Identification and Evaluation of Endophenotypes and Biomarkers of Schizophrenia and Bipolar Disorder: Genomic Dissection of the Psychosis Phenotype

Eirini Zartaloudi

UCL

Division of Psychiatry

Supervisors: Prof Elvira Bramon, Prof Andrew McQuillin,

Dr Johan Hilge Thygesen and Dr Karoline Kuchenbaecker

A thesis submitted to University College London in fulfilment of the requirements for the degree of Doctor of Philosophy

Declaration

I, Eirini Zartaloudi, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature:  Date: 09/02/2021

Acknowledgements

I would like to take this opportunity to reflect on the people who have supported me during those intensive years of my life. Sincere gratitude is hereby extended to the following people, who never ceased in helping until this thesis was completed.

First and foremost, I wish to express, my immense appreciation to my supervisor Professor Elvira Bramon. She has been extremely supportive, and has been motivating and guiding me from the very beginning of my PhD. She has invested a tremendous amount of time teaching me, signposting me to essential reading and providing me with opportunities to expand my knowledge. Her enthusiasm on her role is spectacular and she surely sets an excellent example as a successful woman, researcher and professor. I owe my deepest gratitude to her for her unceasing encouragement and her unwavering moral and emotional support.

My deepest appreciation to my supervisor Professor Andrew McQuillin for his expert, sincere and precious guidance and encouragement extended to me. He has given me remarkably useful advice and insightful comments and discussions during the course of work, which significantly improved my knowledge in the field of genetics. I will always be indebted to him for his invaluable advice and feedback on my research, and for always being so supportive of my work. His devotion and dedication to his work and the people working with him are inspiring. He is an exceptional researcher and I feel very privileged and honoured having worked with him.

My deepest gratitude goes to my supervisor Dr Johan Thygesen. The joy and enthusiasm he has for his research and data science have been extremely motivational. He has shared his expertise with me very generously and has been very patient with my huge knowledge gaps in genetics. He has taught me everything I know about data science and coding, and I would forever be grateful to him. He would always go out of his way to help me and has been dedicatedly involved in every single step. This thesis would never have been completed if it was not for his continuous support, encouragement, motivation and guidance.

I would also like to sincerely thank my supervisor Dr Karoline Kuchenbaecker for her immensely valuable statistical advice and guidance. She has been extremely supportive and helpful and has signposted me to opportunities which helped me

improve my skills in data science. She is a brilliant researcher and person and I feel honoured having worked with her.

My immense gratefulness to Dr Aritz Irizar, Marie Curie Post-doctoral Fellow, who has greatly supported and guided me during my doctorate. He has contributed in every single part of my thesis and has provided patient guidance, enthusiastic encouragement and immeasurably valuable and constructive critiques and suggestions on this research work. He has taught me a great deal about scientific research and data science and it has been a great honour and pleasure learning from him. I am really grateful for his contribution and guidance.

In particular, my sincere thanks are extended to the rest of the research team I have had the honour of working with for the last 3 years: Dr Olga Gianakopoulou, Professor Nick Bass, Dr Stephanie Mueller, Dr Stella Calafato, Ms Anjali Bhat, Ms Isabelle Austin-Zimmerman and Ms Jasmine Harju Seppanen. I have learned so much from them, they have supported me greatly and have always been willing to help me. It has been an extraordinary experience.

Getting through my PhD required much more than academic support. I owe my deepest gratitude to my parents Nikos Zartaloudis and Patricia Masoura and my grandfather Michalis Zartaloudis for their love and unceasing encouragement. They have always been by my side and have unwaveringly supported all my pursuits. I would not be standing where I am today if it was not for those amazing people I have the privilege of calling family.

Abstract

Background: Psychotic disorders affect approximately 3% of the population. Over 100 genetic variants have been associated with schizophrenia and about 50 with bipolar disorder. Each of them individually has a small effect on disease risk but combined in a cumulative polygenic risk score (PRS), they have a major impact. Copy number variants (CNVs) have also been associated with schizophrenia. However, little is known about their functional effects. The investigation of endophenotypes, which fall in the genotype to phenotype pathway, could help us understand the role of genetic variants and their mechanisms.

Methods: In chapter 1 of my thesis, I reviewed the literature on endophenotypes, and genetic variants associated with psychosis, which revealed that the interrelationships between several well-established cognitive, neuroimaging and electrophysiological psychosis endophenotypes, and the joint contributions of CNV burden and polygenic risk scores on psychosis risk have not been studied yet. I investigated those topics in chapters 3 and 4 respectively. In chapter 2 I carried out a scoping review of CNVs associated with neurodevelopmental disorders, psychosis and cognition and carried out a meta-analysis of 16p11.2 distal deletion in schizophrenia. I also investigated the influences of CNV size on schizophrenia risk for 53 CNVs. For all the analyses, I used CNVcatalog, which is a new repository me and my supervisors created, incorporating data from published studies examining associations of CNV loci with several clinical phenotypes, including schizophrenia. Finally, in chapter 5 I summarise the main findings of my thesis and I discuss the strengths, limitations and clinical implications of my research.

Results: *Chapter 2:* The meta-analysis of 16p11.2 distal deletion in schizophrenia revealed that carriers of that CNV had higher risk of developing schizophrenia compared to non carriers. I also found that larger CNV size was associated with larger effect sizes when examining all CNVs together (both deletions and duplications) and CNV deletions. However, the size was not significanly associated with disease risk for CNV duplications. *Chapter 3:* All the cognitive endophenotypes were associated with each other. Endophenotypes across imaging, cognitive and electrophysiological domains did not show a correlation. The relationships between pairs of endophenotypes were consistent in all three participant groups (cases with psychosis,

their unaffected relatives and healthy controls), differing for some of the cognitive pairings only in the strengths of the relationships. *Chapter 4:* I examined the joint contributions of CNV burden and polygenic risk scores on psychosis risk. I analysed two datasets separately and then combined them by meta-analysis. CNV burden and PRS could explain 11.8% and 10.8% of the variance in disease risk in each dataset. The classification accuracy of my models was 81%, 83% and 77% for the comparisons of all psychosis cases vs controls, schizophrenia cases vs controls and bipolar cases vs controls respectively. The addition of CNV burden to the models increased the variance explained only by 0.1% for MPL dataset and by 0.08% in the PEIC dataset.

Discussion: Findings from my thesis contribute to our current knowledge on psychosis endophenotypes and on the genetic influences in psychoses. Deciphering the genetic architecture of psychotic disorders could hopefully in the future improve the lives of affected individuals.

Impact Statement

In this thesis I have identified biological markers of risk for psychosis. Their future potential impact will help us clarify diagnostic boundaries in mental health disorders, which currently are only based on symptoms. My research gets us closer to a mechanistic way to classify mental disorders.

I have performed the first scoping review and meta-analysis of 16p11.2 distal deletion in schizophrenia, a locus that has been repeatedly associated with the disorder and I examined the influences of CNV size on disease risk. For all the analyses I performed, I used CNVcatalog, which is a new biological database. This database facilitates an assortment of meta-analytical procedures and interactive visualizations of the currently included data, while allowing for easy addition of new data. It also provides a comprehensive set of tools assisting the investigation of associations between rare and pathogenic genomic variants and a range of mental health relevant disorders and traits. Once CNVcatalog is available for public use by clinicians and researchers it will provide a comprehensive set of tools facilitating the investigation of CNV effects on adverse clinical phenotypes.

I have investigated the interrelationships between well-established endophenotypes of psychosis related to cognition, electrophysiology and brain structure. Furthermore, I have examined the endophenotype performance across three diagnostic groups, patients with broadly defined psychosis, their unaffected relatives and healthy controls.

I have also investigated the joint contributions of polygenic risk scores and copy number variations on psychosis risk. Expanding our understanding of the mechanisms by which single polymorphisms and larger genetic variations increase the risk of developing psychoses spectrum disorders could lead to the development of biologically informed treatments and even personalised medication.

Although neither CNV burden nor polygenic risk scores are currently ready for clinical use, it is hoped that as they are refined they could help towards risk reduction advice and early interventions for psychosis. Additional research on well-established biomarkers and genetic variants associated with psychosis, and their combination with environmental risk factors could in future help to develop screening tools, with substantially increased accuracy on assessing psychosis risk. The use of such a tool

by clinicians could lead to earlier detection, which ultimately results into earlier treatment of psychotic disorders, and better outcomes.

This thesis contributes to the genomic dissection of the psychosis phenotype by investigating well-characterised endophenotypes of psychosis, exploring genetic variants associated with broadly and narrowly defined psychosis and by contributing to the development of a repository investigating pathogenic genomic variants. My thesis advances our understanding of the causes and mechanisms underlying psychotic disorders.

Table of Contents

Statement of contributions

During my doctoral studies, I have published eight papers, five of them as a first author. Please see publication list in page 13.

For my PhD, I have been working with pre-existing datasets on EEG (electoencephalography) and genetic data, collected by my supervisors and their colleagues. Therefore, I would like to express my gratitude to Prof Elvira Bramon, Prof Andrew McQuillin and the Psychosis Endophenotypes International Consortium (PEIC), for granting me access to their data.

During my studies, I have received supervision and advice from my supervisors and other researchers at UCL (including Rebecca Jones) and King's College (including Siri Ranlund). I wrote the chapter 1 (Introduction) and chapter 5 (Discussion) of the thesis. Below I outline my contribution to the remaining chapters.

Chapter 2

I was involved in the design and development of the database, as well as the selection of suitable phenotypes and their analyses. Dr Thygesen conceived the tool and wrote the code needed for its function as a dynamic database. I designed the search criteria for the inclusion of papers. I carried out the scoping review to identify the papers populating CNVcatalog. I also extracted the data from the literature and formatted them according to the CNVcatalog database criteria. I helped developing the visualisation and analysis approaches for the application, and trailed proof and error checked the final application. Under the supervision of Dr Thygesen and Prof Bramon, I carried out a meta-analysis to demonstrate how CNVcatalog operates and wrote up the manuscript.

Chapter 3

I was involved in the writing of the manuscript, database management, data quality control, the statistical analysis and interpretation of findings. I wrote the response to reviewers' comments and re-analysed data as requested. I edited the manuscript substantially and resubmitted it. I disseminated the paper by presenting it in the half-

day conference of the Neuroscience in Mental Health Module in 2018 and in the in the Neuroscience Symposium held by UCL in March 2018.

Chapter 4

I was involved in the study design discussions together with Prof Elvira Bramon, Prof Andrew McQuillin, Dr Johan Thygesen and Dr Karoline Kuchenbaecker. I set up the dataset, and performed quality control of the data, liaising with those who collected the data when necessary. After reading patients' case notes and completed clinical scales, I extracted and digitised data for approximately 2,000 individuals. I conducted the statistical analysis, interpreted the findings and wrote up the manuscript. I disseminated the paper by presenting it in the Neuroscience Symposium held by UCL in June 2017. I also presented it in the World Congress of Psychiatric Genomics, in 2018 and 2019. I applied for and received a travel grant from Guarantors of Brain and two grants from the Division of Psychiatry, UCL, to attend both conferences.

Further contributions

I have presented posters in the World Congress of Psychiatric Genetics twice, in Glasgow in October, 2018 and then in Los Angeles in October 2019. I have presented twice at the UCL half-day conference, part of the Neuroscience in Mental Health Module for MSc, PhD students and UCL staff. I participated in the 3-minute thesis presentation competition in the Neuroscience symposium held by UCL.

I have also improved substantially my programming skills and I can now confidently code in R and Python. I am also aware of the basic commands in Bash and SQL. I have been practicing my data science skills by participating in competitions in Kaggle, a data science community.

Publications during my PhD

(*joint first author)

1. Blakey, R., Ranlund, S., **Zartaloudi E. *,** Cahn, W., Calafato, S., Colizzi, M., Crespo-Facorro, B., Daniel, C., Díez-Revuelta, A., Di Forti, M., GROUP, Iyegbe, C., Jablensky, A., Jones, R., Hall, M. H., Kahn, R., Kalaydjieva, L., Kravariti, E., Lin, K., McDonald, C., McIntosh, A., PEIC, Picchioni, M., Powell, J., Presman, A., Rujescu, D., Schulze, K., Shaikh, M., Thygesen, J. H., Toulopoulou, T., Van Haren, N., Van Os, J., Walshe, M., WTCCC2, Murray, R. M. and Bramon, E. (2017). Associations between psychosis endophenotypes across brain functional, structural, and cognitive domains. *Psychological medicine*, 1-20.

2. Calafato, M. S., Thygesen, J. H., Ranlund, S., **Zartaloudi, E.,** Cahn, W,. Crespo-Facorro, B., Diez-Revuelta, A., Di Forti, M., GROUP, Hall, M. H., Iyegbe, C., Jablensky, A., Kahn, R., Kalaydjieva, L., Kravariti, E., Lin, K., McDonald, C., McIntosh, A., McQuillin, A., PEIC, Picchioni, M., Rujescu, D., Shaikh, M., Toulopoulou, T., Van Os, J., Vassos, E., Walshe, M., WTCCC2, Powell, J., Lewis, C., M., Murray, R. M. and Bramon, E. (2018). Use of schizophrenia and bipolar disorder polygenic risk scores to identify psychotic disorders. *British Journal of Psychiatry, 213*(3), 535-541.

3. **Zartaloudi E. *,** Polemikou, A. (2018). Functional hemispheric asymmetry and nicotine dependency as variables mediating neurobiological vulnerability to schizotypy in a non-clinical population of college students. Personality and Individual Differences, *137*, 165-172.

4. Polemikou, A., **Zartaloudi, E.*,** & Polemikos, N. (2019). Estimating nonbelief: Translation, cultural adaptation, and statistical validation of the Nonreligious-Nonspiritual Scale in a nationwide Greek sample. *Archive for the Psychology of Religion*, *41*(2), 105-122.

5. **Zartaloudi, E.,** Laws, K. R., & Bramon, E. (2019). Endophenotypes of executive functions in obsessive compulsive disorder? A meta-analysis in unaffected relatives. *Psychiatric Genetics*, *29*(6), 211-219.

6. Polemikou, A., **Zartaloudi, E.*,** & Polemikos, N. (2019). Development of the Greek version of the Spiritual Intelligence Self-Report Inventory-24 (KAPN): factor structure and validation. *Mental Health, Religion & Culture*, 1-15.

7. Calafato, M. S., Austin-Zimmerman, I., Thygesen, J. H., Sairam, M., Metastasio, A., Marston, L., Abad-Santos, F., Bhat, A., Harju-Seppannen, J., Irizar., H., **Zartaloudi, E.** & Bramon, E. (2020). The effect of CYP2D6 variation on antipsychotic-induced hyperprolactinaemia: a systematic review and meta-analysis. *The Pharmacogenomics Journal*, 1-9.

8. Thygesen, J. H., Presman, A., Harju-Seppanen, J., Irizar, H., Calafato, S., **Zartaloudi E.,** Jones, R., McQuillin A, … & Bramon, E. (2019). Genetic copy number variants, cognition and psychosis: a meta-analysis and a family study. Accepted for publication in "*Molecular Psychiatry*".

List of tables

Table S6 Family sizes.

Table S7 Group interactions on associations between endophenotypes

Table S8 Comparison between full models[1] (in the paper, including age, sex and group) and models excluding age and sex[2]

Table S9 Samples per research centre in PEIC dataset

Table S10 Case to control ratios across the polygenic risk score deciles for PEIC dataset

Table S11 Polygenic risk scores (PRS) deciles and odd ratios (OR) of broadly defined psychotic disorder for PEIC dataset

Table S12 Case to control ratios across the polygenic risk score deciles for MPL dataset

Table S13 Polygenic risk scores (PRS) deciles and odd ratios (OR) of broadly defined psychotic disorder for MPL dataset

Table S14 Meta analyses of odd ratios (OR) for schizophrenia and bipolar risk scores combining MPL and PEIC datasets

Table S15 Comparison of the full regression models to the models without the CNV burden for MPL and PEIC datasets

Table S16 Demographics of the combined sample of MPL and PEIC datasets

Table S17 Schizophrenia associated CNV loci identified in the MPL dataset.

Table S18 Schizophrenia associated CNV loci identified in the PEIC dataset.

Table S19 Schizophrenia associated CNV loci identified in both MPL and PEIC datasets.

Table S20 Distribution of schizophrenia associated CNV carriers in sub-diagnosis groups per dataset.

List of figures

Chapter 1. Introduction

1.1. Psychosis

Psychotic disorders have a lifetime prevalence of approximately 3 - 4.5% (Bogren, Mattisson, Isberg, & Nettelbladt, 2009; Perälä et al., 2008; Tandon, Keshavan, & Narsallah, 2008; van Os et al., 2001; Jongsma, Turner, Kirkbride, & Jones, 2019) and are amongst the most severely debilitating psychiatric disorders. Clinical symptoms include hallucinations, delusional thinking and cognitive impairments, severe enough to impair the individual's daily functioning (American Psychiatric Association, 2013). The most common psychotic disorders are schizophrenia, bipolar disorder and schizoaffective disorder amongst others.

The economic and social burden of psychosis is large and multidimensional. It includes reduced productivity of patients due to impairments (Brown, 2011), increased physical morbidity (De Hert et al., 2011), disability (Gureje, Herrman, Harvey, Morgan, & Jablensky, 2017; Lee, Hong, Shin, & Kwon, 2015; Wolf et al., 2015), mortality (Tiihonen et al., 2009; Tiihonen, Suokas, Suvisaari, Haukka, & Korhonen, 2012), the burden imposed to caregivers (Boydell et al., 2014; Cotton et al., 2013; Gómez-de-Regil, Kwapil, & Barrantes-Vidal, 2014), health sector costs (Ekman, Granstrom, Omerov, Jacob, & Landen, 2013; Neil, Carr, Mihalopoulos, Mackinnon, & Morgan, 2014) and aggression and violent offending (Fazel, Gulati, Linsell, Geddes, & Grann, 2009; Large & Nielssen, 2011).

A number of environmental exposures increase psychosis proneness. Several studies have established a strong association between early stressful and traumatic experiences including physical or sexual abuse, maltreatment, neglect and parental death with later manifestation of psychosis (Morgan & Gayer-Anderson, 2016; Varese et al., 2012; Arsenault et al., 2011). Urban upbringing especially in northern European cities (Eaton, Mortensen, & Frydenberg, 2000; Kirkbride, 2017; Pedersen & Mortensen, 2001), has been repeatedly associated with elevated risk of psychosis with a meta-analysis reporting an odds ratio of 2.39 (Vassos et al, 2012).

Other factors including migration (Dapunt, Kluge, & Heinz, 2017; Kirkbride, 2017) and low socioeconomic status (Grant et al., 2005; Kirkbride et al., 2008; March et al., 2008; Merikangas et al., 2007; Morgan et al., 2008) have also been linked to psychosis.

Increased risk for psychosis has been reported to some minority ethnic groups in UK, including people of Pakistani or Bangladeshi origin, and black Caribbean or African ancestry groups (Kirkbride et al., 2012). However, this does not extend to other ethnic minority groups and countries such as people of Turkish origin living in Netherlands (Cantor-Graae & Selten, 2005) or Hispanic origin living in the US (Oh, Abe & Negi, 2015).

Advanced paternal age (Ek, Wicks, Svensson, Idring, & Dalman, 2015; Torrey et al., 2009) has also been linked to increased risk of developing psychosis with de novo mutations in the sperm-producing cells possibly leading to abnormalities in gene expression (Flanagan et al., 2001). Cannabis use is another well-established environmental risk factor. A meta-analysis by Marconi et al. (2016) provided strong evidence of exposure-response relationship between the extent of cannabis use and the risk of developing psychosis.

Another risk factor is season of birth, with births during winter and spring being 5-8% higher in cases with schizophrenia compared to the general population (Davies, Welham, Chant, Torrey, & McGrath, 2003; Tochigi, Nishida, Shimodera, Okazaki, & Sasaki, 2013). However, all studies supporting the seasonality hypothesis have been conducted in developed countries including UK, US, Denmark and Austria. Therefore, this hypothesis cannot be generalised to developing countries (Wang & Zhang, 2017). Obstetric complications including abnormal foetal growth, low birth weight and complicatiions during delivery have also been linked to psychotic disorders (Abel et al. 2010; Suvisaari et al., 2013; Wahlbeck et al., 2001).

An environmental risk score for psychosis incorporating six risk factors (ethnic minority status, urbanicity, birth weight, cannabis use, paternal age and childhood adversities) was developed in order to capture the combined effect of environmental risk factors (Vassos et al., 2018).

Several susceptibility genetic loci (Dahoun, Trossbach, Brandon, Korth, & Howes, 2017; Harrison & Owen, 2003; Psychosis Endophenotypes International Consortium et al., 2014) and variations in DNA sequence (Bassett, Scherer, & Brzustowicz, 2010; Psychosis Endophenotypes International Consortium et al., 2014) have also been

linked to increased risk for developing psychosis, as is described more extensively in the next pages.

Despite extensive research on all the aforementioned aspects, the mechanisms of the aetiological factors of psychosis have not yet been fully characterised (Matheson, Shepherd, & Carr, 2017). Therefore, diagnoses are still being made based on descriptive clinical principles (American Psychiatric Association, 2013; World Health Organisation, 1993) instead of biologically driven diagnostic tests. While antipsychotic medications have demonstrated high efficacy and have transformed the lives of people with psychosis, some patients do not improve sufficiently on them, and they experience common or severe adverse reactions (Leucht, Arbter, Engel, Kissling, & Davis, 2009; Stefan Leucht et al., 2012; Lieberman et al., 2005; Taylor & Perera, 2015). Thus, it is crucial to expand our understanding of the underlying biological mechanisms and genetic architecture of psychosis, and ultimately to develop better diagnostic tools and more effective treatments.

1.2. Genetic epidemiology of psychosis

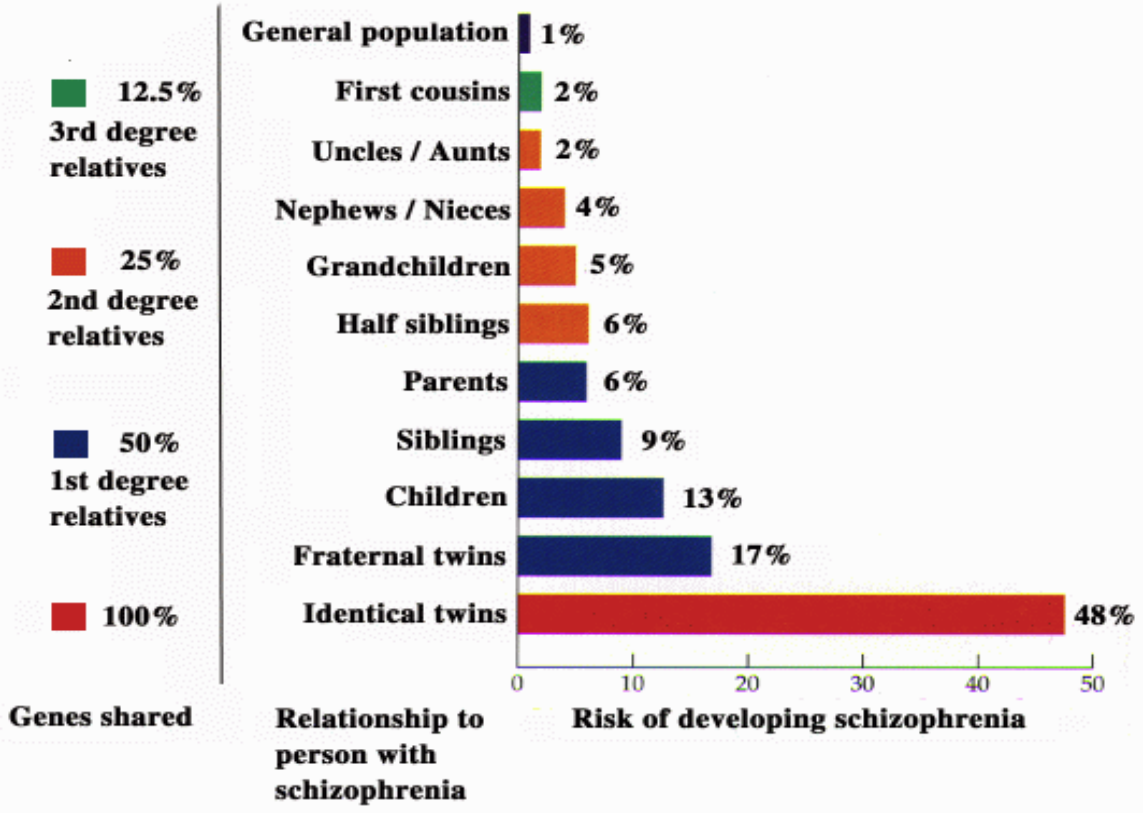
Familial high-risk studies

Familial high-risk (FHR) studies investigate subjects with a severe mental health illness and their unaffected relatives, commonly parents or siblings. The familial aggregation of psychosis is well established and risk for developing a psychotic disorder is increased in patients' unaffected relatives, compared to control families (Braff, Freedman, Schork, & Gottesman, 2006; Goes et al., 2007; Gottesman, 1991; Potash et al., 2003). Family studies have reported that whilst the average lifetime risk for schizophrenia is 1% in the general population, it is 9% for siblings of an individual with schizophrenia, 6% for their parents and 13% for their offspring (Gottesman, 1991; Tandon et al., 2008).

A study of 147 offsprings of healthy individuals and 203 offsprings of patients with psychosis reported that the latter had approximately 6-fold increased risk for developing psychosis (Goldstein, Buka, Seidman, & Tsuang, 2010). A meta-analysis of 33 studies with 3,863 offsprings of patients with schizophrenia or bipolar disorder

and 3,158 offsprings of healthy individuals reported that individuals with a parent with schizophrenia have 12% probability of developing schizophrenia themselves, with a risk ratio of 7.54, and individuals with a parent with bipolar disorder have a 6% probability of developing bipolar disorder, with a risk ratio of 4.06 (Rasic, Hajek, Alda, & Uher, 2014). The risk of developing schizophrenia amongst different degrees of relatedness are demonstrated on figure 1.1.

Figure 1.1 Averaged risks for developing schizophrenia amongst classes of relatives



Lifetime age-adjusted, averaged risks for developing schizophrenia related psychoses amongst classes of relatives of a patient. Data by Gottesman (1991), adapted from Owen and associates (2004).

While higher prevalence of psychotic disorder amongst relatives of patients was corroborated by family studies, this increased risk could be attributable to genetic or shared environmental influences. Twin studies can clarify this matter.

Twin studies

Twin studies compare the concordance in monozygotic (MZ) twins, who share 100% of their genome and dizygotic (DZ) twins who share an average 50% of their genome, while presuming that twin pairs are exposed to the same shared environmental risk factors for psychosis (Cardno & Gottesman, 2000). Studies have reported concordance rates for schizophrenia of approximately 41-65% for monozygotic twins and 0-28% for dizygotic twins (Cardno & Gottesman, 2000; Cardno et al., 1999; Rijsdijk, Gottesman, McGuffin, & Cardno, 2011; Hilker et al., 2018). A meta-analysis of 12 twin studies reported high heritability estimates for schizophrenia, ranging from 73% to 90% (Sullivan, Kendler, & Neale, 2003) and also determined a common familial effect accounting for 11% on liability to schizophrenia, and a shared environmental influence of 8%.

Twin studies consistently report higher concordance rates in MZ rather than DZ twins, providing evidence about the impact of genetic contribution to the liability to schizophrenia. They are important for investigating liability factors between psychosis and other disorders and also for examining the genetic basis of heterogeneity in psychosis. Given that the disease concordance in MZ twins is far lower than 100%, it is also clear that non-genetic factors are equally important. Finally, it should be noted that twin studies assume that both MZ and DZ twins share the same environmental influences. This limitation could be overcomed by conducting adoption studies.

