/10.3389/fnmol.2022.857495

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., Estabrook, R., Bates, T. C., Maes, H. H., & Boker, S. M. (2016). OpenMx 2.0: Extended Structural Equation and Statistical Modeling. *Psychometrika*, *81*(2), 535–549. https://doi.org/10.1007/s11336-014-9435-8

Nenadić, I., Meller, T., Schmitt, S., Stein, F., Brosch, K., Mosebach, J., Ettinger, U., Grant, P., Meinert, S., Opel, N., Lemke, H., Fingas, S., Förster, K., Hahn, T., Jansen, A., Andlauer, T. F. M., Forstner, A. J., Heilmann-Heimbach, S., Hall, A. S. M., … Kircher, T. (2020). Polygenic risk for schizophrenia and schizotypal traits in non-clinical subjects. *Psychological Medicine*. https://doi.org/10.1017/S0033291720002822

Ni, G., Zeng, J., Revez, J. A., Wang, Y., Zheng, Z., Ge, T., Restuadi, R., Kiewa, J., Nyholt, D. R., Coleman, J. R. I., Smoller, J. W., Ripke, S., Neale, B. M., Corvin, A., Walters, J. T. R., Farh, K. H., Holmans, P. A., Lee, P., Bulik-Sullivan, B., … Pedersen, N. L. (2021). A Comparison of Ten Polygenic Score Methods for Psychiatric Disorders Applied Across Multiple Cohorts. *Biological Psychiatry*, *90*(9), 611–620. https://doi.org/10.1016/j.biopsych.2021.04.018

Okayama, T., Hashiguchi, Y., Kikuyama, H., Yoneda, H., & Kanazawa, T. (2018). Next-generation sequencing analysis of multiplex families with atypical psychosis. *Translational Psychiatry*, *8*(1). https://doi.org/10.1038/s41398-018-0272-x

Owen, M. J., Craddock, N., & O'Donovan, M. C. (2005). Schizophrenia: genes at last? *Trends in Genetics*, *21*(9), 518–525. https://doi.org/10.1016/j.tig.2005.06.011

Palmer, D. S., Howrigan, D. P., Chapman, S. B., Adolfsson, R., Bass, N., Blackwood, D., Boks, M. P. M., Chen, C. Y., Churchhouse, C., Corvin, A. P., Craddock, N., Curtis, D., di Florio, A., Dickerson, F., Freimer, N. B., Goes, F. S., Jia, X., Jones, I., Jones, L., … Neale, B. M. (2022). Exome sequencing in bipolar disorder identifies AKAP11 as a risk gene shared with schizophrenia. *Nature Genetics*, *54*(5), 541–547. https://doi.org/10.1038/s41588-022-01034-x

Patsopoulos, N. A., Baranzini, S. E., Santaniello, A., Shoostari, P., Cotsapas, C., Wong, G., Beecham, A. H., James, T., Replogle, J., Vlachos, I. S., McCabe, C., Pers, T. H., Brandes, A., White, C., Keenan, B., Cimpean, M., Winn, P., Panteliadis, I. P., Robbins, A., … de Jager, P. L. (2019). Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science*, *365*(6460). https://doi.org/10.1126/science.aav7188

Patten, S., Svenson, L., Metz, L. (2005). Psychotic disorders in MS: Population-based evidence of an association. *Neurology*, *65*(7), 1123–1125.

Peralta, V., Cuesta, M. J., & de Leon, J. (1991). Premorbid personality and positive and negative symptoms in schizophrenia. *Acta Psychiatrica Scandinavica*, *84*(4), 336–339. https://doi.org/10.1111/j.1600-0447.1991.tb03156.x

Pickrell, J. K. (2014). Joint Analysis of Functional Genomic Data and Genome-wide Association Studies of 18 Human Traits. *The American Journal of Human Genetics*, *94*(4), 559–573. https://doi.org/10.1016/j.ajhg.2014.03.004

Pierce, S. B., Walsh, T., Chisholm, K. M., Lee, M. K., Thornton, A. M., Fiumara, A., Opitz, J. M., Levy-Lahad, E., Klevit, R. E., & King, M. C. (2010). Mutations in the DBP-deficiency protein HSD17B4 cause ovarian dysgenesis, hearing loss, and ataxia of perrault syndrome. *American Journal of Human Genetics*, *87*(2), 282–288. https://doi.org/10.1016/j.ajhg.2010.07.007

Purcell, S. M., Moran, J. L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., O'Dushlaine, C., Chambert, K., Bergen, S. E., Kähler, A., Duncan, L., Stahl, E., Genovese, G., Fernández, E., Collins, M. O., Komiyama, N. H., Choudhary, J. S., Magnusson, P. K. E., Banks, E., … Sklar, P. (2014). A polygenic burden of rare disruptive mutations in schizophrenia. *Nature*, *506*(7487), 185–190. https://doi.org/10.1038/nature12975

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*.