Adoption studies

Adoption studies investigate psychosis in patients and their unaffected relatives, who have been separated by adoption, which minimizes environmental commonalities. Therefore, the increased risk for psychosis is consistent with and can be attributed to genetic influences. Adoption studies have provided evidence for genetic influences in both schizophrenia (Kety et al., 1994; Cardno & Gottesman, 2000; Lichtenstein et al., 2009; Tienari et al., 2003; Wender, Rosenthal, Kety, Schulsinger, & Welner, 1974) and bipolar disorder (Mendlewicz & Rainer, 1977).

A large population register Swedish study of over 2 million families with 35,985 subjects with a diagnosis for schizophrenia and 40,487 subjects with a diagnosis for bipolar disorder investigated familial and adoptive relationships and reported that adopted children with a biological parent or sibling with schizophrenia have a risk of 13.7 (95% CI: 6.1 - 30.8) and 7.6 (95% CI: 0.7 - 87.8) respectively for developing schizophrenia themselves. Increased risk of 4.5 (95% CI: 1.8 - 10.9) and 3.9 (95% CI: 0.2 – 63.3), albeit smaller, was also reported in adopted children with a biological parent or sibling with bipolar disorder (Lichtenstein et al., 2009). Another study reported that adoptees, whose biological mothers had a diagnosis of schizophrenia had a lifetime risk for schizophrenia of 22.46%, in comparison with 4.36% for adoptees whose biological mothers had not received a diagnosis (Tienari et al., 2003).

Adoption studies provide evidence that shared genetic rather than shared environmental influences underlie the elevated risk for relatives of patients to develop schizophrenia. While the high heritability rates of psychosis that have been reported by family, twin and adoption studies indicate the significant role of genetic risk factors, the genetic architecture of psychotic disorders has not yet been fully deciphered.

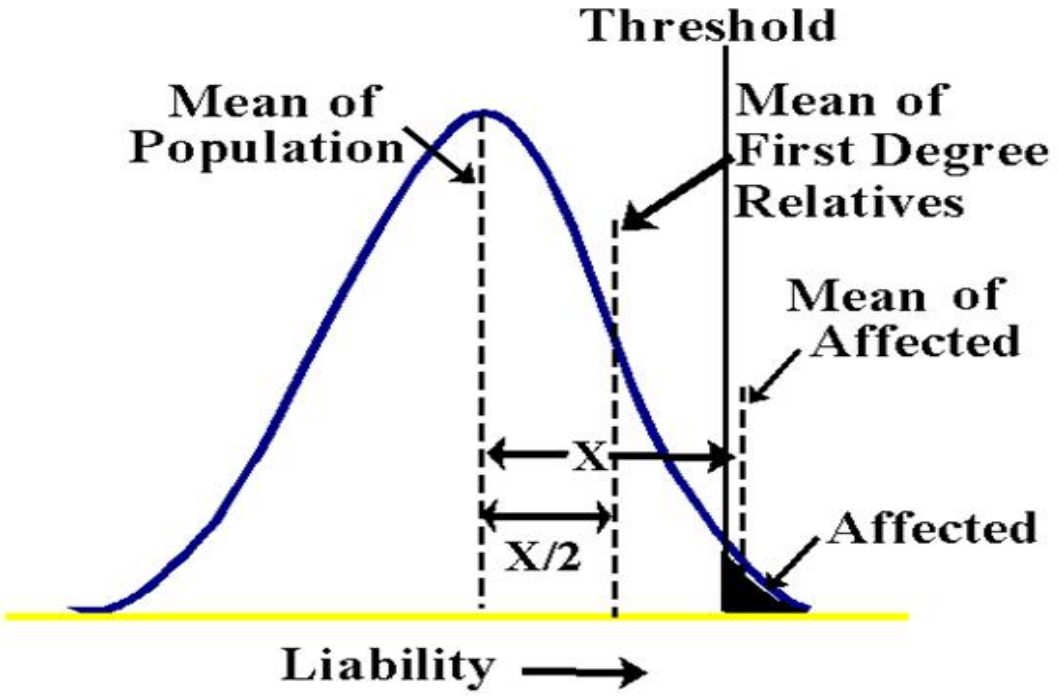
1.3. Mode of inheritance

Research in genetic epidemiology of schizophrenia demonstrates that genetic mechanisms account more than environmental influences for increasing the risk for the disorder. However, even if incomplete penetrance (the condition where some individuals who carry a genetic variant associated with a particular trait, express the associated trait, whereas others do not) and pleiotropy (when one particular gene is associated with multiple unrelated phenotypes) are taken into consideration, schizophrenia does not fit with a single-gene model and classical Mendelian genetics (Gottesman, 1991). The architecture of schizophrenia is highly polygenic involving a large number of risk alleles. Each of those risk factors probably does not suffice to cause schizophrenia on its own, but the more of those factors an individual is burdened with, the higher their susceptibility to the disorder.

A multi-factorial polygenic liability threshold model of schizophrenia was proposed by Gottesman and Shields (1967) supporting that liability for developing schizophrenia has a continuum and several risk factors (both genetic and environmental) act in an

accumulative fashion. As depicted in figure 1.2 once the additive risk factors pass the threshold, one becomes affected. The relatives of affected individuals have an increased liability in comparison with the general population.

Figure 1.2 Graph illustrating the multifactorial threshold model.



Comparisons of the mean liability of affected individuals with their first degree relatives and the general population according to the multifactorial threshold model (Gottesman & Shields, 1967). X represents the difference between the mena liability score of affected individuals and the mean liability score of the general population.

In other multi-factorial illnesses like Alzheimer's disease (Blacker et al., 2003) or breast cancer (Antoniou & Chenevix-Trench, 2010) single genetic mutations causing a Mendelian subform of the illness have been reported in families. However, no Mendelian subform has been reported for schizophrenia or bipolar disorder (Kim, Zerwas, Trace, & Sullivan, 2011, O'Donovan & Owen, 2016).

1.4. Biomarkers of psychosis: The molecular genetics of psychosis

Researchers in molecular genetics focus on identifying DNA risk variants across the genome. They investigate DNA sequences with known chromosomal locations that vary between subjects, which signposts to proximal DNA variants.

Linkage studies

Linkage is the tendency of DNA segments located in close proximity on the same chromosome to be inherited together (Cardno, 2014). Linkage studies have been conducted on families with more than one affected member and examine the inheritance of genes spread along each chromosome, searching for loci with one or more genetic risk variants for schizophrenia (Dawn Teare & Barrett, 2005). These studies begin by investigating possible locations of risk variants rather than focusing on gene functions.

A meta-analysis of 32 linkage studies found loci associated with schizophrenia in 2q, 5q and 8p regions (Ng et al., 2009). Another meta-analysis with a total of 1,286 individuals from 296 families examining 12 regions associated with previously identified schizophrenia endophenotypes, identified several potential genetic loci on chromosomes 3p14, 1p36, 2p25, 16q23, 2p24 2q32, 5p15, 8q24, 10q26, 12p12 and 14q23 (Greenwood et al., 2013).

Association studies

Unlike linkage studies that focus on families with multiple affected individuals, association studies are population-based and typically involve a case-control design. Association studies employ Single Nucleotide Polymorphisms (SNPs), which are small variations in the genome, as genetic markers and examine whether at a certain point in the DNA sequence, the frequency of occurrence of a specific DNA base is different in cases compared to controls. Association studies have identified several candidate genes related to the increased risk of developing psychosis.

Studies have reported associations of several genes with schizophrenia including the DTNBP1 (Weickert et al., 2004; Voisey et al., 2010), the NRG1 (Stefansson et al., 2002; Bousman et al., 2013), the DAOA (formerly known as G72) (Bousman et al., 2013; Chu, 2017), the RGS4 (Chen et al., 2004; Ding, Styblo, Drobna & Hedge, 2016; Williams et al., 2004) and the DISC1 genes (Facal & Costas, 2019; Chen et al., 2004; Dahoun, Trossbach, Brandon, Korth, & Howes, 2017; Hodgkinson et al., 2004).

Polymorphisms in the LRRTM1 gene have also been associated with schizophrenia (Francks, Maegawa, Lauren, Abrahams, Velayos-Baeza et al., 2007). Evidence for allelic methylation in that region has been provided, supporting the notion of this region mediating risk through an interaction of genetic and epigenetic factors (Schalkwyk, Meaburn, Smith, Dempster, Jeffries et al., 2010).

Further genes were positional candidates based on genome-wide linkage or structural variation (*CHRNA7, COMT, DAO, DAOA, NOTCH4, PPP3CC, PRODH,* and *ZDHHC8*) and eight more genes were associated with schizophrenia according to the theory of the aetiology of schizophrenia based on pharmacology (*AKT1, DRD2, DRD3, DRD4, GRM3, HTR2A, SLC6A3,* and *SLC6A4*) (Farrell et al., 2015). However, candidate gene association studies for schizophrenia have a mixed pattern of replication and non-replication (Farrell et al., 2015).

Genome Wide Association Studies

Due to the inconsistent findings yielded by both linkage and association studies, research has been focusing on Genome Wide Association Studies (GWAS) since 2005, following technological advances, including the completion of the Human Genome Project in 2003 and the International HapMap project in 2005. GWAS studies investigate alleles with a higher occurrence in a population with a particular disease compared to unaffected individuals (Craddock, 2013; Fanous, 2010).

The methods employed by linkage studies were not quite accurate for discovering genes associated with schizophrenia, since findings for genes such as the DISC1 could not be supported in subsequent studies and GWAS (Dennison et al., 2020; Mathieson et al., 2012; Sullivan, 2013). Several loci identified by linkage studies,

including 6q23, 6p24, 6q25, 10q24, and 17q21, failed to be replicated by subsequent GWAS (Ripke etl al., 2020; Cariaga-Martinez, Saiz-Ruiz & Alelú-Paz, 2016; Lerer et al., 2003; Williams et al., 2003; Escamilla et al., 2009). Besides, none of the 10 regions that were identified by the latest meta-analysis of linkage studies (Ng et al., 2009) was replicated in the latest schizophrenia GWAS that identified 270 loci (Ripke et al., 2020). To investigate that I annotated the 10 regions identified by Ng et al (2009) to the same genomic build used in the latest GWAS, by using the Human Genome Browser platform.

Therefore, the linkage studies have been superseded by GWAS, that include large samples and can help to identify small effects without any specific knowledge of candidate genes required. This is of extreme importance in the field of psychiatry, considering that the mental illnesses are highly polygenic, being influenced by a large number of genetic variants.

The first significant GWAS finding for schizophrenia was in the zinc finger binding protein gene (ZNF804A) in chromosome 2q (O'Donovan et al., 2008), which was later associated with bipolar disorder as well, suggesting that the ZNF804A gene and its neighbouring genes influence risk to a broader psychosis phenotype (Cardno, 2014). Associations of the ZNF804A gene with both schizophrenia and bipolar disorder were also demonstrated in a large meta-analysis of 18,945 patients with schizophrenia and schizoaffective disorder, 21,274 patients with bipolar disorder and 38,675 healthy subjects (Williams et al., 2011).

Another GWAS of 2,663 schizophrenia patients and 13,498 controls of European ancestry provided evidence for the association of the major histocompatibility complex (MHC) gene on chromosome 6p with schizophrenia and identified SNPs in neurogranin (NRGN) gene in chromosome 11q and the transcription factor 4 (TCF4) gene in chromosome 18q (Stefansson et al., 2009). Associations of the MHC gene with schizophrenia were also reported in another meta-analysis in a European sample of 8,008 cases and 19077 controls (Shi et al., 2009).

Given the polygenic nature of both schizophrenia and bipolar disorder, the risk alleles have rather small effect sizes, lower than 1.5 (Sullivan et al., 2012). The small sample sizes in early GWAS studies could have resulted in lack of statistical power to detect

variants reaching genome-wide significance. In order to increase the sample sizes, and thus their statistical power to identify novel loci, Psychiatric Genomic Consortiums (PGC) have been formed for several illnesses (including schizophrenia and bipolar disorder), combining their multi-site data by mega analyses.

A mega-analysis of over 50,000 individuals from the schizophrenia PGC identified 5 new loci associated with schizophrenia (1p21.3, 2q32.3, 8q21.3, 8p23.2, and 10q24.32-q24.33) and 2 that have been previously implicated (6p21.32-p22.1 and 18q21.2) (Ripke et al., 2011). Additional evidence for loci associated with both schizophrenia and bipolar disorder were identified (CACNA1C, ITIH3-ITIH4 and ANK3). It should be noted that this study failed to replicate associations of ZNF804A with either schizophrenia or bipolar disorder.

Another large GWAS study followed by a meta-analysis of previously schizophrenia associated loci, identified 13 new loci, one of which has also been associated with bipolar disorder (Ripke et al., 2013). Among the 13 loci, it was also the human leukocyte antigen (HLA) locus on chromosome 6, which is also known in mice as the major histocompatibility complex (MHC). HLA proteins mediate the responses of T-lumphocyte cells (the cells that regulate the body's immune response to antigens, including bacteria and viruses) and there is evidence that genetic variability poses a risk factor for several autoimmune and infectious diseases (Mokhtari & Lachma, 2016). Loci in the MHC region have also been linked to schizophrenia by many other studies (Stefansson et al., 2009; Shi et al., 2009; Bergen et al., 2012; Jia et al., 2012). These findings support the hypothesis of schizophrenia resulting from an infectious or an autoimmune disease in a subgroup of cases.

A GWAS in Ashkenazi Jews with 904 schizophrenia patients and 1,640 healthy controls, identified a novel schizophrenia locus near the NDST3 gene, which was replicated in six cohorts with schizophrenia and five cohorts with bipolar disorder (Lencz et al., 2013). This gene is involved in binding affinity to NRG1 gene, which has been repeatedly associated with schizophrenia (Cho et al., 2015) and bipolar disorder (Rolstad et al., 2015). The association of NDST3 gene with schizophrenia was also replicated in a Han Chinese population with two datasets consisting of i) 632 cases with schizophrenia, 654 case swith BD and 684 controls and ii) 2,522 cases with schizophrenia and 547 healthy controls (Zhang et al., 2016).

A further GWAS on schizophrenia with 11,260 cases and 24,542 healthy controls identified a total of 145 loci associated with the disorder, 50 of which were novel (Pardinas, 2018). In the latest PGC mega-analysis of GWA studies on schizophrenia, with the largest sample so far cosisting of 69,369 cases and 236,642 controls, they linked 270 loci with the disorder implicating 130 genes (Ripke et al., 2020). Associations were enriched in genes associated with rare disruptive coding variants in cases with schizophrenia, including the glutamate receptor subunit (GRIN2A) and transcription factor SP4. Several of these genes have also been associated with other clinical phenotypes, including autism spectrum disorders and neurodevelopmental disorders, providing evidence for the pleiotropic effect of these genes.

A systematic review of 22 GWAS on schizophrenia and bipolar disorder, reported that the genes AMBRA1, ANK3, ARNTL, CDH13, EFHD1, MHC, PLXNA2 and UGT1A1 have been associated with both disorders in at least two independent samples, endorsing the theory of a common genetic basis between them (Prata et al., 2019).

The latest PGC mega analysis of GWAS on bipolar disorder included 20,352 cases and 31,358 controls, with a follow up analysis in 9,412 cases and 137,760 controls identified 30 loci associated with bipolar disorder in the combined analysis (Stahl et al., 2018). Twenty of these loci were novel. After performing pathway analysis, they found enrichment in nine gene sets, including regulation of insulin secretion and endocannabinoid signaling (MAP kinase and GABA-A receptor subunit genes).

A further GWAS in bipolar disorder in two datasets consisting of 20,352 and 31,358 controls, and 7,481 cases and 9,250 controls respectively, identified 52 transcription factor binding regions (TFBRs) genes, 44 topologically associated domains (TADs) genes, 55 chromatin interactive regions (CIRs) genes and 21 long non-coding RNA regions (lncRNAs) genes, including the ITIH4, ITIH3, SYNE1 and OPRM1 genes (Qi et al., 2020).

In summary, GWAS have identified over 270 genetic loci associated with schizophrenia (Ripke et al., 2020, Pardinas, 2018, Andreassen et al., 2013; Bramon et al., 2014; Ripke et al., 2014; Rudelfer, 2013; Steinberg et al., 2014); and 30 loci associated with bipolar disorder (Andreassen et al., 2013; Bramon & et al, 2014;

Geschwind & Flint, 2015; Psychiatric GWAS Consortium Bipolar Disorder Working Group, 2011; Rudelfer, 2013; Stahl et al., 2018).

Polygenic risk scores (PRS)

Genome wide association studies have shown that several phenotypes, including schizophrenia and bipolar disorder, are highly polygenic, with their genetic basis comprising of small effects of many genetic variants (Euesden, Lewis & O'Reilly, 2014). However, the odds ratios of each locus range from 1.1 to 1.2 and their predictive power individually is extremely small (Geschwind & Flint, 2015; Harrison, 2015; Purcell et al., 2014).

Therefore, the calculation of a polygenic risk score (PRS), which combines all these loci, has been suggested as a way to investigate their joint effect on disease risk. PRS can be calculated for several traits by using genome-wide genetic data and imputation methods to combine all the variants into a single score, reflecting each individual's personal genetic susceptibility to that particular trait (Dudbridge, 2013).

Several studies have implemented PRSs in their models to predict case-control status, both for schizophrenia and bipolar disorder, and it has shown to be highly predictive (Ohi et al., 2020; Bergen et al., 2019; Calafato et al., 2018; Derks et al., 2012; Tesli et al., 2014; Trotta et al., 2016; Vassos, Forti, et al., 2017). However, despite the extensive evidence for PRSs being able to identify individuals at high risk of developing psychosis, their predictive power is still not high enough to be considered for implementation into clinical practice.

Despite the heritability rates for schizophrenia being as high as 80% (Hilker et al., 2018), the common variants identified by GWAS only account for up to 22.5% of the variance explained on disease risk (Pardinas et al., 2018). This has resulted in a rather challenging problem called missing heritability, which is the gap between the heritability estimates from twin studies, and the heritability estimates from genotype data. A suggestion to overcome this could be the investigation of rare genetic variants.

Chromosomal anomalies and Copy Number Variants (CNVs)

While GWASs focus on detecting genetic markers with a variation in a single DNA base, CNV studies investigate larger genetic variants, which are rarer but have a greater influence on risk of developing psychosis (McCarroll et al., 2006, 2008). Copy number variants (CNVs) are segments of DNA sequence that are deleted or duplicated, altering the diploid status of DNA (Bagshaw et al., 2013). They can range from one kilobase to several megabases in size (Stankiewicz & Lupski, 2010). Whilst many CNVs are benign and part of natural human variation, if the deletion or duplication affects a dosage-sensitive gene, this can result to changes in gene expression and protein function. Some CNVs are implicated in a range of diseases and syndromes (Gordovez & McMahon, 2020; Chen et al., 2016; Li et al., 2016; Marshall et al., 2017; Priebe et al., 2013; Szatkiewicz et al., 2014; Rees et al., 2016).

Several CNVs have been repeatedly associated with increased risk of developing psychosis (Flomen et al., 2006; Kirov et al., 2009; Levinson et al., 2011). Identifying CNVs increasing psychosis risk is quite challenging since they are rare and very large samples are required for potential associations to be investigated (Stranger, Stahl, & Raj, 2011). Furthermore, CNVs associated with mental health disorders are not fully penetrant and are also present in healthy subjects (Morrow, 2010).

Despite the aforementioned difficulties, there is substantial evidence for CNVs being associated with schizophrenia (Chen et al., 2016; Giaroli, Bass, Strydom, Rantell, & McQuillin, 2014; Green et al., 2016; Kirov et al., 2014; Li et al., 2016; Marshall et al., 2017; Priebe et al., 2013; Sriretnakumar et al., 2019; Stefansson et al., 2014; Stone, O'Donovan, Gurling, Kirov, Blackwood, Corvin, Craddock, Sklar, et al., 2008; Sullivan, Daly, & O'Donovan, 2012a; Szatkiewicz et al., 2019; Szatkiewicz et al., 2014; The international Schizophrenia Consortium, 2008; Walsh et al., 2008), and for some, albeit fewer in number, with bipolar disorder (Gordovez & McMahon, 2020; Chen et al., 2016; Green et al., 2016; Karlsson et al., 2012). CNV studies have provided evidence of genetic overlap between schizophrenia, autism spectrum disorders, attention deficit hyperactivity disorder (ADHD) and intellectual disability since those disorders are all associated with CNVs in related chromosomal regions (Burbach & van der Zwaag, 2009; Geschwind, 2011; Moreno-De-Luca et al., 2010; Stefansson et al., 2014; Williams et al., 2010).

The most widely reported CNV associated with schizophrenia is the chromosome 22q11.2 microdeletion (Balan et al., 2014; Bassett, Chow, & Weksberg, 2000; Chow et al., 2011; Forsyth et al., 2019; Goes & Sawa, 2017; Karayiorgou et al., 1995; Stefansson et al., 2008). Further studies have identified CNVs in several loci including 1q21, 3q29, 15q11, 15q13 and 16q11 (Levinson et al., 2011; McCarthy et al., 2009; Stefansson et al., 2008). CNVs disturbing the neurexin1 (NRXN1) gene have also been identified in several studies for schizophrenia risk (Rujescu et al., 2009; Vrijenhoek et al., 2008; Walsh et al., 2008). The probability of schizophrenia patients having exonic CNVs in NRXN1 was higher than healthy subjects, with an odds ratio of 9.97 (Rujescu et al., 2009).

Another study with 13,198 subjects reported deletions in 1q21.1, NRXN1, 15q11.2 and 22q11.2 and duplications at 16p11.2 and the Angelman/Prader-Willi Syndrome (AS/PWS) region in schizophrenia patients (Rees et al., 2014). A recent GWAS of 41,621 subjects also found higher rates of CNVs in schizophrenia patients compared to healthy subjects. They identified deletions in 15q13.2–13.3, 22q11.21 and 1q21.1 and duplications in 16p11.2 and 1q21.1 (Marshall et al., 2017). Table 1.1 presents all the schizophrenia associated CNVs from the studies of Marshall et al. (2017), Kirov et al. (2014) and Stefansson et al. (2014), along with the associated genes, effect sizes and frequencies in the population.

Table 1.1 Schizophrenia associated CNV loci

Locus	Chromosome	Gene of Interest	Odds Ratio	Frequency in controls	Reference
1q21.1.del	chr1		3.8-8.1	0.02-0.07	1,2,3
1q21.1.dup	chr1		2.9-4.2	0.03-0.07	1,2,3
2p25.3.dup	chr2	MYT1L	15.7		3
2p16.del	chr2	NRXN1	10.7-14.4	0.014	1,2,3

3q29.del	chr3	PAK2, DLG1	18-63	0-0.001	1,2,3
7q11.23	chr7		16.1	0.004-0.28	1,2
7q36.3.del	chr7	VIPR2	3.5	0.029	1
7q36.3.dup	chr7	VIPR2	3.2-3.5	0.029	2,3
8q22.2	chr8	VPS13B	14.5	0.004	1
9p24.3.del	chr9	DMRT1	12.4	0.004	1
9p24.3.dup	chr9	DMRT1	12.4	0.004	1
15q11.2.del	chr15	CYFIP1	1.8-2.1	0.25-0.27	1,2,3
15q11.2-13.1.dup	chr15		5.1		2,3
15q13.3.I.del	chr15	CHRNA7	4.7-15.6	0.009	1,2,3
15q13.3.II.del	chr15	CHRNA7	14.9		3
16p13.11.dup	chr16	NTAN1, NDE1	2-2.2	0.13	2,3
16p13.11.del	chr16	NTAN1, NDE1	1.9		3
16p12.1.del	chr16		1.8		3
16p11.2.distal.del	chr16		2.6-20.6	0.004-0.01	1,2,3
16p11.2.del	chr16		0.5-0.9	0.04	2,3
16p11.2.dup	chr16		8-9.4	0.03	1,2,3
17p12.del	chr17		5.7	5.7	3

17q12.del	chr17	HNF1B	4-9.5	0.005	2,3
17q12.dup	chr17	HNF1B	2		3
22q11.21.large.del	chr22		67.7	0.04	1,2
22q11.21.del	chr22		Inf		3
Xq28.distal.dup	chrX		0.35	0.18	1

The loci comprise all schizophrenia associated loci from (1) Marshall et al. (2017), (2) Kirov et al. (2014) and (3) Stefansson et al. (2014), excluding protective loci 22q11.21.dup, 7q11.21.del 7q11.21.dup, 13q12.11.dup, Xq28.dup.

Despite findings being less clear, 30 CNVs have also been associated with bipolar disorder (Chen et al., 2016; Green et al., 2016; Karlsson et al., 2012; Ruderfer et al., 2018; Stahl et al., 2018). Frequencies of de novo CNVs were significantly higher in individuals with bipolar disorder in comparison with healthy subjects, with an odds ratio of 4.8 (Malhotra et al., 2011b). Another study with 6,882 individuals with schizophrenia, 2,591 individuals with bipolar disorder and 8,842 healthy controls reported that three previously schizophrenia associated CNV loci, duplications in 1q21.1 and 16p11.2 and deletions in 3q29, were also associated with bipolar disorder (Green et al., 2016).

1.5. The endophenotype concept in psychiatric illness

Despite the identification of these genetic loci and rare variants, little is known about their functional roles and the mechanisms through which they lead to the disease (Owens, Bachman, Glahn, & Bearden, 2016). This led to the proposal of alternative approaches introducing the concept of investigating endophenotypes rather than simply the presence/absence of disease. Endophenotypes are heritable biological markers that constitute intermediate or mediator traits between genetic factors and clinical phenotypes (Gottesman & Shields, 1973; Gottesman & Gould, 2003) that could

help us gain a better understanding of the underlying neurobiology of psychiatric disorders (Cannon & Keller, 2006; Gottesman & Gould, 2003; Gur et al., 2007).

Endophenotypes are biological markers which are heritable, co-segregate with a disorder within families, are observed in unaffected family members at a higher rate than in the general population and are expressed in an individual whether or not the illness is active (Gottesman & Gould, 2003). They could thus be used to better understand the mechanisms underlying the associations between genetic variants and the disorder of interest (Braff, 2015; Hall & Smoller, 2010).

The notion behind endophenotypes was that even if those traits are determined by multiple genes, their genetic architecture would be simpler than the architecture of the disease (Flint & Munafò, 2007; Lenzenweger, 2013). In the past it had been theorised that endophenotypes should resemble the physiological trait they are associated with and involve the same biochemical pathways but be closer to the level of gene action compared to the psychiatric disorder (Almasy & Blangero, 2001; Flint & Munafò, 2007; Glahn et al., 2014). Therefore, according to that notion, the relationship between genes and the endophenotypes should be stronger than with the disorder itself, since psychiatric disorders result from a combination of genetic and non-genetic abnormalities impacted by environmental and socio-cultural factors.

However, a review on 17 well characterised endophenotypes suggests that endophenotypes are also highly polygenic and could also be influenced by rare genetic variants (Iacono, Vaidyanathan, Vrieze and Malone, 2014). Therefore, even if endophenotypes were theoretically more straightforward than psychiatric disorders and were closer to the underlying biological influences, their genetic architecture is still rather complex for them to be used for gene discovery.

The criteria currently used for a trait to be a successful endophenotype are summarized below in table 1.2.

Table 1.2 Criteria for a trait to be a useful endophenotype for genetic research into a related disorder (Cannon & Keller, 2006; Garver, 1987; Gottesman & Gould, 2003; Kendler & Neale, 2010; Wickham & Murray, 1997).

Criteria for a trait to be an endophenotype for a disorder
• The endophenotype should segregate with the disease in the population
• If the disorder is heritable, the endophenotype should be heritable as well
• If heritable, both the disorder and the endophenotype co-segregate in
• The endophenotype should be more prevalent in unaffected relatives of
• The endophenotype should be assessed reliably in affected and unaffected individuals
• The endophenotype should be congruent with current knowledge of the
• The endophenotype should be non-invasive
• The endophenotype should be prevalent at a higher rate within affected families than in the population

The investigation of endophenotypes is hoped to increase the statistical power to detect more schizophrenia associated genes, improve the nosology by identifying biologically defined subgroups, which are currently diagnosed as “functional psychoses” and also shed light to the mechanisms and function of susceptibility genes for schizophrenia.

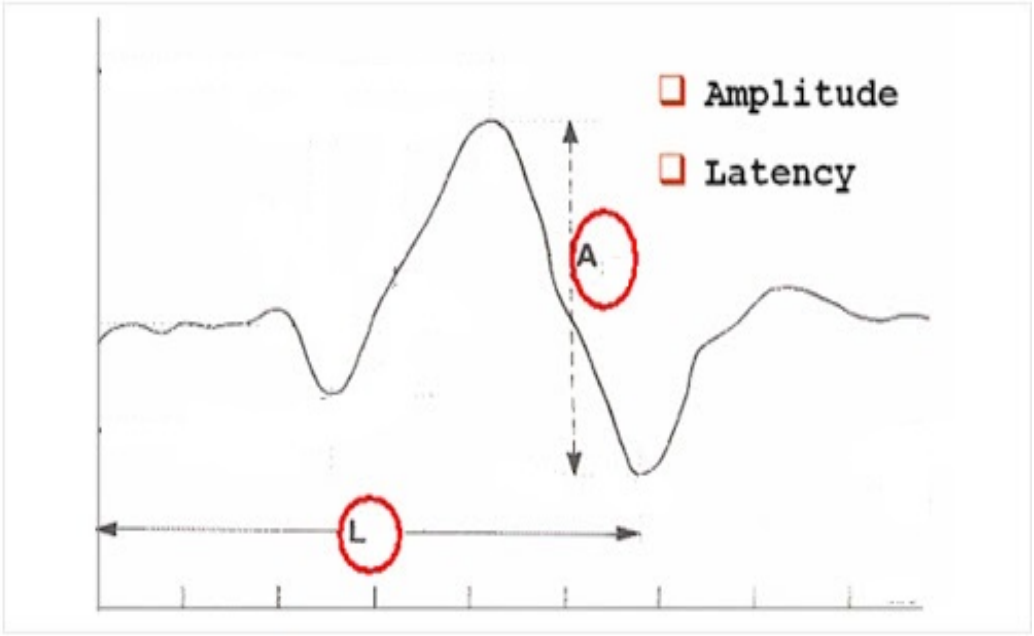
1.6. Endophenotypes associated with psychosis

Potential endophenotypes of psychosis include neuroanatomical, cognitive and electrophysiological measures. Some of the most widely investigated psychosis endophenotypes are:

a) P300 wave amplitude and latency: P300 Amplitude reflects the amount of attention required in a specific task proportionally to the stimulus information. Therefore, greater attention produces larger P300 waveforms. The P300 amplitude is thought to be a correlate of attention and working memory (Ford, 2014; Näätänen, 1990).

P300 Latency can vary according to the difficulty of discriminating the target stimulus from the standard stimuli in the oddball paradigm. Although the latency has been less precisely characterized, it is thought to index classification speed, simply demonstrating how quickly the individual responded to the stimulus (Polich, 2007, 2011). The normal peak latency in a discrimination task is 300ms for a young adult. However, in patients with psychosis the latency is prolonged and occurs later than in age matched healthy controls. Figure 1.3 demonstrates the P300 event related potential and provides a visual presentation of the amplitude and the latency measures.

Figure 1.3 Figure demonstrating the waveform of the P300 event related potential



The A symbolises the P300 amplitude, and the L symbolises the P300 latency. The x axis measures latency in msec, and the Y axis measures amplitude in μv (a unit of electomotive force).

Reduced amplitude and prolonged latency of the P300 wave have consistently been found in patients with psychotic illnesses as well as in unaffected relatives, compared to controls (Blakey et al., 2018; Bodatsch et al., 2015; Earls et al., 2016; Bestelmeyer, Phillips, Crombie, Benson, & Clair, 2009; Blackwood, St Clair, Muir, & Duffy, 1991; Bramon et al., 2005; Díez et al., 2013; Light et al., 2015; Pierson, Jouvent, Quintin, Perez-Diaz, & Leboyer, 2000; Price et al., 2006; Schulze et al., 2008; Turetsky et al., 2015; Weisbrod, Hill, Niethammer, & Sauer, 1999; Winterer et al., 2003).

b) Sensory gating (P50): The P50 measures pre-attentive brain response to stimuli, usually with a paired click task in which two stimuli are presented, separated by an interval of 500ms (Anfred, 2006; Van Tricht et al 2015). The amplitude on the first click is thought to measure the ability of registering salient stimuli, and the second click measures the suppression of irrelevant stimuli (Bramon, 2004). Reduced suppression on the second click have been repeatedly reported in cases with psychosis compared to healthy individuals (Cheng, Chan, Liu, & Hsu, 2016; Earls, Curran, & Mittal, 2016; Gooding, Gjini, Burroughs, & Boutros, 2013; Sánchez-Morla et al., 2008).

c) Mismatch negativity: Mismatch negativity (MMN) is an event related potential, measuring pre-attentive information processing. It occurs when a stimulus deviates from a repetitive pattern of standard stimuli either in frequency, intensity, or duration (Näätänen, 2012). Patients with psychosis and their unaffected relatives have consistently shown smaller MMN compared to controls (Shelley et al, 1991; Näätänen et al, 2012; Bodatsch, Brockhaus-Dumke, Klosterkötter, & Ruhrmann, 2015; Erickson et al, 2016, Ranlund et al., 2016, Earls et al., 2016).

d) Pre-pulse inhibition of the startle response: Pre-pulse inhibition (PPI) is a measure of sensorimotor gating, measured by presenting a weak pre-stimulus, followed by a stronger startle stimulus (Menna et al., 2016). The inclusion of a

pre-stimulus results to a decrement in the startle response. Cases with psychosis and their unaffected relatives have been reported to have deficits in pre-pulse inhibition compared to healthy subjects (Ivleva et al., 2014; Menna et al., 2016; Morales-Muñoz et al., 2017; Notaras, Vivian, Wilson, & van den Buuse, 2017).

e) Antisaccade performance: In the antisaccade paradigm, participants are visually presented with an erroneous stimulus, which they are required to suppress, and instead to make eye movement towards the opposite hemifield (Cutsuridis, Kumari & Ettinger, 2014). Therefore, it measures two decision making processes: inhibition and volitional saccade towards the mirror location. Several studies provide evidence that in comparison with healthy individuals, cases with schizophrenia make more antisaccade errors and their response time is larger, indicating deficits in inhibition (Bodatsch et al., 2015; Ivleva et al., 2014).

f) Lateral ventricular enlargement: Structural brain abnormalities in cases with schizophrenia, especially ventricular dilation, are well established (Vita et al., 2006; Wright et al., 2000). Ventricular volumes are measured by Magnetic Resonance Imaging (MRI) scans. Despite the fact that several studies have reported ventricular enlargement in patients with psychosis (Blakey et al., 2018; Fusar-Poli et al., 2012; Haijma et al., 2013; Kempton, Stahl, Williams, & DeLisi, 2010) it is not clear whether there is a neurogenerative effect or whether this results from the antipsychotic medication. A meta-analysis of longitudinal studies of cases with schizophrenia, provided evidence of progressive ventricular enlargement, which was greater in cases compared to controls (Kempton et al., 2010).

g) Cognitive endophenotypes have been thoroughly characterised in psychotic disorders, with many studies reporting that both patients with broadly defined psychosis and to a lesser extend their unaffected relatives show cognitive impairments compared to controls in a range of cognitive domains, (Calafato et al, 2018, Thygesen et al, 2020, Fusar-Poli et al., 2012, Leeson et al., 2011).

All the aforementioned endophenotypes have exhibited heritability and familial aggregation.

Several studies have examined relationships between psychosis endophenotypes, mainly between cognitive pairings (Dickinson, Iannone, & Gold, 2002; Dickinson, Ragland, Calkins, Gold, & Gur, 2006; Gladsjo et al., 2004; Seidman et al., 2015; Sheffield et al., 2014; Sullivan et al., 2003; Toomey et al., 1998), and also, to a lesser extend, between electrophysiological and cognitive measures (Dong, Reder, Yao, Liu, & Chen, 2015; Fjell & Walhovd, 2001; Hermens et al., 2010; Kaur et al., 2011). These interrelationships between psychosis endophenotypes are presented in chapter 3, which specifically aimed at investigating those relationships, in the first study to examine endophenotype pairs while including not only cases with psychosis but also their unaffected relatives.

1.7. Endophenotypes and polygenic risk scores

The relationship between psychosis endophenotypes and polygenic risk scores has been investigated by several studies. Hubbard et al (2016) calculated polygenic risk scores for both schizophrenia and IQ performance for over 5,000 childen from the Avon Longitudinal Study of Parents and Children birth Cohort. The children had also completed the Wechsler Intelligence Scale for Children (WISC-III; Weschler, Golomboc and Rust, 1992). They reported that schizophrenia PRS was strongly associated with lower performance IQ and lower full IQ. Mcintosh et al (2013) also investigated the relationship between IQ and schizophrenia PRS in the Lothian Birth Cohort 1936, consisting of 937 individuals. They reported schizophrenia PRS being negatively associated with IQ performance at age 70 but not at age 11, and also with greater decline in general cognitive ability between the ages of 11 and 70.
Hall et al. (2014) calculated PRS scores for schizophrenia and bipolar disorder in 271 cases with schizophrenia or bipolar with psychotic features and 128 controls and investigated their relationship with ERP endophenotypes including the P300 ERP, gamma oscilations and the P50. They reported that cases with high schizophrenia PRS exhibited reduced gamma response, whereas cases with high bipolar PRS had reduced P300 amplitude. Another study by Casera et al. (2015) investigated the

relationship between schizophrenia PRS and ventricular volumes in a sample of 274 healthy individuals. They did not find a significant relationship between participants with high polygenic scores and ventricular enlargement, and supported that enlarged ventricles could simply be an epiphenomenon of the illness and not an endophenotype.

A study by Ranlund et al. (2017) investigated the relationship between polygenic risk scores for schizophrenia and bipolar disorder with several well-established psychosis endophenotypes including the P300 event related potential, lateral ventricular volumes and cognitive performance. Their sample consisted of 1,087 cases with psychosis, 822 of their unaffected relatives and 2,333 controls. They reported that higher PRS scores for the schizophrenia but not for bipolar disorder were associated with impaired performance on one of the cognitive measures, the block design. Additionally, the schizophrenia PRS could explain 0.4% of the variance in lateral ventricular volumes although this was not significant ($p = 0.063$).

1.8. Endophenotypes and CNVs

Some studies have also reported significant associations between carrying schizophrenia associated CNVs and psychosis endophenotypes, especially cognition. The study by Stefansson et al (2014) investigated whether CNVs that had been previously associated with schizophrenia and autism could influence cognitive performance in controls, by evaluating controls carrying these CNVs. They reported that in several cognitive domains, including Verbal and performance IQ, visual information processing and spatial working memory, the control CNV carriers were performing somewhere in between the cases with schizophrenia and the control non-carriers. This indicates that, while these CNVs may not have full penetrance for the disease, carriers exhibit some degree of phenotypic change such as impaired cognition.

Another study by Kendall et al (2017) on data from the UK Biobank investigated cognitive performance on 1,087 control carriers of schizophrenia CNVs, 484 control carriers of CNVs associated with neurodevelopmental disorders and 26,628 controls from other datasets. Carriers of either schizophrenia or neurodevelopmental disorders associated CNVs performed significantly worse in nine cognitive tests, and these

results survived correction for multiple testing. They also had lower educational and occupational attainment compared to non CNV carriers. These results further support the effect of neuropsychiatric CNVs impairing cognition in healthy carriers.

A recent study by Thygesen et al (2020) also investigated the influences of CNVs on a range of cognitive domains in a psychosis family study with 769 cases with either schizophrenia or bipolar disorder with psychotic features, 644 of their unaffected relatives and 2,013 unrelated healthy individuals. Carriers of schizophrenia associated CNVs exhibited impaired performance compared to non-carriers in immediate and verbal recall and in working memory and spatial visualisation. These findings provide evidence that those CNVs apart from significantly increasing the risk for schizophrenia they also negatively influence cognitive performance.

1.9. Aims and hypotheses

The aims of this thesis are to:

- Perform a scoping review on schizophrenia and other clinical phenotypes that overlap genetically, and any potential CNV.
- Develop and populate the CNVcatalog, a repository incorporating data from the studies identified by the scoping review.
- Investigate whether carriers of 16p11.2 distal deletion have a higher risk of developing schizophrenia compared to non carriers.
- Investigate whether CNV size influences the risk of developing schizophrenia for total CNVs, CNV deletions and CNV duplications.
- Examine the relationships between different endophenotypes associated with psychosis.
- Examine group differences between cases, unaffected relatives and healthy controls on endophenotype performance.
- Investigate the joint contributions of polygenic risk scores of bipolar disorder and schizophrenia and CNV burden on psychosis risk.
- Investigate whether the addition of CNV burden measures to predictive models including only PRSs can improve the variance explained on psychosis risk.

- Perform an exploratory analysis investigating the predictive accuracy of the models including both the bipolar disorder and schizophrenia PRSs, and CNV burden.
- Investigate whether the inclusion of CNV burden to the models improves the predictive accuracy for both schizophrenia and bipolar risk
- Perform an exploratory analysis exploring whether carrying a schizophrenia associated CNV increases the risk of developing psychosis.

The following hypotheses will be tested:

- Carriers of 16p11.2 distal deletion will have increased risk of developing psychosis, compared to non carriers.
- Larger CNVs will be associated with larger effect sizes for schizophrenia risk when examining total CNVs (deletions and duplications together), CNV deletions and CNV duplications.
- The P300 event related potential will be associated with cognitive markers of psychosis. A poorer cognitive performance will be associated with reduced P300 amplitude and delayed latency.
- Ventricular volumes, which is a measure of brain structure would be associated with other well established neurocognitive and electrophysiological psychosis endophenotypes.
- The unaffected relatives will exhibit worse endophenotype performance than the healthy subjects. The relatives will perform better than the patients. Compared to controls, the patients will show more severe impaired performance in all endophenotypes investigated.
- The addition of CNV burden to the models including only PRS burden will significantly increase the explained variance in the likelihood of diagnosis status (i.e. schizophrenia, bipolar disorder, control) compared to models including only schizophrenia and bipolar PRSs.

Chapter 2. CNV catalog: A database and meta-analysis tool to investigate the influence of copy number variants on neuropsychiatric traits and diseases.

2.1. Abstract

Background and aims: Rare and non-recurrent copy number variants (CNVs) have been consistently associated with adverse clinical phenotypes. In this chapter I perform a scoping review of CNVs associated with a range of clinical phenotypes, including schizophrenia, bipolar disorder, autism spectrum disorders, attention deficit hyperactivity disorder, epilepsy and cognition. I also perform the first random effects meta-analysis of the 16p11.2 distal deletion on schizophrenia, and I investigate the relationship between CNV size and schizophrenia risk, which has not been studied yet.

Methods: The scoping review was conducted in Pubmed and identified 53 studies. For all the analyses that followed, I used CNVcatalog, a repository created by myself and my supervisors. CNV catalog was populated by the data from the studies identified by the review and currently contains information on 485 CNV loci associated with a range of clinical phenotypes, including schizophrenia.

Results: Carriers of the 16p11.2 distal deletion have augmented risk of developing schizophrenia OR: 2.41 [95% CI: 1.30 - 4.44, p = 0.018 (Q = 6.42, p = 0.169)] compared to non carriers. When investigating CNV size and its effect on schizophrenia risk, I found that as CNV size increases the risk for schizophrenia also increases significantly. The same was observed when looking at CNV deletions [$r(53) = .31$, $p = .019$], but this association was not observed for CNV duplications [$r(46) = .18$, $p = .202$].

Discussion: 16p11.2 distal deletions significantly increase the risk of developing schizophrenia. Larger CNVs are associated with larger effect sizes for schizophrenia risk, only for CNV losses. CNVcatalog provides a comprehensive set of tools facilitating the investigation of specific CNV loci and their association with several clinical phenotypes including schizophrenia.

2.2. Introduction

Advances in molecular genetic technologies have enabled the detection of copy number variants (CNVs), which are deletions or duplications of DNA sequence, altering the diploid status of DNA (Bagshaw et al., 2013; Nowakowska, 2017). Those variations can range from one kilobase to several megabases in size (MacDonald, Ziman, Yuen, Feuk, & Scherer, 2014; Stankiewicz & Lupski, 2010) and various CNV maps have been generated to investigate the influence of CNVs on complex clinical phenotypes, including schizophrenia (Conrad et al., 2010; Park et al., 2010; Wellcome Trust Case Control Consortium, 2010; Zarrei, MacDonald, Merico, & Scherer, 2015).

Although the majority of those genomic variations are benign, several rare and non-recurrent CNVs are pathogenic and constitute some of the most significant risk factors for the manifestation of psychiatric and neurological illnesses for instance schizophrenia (Chen et al., 2016; Li et al., 2016; Marshall et al., 2017; Priebe et al., 2013; Szatkiewicz et al., 2014), generalized epilepsy (Kaminsky et al., 2011; Lal et al., 2015) and intellectual disability (Cooper et al., 2011; Kaminsky et al., 2011; Rees et al., 2016).

Rare or de novo CNVs are foud at a rate more than twice in cases with schizophrenia and autism spectrum disorders (ASD) compared to healthy controls (Sebat el al., 2007; Kirov et al., 2012; Marshall et al., 2017) and are also more prevalent in other clinical phenotypes including attention deficit hyperactivity disorder (ADHD; William et al., 2010), developmental delay/intellectual disability (Cooper et al. 2011) and Tourette Syndrome (Huang et al., 2017).

Several CNVs that have been associated with schizophrenia have been reported to have a pleiotropic effect, being associated with several other phenotypes including bipolar disorder, ASD, epilepsy, intellectual disability, ADHD and impaired cognition (St Clair, 2009; van Winkel et al., 2010; Ziats et al., 2016; Hippolyte et al., 2016). In a study with 3,945 schizophrenia cases and 3,611 healthy controls, the authors found evidence for association of several CNVs with schizophrenia, including deletions in chromosomes 1q21.1, NRXN1, 3q29, 15q13.3 and 22q11.2, and duplications in 16p11.2, all of which have also all been previously associated with mental retardation, ASD and epilepsy (Levinson et al., 2011).

Several studies have also reported that 22q11.2 CNVs (both deletions or duplications) increase the risk for several neuropsychiatric disorders including schizophrenia, ASD, intellectual disability and developmental delay (Monks et al., 2014; Niarchou et al., 2014; McDonald-McGill et al. 2015; Hoeffding et al., 2017). Microdeletions in chromosome 15q11.2 have also been reported to be present in individuals with schizophrenia, ASD, ADHD and epilepsy, as well as individuals with intellectual disability, developmental and language delay (Cox and Butler, 2015; Stefansson et al., 2014; Zhao et al., 2013; Burnside et al., 2011).

Several reliable repositories (i.e DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources - DECIPHER, Swaminathan et al., 2012; the Database of Genomic Variants, MacDonald et al., 2014; CNVD, Qiu et al., 2012) have been developed to facilitate the scrutiny of pathogenic genetic variations, their pleiotropic effects on various clinical phenotypes by capturing carrier data at the individual level. Nevertheless, none of these focus on reporting summary statistics to quantify the associations between CNVs and phenotypic data, as has been efficaciously done for Single Nucleotide Polymorphisms (SNPs), in the widely used National Human Genome Research Institute - European Bioinformatics Institute GWAS-catalog (NHGRI-EBI GWAS-catalog, MacArthur et al., 2017). A probable explanation for this could be the lack of consensus over the nomenclature of naming of CNV loci, which impedes their integration across studies. Tested CNVs are given a short loci band name and their exact coordinates are reported. However, different studies use different genotyping platforms with different coverage and mapping to different genomic builds, giving different start and stop postitions. Determining which loci are comparable, or the same, between studies is therefore not a trivial task. Furthermore, valuable samples of well-characterised CNV carriers may result in multiple publications, either alone or as part of meta-analyses. Therefore, sample overlaps need to be taken into account in future meta-analyses.

I have contributed to the development of CNVcatalog, which facilitates accurate meta-analytical procedures and interactive visualizations of the data included in the database, while allowing for easy addition of new data, following input of summary statistics via a standardised format. It also provides a comprehensive set of tools

assisting the investigation of associations between rare and pathogenic genomic variants and clinical phenotypes.

In this chapter of my thesis, I first performed a scoping review of CNVs associated with schizophrenia, bipolar disorder, ASD, ADHD, asperger's syndrome, epilepsy and cognition. A scoping review was deemed appropriate instead of a systematic review due to the broad, complex and highly heterogeneous nature of the topic to be reviewed, which includes several clinical phenotypes and any possible copy number variation (Peters et al. 2015; Munn et al., 2018), with the aim to identify potential research gaps in the current literature.

From the studies identified by the scoping review, I populated the newly developed CVN catalog database. After performing several exploratory analyses with the data, I noticed that there were not enough studies to explore pleiotropic effects of CNVs. The only locus associated with schizophrenia, that I had enough data to perform a meta-analysis, and also that a meta-analysis for that locus had not been conducted before was the 16p11.2 distal deletion.

Several studies have provided evidence that both deletions and duplications at the 16p11.2 locus increase the risk of developing schizophrenia (Kirov, 2010; Vassos et al., 2010; Bergen et al., 2012; Steinberg et al., 2014; Chang et al., 2017). A large study of 13,850 schizophrenia cases and 19,954 healthy controls has reported an odds ratio of 6.25 (95% CI:1.78-21.93) for 16p11.2 distal deletions (Guha et al., 2013). Another study reported the 16p11.2 duplication, but not the distal deletion, being linked to schizophrenia risk with a prevalence of 0.35% (95% CI: 0.27–0.45%) in cases compared to 0.03% in healthy controls (95%CI: 0.02–0.05%) (Rees et al., 2014).

A meta-analysis on proximal duplications and deletions for the 16p11.2 locus in schizophrenia has previously been conducted (Giaroli, Bass, Strydom, Rantell, & McQuillin, 2014), showing an increased risk of developing schizophrenia for carriers of 16p11.2 proximal duplications OR=16.0 (95% CI: 5.4-47.3: $p<0.001$) but not deletions. At that time, only two studies had investigated distal deletions so a meta-analysis was not feasible, and they were excluded from their analysis.

As a result of the scoping review I conducted, I identified another gap in the existing research literature. Schizophrenia associated CNVs with effect sizes ranging from 2

to 30 (Marshall et al., 2017) have been found in various lengths, ranging from over 20 kilobases (Marshall et al., 2017), 100 kb (Walsh et al., 2008; Stone et al., 2008; Szatkiewicz et al., 2014), 200 kb (Stone et al., 2008; Szatkiewicz et al., 2014), 500 kb (Stone et al., 2008; Szatkiewicz et al., 2014) and 1Mb (Kirov et al., 2009). However, the relationship between CNV size and the effect they confer to the disease risk has not been investigated yet.

In this chapter the main aims were i) to perform a scoping review of CNVs associated with schizophrenia, bipolar disorder, ASD, ADHD, Tourette syndrome, epilepsy and cognitive functioning, ii) to perform the first meta-analysis of the 16p11.2 distal deletion on schizophrenia, iii) to perform an exploratory analysis of the relationship between CNV size and the risk they confer to schizophrenia, and iv) to present CNV catalog, the biological database populated by the papers we identified by the scoping review. I hypothesized that carriers of the 16p11.2 distal deletion will have increased risk for developing schizophrenia and that larger CNVs will confer larger effect sizes for schizophrenia risk.

2.3. Methods

Eligibility criteria

The objective of the review was to identify papers comprising information on CNVs associated with psychiatric illnesses, but also with certain key phenotypes with evidence of co-morbidity: epilepsy and cognitive functioning. The following eligibility criteria were employed: Papers should be published in English from 2008 onwards, (the year modern SNP-microarray assays became commercially available for research) and must contain information on either frequency or association statistics for specific CNV loci. Articles should also comprise information on the number of CNV carriers identified and the full sample size tested. All ethnicities were included. Articles not published in peer review journals, non full-text published article and systematic/literature review studies not presenting results from new samples were excluded. Additionally, if a locus is described as both deletion and duplication (i.e. 15q11.1del/dup) and number of carriers are given combined, I excluded that particular locus. Non-human studies and case studies were also excluded.

Scoping review and search strategy

The scoping review was performed in the PubMed database, which is one the most widely accessible and reliable biomedical resources. It has higher sensitivity than Medline-Ovid (Lam & McDiarmid, 2016), and information from additional sources than Medline, including books, articles from life science journals and conference abstracts, making it the most preferred search database for conducting reviews in the biomedical field (Salvador-Olivan, Marco-Cuenca & Arquero-Aviles, 2019).

The review focused on CNVs associated with a range of adverse clinical phenotypes, including psychiatric disorders, neurological disorders and cognitive functioning.

The search terms employed were:

"((DNA Copy Number Variations [MeSH Terms]) OR CNV [MeSH Terms]) AND (schizophrenia OR schizo OR bipolar* OR psychos* OR psychot* OR autism OR autist* OR ASD OR asperg* OR attention deficit hyperactivity disorder OR ADHD OR epilepsy OR epilept* OR cognit*)".*

The time scale covered by my search was from January 2008 up to and including May 2019. Despite including only a specific number of phenotypes in our search, if a paper investigating another clinical phenotype met the inclusion criteria, we decided to include it in the papers populating CNV catalog.

The reference lists of the studies that met the inclusion criteria were assessed for further relevant papers. To minimise single rater risk, I carried out the scoping review process twice, and if uncertain regarding whether or not to include a paper, I seeked advice from my supervisors.

Data extraction

Once the pertinent papers were identified, I performed automated data extraction by converting the tables with results of interest from primary papers on PDF format and exported the data to excel tables using the PDFTables package in R (Persson, 2016). This intended to maximise the fata included whilst minimizing the possibility of human errors during data extraction. To minimise errors at conversion, data were extracted twice and tested for exact likeness using the identical function in R to ensure no typographical mistakes had occurred.

Data from each study were extracted to fit two template excel sheets, one for the CNV loci level data and another containing information on references. Some information was designed to be obligatory (i.e. CNV coordinates, genomic build), to ensure data completion, and functionality of the applications build on top of the catalog and other variables were defined as optional and can be filled if available from the study (i.e. associated genes, age and ethnicity of participants). Tables S1 and S2 demonstrate the columns on each template spreadsheet, accompanied by a description.

For the input template, CNV loci were stored with one row per loci, with certain variables (such as association results from multiple phenotypes) repeated as required (with the addition of an appropriate numeric indicator of additional columns) to comprise all crucial information from the papers. Upon data upload, this information was then computationally checked for consistency, and quality issues. For example, no chromosome numbers outside of 1 to 22, X and Y are accepted. Also, data was rearranged to fit columns of the tables in the defined database schema described earlier. Additionally, if p-values were given as less than values (i.e. $p < 0.01$) I specified the significant value as 0.01 exact and added a note linked to the association, that the exact value given was less than that.