Redin, C., Brand, H., Collins, R. L., Kammin, T., Mitchell, E., Hodge, J. C., Hanscom, C., Pillalamarri, V., Seabra, C. M., Abbott, M.-A., Abdul-Rahman, O. A., Aberg, E., Adley, R., Alcaraz-Estrada, S. L., Alkuraya, F. S., An, Y., Anderson, M.-A., Antolik, C., Anyane-Yeboa, K., … Talkowski, M. E. (2017). The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nature Genetics*, *49*(1), 36–45. https://doi.org/10.1038/ng.3720

Rees, E., & Kirov, G. (2021). Copy number variation and neuropsychiatric illness. *Current Opinion in Genetics & Development*, *68*, 57–63. https://doi.org/10.1016/j.gde.2021.02.014

Rees, E., Han, J., Morgan, J., Carrera, N., Escott-Price, V., Pocklington, A. J., Duffield, M., Hall, L. S., Legge, S. E., Pardiñas, A. F., Richards, A. L., Roth, J., Lezheiko, T., Kondratyev, N., Kaleda, V., Golimbet, V., Parellada, M., González-Peñas, J., Arango, C., … Owen, M. J. (2020). De novo mutations identified by exome sequencing implicate rare missense variants in SLC6A1 in

schizophrenia. *Nature Neuroscience*, *23*(2), 179–184. https://doi.org/10.1038/s41593-019-0565-2

Rees, E., Kirov, G., Sanders, A., Walters, J. T. R., Chambert, K. D., Shi, J., Szatkiewicz, J., O'Dushlaine, C., Richards, A. L., Green, E. K., Jones, I., Davies, G., Legge, S. E., Moran, J. L., Pato, C., Pato, M., Genovese, G., Levinson, D., Duan, J., … Owen, M. J. (2014). Evidence that duplications of 22q11.2 protect against schizophrenia. *Molecular Psychiatry*, *19*(1), 37–40. https://doi.org/10.1038/mp.2013.156

Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2019). CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, *47*(D1), D886–D894. https://doi.org/10.1093/nar/gky1016

Revelle, W. (2021). *psych: Procedures for Psychological, Psychometric, and Personality Research*. 1–457. https://cran.r-project.org/web/packages/psych/psych.pdf

Riley, B. (2004). Linkage studies of schizophrenia. *Neurotoxicity Research*, *6*(1), 17–34. https://doi.org/10.1007/BF03033293

Riley, B., Thiselton, D., Maher, B. S., Bigdeli, T., Wormley, B., McMichael, G. O., Fanous, A. H., Vladimirov, V., O'Neill, F. A., Walsh, D., & Kendler, K. S. (2010). Replication of association between schizophrenia and ZNF804A in the Irish Case-Control Study of Schizophrenia sample. *Molecular Psychiatry*, *15*(1), 29–37. https://doi.org/10.1038/mp.2009.109

Ripke, S., Neale, B. M., Corvin, A., Walters, J. T. R., Farh, K. H., Holmans, P. A., Lee, P., Bulik-Sullivan, B., Collier, D. A., Huang, H., Pers, T. H., Agartz, I., Agerbo, E., Albus, M., Alexander, M., Amin, F., Bacanu, S. A., Begemann, M., Belliveau, R. A., … O'Donovan, M. C. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, *511*(7510), 421–427. https://doi.org/10.1038/nature13595

Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J. L., Kähler, A. K., Akterin, S., Bergen, S. E., Collins, A. L., Crowley, J. J., Fromer, M., Kim, Y., Lee, S. H., Magnusson, P. K. E., Sanchez, N., Stahl, E. A., Williams, S., Wray, N. R., Xia, K., Bettella, F., … Sullivan, P. F. (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genetics*, *45*(10), 1150–1159. https://doi.org/10.1038/ng.2742

Ripke, S., Sanders, A. R., Kendler, K. S., Levinson, D. F., Sklar, P., Holmans, P. A., Lin, D. Y., Duan, J., Ophoff, R. A., Andreassen, O. A., Scolnick, E., Cichon, S., st. Clair, D., Corvin, A., Gurling, H., Werge, T., Rujescu, D., Blackwood, D. H. R., Pato, C. N., … Gejman, P. v. (2011). Genome-wide association study identifies five new schizophrenia loci. *Nature Genetics*, *43*(10), 969–978. https://doi.org/10.1038/ng.940

Risch, N. (1990). Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *American Journal of Human Genetics*, *46*(2), 229–241.