To allow for integration of our database while minimising sample overlap between studies, the catalog include code developed to map these sample relationships between studies, hence making the identification of truly independent studies feasible.

Development of the CNVcatalog database

CNVcatalog is to our knowledge the first CNV repository incorporating data from published studies examining associations of CNV loci with clinical phenotypes, whilst providing a framework for integration of overlapping CNV loci across studies. CNVcatalog is a structured SQLite database containing information on CNVs, loci positions, associated genes and phenotypes, sample sizes, association results, inter-study relations and, when available and demographic information (i.e. age, sex, ethnic group).

The database was built using Sqlite3 (Hipp, 2000) with data stored in eight linked tables each describing various aspects; the phenotypes of interest, their frequencies, the CNV name, their position, the associated genes, association results, inter-study relations and a reference list . Figure S1 demonstrates the database schema.

I used R project for statistical computing, version 3.5.0 (R Core Team, 2013) along with the packages R-shiny (Bailey, 2015), RSQLite (Muller, Wickham, James, & Falcon, 2019), xlsx (Dragulescu & Arendt, 2018), shinyBS (Bailey, 2015), stringdist (van der Loo, 2014), ggplot2 (Wicham, 2016), metaphor (Viechtbauer, 2010), igraph (Csardi & Nepusz, 2006) and RISmed (Kovalchik, 2017) to populate the database, perform quality control checks on input, and help with CNV loci integration across studies and building the user interface and visualisations.

CNV catalog overview

After populating CNVcatalog with the 53 published papers identified by the scoping review, it currently contains information on 485 CNV markers, describing 69 different clinical phenotypes. Some of the key phenotypes that users can query are schizophrenia, bipolar disorder, intellectual disability, neurodevelopmental delay, autism spectrum disorders, major depressive disorder and epilepsy.

CNV catalog can perform a range of analytical procedures, including random effects meta-analyses, comparisons of CNV loci and interactive visualizations.

Meta-analysis of 16p11.2 distal deletion

Since CNV catalog only includes studies up to and including May 2019, I updated the scoping review only for 16p11.2 deletion in association with the schizophrenia phenotype to include studies until April 2020 (Search terms: *"(16p11.2 OR 16p11 2) AND {Schizophrenia OR Schizophr*)"*). No new studies were identified. Only studies with independent samples were included in the analysis.

Since CNVcatalog is a new software, I also performed a manual verification analysis of the random effects meta-analysis in R statistical software, to corroborate my results. The CNVcatalog performs random effects meta-analyses and presents a forest plot with the effect sizes with 95% confidence intervals. A funnel plot to check potential publication bias was produced in R.

I also attempted to perform pleiotropy analyses by investigating the risk of developing bipolar disorder, autism spectrum disorders or intellectual disability for 16p11.2 distal deletion carriers. However, there were no data available from the papers that had met the inclusion criteria of the scoping review.

Investigating the relationship between CNV length and effect size for the schizophrenia phenotype

I performed linear regression analyses with the logged effects size as outcome and the logged variant size as the predictor for total CNVs (deletions and duplications), CNV deletions and CNV duplications in order to investigate their relationship.

2.4. Results

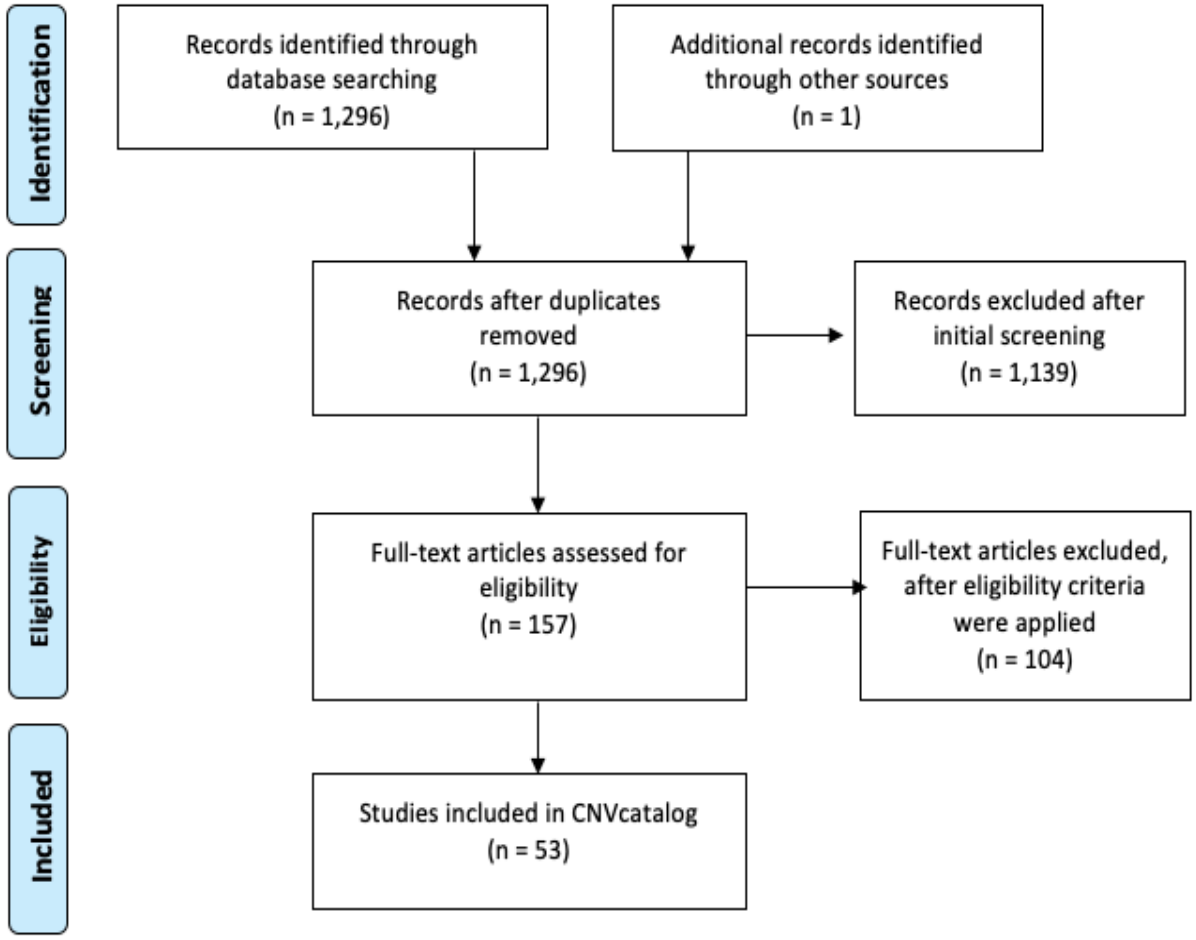
Scoping review

The number of articles yielded by the review was 1,296 and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), guidelines were employed to identify the papers of interest (Figure 2.1). One additional paper was identified by manual search and one duplicate record was removed.

Initially, the titles and abstracts were screened, and 1,139 papers were excluded after the initial screening for being irrelevant, leaving 157 papers. The retrieved set of articles was screened in relation to the eligibility criteria.

I identified 53 studies and the quality assessment tool for diagnostic accuracy studies (QUADAS, Whiting et al., 2011) was employed to evaluate their quality. I screened their bibliographies for additional papers but did not identify any.

Figure 2.1 *PRISMA flow diagram illustrating the study selection procedure*



The PRISMA flowgram illustrates the numbers of records identified, excluded with reasons and included in different phases of the scoping *review process*

The 53 studies identified by the scoping review are presented in table 2.1.

Table 2.1 Table presenting the 53 studies currently included in CNVcatalog

Pubmed ID and title	Sample sizes	CNVs investigated	Related samples (Pubmed ID or sample names)	Main outcome
1. 28030616 Rare CNVs in Suicide Attempt include Schizophrenia-Associated Loci and Neurodevelopmental Genes: A Pilot Genome-Wide and Family-Based Study.	660 offspring of individuals with suicidal behaviour 23,782 healthy controls	NRXN1.del 3q29.del WBS.dup VIPR2.dup 15q11.2.del AS/PWS.dup 15q13.3.del 16p13.11.dup 16p11.2.distal.del 16p11.2.dup 17p12.del 17q12.del 22q11.2.del	5193342; 24776740; 23148125; 22130109	Several rare schizophrenia associated CNVs were found in 9 offspring of individuals with suicidal behaviour, showing the role of such CNVs in suicidal behaviour. Overall, 45 offspring of individuals with suicidal behaviour had CNVs enriched for 65 medically relevant genes previously reported to be affected by CNVs.

2. 27244233 A pilot study on commonality and specificity of copy number variants in schizophrenia and bipolar disorder.	2,127 individuals with schizophrenia 2,491 healthy controls	1q21.1 dup 1q21.1 del NRXN1 del 3q29 del WBS dup VIPR2 dup 15q11.2 del AS/PWS dup 15q13.3 del 16p13.11 dup 16p11.2 dup 17p12 del 17q12 del 22q11.2 del	Independent sample	Variants previously associated with schizophrenia (1q21.1, 15q13.3, 16p11.2 and 22q11.21) were replicated. Disorder-specific CNV aggregated regions (CNVRs) were also found for both schizophrenia: 22q11.21 CNVR (COMT), small CNVRs in 11p15.4 (TRIM5) and 15q13.2 (ARHGAP11B and FAN1), and bipolar disorder: 17q21.2, 9p21.3 and 9q21.13.
3. 23285208 A pilot study on collective effects of 22q13.31 deletions on gray matter concentration in schizophrenia.	151 individuals with schizophrenia 173 healthy controls	1q21.1.del 15q13.3.del 22q11.21.del 15q13.3.dup 22q11.21.dup	Independent sample	22q13.31 deletions were significantly more frequent in patients compared to controls and the deletions load was also significantly associated with reduced gray matter concentration in the peri-limbic

				cortex. The authors concluded that regardless of the size, the 22q13.31 deletion can significantly increase schizophrenia risk.
4. 28096781 A replication study of schizophrenia-related rare copy number variations in a Han Southern Chinese population.	476 individuals with schizophrenia 1,023 healthy controls	1q21.1.del 15q11.2.del 7q11.23.dup 16p11.2.dup	Independent sample	The 16p11.2 duplication, previously associated with schizophrenia, ranging in size from 29.3 Mb to 29.6 Mb was detected in four schizophrenia cases (0.84%) and one healthy control (0.098%).
5. 19404257 Autism genome-wide copy number variation reveals ubiquitin and neuronal genes.	Cohort 1 859 individuals with ADHD 1,409 healthy controls Cohort 2	15q11.2-q13.1.dup 2p16.3.del 3p26.3.del 22q11.21.dup 3p26.3.dup 16p11.2.dup 16p11.2.del	Independent sample	The authors provided support of several previously associated with ASD candidate genes, including NRXN1 and CNTN4, but also for CNV enrichment in the NLGN1 and ASTN2 genes.

	1,136 individuals with ADHD 1,110 healthy controls	7q11.22.dup Xq13.1.dup 22q13.33.del 6q26.del 1q25.2.dup 2p24.3.dup 3p26.2.del 10q23.2.del 3q26.31.dup 4q31.21.dup		
6. 26390827 Common alleles contribute to schizophrenia in CNV carriers.	5,423 individuals with schizophrenia 6,005 healthy controls	1q21.1.del 1q21.1.dup 2p16.3.del 3q29.del 7q11.23.dup 15q11.2.del 15q11.2-q13.1.dup 15q13.3.del 16p13.11.dup 16p11.2.dup 22q11.21.del	Independent sample	The authors reported that cases with schizophrenia who carry known schizophrenia associated CNVs have an excess burden of common risk alleles compared to healthy controls.

7. 21285140 Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications.	3,945 individuals with schizophrenia 3,611 healthy controls	1p34.1.dup 3q26.1.del 3q29.del 4p16.3.del 6q26.del 6q26.dup 7q31.1.del 7q35-q36.1.del 7q36.3.dup 11q22.3.dup 14q11.2.dup 15q21.3.dup 16p13.2.dup 16p13.11.del 16p12.3.del 16q23.1.dup 18q21.31.dup 19p13.2-q13.31.dup	Independent sample	Evidence for previously schizophrenia associated CNVs, including deletions in chromosomes 1q21.1, NRXN1, 15q13.3 and 22q11.21, and duplications in 16p11.2 was found. They also found evidence for association of 3q29 deletions and VIPR2 duplications with schizophrenia.
8. 25055870	78 individuals with schizophrenia	1q21.1.del 1q21.1.dup	Independent sample	The authors identified 15 de novo CNVs in individuals with bipolar

De novo CNVs in bipolar affective disorder and schizophrenia.	371 individuals with bipolar disorder	2p25.2.del 15q11.2.del 16p13.11.dup 16p11.2.dup 17p12.del 22q11.2.del		disorder and 6 de novo CNVs in individuals with schizophrenia. One of the de novo CNVs in bipolar disorder was in the previously schizophrenia associated locus 16p11.2. The median size of de novo CNVs in bipolar disorder was 448 kb, which was intermediate between the size for schizophrenia CNVs (613 kb) and healthy controls (338 kb).
9. 21346763 Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia.	7,488 individuals with schizophrenia 6,689 healthy controls	22q11.2.del 7q36.3.dup 16p11.2.dup 15q13.3.del 7q36.3.II.dup 15q13.3.II.del 3q29.dup 6q26.dup	Independent sample	The authors reported a significant relationship between 7q36.3 microduplications and schizophrenia. All duplications overlapped with the vasoactive intestinal peptide receptor gene VIPR2. The authors believed that increases in VIPR2 transcription

				and VPAC2-mediated cyclic-AMP signalling could possibly be due to the microduplications at 7q36.3.
10.25585696 Genome-wide analysis identifies a role for common copy number variants in specific language impairment.	127 individuals with Specific language impairment 385 first degree relatives 269 healthy controls	11q24.1.dup 2p14.del 8p23.1.dup 4q26.del 7q21.11.del 13q31.3.del 5q23.2.del Xq11.2-q12.dup 9p23.del 2q33.1.dup 2q33.1.II.dup 11q14.3.del 10p13.dup 17p11.2.del Xq21.31.dup 2q11.2.dup 13q14.13.dup 5p15.1.dup	Independent sample	There was a significantly increased CNV burden in patients with specific language impairment (SLI) compared to healthy controls, with larger total CNV length per person, larger average CNV size and higher number of genes affected.

		Xq13.2.dup Xp21.1.dup 5q32.del 22q11.21.dup 15q22.2.dup 4q21.1.dup		
11.24269040 Genome-wide copy number variation analysis in adult attention-deficit and hyperactivity disorder.	400 individuals with ADHD 526 healthy controls	15q11.2.dup 3p26.3.dup 1q42.2.dup 3p26.2.dup 15q24.1-q24.2.dup 2q21.dup 6q14.1.dup 20p12.3-p12.2.dup 8p21.3.dup 16p13.3.dup 22q11.21.dup 4p16.1-p15.33.dup 7q31.1.del 15q11.2.del 11q14.1.del	Independent sample	The rate of CNVs larger than 100 kb was higher in subjects with ADHD compared to healthy controls. The differences remained significant, even after the authors considered CNVs that overlap genes or when structural variants spanning candidate genes for psychiatric disorders were assessed. The biggest differences were found in CNV duplications. However, no significant enrichment was detected in our ADHD cohort for childhood ADHD-associated

		12p11.23.del 22q11.21.del 11q22.1.del 1p13.2.del 14q31.1.del		CNVs, CNVs previously associated with ADHD, autism or schizophrenia.
12. 19966786 Large, rare chromosomal deletions associated with severe early-onset obesity.	300 individuals with severe early onset obesity 7,366 healthy controls	3p11.2.dup 6p12.1.del 8q24.3.dup 10p15.3.dup 11q22.2.del 11q13.4.dup 15q13.2-q13.3.dup 16p11.2.del 17p13.3.dup 22q13.dup	Independent sample	Three patients were identified with deletions on chromosome 16p11.2. The 16p11.2 deletions affect several genes, including SH2B1, which is involved in leptin and insulin signalling. Carriers of that deletion manifested hyperphagia and severe insulin resistance inconsistent for the degree of obesity.
13. 22169095 Rare copy number variants in tourette syndrome disrupt genes in	460 individuals with Tourette syndrome 1,131 healthy controls	3p14.2.del 3q26.del 5p15.2.del 5p15.2.II.del	Independent sample	The authors reported no significant difference in the number of either de novo or transmitted CNVs in cases with Tourette syndrome versus

histaminergic pathways and overlap with autism.		1p33.del 4p15.31.del 5p15.1.del 5q31.3.del 5q32.del 7q35-q36.1.del 8p23.1.del 8p22.del 9p21.2.del 9q33.1.del 22q11.23.del 3p14.2.dup 5p15.2.dup 8p11.21.dup 8p11.1.dup 5p15.2.II.dup 3p26.2.dup 3p26.2.II.dup 5p15.1.dup 5q31.3.dup 5q32.dup		healthy controls. Genes mapping within rare CNVs associated with Tourette syndrome, overlapped with CNVs previously identified in autism spectrum disorders. Three de novo CNVs were identified, a duplication at 6p25.3, a deletion at 20p13 and a duplication at 22q11.21.

		7q36.3.dup 8p22.dup 9p21.2.dup 9q33.1.dup 9q34.3.dup 12q24.33.dup 13q14.11.dup 22q11.23.dup		
14.27569545 Rare Inherited and De Novo CNVs Reveal Complex Contributions to ASD Risk in Multiplex Families.	Cohort 1 1,359 individuals with ASD 521 unaffected siblings Cohort 2 2,100 individuals with ASD 2,100 unaffected siblings	1p35.2.del 1p35.2.dup 1p33.dup 1q43.del 1q43.dup 2q22.3.del 2q22.3.dup 2q32.3.del 2q32.3.dup 3p14.1.del 3p14.1.dup 5p14.3.del 5p14.3.dup	Independent sample	The authors reported an increased burden of large and rare CNVs in cases with autism spectrum disorders compared to their unaffected first-degree relatives. They also identified 49 CNVs associated with autism, and a higher enrichment in cases versus healthy controls.

		5q15.dup 7p21.3.del 7p21.3.dup 7q36.2.dup 8p21.3.del 8p21.3.dup 10q25.1.del 11p11.2.del 11p11.2.dup 13q21.31.del 13q21.31.dup 15q26.3.dup 16q23.1.del 16q23.1.dup 16q23.3.del 16q23.3.dup 20p12.1.dup 22q13.32- q13.33.del 22q13.32- q13.33.dup		

15. 20489179 Strong synaptic transmission impact by copy number variations in schizophrenia.	Cohort 1 977 individuals with schizophrenia 2,000 healthy controls Cohort 2 758 individuals with schizophrenia 1,485 healthy controls	16q22.1.del 16p11.2.dup 22q11.21.del 9q34.3.del 10q11.21.del 4p16.1.del 18q12.3.del 3p26.2.del	Independent sample	The authors reported larger enrichment for calcium signalling genes, CACNA1B and DOC2A. This finding was replicated in the second cohort. They also reported that the RET and RIT2 genes, both RAS related genes were affected by CNVs.
16. 25217958 Refining analyses of copy number variation identifies specific genes associated with developmental delay.	13,318 individuals with ID/DD/ASD 11,255 healthy controls	1q24 del 2q33.1 del 2p16.1 del 2p15-16.1 proximal dup 3p25.3 dup 3p11.2 del 3q13 del 3q28-29 del	Independent sample	An analysis combining CNV and single nucleotide polymorphism (SNP) data identified 10 genes enriched for putative loss of function. Follow-up research in a subset of the affected individuals identified disease associated CNVs affecting the SETBP1 and

		4q21 del 5q14 del 9p13 dup 10q11 dup 10q23.1 del 12p13 dup		ZMYND11 genes, associated with intellectual disability, loss of expressive language, autism and aggressive behaviour.
17. 21841781 A copy number variation morbidity map of developmental delay.	15,767 individuals with ID/DD/ASD 8,329 healthy controls	2q13 del 10q23.1 del 2p21 dup 2q13 dup 4p16.1 dup 4q21.21 - q21.22 del 2p25.3 dup 2q24.3 del 21q21.1 del 8q11.23 dup 1q24.3 del 12p13.31 dup 15q25 del 6p22.3 del	20841430; 18292342; 17116639; 16516587; 18464913; 20360734; 19166990; 11434828	A larger CNV enrichment was found in individuals with craniofacial anomalies and cardiovascular defects in comparison with subjects with epilepsy or autism spectrum disorders. The authors identified 59 CNVs in total, 14 of which were novel.

18. 22138692 Genome-wide copy number variation study associates metabotropic glutamate receptor gene networks with attention deficit hyperactivity disorder.	3,506 individuals with ADHD 13,327 healthy controls	11q14.3 del 7q31.33 del 3p26.1 del 6q24.3 dup 1p31.1 dup 7q36.2 dup 5q12.3 del 1p32.3 del 19q13.11 del 3p26.3 del 2p12 dup 4q25 dup	Independent sample	The authors reported an enrichment of CNVs affecting metabotropic glutamate receptor genes, specifically GRM1, GRM5, GRM7 and GRM8, in cases with ADHD against healthy controls across all cohorts.
19. 22970919 Phenotypic heterogeneity of genomic disorders and rare copy-number variants.	32,587 individuals with DD 8,329 healthy controls	1p36 del 1q21.1 del 10q23 del 15q11.2 del Prader-Willi/Angelman del 15q13.3 del	21841781; 18471269; 18811697; 20588305; 22495309; 21658581	The authors reported that 10.1% of children with developmental delay carried a second large CNV. Children with two large CNVs were eight time more likely to have

		15q13.3.II del 15q24 del 15q24.2q24.3 del 15q25 del Rubinstein-Taybi del 16p13.11 del 16p11.2p12.1 del 16p12.1 del 16p11.2 (SH2B1) del 16p11.2.III del 17p13.3.III del 17p13.3.II (YWHAE) del 17p13.3 (PAFAH1B1) del Smith-Magenis syndrome del NF1 syndrome del 17q12 del 17q21.31 del 17q23 del 19p13.12 del 2q23.1 del 2q37 del		developmental delay compared to healthy controls.

		DiGeorge/VCFS del 22q11.2 distal del Phelan-McDermid syndrome del 3q29 del Wolf-Hirschhorn del Sotos syndrome del 6p25 del 6q16 del Williams syndrome del 8p23.1 del 9q34 del PLP1 del 1q21.1 dup 10q23 dup PWS dup 15q13.3 dup 15q13.3.II dup 15q24 dup 15q24.2q24.3 dup 15q25 dup		

		16p13.11 dup 16p11.2p12.1 dup 16p12.1 dup 16p11.2 (SH2B1) dup 16p11.2.II dup 17p13.3 (YWHAE) dup 17p13.3.II (PAFAH1B1) dup Potocki-Lupski syndrome dup NF1 dup 17q12 dup 17q21.31 dup 17q23 dup 2q37 dup 22q11.2 dup 22q11.2.II distal dup 22q13 dup 3q29 dup WHS dup 5q35 dup 6p25 dup 6q16 dup		

		WBS dup 8p23.1 dup 9q34 dup PLP1 dup		
20. 20691406 Microdeletions of 3q29 confer high risk for schizophrenia.	245 individuals with schizophrenia 490 healthy controls	3p26.3.del 3q29.del 10q11.23-q21.1.del 16p11.2-p12.1.del 22q11.21.del 3q12.3.del 5p15.2.del 9p21.1.del 10q11.21-q11.23.del	Independent sample	Deletions at 3q29 were significantly higher in cases with schizophrenia compared to healthy controls. Twenty genes, including the PAK2 and DLG1, were implicated in schizophrenia.
21. 20368508 Rare copy number variants: a point of rarity in genetic risk for bipolar	1,697 individuals with bipolar disorder 2,806 healthy controls	1q25.1 dup 12p11.21 dup 18p11.21-11.1 dup 19p12 dup	17554300	The authors reported that CNV burden in schizophrenia cases was higher compared to cases with bipolar disorder. There was not a significant difference in the comparison of CNV burden in

disorder and schizophrenia.				bipolar cases versus healthy controls. Schizophrenia associated CNVs were not found to be more common in cases with bipolar disorder compared to healthy individuals.
22.22130109 Independent estimation of the frequency of rare CNVs in the UK population confirms their role in schizophrenia.	11,863 individuals with schizophrenia 60,367 healthy controls	1q21.1 del 3q29 del 15q11.2 del 15q13.3 del 16p11.2 dup 16p13.1 dup 17p12 del 17q12 del 22q11.2 del	Wellcome Trust Case Control Consortium (WTCCC)	CNV deletions at 1q21.1, 3q29, 15q11.2, 15q13.1 and 22q11.2, were found significantly more frequently in cases with schizophrenia cases compared to healthy controls. When focusing on the healthy controls, the authors reported that frequencies of CNVs deletions at 17p12 and CNV duplications at 15q11.2, were higher compared to previously reported findings in controls populations.

23. 23325106 Implication of a rare deletion at distal 16p11.2 in schizophrenia.	790 individuals with schizophrenia 1,347 healthy controls	1q21.1 del 2p16.3 del 3q29 del 7q11.23 dup 7q36.3 dup 15q11.2 del 15q11-13 dup 15q13.3 del 16p13.11 dup 16p11.2 dup 17p12 del 17q12 del 22q11.21 del	Independent sample	A novel locus, a distal deletion at 16p11.2, was more frequent in schizophrenia cases compared to healthy controls. That locus has previously been associated with developmental delay and obesity.
24. 24927284 The impact of the metabotropic glutamate receptor and other gene family interaction networks on autism.	6,742 individuals with ADHD/schizophrenia 12,544 healthy controls	9q34.3 dup 6q15 dup 10q26.3 dup 19p13.11 dup 2q37.1 del 22q11.21 dup 1q43 del 15q13.1 del	20531469; 19404257; 19404256; 20663923	The authors reported significant enrichment in the metabotropic glutamate receptor (GRM) GFIN, previously associated with both schizophrenia and ADHD, in the MXD-MYC-MAX network of genes, previously associated with cancer, and in the calmodulin 1

		9q34.3.large dup 13q12.11 dup 6q24.3 del 7q21.12 del 6p21.31 del 11q14.3 del 5q35.3 dup 3p26.1 del 7q31.33 del		(CALM1) gene interaction network.
25.27185616 Gender differences in CNV burden do not confound schizophrenia CNV associations.	13,276 individuals with schizophrenia 17,863 healthy controls	15q11.2 del 15q13.3 del 16p11.2 distal del 16p11.2 dup 16p13.11 dup 17p12 del 17q12 del 1q21.1 del 1q21.1 dup 3q29 del	18668038; 24311552; 21285140	Authors investigated gender differences in CNV burden and reported that 11 schizophrenia associated CNV loci had a higher burden in female schizophrenia cases compared to male cases. However, none of these differences remained significant after accounting for the rates of CNVs in the control group.

		22q11.2 del 22q11.2 dup NRXN1 del PWS/AS dup VIPR2 dup WBS dup		
26.30411505 Rare copy number variation in extremely impulsively violent males.	120 individuals with Dissocial Personality Disorder 182 healthy controls	2q23.1.dup 3p26.3.del 4p16.3.dup 17q11.2.dup 7q35.del 3q26.del 11q14.3.dup 17q25.3.del 5p15.31.dup 3p24.3.del 1p33.dup 6p21.32.dup 10q26.3.dup 6p25.1.del 4q22.1.del	Independent sample	828 CNVs affecting 754 genes were identified. Many of these genes are associated with cognition, learning, intelligence, neurodevelopment, neurodegeneration, obesity and neuropsychiatric phenotypes.

		10q11.21.dup 16p13.11.dup		
27.28756411 Heterogeneous contribution of microdeletions in the development of common generalised and focal epilepsies.	2,424 individuals with epilepsy 6,746 healthy controls	16p13.3.del 2p16.3.del 2p24.1.del 4p15.1.del 6q26.del 8q24.3.del 12q21.1.del 1p36.33.del 1p13.3.del 14q22.2.del	Independent sample	The authors reported an enrichment in microdeletions in the sample combining all patients with epilepsy compared to healthy controls. Sub-group analysis demonstrated that most of the signal was coming from the cases with generic generalised epilepsy. Four genes (NRXN1, RBFOX1, PCDH7 and LOC102723362) were also identified in all the case subgroups.
28.29890507 Genome wide analysis of rare copy number	712 individuals with alcohol misuse diagnosis 804 healthy controls	1q21.1.del 1q21.1.dup 2p16.3.del 3q29.del 7.11.23.dup	Independent sample	The authors reported higher frequencies of CNV deletions, and more genes affected by deletions in cases with alcohol abuse or

variations in alcohol abuse or dependence.		15q11.2.del 15q11.2-13.1.dup 15q13.3.del 16p13.11.dup 16p11.2.distal.del 16p11.2.dup 17p12.del 17q12.del 22q11.2.del		dependence in comparison with the healthy controls.

29.23164820 Genome-wide analysis of rare copy number variations reveals PARK2 as a candidate gene for attention-deficit/hyperactivity disorder.	Cohort 1 489 individuals with ADHD 1,285 healthy controls Cohort 2 386 individuals with ADHD 781 healthy controls	1q32.3 del 1q32.3 dup 4q24 del 6q22.31 del 6q22.31 dup 6q26 del 6q26 dup 7p14.3 dup 7q31.1 del 7q36.2 del 9p23 del 9p23 dup 16p13.3 del 16p13.3 dup 4p16.del 6p24.2 large del 6p24.2.dup 6p24.2.del 7q36.3.dup 15q13.del 15q13.dup	Independent sample	The authors reported that rare CNVs were more frequent in cases with ADHD compared to controls. They provided evidence for validation of 11 out of the 12 CNVs that were found in ADHD cases. These findings were also replicated in a second smaller sample. Rare CNVs within the parkinson protein 2 gene (PARK2) were more frequent in cases than in healthy controls.

		16p11.2.del 20p12.1 large del 20p12.1.del		
30.21844811	15,749 individuals with DD/ID/ASD/MCA	22q11.2 del 16p11.2 del 1q21.1 del	20466091; 16175506; 17637806;	The authors provided evidence that fourteen CNV deletions and seven CNV duplications were

An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities.	10,118 healthy controls	15q13.2-q13.3 del 15q11.2-q13 del 7q11.23 del 16p13.11 del 17q21.31 del 17q12 del 1q21 del 17p11.2 del 8p23.1 del 3q29 del 5q35 del 16p13.11 dup 16p11.2 dup 15q11.2-q13 dup 22q11.2 dup 1q21.1 dup 17q12dup 7q11.23 dup 17p11.2 dup 15q13.2-q13.3 dup 1q21 dup	14628292; 15060094; 16141005; 16283669; 15834244; 17103431; 16909388; 17124408; 16490798; 16619270; 19166990; 16419101; 15980116; 16906162; 16860135; 17568414; 17910064; 17309648; 17621639; 17389918; 17309648;	significantly more prevalent in cases in comparison with the healthy controls.

		3q29 dup 8p23.1 dup 5q35 dup 17q21.31 dup	17910076; 17901113; 17901693; 17847001; 18178633; 18496225; 18414209; 18929052; 18627053; 18698622; 19047251	
31. 25950944 Burden analysis of rare microdeletions suggests a strong impact of neurodevelopmental genes in genetic generalised epilepsies.	1,366 individuals with genetic generalised epilepsy 5,234 healthy controls	1q21.1.del 15q11.2.del 15q13.3.del 16p13.11.del 16p12.del 16p11.2.del 22q11.2.del	19136953; 19592580; 16032514; 16490960; 19843651; 19136953	The authors reported a higher rate of microdeletions in patients with generic generalised epilepsy compared to healthy controls. Microdeletions in cases harboured several genes previously associated with epilepsy or neuropsychiatric disorders including the NRXN1, RBFOX1,

				PCDH7, KCNA2, EPM2A, RORB and PLCB1 genes.
32.26795442 Genome-wide Analysis of the Role of Copy Number Variation in Schizophrenia Risk in Chinese.	6,588 individuals with schizophrenia 11,904 healthy controls	22q11.2 del 22q11.2 dup 1q21.1-21.2 del 16p11.2 dup 15q11.2-13.1 dup VIPR2 dup 7q11.23 dup NRXN1 del 15q11.2 del 17q12 del 16p13.11 dup 16p11.2 distal del 1q21.1 dup 15q13.3 del 17p12 del 3q29 del 1p36.32 dup 10p12.1 dup 13q13.3 dup	4393692; 26206863	The authors reported higher CNV burden in schizophrenia cases compared to controls, especially when focusing on CNVs larger than 1 Mb. They also found evidence of association with schizophrenia for six previously associated CNV loci: duplications at 16p11.2, 15q11.2-13.1, 7q11.23, and VIPR2 and deletions at 22q11.2, 1q21.1-q21.2, and NRXN1, and three novel loci: duplications at 1p36.32, 10p12.1, and 13q13.3.

33. 23341896 Identification of rare recurrent copy number variants in high-risk autism families and their prevalence in a large ASD population.	3,000 individuals with ASD 6,000 healthy controls	1q21.1 dup 1q41 del 2p16.3 del 3q26.31 dup 4q35.2 del 6p24.3 del 6q11.1 dup 6q24.3 large del 7p22.1 dup 7q21.3 del 9p21.1 large del 9p21.1 del 10q23.1 del 10q23.31 dup 12q23.2 dup 13q13.3 del 14q32.2 dup 14q32.31 dup 14q32.31 del 14q32.31.II dup 15q11.2-q13.1 dup	23341896	The authors identified 15 CNVs in families with high-risk ASD, and those CNVs were also more frequent in cases with ASD compared to healthy individuals. The authors identified 25 CNVs with higher frequencies in ASD cases in comparison to controls, 18 of which were novel.

		15q13.2–15q13.3 del 15q13.3 dup 20q11.22 dup		
34. 28726807 When genotype is not predictive of phenotype: implications for genetic counseling based on 21,594 chromosomal microarray analysis examinations.	2,791 high-risk prenatal-women whose fetuses had MCA 3,588 postnatal-individuals with unexplained DD/ID, ASD, or MCA 15,215 low-risk prenatal-women with uneventful pregnancy (control group)	1q21.1.distal.del 7q11.23.del 7q11.23.dup 15q13.3.del 16p11.2.proximal.del 17q12.del 22q11 proximal.del 22q11 distal.del	23258348	The frequency of high-penetrance CNVs higher in the group of individuals with unexplained DD/ID, ASD, or MCA (2.6%) compared to individuals with high (0.9%) and low (0.1%) prenatal risk. The differences on the frequency of low-penetrance CNVs were not significant among the three groups. The authors concluded that solely the low-penetrance CNVs do not contribute to the risk of DD/ID, ASD, or MCA.
35. 21982423	197 individuals with psychosis	3p26.1.del 4p16.1.dup	Independent sample	Previously schizophrenia associated CNVs 15q11.2 and

Copy number variants for schizophrenia and related psychotic disorders in Oceanic Palau: risk and transmission in extended pedigrees.	185 unaffected relatives 159 unrelated healthy controls	6q25.2.del 8p23.2.del 8p23.2.dup 9p24.2.del 10q21.3.del 11q23.3.dup 17p12.dup 17p12-p11.2.dup 19p13.3.dup		1q21.1 deletions and Xp21.3 duplication were also identified in the Palauan sample of schizophrenia cases examined here. Duplications within A2BP1 were found to have an eightfold increased risk in male subjects but not in females.
36. 23843933 Copy number variants in German patients with schizophrenia.	1,637 individuals with schizophrenia/schizoaffective disorder 1,627 healthy controls	1q21.1.del 2p16.3.del 2p16.3.dup 7q36.3.dup 15q11.2.del 16p13.11.del 16p13.11.dup 16p11.2.dup 22q11.21.del	22344817	The study reported higher prevalence of previously known schizophrenia associated CNVs in the schizophrenia sample compared to healthy controls.
37. 24311552	Cohort 1	1p36.33 dup 2q37.3 dup	24163246; 19675094;	Thirteen CNVs previously associated with schizophrenia

Analysis of copy number variations at 15 schizophrenia-associated loci.	6,882 individuals with schizophrenia 6,313 healthy controls Cohort 2 21,450 individuals with schizophrenia 26,529 healthy controls	4q25 dup 4q35.1 dup 4q35.2 del 5q33.1 del 6q24.2 dup 9p24.2.large del 9p24.2 del 15q21.3 dup 16p12.1 del 18q23 dup 1q21.1 del 1q21.1 dup NRXN1 del 3q29 del WBS dup VIPR2 dup 15q11.2 del AS/PWS dup 15q13.3 del 16p13.11 dup	18945720; 21346763; 23992924; 22614287; 23871472; 21285140; 22424231	were found to have significantly higher rates in the schizophrenia sample compared to controls in the cohort comprising of new data. When this cohort was combined with additional published data, eleven of these loci were found to be associated with schizophrenia.

38. 24163246 CNV analysis in a large schizophrenia sample implicates deletions at 16p12.1 and SLC1A1 and duplications at 1p36.33 and CGNL1.	21,450 individuals with schizophrenia 26,529 healthy controls	1p36.33 dup 2q37.3 dup 4q25 dup 4q35.1 dup 4q35.2 del 5q33.1 del 6q24.2 dup 9p24.2 large del 9p24.2 del 15q21.3 dup 16p12.1 del 18q23 dup	19675094; 18945720; 21346763; 23992924; 22614287; 23871472; 21285140; 22424231; 24311552	The authors found 12 CNV loci enriched in schizophrenia cases including deletions at 16p12.1 and duplications at 1p36.33. However, none survived correction for multiple testing.
39. 24939913 16p11.2 600 kb Duplications confer risk for typical and atypical Rolandic epilepsy.	281 individuals with Rolandic epilepsy 1,512 healthy controls	15q11.2 del 15q11.2 dup 15q13.3 del 15q13.3 dup 16p11.2 dup 16p13.11 del 16p13.11 dup 22q11.2 dup	Independent sample	The authors reported higher frequencies of duplications at 16p11.2 in cases with both typical and atypical Rolandic epilepsy compared to healthy controls. However, duplications at 16p11.2 were not identified in either cases with temporal local epilepsy or generic generalised epilepsy,

				suggesting that this duplication poses a significant risk factor solely for Rolandic epilepsy.
40.24776740 Copy number variation in schizophrenia in Sweden.	4,719 individuals with schizophrenia 5,917 healthy controls	1q21.1.del 2p16.3.del 3q29.del 7q11.23.del 7q36.3.del 15q11.2.del 15q13.3.del 16p13.11.del 16p11.2 distal.del 16p11.2.del 17q12.del 22q11.2.del 1q21.1.dup 2p16.3.dup 3q29.dup 7q11.23.dup 7q36.3.dup	12395142; 9974454; 14638593; 11248156; 8657240	The authors reported higher frequencies of duplications at 16p11.2 and 17q12, and deletions at 22q11.2 and 3q29 in schizophrenia cases in comparison with healthy controls. Increased burden of large CNVs (above 500 kb) was found in genes present in the postsynaptic density and in gene products localized to mitochondria and cytoplasm.

		15q11.2.dup 15q13.3.dup 16p13.11.dup 16p11.2 distal.dup 16p11.2.dup 17q12.dup 22q11.2.dup		
41.24445990 Copy number variants and therapeutic response to antidepressant medication in major depressive disorder.	1,565 individuals with Major Depressive Disorder	15q13.3 large dup 4q28.3 large del 6q12 del 3q26.2 del 20p12.1-12.2 dup 8p23.2 dup 2p16.3 del 17q25.1 dup 18p11.32 del 9p23 del 4q28.3 del 15q13.3 dup 15q13.2 dup	20360315; 18498636; 21449676	There was no relationship between the presence of either rare or common CNVs, the number of CNVs or the CNV burden and response to antidepressant medication. Only two CNVs were associated with poor antidepressant response: duplications at 16p13.3 and deletions at 2p16.3.

		3q23 dup		
42.27042285 Genome-wide analysis of copy number variations identifies PARK2 as a candidate gene for autism spectrum disorder.	335 individuals with ASD 1,093 healthy controls	1p36.21 dup 1p36.13 dup 1q25.1 del 2p11.2-p11.1 dup 3p12.3 dup 3q22.1 dup 3q22.1 del 4p16.1 del 4p16.1 dup 4p16.3 dup 4q13.2 dup 6q26 dup 6q26 del 8p23.1 dup 9q13 dup 12p13.31 del 14q11.2 dup 14q11.2 del 19q13.42 dup 21q11.2 dup	Independent sample	Six CNVs (both deletions and duplications) at 6p26 were identified in cases with ASD. The expression level of PARK2 was down regulated in ASD cases with CNVs at 6p26.

		21q11.2 del 22q11.23 del		
43. 24667286 Assessing the impact of copy number variants on miRNA genes in autism by Monte Carlo simulation.	197 individuals with ASD	1p21.3.del 1p36.33.del 22q11.21.del 22q11.21.dup 22q11.22.dup 2q13.del 2q37.3.del	20531469; 21658582	The authors reported that the number of miRNA loci that were affected by de novo CNVs at chromosomes 1, 2 and 22 in patients with ASD was significantly higher that the estimation provided by the Monte Carlo simulation.
44. 27869829 Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects.	21,094 individuals with schizophrenia 20,227 healthy controls	22q11.21.del 16p11.2.proximal.dup 2p16.3.(NRXN1).del 15q13.3.del 3q29.del 16p11.2.distal.del 7q11.23.dup Xq28.distal.dup 22q11.21.dup	22424231; 22688191; 20368508; 18945720; 21346763; 23992924; 24163246	The authors reported enrichment of CNV burden in schizophrenia cases compared to healthy controls. They also found evidence for eight loci associated with schizophrenia: 1q21.1, 2p16.3 (NRXN1), 3q29, 7q11.2, 15q13.3, distal 16p11.2, proximal 16p11.2 and 22q11.2.

		13q12.11.(ZMYM5).dup Xq28.(MAGEA11).dup 15q11.2.del 8q22.2(VPS13B).del		
45. 27773354 Cognitive Performance Among Carriers of Pathogenic Copy Number Variants: Analysis of 152,000 UK Biobank Subjects.	151,619 healthy controls	1p36 del (*GABRD*) 1p36 dup (*GABRD*) TAR del TAR dup 1q21.1 del 1q21.1 dup *NRXN1* del 2q11.2 del (*LMAN2L, ARID5A*) 2q11.2 dup (*LMAN2L,ARID5A*) 2q13 del (*NPHP1*) 2q13 dup (*NPHP1*) 2q13 del 2q13 dup 2q21.1 del 2q21.1 dup	UK-biobank	Carriers of CNVs associated with schizophrenia or neurodevelopmental disorders had impaired performance in cognitive tasks compared to non carriers. CNV carriers also had lower educational and occupational attainment.

		2q37 del (*HDAC4*) 2q37 dup (*HDAC4*) 3q29 del 3q29 dup Wolf-Hirschhorn del Wolf-Hirschhorn dup Sotos syndrome del 5q35 dup 6q16 del (*SIM1*) 6q16 dup (*SIM1*) Williams-Beuren syndrome (WBS) del WBS dup 7q11.23 distal del (1.2-Mb) 7q11.23 distal dup (1.2-Mb) 8p23.1 del 8p23.1 dup 9q34 del (*EHMT1*) 9q34 dup (*EHMT1*) 10q11.21q11.23 del 10q11.21q11.23 dup 10q23 del (*NRG3, GRID1*)		

		10q23 dup (*NRG3, GRID1*) Potocki-Shaffer syndrome del (*EXT2*) 11p11.2 dup (*EXT2*) 13q12 del (*CRYL1*) 13q12 dup (*CRYL1*) 13q12.12 del 13q12.12 dup 15q11.2 del BP1-BP2 15q11.2 dup BP1-BP2 Prader-Willi syndrome/Angelman syndrome (PWS/AS) del PWS/AS dup 15q11q13 del BP3-BP4 (*APBA2, TJP1*) 15q11q13 dup BP3-BP4 (*APBA2, TJP1*) 15q11q13 del BP3-BP5 15q11q13 dup BP3-BP5 15q13.3 del BP4-BP5 15q13.3 dup BP4-BP5 15q13.3 del (*CHRNA7*) 15q13.3 dup (*CHRNA7*) 15q24 del		

		15q24 dup 15q25 del 15q25 dup Rubinstein-Taybi del (*CREBBP*) Rubinstein-Taybi dup (*CREBBP*) 16p13.11 del 16p13.11 dup 16p12.2-p11.2 del (7.1-8.7Mb) 16p12.2-p11.2 dup (7.1-8.7Mb) 16p12.1 del (520kb) 16p12.1 dup (520kb) 16p11.2 distal del (220kb) 16p11.2 distal dup (220kb) 16p11.2 del (593kb) 16p11.2 dup (593kb) 17p13.3 del (*YWHAE*) 17p13.3 dup (*YWHAE*) 17p13.3 del (*PAFAH1B1*) 17p13.3 dup (*PAFAH1B1*) Hereditary Neuropathy with Pressure Palsies del (HNPP)		

		Charcot-Marie-Tooth disease type 1A dup (CMT1A) Smith-Magenis syndrome del Potocki-Lupski syndrome dup 17q11.2 del (*NF1*) 17q11.2 dup (*NF1*) Renal cysts and diabetes syndrome del (RCAD) 17q12 dup 17q21.31 del 17q21.31 dup 17q23.1q23.2 del 17q23.1q23.2 dup 22q11.2 del 22q11.2 dup 22q11.2 distal del 22q11.2 distal dup *SHANK3* del *SHANK3* dup		
46.25560756 Copy number variation in bipolar disorder.	8,968 individuals with bipolar disorder	1q21.1 del 1q21.1 dup NRXN1 del 3q29 del	22424231; 22688191; 20368508; 24163246	The authors provided evidence of three previously schizophrenia associated CNVs (duplications at 1q21.1 and 16p11.2 and deletions

	81,121 healthy controls	WBS dup VIPR2 dup 15q11.2 del AS/PWS dup 15q13.3 del 16p13.11 dup 16p11.2 distal del 16p11.2 dup 17p12 del 17q12 del 22q11.2 del		at 3q29) being also prevalent in cases with bipolar disorder. 55 genes were found enriched in bipolar cases compared to controls but none survived multiple testing.
47.29213072 Characterization of Large Copy Number Variation in Mexican Type 2 Diabetes subjects.	686 individuals with diabetes type 2 194 healthy controls	4p16.3.del 4p16.3.dup 1p21.1.dup 3q28.dup 5p15.2.dup 6p22.1.del 6p22.3.del 15q14.dup	Independent sample	Five loci previously associated with type 2 diabetes had duplications or deletions in the type 2 diabetes sample. A gene-set analysis comprising genes with CNVs observed in the type 2 diabetes sample highlighted gene-sets related with sensory perception and dopachrome

				isomerase activity (MIF and DDT genes).
48.28963451 CNV-association meta-analysis in 191,161 European adults reveals new loci associated with anthropometric traits.	191,161 adult samples: 175,183 had information on weight 191,161 had information on Body Mass Index (BMI) 181,965 had information on height 161,244 had information on waist-hip ratio	1q21.1.del 1q21.1.dup 3q22.2.del 3q22.2.dup 3q29.del 3q29.dup 7q11.22.del 7q11.22.dup 11p14.2.del 11p14.2.dup 16p11.2.del 16p11.2.dup 16p11.2.II.del 16p11.2.II.dup 18q21.2.del 18q21.2.dup	3D; B58C; COROGENE CASE; COROGENE CTRL, DNBC; EGCUT; FAMHS; FAMHS1M; FINRISK; Generation Scotland; H2000; HBCS; Hypergenes, InCHIANTI; LifeLine; LLFS; LOLIPOP; Mt Sinai BioMe; PREDICTCVD; QIMR, SCCS;	Five novel CNV loci (1q21.1, 3q29, 7q11.23, 11p14.2, and 18q21.32) and two previously implicated (16p11.2 and 22q11.21) were found to have large effects on several anthropometric traits: height (>2.4 cm), weight (>5 kg), and body mass index (BMI) (>3.5 kg/m^2). This study provides evidence that anthropometric traits share genetic loci with developmental and psychiatric disorders.

			SSC; TRAILSPOP; Twins UK; UKBB; YFS;	
49. 29649218 Global characterization of copy number variants in epilepsy patients from whole genome sequencing.	Cohort 1 198 individuals with epilepsy 301 healthy controls Cohort 2 325 individuals with epilepsy 380 healthy controls	2p22.3.dup 12p13.31.del 15q26.1.dup 16p13.11.del 16p13.11.dup	Independent sample	The authors reported an enrichment of rare exonic variants in patients with epilepsy, particularly in genes with low loss of function tolerance. They also identified rare non-coding CNVs near genes that have been previously associated with epilepsy.
50. 24352232	927 CNV carriers	2p16.3.del 22q11.21.dup	Independent sample	The authors demonstrate that control subjects carrying

CNVs conferring risk of autism or schizophrenia affect cognition in controls.	75,657 CNV non carriers	1q21.1.dup 17q12.dup 17p12.del 16p13.11.dup 16p12.1.del 16p11.2.dup 16p11.2.del 15q11.2.dup 15q11.2.del 13p31.3.dup		schizophrenia associated CNVs perform somewhere between schizophrenia patients and population controls in a range of cognitive tasks. They also report that CNVs differ significantly in terms of the cognitive domains they affect.
51. 23992924 The penetrance of copy number variations for schizophrenia and developmental delay.	32,587 individuals with MR/DD/CM/ASD 13,465 individuals with schizophrenia 81,821 healthy controls	1p36 del 1q21.1 del 1q21.1 dup NRXN1 del 2q23.1 del 2q37 del 2q37 dup 3q29 del 3q29 dup Wolf-Hirschhorn del Wolf-Hirschhorn dup	22970919; 23258348; 23472757; 22424231; 23871472; 23325106; 21285140; 19675094	The authors investigated the penetrance of schizophrenia associated CNVs on schizophrenia, developmental delay, autism spectrum disorders and congenital malformations. They reported that penetrance of almost all CNVs was higher in developmental delay, autism spectrum disorders and congenital malformations

		Sotos syndrome del Sotos syndrome dup 6p25 del 6p25 dup 6q16 (SIM1) del 6q16 (SIM1) dup WBS del WBS dup 8p23.1 del 8p23.1 dup 9q34 del 9q34 dup 10q23 del 10q23 dup 15q11.2 del 15q11-13 (PWS/AS) any del 15q11-13 (PWS/AS) any dup 15q13.3 del 15q13.3 dup 15q13.3 smaller (CHRNA7) del 15q13.3 smaller (CHRNA7) dup		compared with schizophrenia. The overall penetrance of SCZ-associated CNVs for developing any disorder was high (range: 10.6% - 100%).

		15q24 del 15q24 dup 15q24.2q24.5 del 15q24.2q24.5 dup 15q25 del 15q25 dup 16p13.11 del 16p13.11 dup 16p11.2p12.1 del 16p11.2p12.1 dup 16p12.1 del 16p12.1 dup 16p11.2 distal del 16p11.2 distal dup 16p11.2 del 16p11.2 dup Rubinstein-Taybi del 17p13.3 (YWHAE) del 17p13.3 (YWHAE) dup 17p13.3 (PAFAH1B1) del 17p13.3 (PAFAH1B1) dup		

		Smith-Magenis (del) Potocki-Lupski (dup) NF1 del NF1 dup 17q12 del 17q12 dup 17q21.31 del 17q21.31 dup 17q23 del 17q23 dup 19p13.12 del 22q11.21 del (VCFS) 22q11.21 dup 22q11.2 distal del 22q11.2 distal dup Phelan-McDermid del 22q13 dup		
52. 29225144 Copy number variants in people with autism	116 individuals with schizophrenia	10p11.21.dup 10p12.33.dup 10q11.22.dup 10q21.3.del	Independent sample	27 novel CNVs were identified. 49 rare CNVs (prevalence less than 1.5% rate in the general population) were also identified at

spectrum disorders and co-morbid psychosis.		10q22.3.dup 10q23.2.dup 10q23.33.dup 10q24.33.dup 11p15.1.del 11p15.4.dup 11q11.dup 11q14.1.del 12p12.2.dup 12q24.33.dup 13q31.3.dup 15q11.2.dup 16p11.2.dup 16p13.3.del 16q22.1.dup 16q24.3.del 17q22.dup 18q21.1.dup 18q22.3.dup 19p13.2.del 1p36.33.del		significantly higher frequencies than anticipated.

		1q42.12.del 20p13.dup 20q13.31.dup 21q21.3.dup 2p25.3.dup 2q11.2.dup 2q14.2.dup 3p14.2.dup 3p14.2.II.dup 3q25.32.dup 4q21.1.dup 4q21.3-q22.1.dup 4q32.3.dup 4q35.1.dup 4q35.2.dup 6p11.2.dup 7p21.3.dup 7p22.1.dup 7q22.1.dup 9p23.del		

53.27602560 Analysis of Intellectual Disability Copy Number Variants for Association With Schizophrenia	20,403 individuals with schizophrenia 26,628 healthy controls	1p36.del 1p36.dup Thrombocytopenia absent radius (TAR).del TAR.dup 1q21.1.del 1q21.1.dup 1q24 (FMO andDNM3).del NRXN1.del 2p15-16.1proximal (PEX13to AHSA2).dup 2q11.2.del 2q13.del 2q13.dup 2q33.1 (SATB2).del 2q37 (HDAC4).del 3p25.3 (JAGN1 toTATDN2).dup 3p11.2 (CHMP2Bto POU1F1).del 3q13 (GAP43).del 3q28-29 (FGF12)del 3q29.del Wolf-Hirschhorn.del	22424231; 19675094; 21285140; syndrome 23871472	Deletions at 16p12.1 and 2q11.2 and duplications at 10q11.2 and 1q11.23 were significantly associated with schizophrenia. Only the deletion at 16p12.1 survived correction for multiple testing. The study also provided evidence for the protective effects of the 22q11.2 duplication.