Rivas, M. A., Beaudoin, M., Gardet, A., Stevens, C., Sharma, Y., Zhang, C. K., Boucher, G., Ripke, S., Ellinghaus, D., Burtt, N., Fennell, T., Kirby, A., Latiano, A., Goyette, P., Green, T., Halfvarson, J., Haritunians, T., Korn, J. M., Kuruvilla, F., … Daly, M. J. (2011). Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nature Genetics*, *43*(11), 1066–1073. https://doi.org/10.1038/ng.952

Rødevand, L., Bahrami, S., Frei, O., Lin, A., Gani, O., Shadrin, A., Smeland, O. B., Connell, K. S. O., Elvsåshagen, T., Winterton, A., Quintana, D. S., Hindley, G. F. L., Werner, M. C. F., Djurovic, S., Dale, A. M., Lagerberg, T. v., Steen, N. E., & Andreassen, O. A. (2021). Polygenic overlap and shared genetic loci between loneliness, severe mental disorders, and cardiovascular disease risk factors suggest shared molecular mechanisms. *Translational Psychiatry*, *11*(1). https://doi.org/10.1038/s41398-020-01142-4

Roger Bobowick, A., Kurtzke, J. F., Brody, J. A., Hrubec, Z., & Gillespie, M. (1978). Twin study of multiple sclerosis: An epidemiologic inquiry. *Neurology*, *28*(10), 978–987. https://doi.org/10.1212/wnl.28.10.978

Ruan, Y., Lin, Y.-F., Feng, Y.-C. A., Chen, C.-Y., Lam, M., Guo, Z., Ahn, Y. M., Akiyama, K., Arai, M., Baek, J. H., Chen, W. J., Chung, Y.-C., Feng, G., Fujii, K., Glatt, S. J., Ha, K., Hattori, K., Higuchi, T., Hishimoto, A., … Ge, T. (2022). Improving polygenic prediction in ancestrally diverse populations. *Nature Genetics*, *54*(5), 573–580. https://doi.org/10.1038/s41588-022-01054-7

Ruderfer, D. M., Ripke, S., McQuillin, A., Boocock, J., Stahl, E. A., Pavlides, J. M. W., Mullins, N., Charney, A. W., Ori, A. P. S., Loohuis, L. M. O., Domenici, E., di Florio, A., Papiol, S., Kalman, J. L., Trubetskoy, V., Adolfsson, R., Agartz, I., Agerbo, E., Akil, H., … Kendler, K. S. (2018). Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell*, *173*(7), 1705-1715.e16. https://doi.org/10.1016/j.cell.2018.05.046

Sadovnick, A. D., Armstrong, H., Rice, G. P. A., Bulman, D., Hashimoto, L., Party, D. W., Hashimoto, S. A., Warren, S., Hader, W., Murrary, T. J., Seland, T. P., Metz, L., Bell, R., Duquette, P., Gray, T., Nelson, R., Weinshenkar, B., Brunt, D., & Ebers, G. C. (1993). A population-based study of multiple sclerosis in twins: Update. *Annals of Neurology*, *33*(3), 281–285. https://doi.org/10.1002/ana.410330309

Saha, S., Chant, D., & McGrath, J. (2007). A systematic review of mortality in schizophrenia: Is the differential mortality gap worsening over time? *Archives of General Psychiatry*, *64*(10), 1123–1131. https://doi.org/10.1001/archpsyc.64.10.1123

Saitsu, H., Watanabe, M., Akita, T., Ohba, C., Sugai, K., Ong, W. P., Shiraishi, H., Yuasa, S., Matsumoto, H., Beng, K. T., Saitoh, S., Miyatake, S., Nakashima, M., Miyake, N., Kato, M., Fukuda, A., & Matsumoto, N. (2016). Impaired neuronal KCC2 function by biallelic SLC12A5 mutations in migrating focal seizures and severe developmental delay. *Scientific Reports*, *6*. https://doi.org/10.1038/srep30072

Sanchez-Roige, S., Palmer, A. A., Fontanillas, P., Elson, S. L., Adams, M. J., Howard, D. M., Edenberg, H. J., Davies, G., Crist, R. C., Deary, I. J., McIntosh, A. M., & Clarke, T.-K. (2019). Genome-Wide Association Study Meta-Analysis of the Alcohol Use Disorders Identification Test (AUDIT) in Two Population-Based Cohorts. *American Journal of Psychiatry*, *176*(2), 107–118. https://doi.org/10.1176/appi.ajp.2018.18040369

Saunders, J. B., Aasland, O. G., Babor, T. F., de La Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. *Addiction*, *88*(6), 791–804. https://doi.org/10.1111/j.1360-0443.1993.tb02093.x

Shen, Y.-C., Fan, J.-H., Edenberg, H. J., Li, T.-K., Cui, Y.-H., Wang, Y.-F., Tian, C.-H., Zhou, C.-F., Zhou, R.-L., Wang, J., Zhao, Z.-L., & Xia, G.-Y. (1997). Polymorphism of ADH and ALDH Genes among Four Ethnic Groups in China and Effects upon the Risk for Alcoholism. *Alcoholism: Clinical and Experimental Research*, *21*(7), 1272–1277. https://doi.org/10.1111/j.1530-0277.1997.tb04448.x

Shergill, S. S., Brammer, M. J., Williams, S. C. R., Murray, R. M., & McGuire, P. K. (2000). Mapping auditory hallucinations in schizophrenia using functional magnetic resonance imaging. *Archives of General Psychiatry*, *57*(11), 1033–1038. https://doi.org/10.1001/archpsyc.57.11.1033

Siever, L. J., & Gunderson, J. G. (1983). The search for a schizotypal personality: Historical origins and current status. *Comprehensive Psychiatry*, *24*(3), 199–212. https://doi.org/10.1016/0010-440X(83)90070-6

Singh, T., Kurki, M. I., Curtis, D., Purcell, S. M., Crooks, L., McRae, J., Suvisaari, J., Chheda, H., Blackwood, D., Breen, G., Pietiläinen, O., Gerety, S. S., Ayub, M., Blyth, M., Cole, T., Collier, D., Coomber, E. L., Craddock, N., Daly, M. J., … Barrett, J. C. (2016). Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nature Neuroscience*, *19*(4), 571–577. https://doi.org/10.1038/nn.4267

Singh, T., Poterba, T., Curtis, D., Akil, H., al Eissa, M., Barchas, J. D., Bass, N., Bigdeli, T. B., Breen, G., Bromet, E. J., Buckley, P. F., Bunney, W. E., Bybjerg-Grauholm, J., Byerley, W. F., Chapman, S. B., Chen, W. J., Churchhouse, C., Craddock, N., Cusick, C. M., … Daly, M. J. (2022). Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature*, *604*(7906), 509–516. https://doi.org/10.1038/s41586-022-04556-w

Smeland, O. B., Bahrami, S., Frei, O., Shadrin, A., O'Connell, K., Savage, J., Watanabe, K., Krull, F., Bettella, F., Steen, N. E., Ueland, T., Posthuma, D., Djurovic, S., Dale, A. M., & Andreassen, O. A. (2020). Genome-wide analysis reveals extensive genetic overlap between schizophrenia, bipolar disorder, and intelligence. *Molecular Psychiatry*, *25*(4), 844–853. https://doi.org/10.1038/s41380-018-0332-x

Smeland, O. B., Frei, O., Dale, A. M., & Andreassen, O. A. (2020). The polygenic architecture of schizophrenia — rethinking pathogenesis and nosology. *Nature Reviews Neurology*, *16*(7), 366–379. https://doi.org/10.1038/s41582-020-0364-0

Smeland, O. B., Frei, O., Shadrin, A., O'Connell, K., Fan, C. C., Bahrami, S., Holland, D., Djurovic, S., Thompson, W. K., Dale, A. M., & Andreassen, O. A. (2020). Discovery of shared genomic loci using the conditional false discovery rate approach. *Human Genetics*, *139*(1), 85–94. https://doi.org/10.1007/s00439-019-02060-2

Smeland, O. B., Shadrin, A., Bahrami, S., Broce, I., Tesli, M., Frei, O., Wirgenes, K. v., O'Connell, K. S., Krull, F., Bettella, F., Steen, N. E., Sugrue, L., Wang, Y., Svenningsson, P., Sharma, M., Pihlstrøm, L., Toft, M., O'Donovan, M., Djurovic, S., … Andreassen, O. A. (2021). Genome-wide Association Analysis of Parkinson's Disease and Schizophrenia Reveals Shared Genetic Architecture and Identifies Novel Risk Loci. *Biological Psychiatry*, *89*(3), 227–235. https://doi.org/10.1016/j.biopsych.2020.01.026

Smigielski, L., Papiol, S., Theodoridou, A., Heekeren, K., Gerstenberg, M., Wotruba, D., Buechler, R., Hoffmann, P., Herms, S., Adorjan, K., Anderson-Schmidt, H., Budde, M., Comes, A. L., Gade, K., Heilbronner, M., Heilbronner, U., Kalman, J. L., Klöhn-Saghatolislam, F., Reich-Erkelenz, D., … Grünblatt, E. (2021). Polygenic risk scores across the extended psychosis spectrum. *Translational Psychiatry*, *11*(1). https://doi.org/10.1038/s41398-021-01720-0

Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M., & Smoller, J. W. (2013). Pleiotropy in complex traits: Challenges and strategies. *Nature Reviews Genetics*, *14*(7), 483–495. https://doi.org/10.1038/nrg3461

Speed, D., Hemani, G., Johnson, M. R., & Balding, D. J. (2012). Improved heritability estimation from genome-wide SNPs. *American Journal of Human Genetics*, *91*(6), 1011–1021. https://doi.org/10.1016/j.ajhg.2012.10.010

Spielman, R. S., McGinnis, R. E., & Ewens, W. J. (1993). Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American Journal of Human Genetics*, *52*(3), 506–516.