		Wolf-Hirschhorn.dup 4q21 (BMP3).del 5q14 (MEF2C).del Sotos syndromedel Williams-Beurensyndrome (WBS).del WBS.dup 8p23.1.del 8p23.1.dup 9p13.dup 9q34.dup 10q11.21q11.23.dup 10q23.del Potocki-Shaffersyndrome.del 12p13 (SCNN1Ato PIANP).dup Prader-Willisyndrome/Angelman syndrome(PWS/AS).del PWS/AS.dup 15q11.2 BP1-BP2.del 15q13.3.del 15q24.del 15q24.dup		

		15q25.del 16p13.11.del 16p13.11.dup 16p12.1.del 16p11.2 distal.del 16p11.2 distal.dup 16p11.2.del 16p11.2.dup 17p13.3 (YWHAEand PAFAH1B1).del 17p13.3 (YWHAEand PAFAH1B1).dup Smith-Magenissyndrome.del Potocki-Lupskisyndrome.dup 17q11.2.del 17q11.2.dup 17q12.del 17q12.dup 17q21.31 (Koolen-de Vriessyndrome).del 22q11.2(DiGeorge/ VCFSsyndrome).del 22q11.2.dup		

		distal 22q11.2.del distal/22q11.2.dup Phelan-McDermidsyndrome.del Phelan-McDermidsyndrome.dup		

ADHD: Attention deficit hyperactivity disorder; ASD: Autism spectrum disorder; CM: Congenital malformations; DD: Developmental delay; ID: Intellectual disability; MCA: Multiple Congenital Anomalies; MR: Mental retardation. The quality of all the studies has been assessed using the the quality assessment tool for diagnostic accuracy studies *(QUADAS).*

Random effects meta-analysis of the 16p11.2 distal deletion literature in schizophrenia

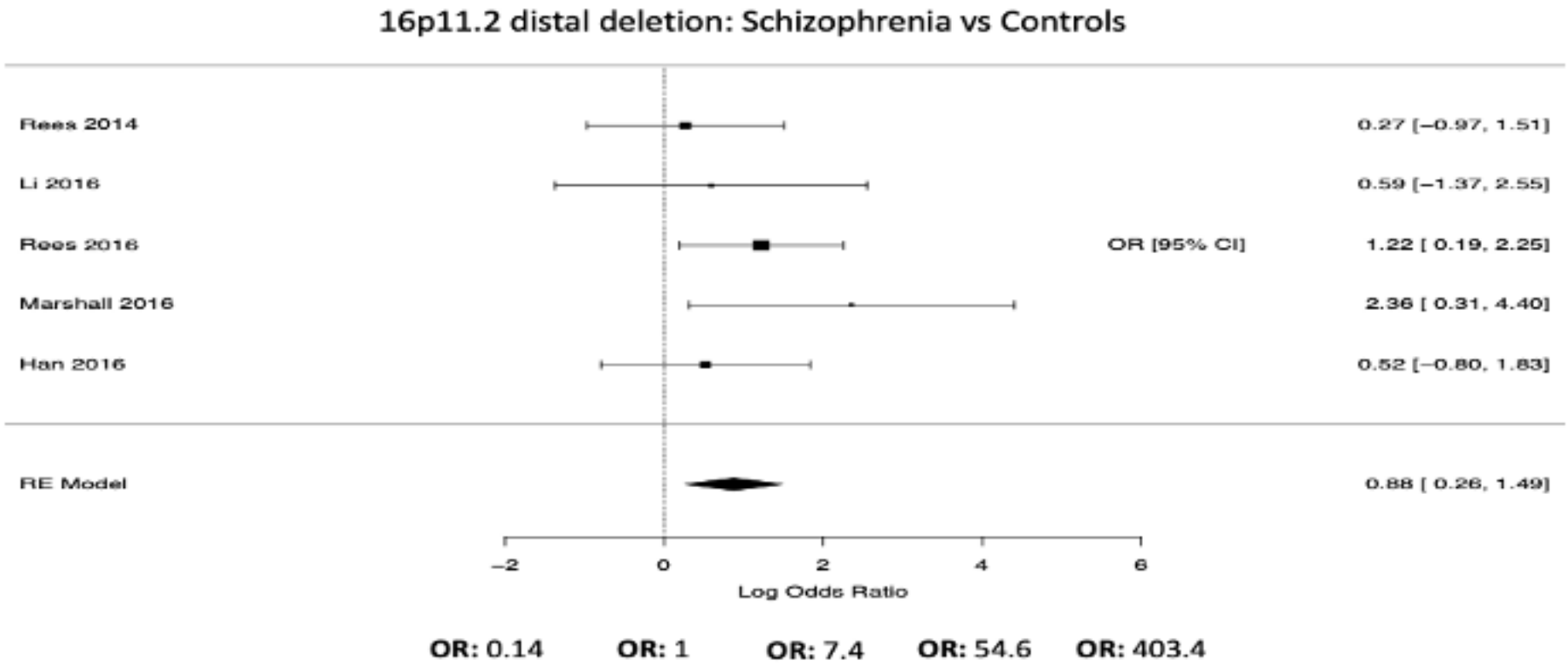
In total there were 5 published studies investigating the 16p11.2 distal deletion in schizophrenia. The total sample size was 75,929 patients with schizophrenia and 91,896 healthy controls. The sample included 53 carriers of the CNV and 167,772 non-carriers.

The five studies eligible for this meta-analysis reported inconsistent results regarding the influence of the 16p11.2 distal deletion on schizophrenia risk.The study by by Rees et al (2014) with a sample of 14,568 cases and 15,274 healthy controls did not find a significant effect of the 16p11.2 distal deletion on schizophrenia risk (OR: 1.7 (95% CI, 0.37-7.6, p = .51). In the study by Han et al (2016) with a sample of 13,276 schizophrenia cases and 17,863 controls, the 16p11.2 distal deletion was not found to be significantly associated with schizophrenia risk either (OR: 2.12, 95% CI, 0.44-10.95, p = .30). The study by Li et al (2016) with 6,588 individuals with schizophrenia and 11,904 controls, is in agreement with the two previous studies, reporting no association between carrying a distal deletion on 16p11.2 and increased schizophrenia risk (OR: 1.83, 95% CI, 0.13-25.30, p = .617).

However, the study by Rees et al (2016) with 20,403 schizophrenia cases and 26,628 healthy controls showed that distal deletions at the 16p11.2 locus were associated with schizophrenia with an effect size of 3.3 (95% CI, 1.61-7.05, p = .017). Finally, the study by Marshall et al (2017) with the largest sample so far to have investigated 16p11.2 distal deletions, consisting of 21,094 schizophrenia cases and 20,227 healthy controls also reported a significant relationship with an effect size of 20.6 (95% CI 2.6-162.2 $p = 5.52 \times 10^{-4}$).

I performed a random effects meta-analysis of schizophrenia versus controls for the 16p11.2 distal deletion locus as shown in figure 4.2. The odds ratio was 2.41 (95% CI: 1.30 - 4.44, p = 0.018) indicating that carriers of this deletion are more than twice as likely to develop schizophrenia compared to non-carriers. The heterogeneity between the studies was found to be high (Q = 6.42, p = 0.169).

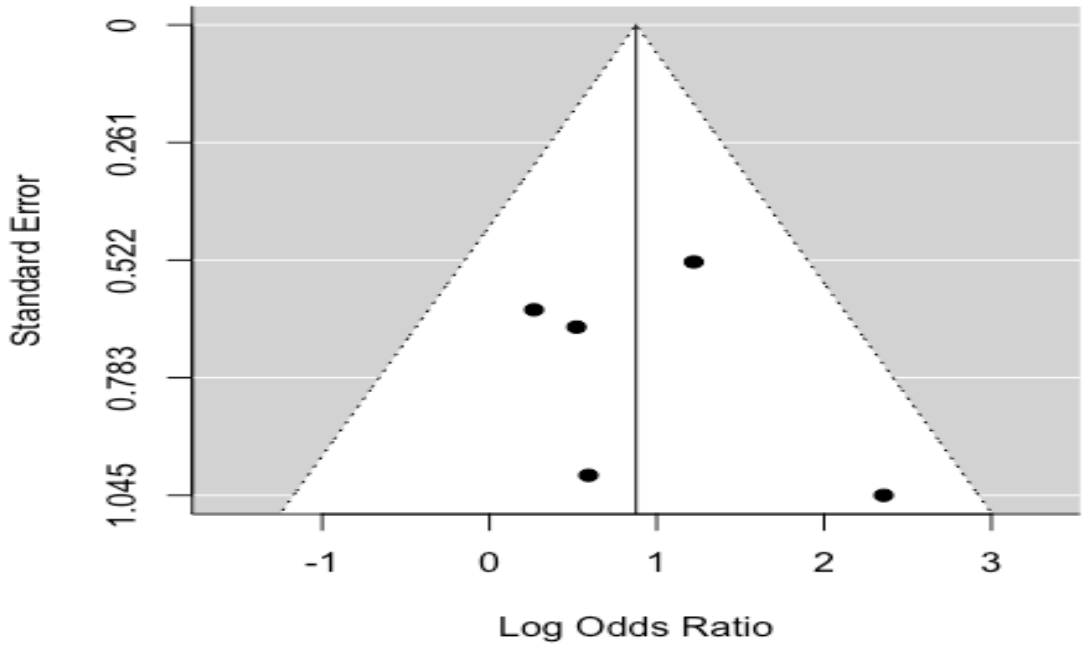
Figure 2.2 *Performing a meta-analysis of the literature investigating 16p11.2 distal deletion in schizophrenia using CNVcatalog.*



Forest plot of the meta-analysis of 5 studies investigating the effect of 16p11.2 distal deletion in schizophrenia risk. Results are presented in logarithmic scale. The linear scale is also provided below the log scale to aid interpretation.

I produced a funnel plot, as shown in figure 4.3, to investigate any possible effects of small studies. There was no evidence for publication bias. However, with only 5 studies the power to detect publication bias is limited since a minimum of 10 studies has been proposed (Higgins et al, 2019).

Figure 2.3 *Funnel plot investigating publication bias in the five studies included in the meta-analysis.*



Funnel plot demonstrating no evidence for publication bias. Eachpoint represents a study. The x axis shows the standard error of te effect estimate in a reversed scale and the y axis represents the study results (odds ratios) The dotted lines represent the 95% confidence intervals. The tool is a visual assessment, and the symmetry of the studies distribution suggests there was no evidence of publication bias (though power is limited given only 5 studies are available).

Investigating the relationship between CNV length and effect size for the schizophrenia phenotype

From the 53 primary studies included in the catalogue, I selected 9 studies with 53 loci in 1,643 carriers, focusing on schizophrenia and performed an analysis examining whether the size of schizophrenia associated CNVs (measured as the average CNV length in kilobases) had an influence on the risk for schizophrenia. These analyses included 53 loci in total for this analysis of CNV size upon schizophrenia odds ratio.

Table 2.2 describes the studies included in the analyses. For more information regarding the CNVs included in the analyses, please refer to table S3.

Table 2.2 Table demonstrating the descriptives of the total CNVS, CNV deletions and CNV duplications for the schizophrenia phenotype

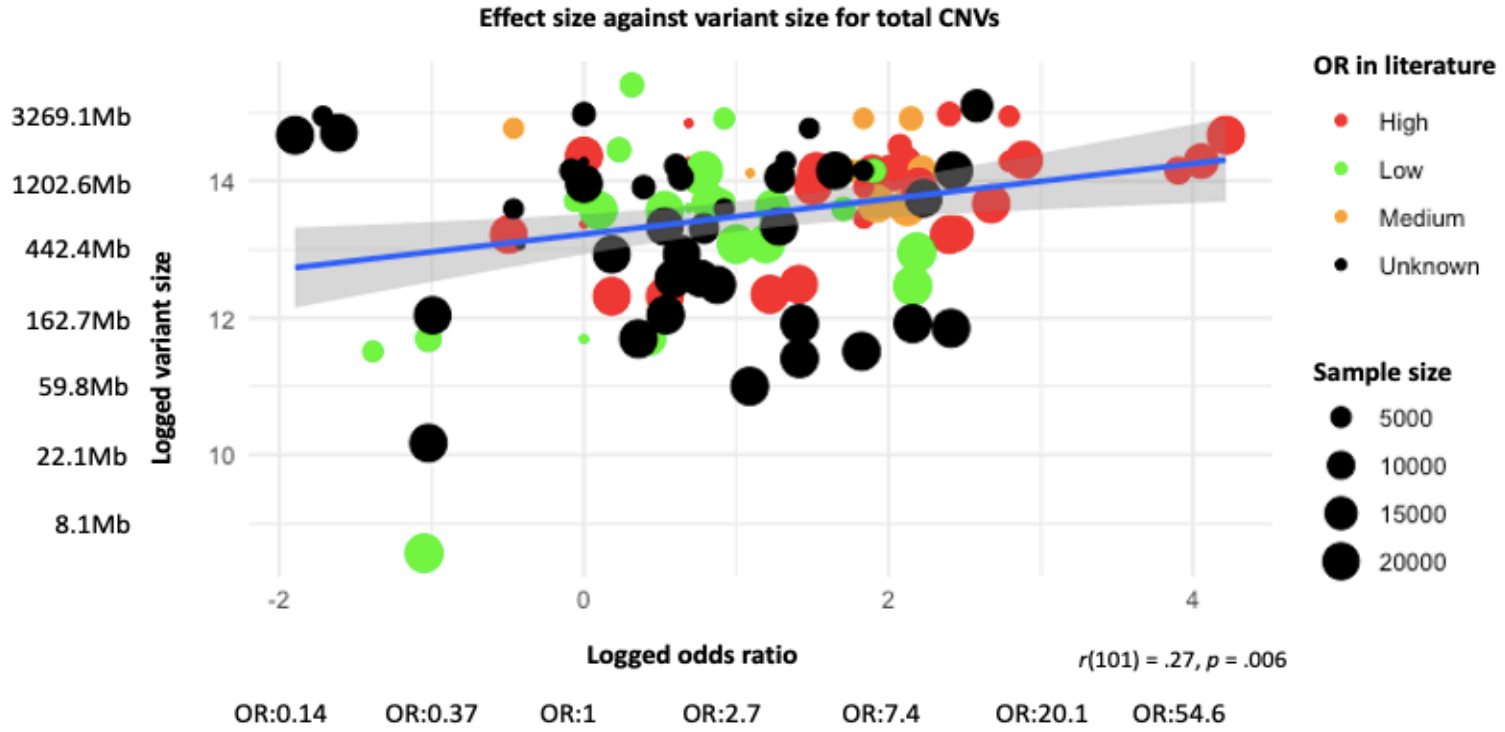
	Total	**Deletions**	**Duplications**
n of loci	53	27	26
n of studies	9	9	8
n of carriers	1,643	927	716
Sample size	102,440	102,440	90,553

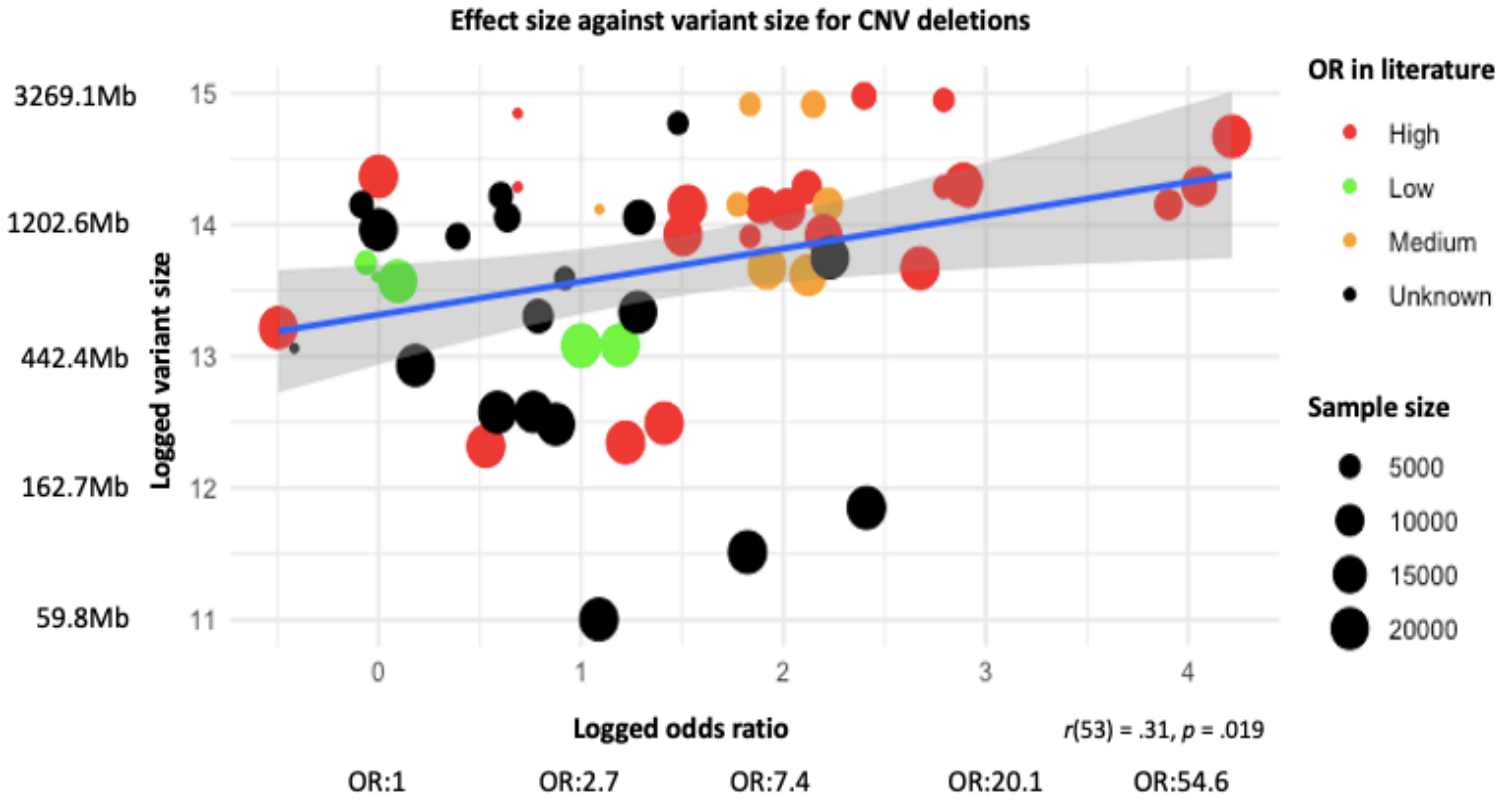
When I examined all 53 CNVs in the CNVcatalog (both deletions and duplications) associated with schizophrenia there was a highly significant positive association between the CNV size and its effect on disease risk measured as the odds ratio for schizophrenia [$r(101) = .27$, $p = .006$]. When looking at deletions and duplications separately, CNV size was positively associated with effect size for the deletions [$r(53) = .31$, $p = .019$], with larger deletions being significantly associated

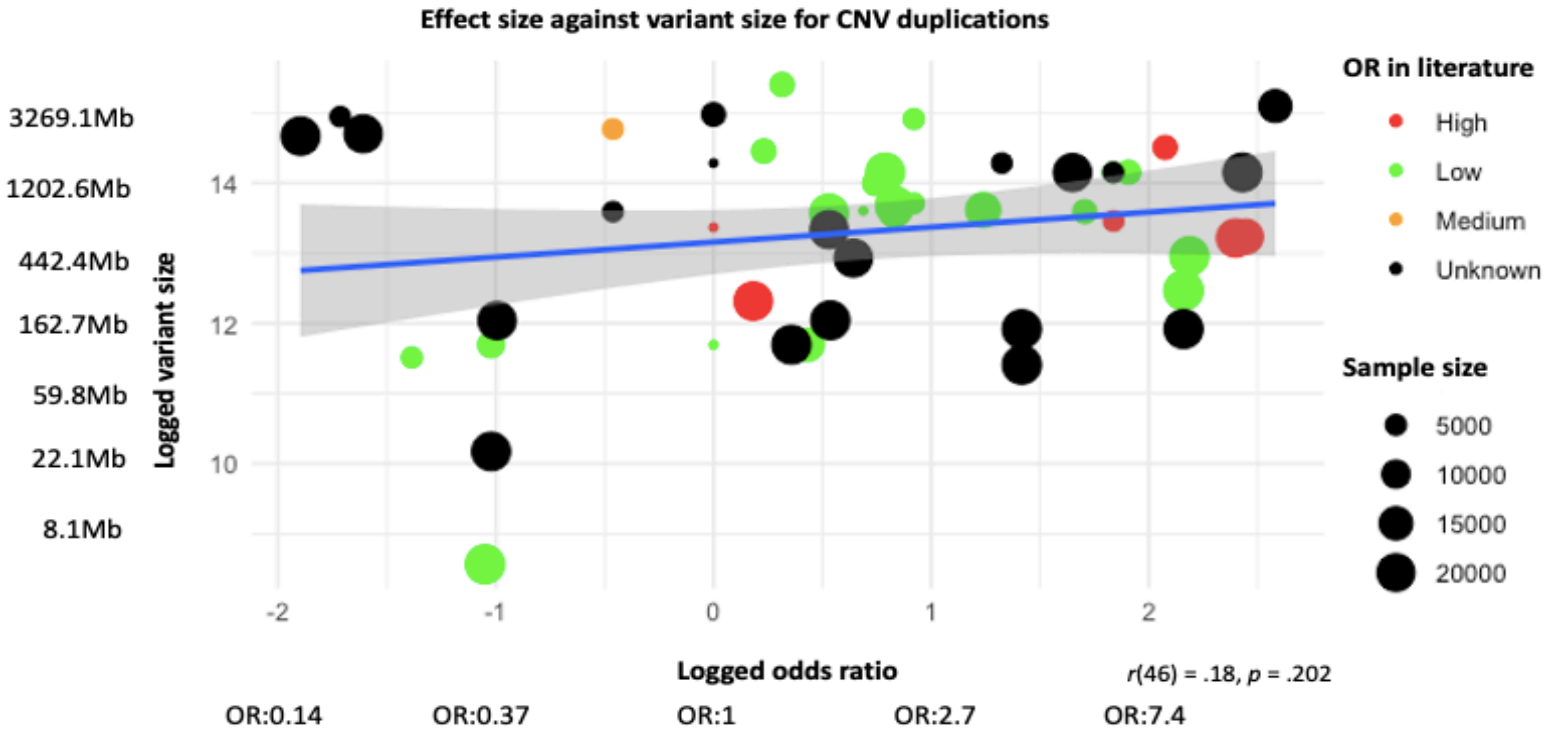
with larger odds ratios for schizophrenia. However, the size of the duplications did not have a significant influence on disease risk [$r(46) = .18$, $p = .202$].

Figure 2.4 shows the scatterplots demonstrating the effect size of schizophrenia risk against CNV size for total CNVs, CNV deletions and CNV duplications

Figure 2.4 *Scatterplots demonstrating the effect size of schizophrenia risk against CNV size for total CNVs, CNV deletions and CNV duplications*




Effect size against variant size for CNV deletions
3269.1Mb
1202.6Mb
442.4Mb
162.7Mb
59.8Mb
Logged variant size
15
14
13
12
11
0
1
2
3
4
Logged odds ratio
r(53) = .31, p = .019
OR:1
OR:2.7
OR:7.4
OR:20.1
OR:54.6
OR in literature
High
Low
Medium
Unknown
Sample size
5000
10000
15000
20000


Effect size against variant size for CNV duplications
3269.1Mb
1202.6Mb
442.4Mb
162.7Mb
59.8Mb
22.1Mb
8.1Mb
Logged variant size
14
12
10
-2
-1
0
1
2
Logged odds ratio
r(46) = .18, p = .202
OR:0.14
OR:0.37
OR:1
OR:2.7
OR:7.4
OR in literature
High
Low
Medium
Unknown
Sample size
5000
10000
15000
20000

Scatterplots demonstrating the relationships between the variant size and the schizophrenia odds ratio for total CNVs, CNV deletions and CNV duplications. Each point represents one locus and its size is adjusted by the sample size of that study. Loci in my data that have been reported to have high schizophrenia ORs in previous studies (15q13.3del, 22q11.2del, 16p11.2del, 16p11.2dup, 17q12del, 1q21.1del, 3q29del, 8q22.2del and 2p16del) are colour coded in red, whereas the loci that have been reported to have moderate (1q21.1del, 15q11.3dup and 17p12del) and low schizophrenia ORs (16p12.1del, 16p13.11 del, 16p13.11 dup, 17q12dup, 1p36dup, 1q21.1dup, 7q36.3dup, VIPR2dup and Xq28dup) are colour coded in amber and green respectively. Loci with unknown risk to schizophrenia are colour coded in black. The plots are on the common logarithmic scale due to skewness of the data. The linear scale for both the odds ratio and the variant size measured in mega bases (Mb) are also provided to aid interpretation.

2.5. Discussion

In conclusion, in this chapter I performed a scoping review of the literature investigating associations of several clinical phenotypes, including schizophrenia, and any possible CNV. By performing exploratory analysis with the data, I identified two gaps in the current literature: i) no meta-analysis of the effect of 16p11.2 distal deletion on schizophrenia risk had been conducted before, ii) no study had investigated the relationship between CNV size and the risk they confer on schizophrenia risk.

I conducted the first meta-analysis of the literature investigating on 16p11.2 distal deletion on schizophrenia risk using the database CNVcatalog. Inconsistencies existed in literature, with three out of the five included studies reporting no significant association of that locus with schizophrenia risk (Rees et al., 2014; Li et al., 2016; Han et al., 2016). However, the other two studies by Rees et al (2016) and Marshall et al (2016), which included much larger samples than the other three studies, reported a significant positive effect of a 16p11.2 distal deletion on schizophrenia risk with odds ratios of 3.3 and 20.6 respectively. A previous meta-analysis exploring 16p11.2 proximal deletions did not find a significant effect on schizophrenia risk (Giaroliwt al., 2014). However, in that meta-analysis they did not investigate distal deletions. The combined meta-analysis of the 5 studies with 53 carriers and 167,772 non-carriers demonstrated a significant effect of carrying this CNV on disease risk with an odds ratio of 2.41; therefore carriers were 2.4 times more likely to develop schizophrenia than non-carriers. My findings render carrying a 16p11.2 distal deletion a significant risk factor for schizophrenia.

To corroborate these results, additional studies are certainly required. Considering that the 16p11.2 distal deletion is quite rare (less than 0.0001% in the general population), the best approach would be the combination of multiple samples in a Psychiatric Genomic Consortium (PGC) type mega-analysis. In addition, it should be noted that the meta-analysis I have conducted is restricted by the inclusion criteria I have set. For instance, all studies should report the CNV coordinates since I am aiming at combining data from different genomic builds. This could have

resulted in the exclusion of studies reporting only association results or frequencies but not the CNV start and stop positions.

The second analysis I performed was an exploratory analysis investigating whether the CNV size is associated with the effect sizes for schizophrenia risk. I found that there was a moderate positive relationship when looking at both CNV deletions and duplications together and that the larger CNVs had larger odds ratios for schizophrenia. The same was found for the CNV deletions, with larger CNVs also associated with larger odds ratios for developing schizophrenia. However, the size of CNV duplications was not associated with increased disease risk.

In the study by Marshall et al (2017) it is reported that the CNV burden measure with the strongest enrichment was the number of genes affected by CNVs, with deletions having a greater schizophrenia risk compared to duplications. The CNV size is undoubtedly important since larger CNVs would result in the disruption of more genes. Several studies have reported that specific CNVs result to a significant risk for schizophrenia as well as cognitive deficits and intellectual disabilities (Thygesen et al., 2020; Guyatt at el., 2018, Clifton et al., 2017; Kendall et al., 2016). Huguet et al. (2018) reported in their study that performance IQ was negatively associated with CNV size for the rare deletions but not for the duplications. Another study by Guyatt et al. (2018) also reported that CNV deletion burden, measured as both number of genes affected, and CNV length, resulted in lower IQ.

The complexity of structural variations in the human genome makes the development of new databases and bioinformatics tools highly important. To the best of my knowledge, CNVcatalog is the first database incorporating sample frequency and association results from published studies examining the phenotypic effects of CNVs.

CNVcatalog can be used in multiple ways to query either by genomic features (such as a specific CNV locus) or by phenotype. All entries are linked to other tables, containing more detailed information about the genomic feature or phenotype in question (genomic built, cytogenetic location, frequencies etc.).