Squires-Wheeler, E., Skodol, A. E., Friedman, D., & Erlenmeyer-Kimling, L. (1988). The specificity of DSM-III schizotypal personality traits. *Psychological Medicine*, *18*(3), 757–765. https://doi.org/10.1017/S0033291700008461

St Clair, D. (2005). Rates of Adult Schizophrenia Following Prenatal Exposure to the Chinese Famine of 1959-1961. *JAMA*, *294*(5), 557. https://doi.org/10.1001/jama.294.5.557

Stefansson, H., Ophoff, R. A., Steinberg, S., Andreassen, O. A., Cichon, S., Rujescu, D., Werge, T., Pietiläinen, O. P. H., Mors, O., Mortensen, P. B., Sigurdsson, E., Gustafsson, O., Nyegaard, M., Tuulio-Henriksson, A., Ingason, A., Hansen, T., Suvisaari, J., Lonnqvist, J., Paunio, T., … Collier, D. A. (2009). Common variants conferring risk of schizophrenia. *Nature*, *460*(7256), 744–747. https://doi.org/10.1038/nature08186

Storey J. (2015). qvalue: Q-Value estimation for false discovery rate control. *R Package Version 2.0.0*.

Stover, D. A., Housman, G., Stone, A. C., Rosenberg, M. S., & Verrelli, B. C. (2022). Evolutionary Genetic Signatures of Selection on Bone-Related Variation within Human and Chimpanzee Populations. *Genes*, *13*(2), 183. https://doi.org/10.3390/genes1302018

Straub, R. E., Jiang, Y., MacLean, C. J., Ma, Y., Webb, B. T., Myakishev, M. v., Harris-Kerr, C., Wormley, B., Sadek, H., Kadambi, B., Cesare, A. J., Gibberman, A., Wang, X., O'Neill, F. A., Walsh, D., & Kendler, K. S. (2002). Genetic Variation in the 6p22.3 Gene DTNBP1, the Human Ortholog of the Mouse Dysbindin Gene, Is Associated with Schizophrenia. *The American Journal of Human Genetics*, *71*(2), 337–348. https://doi.org/10.1086/341750

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine*, *12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

Sun, Y., Zhao, J., Sun, X., & Ma, G. (2021). Identification of TNFAIP8 as an Immune-Related Biomarker Associated With Tumorigenesis and Prognosis in Cutaneous Melanoma Patients. *Frontiers in Genetics*, *12*. https://doi.org/10.3389/fgene.2021.783672

Surakka, I., Horikoshi, M., Mägi, R., Sarin, A. P., Mahajan, A., Lagou, V., Marullo, L., Ferreira, T., Miraglio, B., Timonen, S., Kettunen, J., Pirinen, M., Karjalainen, J., Thorleifsson, G., Hägg, S., Hottenga, J. J., Isaacs, A., Ladenvall, C., Beekman, M., … Ripatti, S. (2015). The impact of low-frequency and rare variants on lipid levels. *Nature Genetics*, *47*(6), 589–597. https://doi.org/10.1038/ng.3300

Suzuki, Y., Jiang, L. L., Souri, M., Miyazawa, S., Fukuda, S., Zhang, Z., Une, M., Shimozawa, N., Kondo, N., Orii, T., & Hashimoto, T. (1997). D-3-hydroxyacyl-CoA dehydratase/D-3-hydroxyacyl-CoA dehydrogenase bifunctional protein deficiency: A newly identified peroxisomal disorder. *American Journal of Human Genetics*, *61*(5), 1153–1162. https://doi.org/10.1086/301599

Szatkiewicz, J., Crowley, J. J., Adolfsson, A. N., Åberg, K. A., Alaerts, M., Genovese, G., McCarroll, S., Del-Favero, J., Adolfsson, R., & Sullivan, P. F. (2019). The genomics of major psychiatric disorders in a large pedigree from Northern Sweden. *Translational Psychiatry*, *9*(1). https://doi.org/10.1038/s41398-019-0414-9

Szustakowski, J. D., Balasubramanian, S., Kvikstad, E., Khalid, S., Bronson, P. G., Sasson, A., Wong, E., Liu, D., Wade Davis, J., Haefliger, C., Katrina Loomis, A., Mikkilineni, R., Noh, H. J., Wadhawan, S., Bai, X., Hawes, A., Krasheninina, O., Ulloa, R., Lopez, A. E., … Ye, Z. (2021). Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. *Nature Genetics*, *53*(7), 942–948. https://doi.org/10.1038/s41588-021-00885-0

Tansey, Katherine., & Hill, Matthew. (2018). Enrichment of schizophrenia heritability in both neuronal and glia cell regulatory elements. *Translational Psychiatry*, *8*(1), 7. https://doi.org/10.1038/s41398-017-0053-y

The "All of Us" Research Program. (2019). *New England Journal of Medicine*, *381*(7), 668–676. https://doi.org/10.1056/NEJMsr1809937

The International Schizophrenia Genomics Consortium. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, *460*(7256), 748–752. https://doi.org/10.1038/nature08185

The Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, *447*(7145), 661–678. https://doi.org/10.1038/nature05911

Therneau, T. M. (2015). coxme: Mixed Effects Cox Models. *R-Package Description.*, 1–14. http://cran.rproject.org/package=coxme%5Cnftp://ftp5.us.postgresql.org/pub/CRAN/web/packages/coxme/vignettes/coxme.pdf

Thomasson, H. R., Crabb, D. W., Edenberg, H. J., Li, T.-K., Hwu, H.-G., Chen, C.-C., Yeh, E.-K., & Yin, S.-J. (1994). Low Frequency of the ADH2*2 Allele among Atayal Natives of Taiwan with Alcohol Use Disorders. *Alcoholism: Clinical and Experimental Research*, *18*(3), 640–643. https://doi.org/10.1111/j.1530-0277.1994.tb00923.x

Tienari, P., Wynne, L. C., Moring, J., Läksy, K., Nieminen, P., Sorri, A., Lahti, I., Wahlberg, K. E., Naarala, M., Kurki-Suonio, K., Saarento, O., Koistinen, P., Tarvainen, T., Hakko, H., & Miettunen, J. (2000). Finnish adoptive family study: Sample selection and adoptee DSM-III-R diagnoses. *Acta Psychiatrica Scandinavica*, *101*(6), 433–443. https://doi.org/10.1034/j.1600-0447.2000.101006433.x

Timms, A. E., Dorschner, M. O., Wechsler, J., Choi, K. Y., Kirkwood, R., Girirajan, S., Baker, C., Eichler, E. E., Korvatska, O., Roche, K. W., Horwitz, M. S., & Tsuang, D. W. (2013). Support for the N -Methyl-D-Aspartate Receptor Hypofunction Hypothesis of Schizophrenia From Exome Sequencing in Multiplex Families. *JAMA Psychiatry*, *70*(6), 582. https://doi.org/10.1001/jamapsychiatry.2013.1195

Toma, C., Torrico, B., Hervás, A., Valdés-Mas, R., Tristán-Noguero, A., Padillo, V., Maristany, M., Salgado, M., Arenas, C., Puente, X. S., Bayés, M., & Cormand, B. (2014). Exome sequencing in multiplex autism families suggests a major role for heterozygous truncating mutations. *Molecular Psychiatry*, *19*(7), 784–790. https://doi.org/10.1038/mp.2013.106

Toyo-Oka, K., Shionoya, A., Gambello, M. J., Cardoso, C., Leventer, R., Ward, H. L., Ayala, R., Tsai, L. H., Dobyns, W., Ledbetter, D., Hirotsune, S., & Wynshaw-Boris, A. (2003). 14-3-3ε is important for neuronal migration by binding to NUDEL: A molecular explanation for Miller-Dieker syndrome. *Nature Genetics*, *34*(3), 274–285. https://doi.org/10.1038/ng1169

Trubetskoy, V., Pardiñas, A. F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T. B., Bryois, J., Chen, C.-Y., Dennison, C. A., Hall, L. S., Lam, M., Watanabe, K., Frei, O., Ge, T., Harwood, J. C., Koopmans, F., Magnusson, S., Richards, A. L., Sidorenko, J., … Bertolino, A. (2022). Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*, *604*(7906), 502–508. https://doi.org/10.1038/s41586-022-04434-5

Tsuang, M. (2000). Schizophrenia: genes and environment. *Biological Psychiatry*, *47*(3), 210–220. https://doi.org/10.1016/S0006-3223(99)00289-9

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I. M., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A. K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., … Pontén, F.