CNVcatalog can be queried for lists of disease related CNVs grouped by associated phenotypes, phenotypes grouped by associated CNVs, comparison of different loci and meta-analytic procedures.

All publicly available studies matching the given search criteria and date restrictions and containing information on either CNV frequencies or associations with clinical phenotypes have been included in the database. My aim is to keep CNVcatalog as a well-curated database that is updated with data from new studies, as it becomes available. My supervisors and I have secured some funding to be able to do so and we are going to apply for additional grants in the future.

To date, a challenge on carrying out meta-analytical procedures on different CNV loci, has been the fact that different studies use different genomic builds to report loci positions, making them not directly comparable. In the CNVcatalog I am currently working on annotating all the genomic positions in build 37 (hg19) to address this major issue. Once this is accomplished, CNV catalog will provide the tools to accurately integrate data from various genomic builds and to easily conduct meta-analyses of several loci and associated phenotypes that have not been feasible before.

One limitation of the CNV catalog is that currently when conducting a meta-analysis it solely reports the logarithm of the odds ratio instead of the odds ratio in the linear scale and does not give a p value, which makes the interpretation of the results challenging. I am working on addressing this before making the database publicly available. For the analyses presented in this thesis, I ran the scripts in the R statistical software in order to obtain the odds ratios and p values reported in the results section.

A further ilmitation is that despite the fact that I carried out the scoping review twice and discussed whether certain papers should be included or not with my supervisors, the possibility of bias due to single rater screening of the papers cannot be ruled out with certainty.

CNVcatalog can also easily be expanded to include additional phenotypes. The scoping review I conducted, focused on CNVs that have been associated with schizophrenia and other disorders with a known genetic overlap with schizophrenia. However, in the future studies exploring CNV associations with any potential phenotype could be added to the database and researchers could employ the tools provided, to analyse the data and visualise the findings. CNVcatalog is a user-friendly interface allowing users to query the database, perform complex meta-analytic procedures and visualise the results, without requiring extensive computational and programming skills. Therefore, it is aimed not only at researchers but also at clinical practitioners.

I anticipate that the CNVcatalog will result in identification of new associations between genomic variations and clinical phenotypes and traits. Additionally, the identification of CNVs linked to multiple phenotypes, could contribute to the deciphering of the shared genetic architecture between different disorders, such as schizophrenia and bipolar disorder, and shed light on the molecular mechanisms underlying the associations between them.

As I add new data, as they become available, this resource could help to explore the pleiotropic effects of CNVs, potentially helping to elucidate associations with new phenotypes. By combining the data from multiple studies, CNVcatalog may also help to quantify and compare the penetrance of CNVs for multiple diseases.

Chapter 3. Associations between psychosis endophenotypes across brain functional, structural and cognitive domains

Part of this chapter has been adapted from the following published article, to which I am a joint first author:

Blakey, R., Ranlund, S., **Zartaloudi, E.,** Cahn, W., Calafato, S., Colizzi, M., Crespo-Facorro B., Daniel C., Díez-Revuelta A., Di Forti M., GROUP, Iyegbe C., Jablensky A., Jones R., Hall M. H., Kahn R., Kalaydjieva L., Kravariti E., Lin K., McDonald C., McIntosh A., PEIC, Picchioni M., Powell J., Presman A., Rujescu D., Schulze K., Shaikh M., Thygesen J. H., Toulopoulou T., Van Haren N., Van Os J., Walshe M., WTCCC2, Murray R. M., Bramon E. (2018). Associations between psychosis endophenotypes across brain functional, structural, and cognitive domains. *Psychological medicine*, 1-20.

3.1. Abstract

Background: A range of endophenotypes characterise psychosis, however, there has been limited work understanding if and how they are inter-related.

Aims: I aimed to examine the relationships between several neurocognitive, brain structural and electophysiological endophenotypes associated with psychosis and to examine group differences between cases, unaffected relatives and healthy controls on endophenotype performance.

Methods: This multi-centre study includes 8,754 participants: 2,212 people with a psychotic disorder, 1,487 unaffected relatives of probands, and 5,055 healthy controls. I investigated cognition [digit span (N=3,127), block design (N=5,491), and the Rey Auditory Verbal Learning Test (N=3,543)], electrophysiology [P300 amplitude and latency (N=1,102)], and neuroanatomy [lateral ventricular volume (N=1,721)]. I used linear regression to assess the interrelationships between endophenotypes.

Results: The P300 amplitude and latency were not associated (regression coef. -0.06, 95% CI -0.12–0.01, p=0.060), and P300 amplitude was positively associated with block design (coef. 0.19, 95% CI 0.10–0.28, $p<0.001$) and digit span coef. 0.15, 95% CI 0.04–0.26, $p=0.009$). There was no evidence of associations between lateral ventricular volume and the other measures (all p>0.38). All the cognitive endophenotypes were associated with each other in the expected directions (all p<0.001). Lastly, the relationships between pairs of endophenotypes were consistent in all three participant groups, differing for some of the cognitive pairings only in the strengths of the relationships.

Conclusions: The P300 amplitude and latency are independent endophenotypes; the former indexing spatial visualisation and working memory, and the latter is hypothesised to index basic processing speed. Individuals with psychotic illnesses, their unaffected relatives, and healthy controls all show similar patterns of associations between endophenotypes, endorsing the theory of a continuum of psychosis liability across the population.

3.2. Introduction

As described in the general introduction of this thesis, endophenotypes are biological markers which are heritable, co-segregate with a disorder within families, and are observed in unaffected family members at a higher rate than in unrelated healthy individuals (Gottesman & Gould, 2003).

Although there is an extensive literature identifying and validating endophenotypes for psychosis, as described in chapter 1, fewer studies have examined the relationships between different endophenotypes. Studies conducted so far have mainly analysed the associations between different cognitive measures. The study by Dickinson, Iannone and Gold (2002), with 120 cases with schizophrenia and 200 healthy controls, investigated the interrelationships between several cognitive measures as evaluated by the Wechsler Adult Intelligence Scale-III (WAIS-III, Weschler, 1997), and clustered them into 4 factors, measuring verbal

comprehension, perceptual organisation, working memory and processing speed. All factors were positively associated with each other, with the relationships between endophenotypes being stronger among the cases with psychosis than among healthy controls. Dickinson and associates conducted another study in 2006 with 148 schizophrenia patients and 157 healthy controls using a battery of sets examining verbal comprehension, perceptual organization, verbal memory, visual memory, information processing speed and working memory. They reported moderate to strong intercorrelations for all cognitive functions in the schizophrenia group and small to moderate intercorrelations in the healthy control group.

Another study by Seidman et al. (2015) assessed cognitive functioning in 83 schizophrenia cases, 151 unaffected siblings and 209 healthy controls, using 12 neurocognitive tests assessing several cognitive functions. After performing a factor analysis, they clustered the cognitive functions into 5 factors: episodic memory, working memory, perceptual vigilance, visual abstraction and inhibitory processing. All factors, apart from perceptual vigilance, were strongly associated with each other with correlations ranging from small to moderate. Besides, a study by Sheffield et al (2014) also investigated cognitive pairings in 104 schizophrenia cases and 132 healthy controls. The authors reported that there were strongly significant interrelationships between episodic memory, goal maintenance, processing speed and verbal learning both in cases and controls. However, no significant relationships were found in visual processing with any of the other cognitive functions.

Despite the studies described, that examine cognitive pairs, there is still a lack of literature examining brain structural–cognitive pairs and limited literature, constrained to small sample sizes investigating electrophysiological–cognitive pairings. Hermens et al (2010) investigated the relationship between P300 amplitude and cognition in 17 cases with first episode psychosis and 17 healthy controls. The authors reported strong relationships between reduced P300 amplitude and cognitive deficits in processing speed, attention switching, simple attention and verbal learning and memory amongst the cases with first episode psychosis. An additional study by Kim et al (2018) investigated P300 event related

potential (ERP) and neurocognitive performance as assessed by the Global Assessment of Functioning (GAF) in 42 cases with schizophrenia, 32 individuals at genetic high risk (GHR), 32 individuals with clinical high risk (CHR) and 52 healthy subjects. There was a significant association between smaller P300 amplitudes and deficits in all neurocognitive tasks in the schizophrenia group.

Another study by Kaur and associates (2011) reported significant relationships between reduced P300 amplitude and impaired cognitive performance in Rey Auditory Verbal Learning Task (RAVLT) and Trial Making Task (TMT) as measured by WAIS-III, in a sample of 17 cases with first episode psychosis - affective spectrum, 17 cases with first episode psychosis - schizophrenia spectrum and 18 healthy age matched controls. Further, a study by Dong et al (2015) examined the relationship between P300 ERP and working memory in 28 undergraduate students and reported that individuals who exhibited deficits in working memory produced reduced P300 amplitudes compared to individuals without deficits in working memory. A study by Fjell and Walhovd (2001) also examined associations of P300 measures and the digit span and block design tasks as measured by the WAIS-III scale in 72 healthy volunteers. They reported associations of reduced P300 amplitude and impaired cognitive performance in both digit span and block design tasks.

Besides, the inclusion of unaffected relatives in studies investgating endophenotype pairs has been rather rare. Yet, examining relatives – who carry increased genetic risk but have no illness or treatment confounding factors – is crucial for establishing the utility of these markers for genetic research.

In this chapter, I am investigating the relationships between the following electrophysiological, neurocognitive, and neuroanatomical endophenotypes of psychosis in caseswith psychosis, their unaffected relatives and healthy controls:

- P300 event-related potential: A detailed description of the P300 ERP is presented in the general introduction. Briefly, the amplitude reflects the attention required in a task, in proportion to the information by the stimulus, whereas latency is an index of classification speed. Reduced amplitude and

prolonged latency have been reported in cases with psychosis and their unaffected relatives, when compared to unrelated healthy individuals (Blakey et al., 2018; Bodatsch et al., 2015; Earls et al., 2016).

- Cognitive performance: Deficits on cognitive tests such as the digit span, which assesses working memory by requiring individuals to recall a series of digits as presented, the block design, which measures working memory and spatial visualisation by requiring individuals to reconstruct specific shapes with blocks, and the Rey Auditory Verbal Learning Task (RAVLT) immediate and delayed recall, which measures short and long term verbal memory, respectively, by asking individuals to recall a list of words presented to them either right away or after a certain amount of time, are common and persistent across psychotic disorders (Bora & Pantelis, 2015; Bora, Yucel, & Pantelis, 2009; Gur et al., 2007; Heinrichs & Zakzanis, 1998; Kim et al., 2015; Lee et al., 2015; Stone et al., 2011). Abnormalities are often observed before the onset of the illness as well as in unaffected relatives (Birkett et al., 2008; Forbes, Carrick, McIntosh, & Lawrie, 2009; Glahn et al., 2006; Gur et al., 2015; Horan et al., 2008; Ivleva et al., 2012; Park & Gooding, 2014; Reichenberg et al., 2010; Saperstein et al., 2006; Snitz, Macdonald, & Carter, 2006).

- Lateral Ventricular Volume: Increased ventricular volume is a highly replicated finding in patients with psychosis compared to controls (Boos, Aleman, Cahn, Hulshoff Pol, & Kahn, 2007; Crespo-Facorro et al., 2009; Fannon et al., 2000; Fusar-Poli et al., 2013; Haijma et al., 2013; Kempton, Stahl, Williams, & DeLisi, 2010; Kumra et al., 2014; McDonald et al., 2002, 2006; Sharma et al., 1998; Shenton, Dickey, Frumin, & McCarley, 2001; Strasser et al., 2005; Wright et al., 2000). This enlargement has been attributed to neurodevelopmental difficulties, disease progression, or the effects of antipsychotic medications (Gogtay et al., 2003; McDonald et al., 2006; Pilowsky, Kerwin, & Murray, 1993).

I conducted a mega-analysis, which is a statistical analysis that comprises and analyzes data from multiple studies, seeking to investigate the relationships

between multi-modal endophenotypes. It includes the largest sample yet of individuals with psychosis, their unaffected first-degree relatives, and controls. The main objective is to facilitate the use of endophenotypes for genetic research into psychosis, which requires well defined and characterised measures. The aim of this study was therefore to examine the relationships between different endophenotype pairs, and in particular, to characterise the P300 event related potential in the context of well-defined cognitive markers. This is the largest sample so far with both ERP and cognitive data that also includes unaffected relatives of patients with psychosis. I hypothesize that a poorer cognitive performance will be associated with reduced P300 amplitude and delayed latency, and that ventricular volumes, which is a measure of brain structure would be associated, with other psychosis endophenotypes of functional nature. I also expect that unaffected relatives will exhibit worse endophenotype performance compared to healthy subjects but better than the patients.

3.3. Methods

Sample and clinical assessments

The total sample included 8,754 participants: 2,212 individuals with a diagnosis of a psychotic disorder (see table 3.1 for a breakdown of diagnoses), 1,487 of their unaffected first-degree relatives (with no personal history of psychosis), and 5,055 healthy controls (with no personal or family history of psychosis). Relatives and controls were not excluded if they had a personal history of non-psychotic disorders (such as depression or anxiety), provided they were well and off psychotropic medication at the time of testing and for the preceding 12 months.

To confirm or rule out a DSM-IV (APA, 1994) diagnosis, all participants underwent a structured clinical interview with either the Comprehensive Assessment of Symptoms and History (Andreasen, Flaum, & Arndt, 1992), the Structured Clinical Interview for DSM Disorders (Spitzer, Williams, Gibbon, & First, 1992), the Schedule for Affective Disorders and Schizophrenia (Endicott & Spitzer, 1978) or

the Schedule for Clinical Assessment in Neuropsychiatry, Version 2.0 (Wing, Babor, Brugha, Burke, et al., 1990). Participants were excluded if they had a history of neurologic disease or a loss of consciousness due to a head injury.

Recruitment took place across 11 locations in Australia and Europe (Germany, Holland, Spain, and the United Kingdom) (see table S4). Participants provided written informed consent, and the study was approved by the respective ethical committees at each of the 11 participating centres. Some centres have previously published comparisons in endophenotype performance between groups (patients, relatives and controls) (Bramon et al., 2005; Collip et al., 2013; Crespo-Facorro et al., 2009; González-Blanch et al., 2007; Hall et al., 2006; Hulshoff Pol et al., 2002; Johnstone, Ebmeier, Miller, Owens, & Lawrie, 2005; McDonald et al., 2002; Price et al., 2006; Schulze et al., 2006; Steel et al., 2002; Toulopoulou et al., 2010; Waters, Price, Dragović, & Jablensky, 2009; Weisbrod et al., 1999; Wobrock et al., 2009). Here, I also present results of a mega-analysis of the combined multi-centre sample in table 3.1. For information regarding the family sizes participating in the study refer to table S6.

Table 3.1 Sample characteristics (N=8,754).

	Patients with psychosis	**Unaffected relatives**	**Controls**	**Total sample**
Sample size, N (%)	2,212 (25.3%)	1,487 (17.0%)	5,055 (57.7%)	8754
Age, mean years (SD)†	33.6 (10.6)	46.0 (15.8)	45.5 (16.2)	42.6 (15.8)
Age range (years)	16 – 79	16 – 85	16 – 89	16 – 89
Gender (% female)†	32.1%	58.0%	51.5%	47.7%
Diagnoses; N (%)				
Schizophrenia	1396 (63.1%)	-	-	1396 (15.9%)
Bipolar I Disorder	135 (6.1%)	-	-	135 (1.5%)
Psychosis NOS	168 (7.6%)	-	-	168 (1.9%)
Schizophreniform Disorder	158 (7.1%)	-	-	158 (1.8%)
Schizoaffective Disorder	124 (5.6%)	-	-	124 (1.4%)
Brief Psychotic Disorder	56 (2.5%)	-	-	56 (0.6%)
Other psychotic illness	175 (7.9%)	-	-	175 (2.0%)
Depression		246 (16.5%)	232 (4.6%)	478 (5.5%)

Anxiety		47 (3.2%)	24 (0.5%)	71 (0.8%)
Other non-psychotic illness		62 (4.2%)	106 (2.1%)	168 (1.9%)
No psychiatric illness		1,132 (76.1%)	4,693 (92.8%)	5,825 (66.5 %)
Endophenotypes; N (sample size), Mean (SD) of raw scores, unadjusted for covariates				
P300 amplitude	N=397	N=379	N=313	N=1,089
(μV)	10.5 (6.1)	11.0 (6.7)	13.7 (7.0)	11.6 (6.7)

Neuropsychological assessments

The Wechsler Adult Intelligence Scale, revised version (Wechsler, 1981) or third edition (Wechsler, 1997), were administered to participants. Performance on two subtests was used for analyses: the combined forward and backward digit span (measuring attention and working memory) and block design (measuring spatial visualisation). The Rey Auditory Verbal Learning Test (Rey, 1964), including both immediate and delayed recall (assessing short-and long-term verbal memory, respectively), was also administered. Higher scores on the cognitive tasks indicate better performance. The full methodology for each contributing site has been previously published (Crespo-Facorro et al., 2007; González-Blanch et al., 2007; Johnstone et al., 2005; Korver, Quee, Boos, Simons, & de Haan, 2012; Toulopoulou et al., 2010; Walters et al., 2010; Waters et al., 2009).

EEG data collection and processing

Electrophysiological data were obtained from three sites (table S4). EEG was collected from 17 to 20 electrodes placed according to the International 10/20 system (Jasper, 1958) during the delivery of an oddball task.

Although the oddball task is very well established in psychosis, the way of delivery differs significantly between studies. Despite the fact that both Picton et al. (2000) and Luck (2014) have attempted to provide guidelines for aquiring ERP data, there has been no attempt to enforce those suggestions. Therefore, to date, there is not a standard way of acquiring ERP data using the oddball paradigm. A study by Collier et al (2016) provided evidence of a significant difference between auditory and visual oddball tasks and reported that activation abnormalities were more pronounced in the auditory tasks. However, in this chapter all stimuli used to measure the P300 response in the oddball paradigm by all sites, used solely auditory stimuli and therefore, it cannot have confounded my results.

Additional differences in the delivery of the oddball paradigm across studies involve differences by tone (either two or three-tone experiment), level of difficulty of the task (how much the deviant stimulus differs from the standard stimuli) and differences in the pitch and the duration that stimuli are presented to the participants (Krigolson, 2018). EEG data acquisition and processing methods in my study varied slightly between sites as summarised below.

In my study, the P300 event related potential was obtained using a standard two-tone frequency deviant auditory oddball paradigm, with standard (‘non target’) tones of 1000Hz and rare (‘target’) tones of 1500Hz. The number of tones presented varied from 150 to 800, the tones were 80dB or 97dB, lasted for 20-50ms, and the inter-stimulus interval was between 1 and 2 seconds. The majority of participants (93.4%) were asked to press a button in response to ‘target’ stimuli, but a subset was asked to close their eyes and count ‘target’ stimuli in their head instead.

The data were continuously recorded in one of three ways: 500Hz sampling rate and 0.03-120Hz band pass filter; 200Hz sampling rate and 0.05-30Hz band pass filter; or 400Hz sampling rate and 70Hz low-pass filter. Linked earlobes or mastoids were used as reference and vertical, and in most cases also horizontal, electro-oculographs were recorded at each site and used to correct for eye-blink artefacts using regression based weighting coefficients (Semlitsch, Anderer, Schuster, & Presslich, 1986). After additional manual checks, artefact-free epochs were included and the baseline was corrected before averaging. The averaged waveforms to correctly detected targets were then filtered using 0.03 or 0.05 Hz high-pass and 30 or 45 Hz low-pass filters. The peak amplitude and latency of the P300 were measured at electrode location PZ (parietal midline), within the range of 250-550ms post-stimulus.

Magnetic Reasonace Imaging (MRI) data collection and processing

MRI data acquisition and image processing varied between sites (Barta, Dhingra, Royall, & Schwartz, 1997; Collip et al., 2013; Crespo-Facorro et al., 2009; Dutt et al., 2009; Frangou et al., 1997; Habets, Marcelis, Gronenschild, Drukker, & Van Os, 2011; Hulshoff Pol et al., 2002; Mata et al., 2009; McDonald et al., 2006, 2002; McIntosh et al., 2004; McIntosh, Job, et al., 2005; McIntosh, Harrison, Forrester, Lawrie, & Johnstone, 2005; Schulze et al., 2006; Wobrock et al., 2009) and are presented in detail in table S5. Field strengths included 1, 1.5 or 3 Tesla. Lateral ventricular volumes were measured using automatic or semi-automatic region of interest analyses, and included the body, frontal, occipital and temporal horns.

Statistical methods

Mega-analysis of group comparisons: Endophenotype measures were first standardised for each site separately using the mean and standard deviation within each site. Linear regression analyses for each measure were used to establish whether the endophenotype performance differed according to group (patients,

relatives, and controls). The outcome in each regression model was the endophenotype measure and the main predictor was the group. These analyses were adjusted for age, gender, clinical group, study site and, where significant, group by site interactions.

Associations between endophenotypes: Linear regression models were used to investigate associations between each pair of endophenotypes. The potential effect modification by group membership was assessed by specifying in the statistical model a term for the interaction between the predictor of the endophenotype pair and group (patient, relative, control). Where I found evidence that the relationship between a pair of endophenotypes differed significantly according to group, associations are reported separately for patients, relatives and controls. Where there was no evidence of effect modification, the interaction term was dropped from the model, and associations are reported for the whole sample adjusted for group. These analyses were adjusted for age, gender, clinical group and study site.

In all analyses, I accounted for correlations between individuals within families using robust standard errors. In this study, 63% of the participants had no other family member taking part, but the study also included 1,056 families of 2-11 members each (85% of the families had only two members included in the sample). This family clustering violates the independence of observations assumption in linear regression. To account for this clustered structure in the dataset I created a new variable “family ID” that was shared by all individuals in each family. Then, I used the variance estimator with the robust cluster option in all the linear regression models. This allowed us to account for the within-family correlations and maintain correct type-1 error rates. This is a standard approach in family studies (Bramon et al, 2014; Ranlund et al., 2014; Shaikh et al., 2013).

I examined the distribution of residuals and plots of residuals versus fitted values for all models and was able to rule out departures from normality and heteroscedasticity. Lateral ventricular volume showed a positively skewed distribution and to account for this I used bootstrap methods for analyses where this is the outcome variable. Heteroscedasticity was not found to be a concern for

ventricular volumes. P values are not presented for the models which used bootstrapping; instead, I examined the 95% bias-corrected confidence intervals (CI) to check for statistical significance at the 5% level (p=0.05).

Although I tested 7 endophenotypes, I expect measurements within domains to be correlated and thus a correction of p-values by 7 tests through Bonferroni was deemed too stringent for a hypothesis-driven study such as this (Perneger, 1998; Rothman, 1990; Savitz & Olshan, 1995). I therefore corrected for associations between 3 domains (EEG, MRI, cognition), with a corrected significance threshold of 0.05/3 = 0.0167, that I rounded to the slightly more stringent cut-off of p<0.01. Statistical analyses were conducted using STATA version 13.

3.4. Results

Sample characteristics

The sample characteristics are summarised in table 3.1. Patients were on average 12.4 years younger than relatives (95% CI: 11.4 to 13.4; p<0.001) and 11.9 years younger than controls (95% CI: 11.1 to 12.7; p<0.001). There was no evidence of any age difference between relatives and controls. There was a lower proportion of females than males among patients than among relatives and controls (32.1%, 58.0% and 51.5% respectively; global p<0.001).

Group comparisons on endophenotype performance

As shown in figure 3.1 and table 3.2, differences between the three participant groups on the endophenotypes followed the expected pattern with performance improving from patients through to relatives and controls. I found evidence that patients' scores differed significantly from those of controls with smaller P300 amplitudes, delayed P300 latency, larger lateral ventricular volumes and deficits in digit span, block design and RVLT immediate recall. When compared to controls,

the unaffected relatives showed reduced P300 amplitude, delayed P300 latency and poorer performance in digit span and block design.

Table 3.2 Endophenotype performance comparison across clinical groups.

	Total Sample	**Patients – Controls**	**Patients – Relatives**	**Relatives – Controls**
Endophenotype	**Global p-value***	**Mean difference (95% CI)**	**Mean difference (95% CI)**	**Mean difference (95% CI)**
P300 amplitude	< 0.001	-0.50 (-0.71 to -0.29) p < 0.001	-0.16 (-0.32 to -0.01) p = 0.061	-0.34 (-0.54 to -0.14) p = 0.001
P300 latency	< 0.001	0.47 (0.33 to 0.61) p < 0.001	0.03 (-0.14 to 0.19) p = 0.749	0.44 (0.29 to 0.60) p < 0.001
Lateral Ventricular Volume		0.20 (0.08 to 0.32)	0.09 (-0.06 to 0.23)	0.11 (-0.04 to 0.25)
Digit Span	< 0.001	-0.72 (-0.88 to -0.55) p < 0.001	-0.14 (-0.32 to 0.05) p = 0.141	-0.58 (-0.77 to -0.39) p < 0.001

Block Design	< 0.001	-0.91 (-1.07 to -0.75) p < 0.001	-0.08 (-0.21 to 0.04) p = 0.190	-0.83 (-0.97 to -0.69) p < 0.001
RAVLT **immediate recall**	< 0.001	-1.32 (-2.29 to -0.37) p = 0.007	-1.24 (-2.22 to -0.27) p = 0.012	-0.08 (-0.24 to 0.07) p = 0.286
RAVLT **delayed recall**	=0.123	-0.98 (-2.21 to 0.25) p =0.118	-0.94 (-2.18 to 0.30) p =0.136	-0.03 (-0.20 to 0.13) p = 0.669

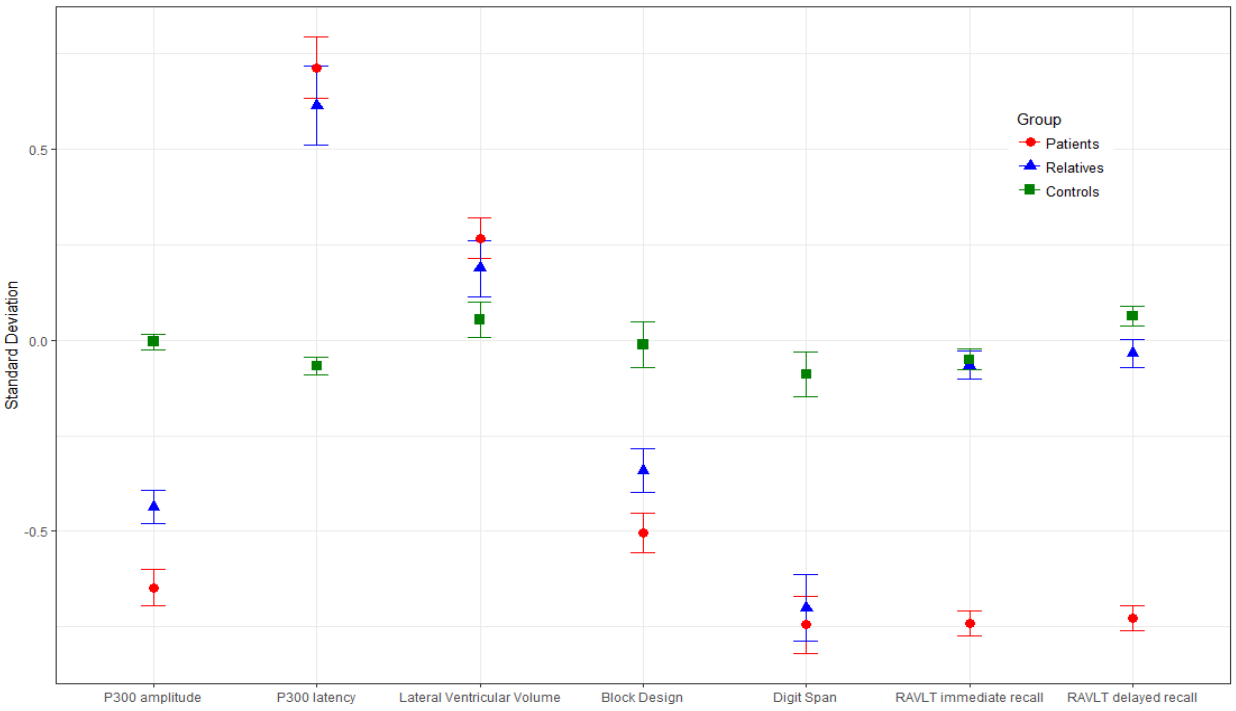
Linear regression models investigating group differences on endophenotype performance. Endophenotype data were standardised for each site using the mean and standard deviation within each site. The main predictor was clinical group (patients, relatives and controls). All models included age, gender, study site and, where significant, group by centre interactions. I used robust standard errors to account for correlations within families in all models.