(2015). Tissue-based map of the human proteome. *Science*, *347*(6220). https://doi.org/10.1126/science.1260419

van Rheenen, W., Peyrot, W. J., Schork, A. J., Lee, S. H., & Wray, N. R. (2019). Genetic correlations of polygenic disease traits: from theory to practice. *Nature Reviews Genetics*, *20*(10), 567–581. https://doi.org/10.1038/s41576-019-0137-z

Vassos, E., Sham, P. C., Cai, G., Deng, H., Liu, X., Sun, X., Zhao, J., Murray, R. M., Collier, D. A., & Li, T. (2008). Correlation and familial aggregation of dimensions of psychosis in affected sibling pairs from China. *British Journal of Psychiatry*, *193*(4), 305–310. https://doi.org/10.1192/bjp.bp.107.046037

Verhulst, B., Neale, M. C., & Kendler, K. S. (2015). The heritability of alcohol use disorders: a meta-analysis of twin and adoption studies. *Psychological Medicine*, *45*(5), 1061–1072. https://doi.org/10.1017/S0033291714002165

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *American Journal of Human Genetics*, *101*(1), 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Wainschtein, P., Jain, D., Zheng, Z., Aslibekyan, S., Becker, D., Bi, W., Brody, J., Carlson, J. C., Correa, A., Du, M. M., Fernandez-Rhodes, L., Ferrier, K. R., Graff, M., Guo, X., He, J., Heard-Costa, N. L., Highland, H. M., Hirschhorn, J. N., Howard-Claudio, C. M., … Visscher, P. M. (2022). Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. *Nature Genetics*, *54*(3), 263–273. https://doi.org/10.1038/s41588-021-00997-7

Walder, D. J., Faraone, S. v., Glatt, S. J., Tsuang, M. T., & Seidman, L. J. (2014). Genetic liability, prenatal health, stress and family environment: Risk factors in the Harvard Adolescent Family High Risk for Schizophrenia Study. *Schizophrenia Research*, *157*(1–3), 142–148. https://doi.org/10.1016/j.schres.2014.04.015

Walsh, T., McClellan, J. M., McCarthy, S. E., Addington, A. M., Pierce, S. B., Cooper, G. M., Nord, A. S., Kusenda, M., Malhotra, D., Bhandari, A., Stray, S. M., Rippey, C. F., Roccanova, P., Makarov, V., Lakshmi, B., Findling, R. L., Sikich, L., Stromberg, T., Merriman, B., … Sebat, J. (2008). Rare Structural Variants Disrupt Multiple Genes in Neurodevelopmental Pathways in Schizophrenia. *Science*, *320*(5875), 539–543. https://doi.org/10.1126/science.1155174

Walters, R. K., Polimanti, R., Johnson, E. C., McClintick, J. N., Adams, M. J., Adkins, A. E., Aliev, F., Bacanu, S.-A., Batzler, A., Bertelsen, S., Biernacka, J. M., Bigdeli, T. B., Chen, L.-S., Clarke, T.-K., Chou, Y.-L., Degenhardt, F., Docherty, A. R., Edwards, A. C., Fontanillas, P., … Agrawal, A. (2018). Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nature Neuroscience*, *21*(12), 1656–1669. https://doi.org/10.1038/s41593-018-0275-1

Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F. C. P., Clarke, D., Gu, M., Emani, P., Yang, Y. T., Min, X., Gandal, M. J., Lou, S., Zhang, J., Park, J. J., Yan, C., KyongRhie, S., Manakongtreecheep, K., Zhou, H., … Gerstein, M. B. (2018). Comprehensive functional genomic resource and integrative model for the human brain. *Science*, *362*(6420). https://doi.org/10.1126/science.aat8464

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, *38*(16). https://doi.org/10.1093/nar/gkq603

Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-01261-5

Watson, H. J., Yilmaz, Z., Thornton, L. M., Hübel, C., Coleman, J. R. I., Gaspar, H. A., Bryois, J., Hinney, A., Leppä, V. M., Mattheisen, M., Medland, S. E., Ripke, S., Yao, S., Giusti-Rodríguez, P., Hanscombe, K. B., Purves, K. L., Adan, R. A. H., Alfredsson, L., Ando, T., … Bulik, C. M. (2019). Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nature Genetics*, *51*(8), 1207–1214. https://doi.org/10.1038/s41588-019-0439-2

Weiner, D. J., Nadig, A., Jagadeesh, K. A., Dey, K. K., Neale, B. M., Robinson, E. B., Karczewski, K. J., & O'Connor, L. J. (2022). Polygenic architecture of rare coding variation across 400,000 exomes. *MedRxiv*, 2022.07.06.22277335. https://doi.org/10.1101/2022.07.06.22277335

Weiner, D., Wigdor, E., Ripke, S., Walters, R., Kosmicki, J., Grove, J., Samocha, K., Goldstein, J., Okbay, A., Bybjerg-Gauholm, J., Werge, T., Hougaard, D., Taylor, J., Skuse, D., Devlin, B., Anney, R., Sanders, S., Bishop, S., Bo Mortensen, P., … Robinson, E. (2016). Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *BioRxiv*, 089342. https://doi.org/10.1101/089342

Weissbrod, O., Hormozdiari, F., Benner, C., Cui, R., Ulirsch, J., Gazal, S., Schoech, A. P., van de Geijn, B., Reshef, Y., Márquez-Luna, C., O'Connor, L., Pirinen, M., Finucane, H. K., & Price, A. L. (2020). Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nature Genetics*, *52*(12), 1355–1363. https://doi.org/10.1038/s41588-020-00735-

Wickham, H., Walsh, C., Asherson, P., Taylor, C., Sigmundson, T., Gill, M., Owen, M. J., McGuffin, P., Murray, R., & Sham, P. (2001). Familiality of symptom dimensions in schizophrenia. *Schizophrenia Research*, *47*(2–3), 223–232. https://doi.org/10.1016/S0920-9964(00)00098-0

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., Adams, M. J., Agerbo, E., Air, T. M., Andlauer, T. M. F., Bacanu, S. A., Bækvad-Hansen, M., Beekman, A. F. T., Bigdeli, T. B., Binder, E. B., Blackwood, D. R. H., Bryois, J., Buttenschøn, H. N., Bybjerg-Grauholm, J., … Sullivan, P. F. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*, *50*(5), 668–681. https://doi.org/10.1038/s41588-018-0090-3

Xu, B., Ionita-Laza, I., Roos, J. L., Boone, B., Woodrick, S., Sun, Y., Levy, S., Gogos, J. A., & Karayiorgou, M. (2012). De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nature Genetics*, *44*(12), 1365–1369. https://doi.org/10.1038/ng.2446

Xu, B., Roos, J. L., Dexheimer, P., Boone, B., Plummer, B., Levy, S., Gogos, J. A., & Karayiorgou, M. (2011). Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nature Genetics*, *43*(9), 864–868. https://doi.org/10.1038/ng.902

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, *88*(1), 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Yang, J., Lee, S. H., Wray, N. R., Goddard, M. E., & Visscher, P. M. (2016). GCTA-GREML accounts for linkage disequilibrium when estimating genetic variance from genome-wide SNPs. *Proceedings of the National Academy of Sciences*, *113*(32). https://doi.org/10.1073/pnas.1602743113

Yang, J., Visscher, P. M., & Wray, N. R. (2010). Sporadic cases are the norm for complex disease. *European Journal of Human Genetics*, *18*(9), 1039–1043. https://doi.org/10.1038/ejhg.2009.177

Yengo, L., Vedantam, S., Marouli, E., Sidorenko, J., Bartell, E., Sakaue, S., Graff, M., Eliasen, A. U., Jiang, Y., Raghavan, S., Miao, J., Arias, J. D., Graham, S. E., Mukamel, R. E., Spracklen, C. N., Yin, X., Chen, S.-H., Ferreira, T., Highland, H. H., … Hirschhorn, J. N. (2022). A saturated map of common genetic variants associated with human height. *Nature*. https://doi.org/10.1038/s41586-022-05275-y

Zhang, J.-P., Robinson, D. G., Gallego, J. A., John, M., Yu, J., Addington, J., Tohen, M., Kane, J. M., Malhotra, A. K., & Lencz, T. (2015). Association of a Schizophrenia Risk Variant at the *DRD2* Locus With Antipsychotic Treatment Response in First-Episode Psychosis. *Schizophrenia Bulletin*, *41*(6), 1248–1255. https://doi.org/10.1093/schbul/sbv116

Zheng, J., Erzurumluoglu, A. M., Elsworth, B. L., Kemp, J. P., Howe, L., Haycock, P. C., Hemani, G., Tansey, K., Laurin, C., Pourcain, B. S., Warrington, N. M., Finucane, H. K., Price, A. L., Bulik-Sullivan, B. K., Anttila, V., Paternoster, L., Gaunt, T. R., Evans, D. M., & Neale, B. M. (2017). LD Hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*, *33*(2), 272–279. https://doi.org/10.1093/bioinformatics/btw613

Zhou, H., Sealock, J. M., Sanchez-Roige, S., Clarke, T.-K., Levey, D. F., Cheng, Z., Li, B., Polimanti, R., Kember, R. L., Smith, R. V., Thygesen, J. H., Morgan, M. Y., Atkinson, S. R., Thursz, M. R., Nyegaard, M., Mattheisen, M., Børglum, A. D., Johnson, E. C., Justice, A. C., … Gelernter, J. (2020). Genome-wide meta-analysis of problematic alcohol use in 435,563 individuals yields insights into biology and relationships with other traits. *Nature Neuroscience*, *23*(7), 809–818. https://doi.org/10.1038/s41593-020-0643-5

Zuk, O., Schaffner, S. F., Samocha, K., Do, R., Hechter, E., Kathiresan, S., Daly, M. J., Neale, B. M., Sunyaev, S. R., & Lander, E. S. (2014). Searching for missing heritability: Designing rare variant association studies. *Proceedings of the National Academy of Sciences*, *111*(4). https://doi.org/10.1073/pnas.1322563111