* p-value for the overall test of a group effect; Note that P values were not produced for the models that include lateral ventricular volume since I used bootstrapping, which is a percentile-based method; therefore I looked at the bias-corrected confidence intervals to check for significance.

RAVLT = Rey Auditory Verbal Learning Task; CI = Confidence Interval.

As shown in table S8, there was no evidence of model instability based on the estimates and confidence interval width between the models with and without age and sex.

Figure 3.1 Estimated marginal means (adjusted for average age, gender, and study site) of standardised endophenotype scores by group (patients, relatives, and controls).



Error bars represent standard errors of the means. RAVLT = Rey Auditory Verbal Learning Task.

Associations between endophenotype pairs

Associations which do not differ according to clinical group

Associations between endophenotype pairs where there was no evidence of effect modification by group are reported in Table 3.3. There was no evidence of an association between the P300 amplitude and latency at the 1% level of statistical significance (coef. -0.06, 95% CI -0.12 to 0.01, p=0.06). The P300 amplitude was positively associated with digit span (coef. 0.15, 95% CI 0.04 to 0.26, p=0.009) and block design (coef. 0.19, 95% CI 0.10 to 0.28, p<0.001) performances, but not with either of the RAVLT measures. The P300 latency showed weak evidence of a negative association with digit span (coef. -0.15, 95& CI -0.28 to -0.03, p=0.017). Lateral ventricular volume showed no evidence of an association with any of the other measures. All cognitive pairings were significantly positively associated (all p<0.001).

Associations which differ according to clinical group

For three pairs of cognitive endophenotypes, I found evidence of an interaction with group. This indicates that the association between these endophenotype pairs differs between patients, relatives and controls, as shown in figure 3.2. In all three cases, the relationship between endophenotype pairs was in the same direction for the three groups, differing only in magnitude.

There was strong evidence that the digit span and RAVLT immediate and delayed recall were positively associated with scores on the block design task in all three groups (patients, relatives and controls). The magnitude of each association was greater among patients than controls in the associations of block design with digit span (0.28, 95% CI 0.19 to 0.38, p < 0.001), with RAVLT delayed recall (0.19, 95% CI 0.09 to 0.29, p < 0.001) and with RAVLT immediate recall (0.12, 95% CI 0.02 to 0.23, p = 0.018). There was no evidence that the strength of the relationship among relatives was different from that among controls (all p>0.03). Full results are presented in table S7.

Table 3.3 Adjusted associations between endophenotypes in the whole sample

	P300 latency	**Lateral Ventricular Volume**	**Digit Span**	**Block Design**	**RAVLT immediate recall**	**RAVLT Delayed recall**
P300 amplitude	N=1,083 -0.06 (-0.12 to 0.01) p = 0.060	N=428 0.05 (-0.07 to 0.15)	N=340 0.15 (0.04 to 0.26) p = 0.009	N=426 0.19 (0.10 to 0.28) p < 0.001	N=255 0.11 (-0.02 to 0.25) p = 0.102	N=255 0.08 (-0.06 to 0.22) p = 0.281
P300 latency	-	N=434 0.02 (-0.08 to 0.15)	N=346 -0.15 (-0.28 to -0.03) p = 0.017	N=437 -0.04 (-0.12 to 0.04) p = 0.333	N=254 0.03 (-0.09 to 0.15) p = 0.699	N=254 0.03 (-0.07 to 0.14) p = 0.501
Lateral Ventricular Volume		-	N=468 -0.01 (-0.09 to 0.09)	N=1001 0.02 (-0.04 to 0.09)	N=498 -0.04 (-0.14 to 0.06)	N=492 -0.02 (-0.11 to 0.09)

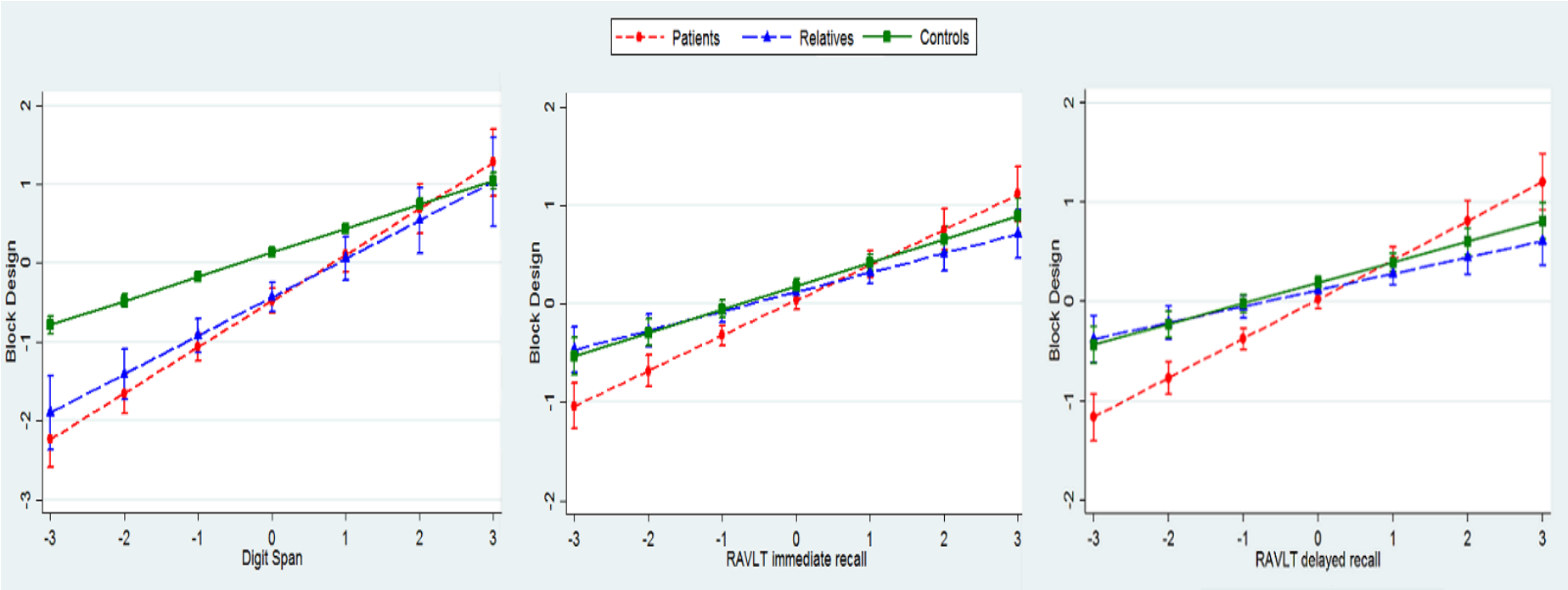
Digit Span	-	N=2754 0.33 (0.30 to 0.36) p < 0.001	N=291 0.39 (0.28 to 0.49) p < 0.001	N=291 0.31 (0.20 to 0.42) p < 0.001
Block Design	-	-	N=2169 0.26 (0.21 to 0.30) p < 0.001	N=2137 0.24 (0.20 to 0.29) p < 0.001
RAVLT immediate recall	-	-	-	N=3505 0.76 (0.74 to 0.78) p < 0.001

RAVLT = Rey Auditory Verbal Learning Task.

Regression models using standardised scores, adjusted for age, gender, study site and group using robust standard errors to account for correlations within families and, where significant, group by centre interactions.

Statistics reported are sample sizes, regression coefficients (95% confidence intervals), and p-values. Note that P values were not produced for the models that include lateral ventricular volume since I used bootstrapping, which is a percentile-based method; therefore I looked at the bias-corrected confidence intervals to check for significance.

Figure 3.2 Interactions between group (patient, relative and control) and endophenotype pairs (standardised scores).



Graphs are adjusted for covariates (age, gender and study site), and include 95% confidence intervals. RAVLT = Rey Auditory Verbal Learning Task.

3.5. Discussion

This study examined the relationships between different multi-modal psychosis endophenotypes in a large multi-centre sample of patients, their unaffected first-degree relatives, and controls.

I found no evidence of association between the P300 amplitude and latency, supporting that these are independent measures. To examine whether variability on P300 amplitude and latency could potentially affect the correlations between these, I tested for heteroscedasticity between clinical groups. The standard deviations between the patient, relative and control groups did not vary significantly and are thus unlikely to explain the lack of correlation between P300 amplitude and latency performance.

In contrast to my results, Hall et al (Hall et al., 2006) and Polich and colleagues (Polich, 1992; Polich et al., 1997) found a negative correlation between the amplitude and latency. Notably however, these past studies included only small samples (up to 128 participants) compared to my study (N=1,083), and they did not take into account covariates such as age and gender that are known to influence both P300 parameters (Chen et al., 2013; Conroy & Polich, 2007; Goodin, Squires, Henderson, & Starr, 1978; Polich, Howard, & Starr, 1985). Furthermore, in the studies by Polich et al (Polich, 1992; Polich et al., 1997) the amplitude – latency correlation was strongest over frontal electrodes, and not parietal as investigated in the current study.

More recently, Hall et al (Hall et al., 2014) found a negative correlation between the amplitude and latency in a sample of 274 patients with psychosis and controls after controlling for age and gender effects. Further research is therefore required to clarify the relationship between the P300 amplitude and latency, since in my study I only report absence of evidence and not evidence of absence.

I found associations between the P300 amplitude and both digit span and block design, as in previous smaller studies (Dong, Reder, Yao, Liu, & Chen, 2015; Fjell & Walhovd, 2001; Hermens et al., 2010; Kaur et al., 2011). According to the

context-updating theory (Heslenfeld, 2003; Kujala & Naatanen, 2003), the P300 amplitude is an attention-driven, context-updating mechanism, which subsequently feeds into memory stores (Polich, 2007; 2011). Hence, one would expect the amplitude to be associated with cognitive tasks that require attention and working memory, such as digit span and block design (Baddeley, 1992; Ford, 2014; Näätänen, 1990). The context-updating theory provides a possible explanation for the association between P300 amplitude and block design, since this task requires a constant update of the mental representation of the blocks, in order to complete the target pattern (John Polich, 2007, 2011). The lack of evidence for associations between P300 amplitude and the RAVLT tests support the idea that the neurobiology of verbal memory is distinct from the attentional and working memory processes linked to the P300 amplitude (Polich et al., 2011).

The P300 latency showed evidence of a trend-level association with digit span, and no evidence of an association with the other measures. Previous studies have provided conflicting results, with some reporting associations with attention and working memory (Polich, Howard, & Starr, 1983), while others have not (Dong et al., 2015; Fjell & Walhovd, 2001; Walhovd & Fjell, 2003). The P300 latency has been conceptualised as a measure of classification speed (Polich, 2011; van Dinteren, Arns, Jongsma, & Kessels, 2014). Investigating the relationship between behavioural reaction times (i.e. the speed of button press in the task) and the P300 latency, some have found associations (Bashore, Wylie, Ridderinkhof, & Martinerie, 2014) while others have not (Ramchurn, de Fockert, Mason, Darling, & Bunce, 2014). Furthermore, there is a substantial body of research showing that the P300 latency, as well as reaction times increase (that is they slow down) with ageing in healthy participants (Chen et al., 2013; Polich, 1996). Based on my findings I hypothesise that the P300 latency is a specific measure of processing speed at a basic neuronal level. In contrast, block design and the RAVLT task – while influenced by processing speed – reflect wider cognition including spatial abilities and verbal memory. The more complex elements to these tasks may therefore obscure effects of a simple processing speed, and hence explain the lack of association with P300 latency. The trend-level association with digit span

performance – a task dependent on attention and short-term working memory – is in line with this interpretation too.

In terms of lateral ventricular volume, there was no evidence of a relationship with any other endophenotype investigated. Enlargement of cerebral ventricles remains the best replicated biological marker in schizophrenia and bipolar disorder, according to several meta-analyses (De Peri et al., 2012; Fraguas, Díaz-Caneja, Pina-Camacho, Janssen, & Arango, 2016; Fusar-Poli et al., 2013; Huhtaniska et al., 2017; Kempton et al., 2010; Moberget et al., 2017; Olabi et al., 2011; van Erp et al., 2016). My hypothesis that ventricular volumes would correlate with other endophenotypes of a functional nature was not confirmed by my data. Of course, for such analyses my sample size was modest ranging 428 to 1,001 and lack of statistical power could be a potential reason. Keilp et al (Keilp et al., 1988) found an association with verbal memory and others have found enlarged lateral ventricles to be associated with poorer motor speed (Antonova, Sharma, Morris, & Kumari, 2004; Dong et al., 2015; Hartberg et al., 2011). A limitation of my study is the heterogeneity of the MRI methodology between study sites, which might have obscured any true associations. I concluded that ventricular volumes do not seem to exert a detectable influence on brain function in terms of cognition or cortical neurophysiology, however association studies of structural-functional biomarkers in larger samples are still needed.

With regards to group comparisons, our mega-analysis confirms that both patients and relatives showed reduced amplitudes and prolonged latencies of the P300, compared to controls, replicating past findings and providing further evidence that these are endophenotypes for psychosis (Bestelmeyer et al., 2009; Bramon et al., 2005; Díez et al., 2013; Price et al., 2006; Schulze et al., 2008; Thaker, 2008; Turetsky, Cannon, & Gur, 2000).

Although patients showed enlarged lateral ventricles compared to controls, a very well supported finding (Cahn et al., 2009; Kempton et al., 2010; Steen, Mull, McClure, Hamer, & Lieberman, 2006; Wright et al., 2000), having adjusted by age and sex I observed no volume differences between relatives and controls. This is consistent with the latest meta-analysis of brain structure in relatives of patients

with schizophrenia (Boos et al., 2007), and suggests that enlarged ventricles in patients are less heritable than previously thought. Instead, they might be related to illness progression, or to environmental effects or antipsychotic medication, as seen in both animal models of antipsychotic exposure (Dorph-Petersen et al., 2005; Konopaske et al., 2007), and in human studies (Fusar-Poli et al., 2013; Ho, Andreasen, Ziebell, Pierson, & Magnotta, 2011; Van Haren, Cahn, Hulshoff Pol, & Kahn, 2013).

For all cognitive measures, patients performed less well than controls, consistent with extensive literature (Ayres et al., 2007; Bora et al., 2014; Bora & Murray, 2014; Bora, Yücel, & Pantelis, 2010; Fatouros-Bergman, Cervenka, Flyckt, Edman, & Farde, 2014; Fusar-Poli et al., 2012; Horan et al., 2008; Stone et al., 2015). For the digit span and block design, there were also statistically significant differences between relatives and controls, suggesting a possible effect of increased genetic risk for psychosis. However, this was not the case for the immediate or delayed recall of the RAVLT task, where controls and relatives had similar performance. While some studies have reported verbal memory impairments in relatives of patients (Massuda et al., 2013; Sitskoorn, Aleman, Ebisch, Appels, & Kahn, 2004; Wittorf, Klingberg, & Wiedemann, 2004), other studies have not (Kim, Kim, Koo, Yun, & Won, 2015; Üçok et al., 2013). These findings suggest that working memory and spatial visualisation might represent more promising endophenotypes for genetic research into psychosis than verbal memory.

The associations between pairs of cognitive measures were strong and in the expected directions, as per previous findings (Dickinson et al., 2002; Gladsjo et al., 2004; Seidman et al., 2015; Sheffield et al., 2014; Sullivan et al., 2003). It is notewhorthy that for some cognitive measures, the relationships interacted with group; however, the direction of the effect remained the same across patients, relatives and controls. The interaction effects with group were found exclusively amongst the cognitive measures, and not in any of the other domains. This is possibly due to the larger sample sizes for the cognitive measures, yielding greater statistical power and enabling the detection of subtle interaction effects.

Both the lack of interaction effects for most associations investigated, and the gradient effects identified (where there was an interaction), are consistent with the notion that endophenotype impairments characterising psychosis represent a continuum that includes both relatives and the general population. Ultimately this continuum reflects the underlying variation in genetic liability of developing the disease (Allardyce, Suppes, & van Os, 2007; DeRosse & Karlsgodt, 2015; Esterberg & Compton, 2009; Ian, Jenner, & Cannon, 2010; Johns & van Os, 2001; Wiles et al., 2006).

This study has several limitations that merit discussion. Firstly, association analyses could only be performed for those participants with data available for pairs of endophenotypes and this led to relatively smaller samples for some of the associations. Secondly, there was a mismatch in age and gender between patients and relatives. The group of relatives has older individuals and more females compared to the group of patients who are younger and include more males. This is a common occurrence in psychosis family studies because the onset of psychosis in typically in youth. Most of the families who participated in the study include unaffected parents (with greater participation of mothers) and their affected and unaffected offspring. Family studies in psychosis are less likely to recruit affected parents. Because of this, I recruited a control group with a wider age range than either the other groups and with a balanced gender distribution so as to improve the age and sex matching across the two key comparisons (controls versus patients, controls versus relatives).

Another limitation of this study is that I was unable to account for potential moderators such as tobacco, other drug use and medication. Also, information about participants' socioeconomic status was not available. These clinical and demographic variables could have a potentially important influence on how the three clinical groups perform on endophenotypes. However, the main analyses, which was to investigate associations between endophenotypes, were all done within-individuals and were thus less likely to be influenced by exposure to drugs and medication. As for clinical variables such as depression, the sample included 5.5% of individuals with a history of depression. Depression did not constitute an

exclusion criterion for my study because it is such a prevalent disorder that if excluded it would probably make my findings hard to generalize. I have re-analyzed the group comparisons excluding all participants with a history of depression and the overall findings are unchanged.

A further potential limitation was the heterogeneity of methods between study sites; differences in cognitive test versions and variation on the EEG and MRI protocols all introduced greater variability into the data. To overcome this, all measures were standardised within centres to minimise this variability. Despite this challenge, it is precisely through this multi-centre effort that I was able to achieve a very large sample, the key strength of this study. As the Psychiatric Genomics Consortium's work shows, large international collaborations are essential in genetic studies of common diseases and traits (Lee et al., 2013; Ripke et al., 2014; Sklar et al., 2011; Smoller et al., 2013). A further strength of this study is the use of regression models as opposed to the correlation approach frequently seen in the literature (Breteler et al., 1994; Brewer, Campbell, & Crano, 1970; Brillinger, 2001; Kim et al., 2003; Polich et al., 1997, 1983), which allowed us to account for some important confounding factors, such as ageing effects. Not only did this approach reduce vulnerability to spurious correlations, but it allowed the examination of interesting interaction effects across groups.

In summary, this study has investigated the relationships between endophenotypes for psychosis, including measures of cognition, electrophysiology, and brain structure. I have shown that cognitive measures are associated with each other as expected, and I have provided support for the notion that the amplitude and latency of the P300 are independent endophenotypes. The P300 amplitude is an index of spatial visualisation and working memory, while the latency is hypothesised to be a correlate of basic speed of processing. Individuals with psychotic illnesses, their unaffected relatives, and healthy controls, all have similar patterns of associations between all pairs of tested endophenotypes, endorsing the theory of a continuum of liability of developing psychosis across the population.

Chapter 4. Influences of polygenic risk scores and CNV burden on psychosis

4.1. Abstract

Background: Over 100 single nucleotide polymorphisms (SNPs) have been associated with schizophrenia and 30 with bipolar disorder. Individually their predictive power is extremely small, however their combination into a polygenic risk score (PRS) has been proposed as a better alternative to estimate disease risk. Several copy number variants (CNVs) have also been reported to increase the risk of developing psychosis. Their joint contributions on psychosis risk has not been studied yet.

Aim: In this chapter the main aim was to explore the joint contributions of schizophrenia and bipolar disorder PRSs and CNV burden on psychosis liability. I hypothesize that the addition of CNV burden to the models including only PRS burden will significantly increase the explained variance in the likelihood of diagnosis status (i.e. schizophrenia, bipolar disorder, control). I also performed an exploratory analysis examining the classification accuracy of my models to investigate whether they could be accurate enough (>90% predictive accuracy) to potentially be considered for application to a clinical setting.

Method: My sample consisted of 3,695 individuals from the UCL Molecular Psychiatry Lab and the Psychosis Endophenotypes International Consortium, resulting in a total of 1,302 people with schizophrenia, 348 with bipolar disorder, 100 with other psychotic disorders and 1,945 healthy subjects. PRSs were calculated for both schizophrenia and bipolar disorder following standard methods and using the latest data from the Psychiatric Genomics Consortium. CNVs were identified with the PennCNV algorithm. CNV burden was defined as the number of genes affected. I analysed the two datasets separately and then combined them by meta-analysis.

Results: The two PRSs for bipolar and schizophrenia, and CNV burden could explain 11.8% and 10.9% of the variance in disease risk in MPL and PEIC datasets respectively, according to Nagelkerker's pseudo R^2. The addition of CNV burden to the models increased the variance explained only by 0.1% for MPL dataset and by 0.08% in the PEIC dataset. In the meta-analyses, the classification accuracy of my models according to the area under the ROC curve were 81%, 83% and 77% for the comparisons of psychosis vs controls, schizophrenia vs controls and bipolar disorder vs controls respectively.

Discussion: CNV burden significantly contributes to the variance explained but only by a small percentage. a A better understanding of PRS and CNV influences on the risk of developing psychosis is crucial for developing new treatments and could be useful towards early detection and treatment of psychosis. I provide evidence that they are a powerful research tool, albeit not yet accurate enough for clinical use.

4.2. Introduction

Psychotic disorders, including schizophrenia and bipolar disorder, have a lifetime prevalence of over 3% (Bogren, Mattisson, Isberg, & Nettelbladt, 2009; Jongsma, Turner, Kirkbride, & Jones, 2019; Jonna Perälä et al., 2007). Clinical symptoms include hallucinations, delusions and cognitive impairments, severe enough to impair the individual's daily functioning (American Psychiatric Association, 2013). The aetiology of psychotic disorders is attributed to an assortment of factors including environmental and genetic influences (Gratten, Wray, Keller, & Visscher, 2014).

Despite the high heritability estimates, ranging between 60-85%, (Hilker et al., 2018; Johansson, Kuja-Halkola, Cannon, Hultman, & Hedman, 2019; Sullivan et al., 2012b) the genetic architecture of psychotic disorders has not yet been fully deciphered, although it is clear that they are highly polygenic (Matheson et al.,

2017; Owens et al., 2016). Recent Genome Wide Association Studies (GWAS) have revealed 270 loci associated with schizophrenia (Ripke et al., 2020 Andreassen et al., 2013; Pardiñas et al., 2018; Psychosis Endophenotypes International Consortium et al., 2014; Stephan Ripke, Neale, Corvin, Walters, et al., 2014; Rudelfer, 2013; Steinberg et al., 2014) and 30 loci have been associated with bipolar disorder (Andreassen et al., 2013; Bramon & et al, 2014; Geschwind & Flint, 2015; Psychiatric GWAS Consortium Bipolar Disorder Working Group, 2011; Rudelfer, 2013) (Stahl et al., 2019). The odds ratios of each genetic locus range from 1.1 to 1.2 and their predictive power individually is exceedingly small (Geschwind & Flint, 2015; Harrison, 2015; Purcell et al., 2014). However, the calculation of a cumulative polygenic risk score (PRS) has been proposed as a better alternative for exploring the distribution of genetic risk within different samples (Dudbridge, 2013; Wray et al., 2014). Indeed, the PRS for both schizophrenia and bipolar disorder has been repeatedly reported to be highly predictive of case-control status (Ohi et al., 2020; Bergen et al., 2019; Calafato et al., 2018; Derks et al., 2012; Tesli et al., 2014; Trotta et al., 2016; Vassos, Forti, et al., 2017).

Copy number variants (CNVs), which are duplications or deletions of DNA sequence, altering the diploid status of DNA (Bagshaw et al., 2013; Nowakowska, 2017), have also been associated with increased risk of developing schizophrenia (Flomen et al., 2006; Kirov et al., 2009; Levinson et al., 2011; Marshall et al., 2017). Several rare and non-recurrent CNVs show evidence of association with schizophrenia, with odds ratios varying from 2 to 30 (Bergen et al., 2019; Chen et al., 2016; Giaroli et al., 2014; Green et al., 2016; Kirov et al., 2014; Li et al., 2016; Marshall et al., 2017; Priebe et al., 2013; Stefansson et al., 2014; Stone, O'Donovan, Gurling, Kirov, Blackwood, Corvin, Craddock, Sklar, et al., 2008; Szatkiewicz et al., 2014; The international Schizophrenia Consortium, 2008; Walsh et al., 2008). Regarding bipolar disorder, CNVs seem to play a smaller role in the risk of developing the illness (Grozeva et al., 2010). However, there is evidence of association of 30 CNVs with bipolar disorder (Gordovez & McMahon, 2020, Green et al., 2016) with a duplication on chromosome 16p11.2 having an OR of 4.37.

Defining specific CNVs as pathogenic and potentially accountable for psychosis is relatively challenging since large samples are required (Stranger et al., 2011) to detect these rare variants. Moreover, CNVs associated with an assortment of mental illnesses, including psychosis, are not fully penetrant and can be present in healthy individuals (Morrow, 2010). CNV burden has been proposed as an alternative measure to explore the contribution of CNVs to disease risk and has been observed to be significantly increased in patients with schizophrenia (Marshall et al., 2017) and schizoaffective disorder bipolar type (Charney et al., 2019).

Despite the wealth of evidence for both SNPs and CNVs contributing to the genetic liability for psychosis, their joint influence on disease risk is less clear. Bergen et al (2019) investigated the joint contributions of schizophrenia associated CNVs and PRSs in a large sample of 21,094 schizophrenia cases and 20,227 healthy controls. Schizophrenia PRSs were found to be lower in schizophrenia patients who are carriers of known specific risk CNVs compared to non-carriers. The authors concluded that the PRS was reduced in proportion to the effect size of the CNV, thus the higher the CNV odds ratio, the lower the PRS needed to be to develop schizophrenia.

Taniguchi et al. (2019) also investigated the joint contributions of CNV burden and PRS in a sample of 724 schizophrenia patients and 1,178 healthy controls. They reported no significant difference in PRS scores in cases with and without a schizophrenia associated CNV. However, they found significant enrichment of the 22q11.2 deletion in the lowest decile of PRS in schizophrenia patients carrying this variant. To my knowledge, the joint effect of SNPs and CNVs on the liability for bipolar disorder and for psychosis broadly defined has not been studied yet.

Thus, in the present study I aimed at exploring the joint contributions of schizophrenia and bipolar disorder PRSs and CNV burden, on psychosis liability, making use of two datasets with cases of psychosis (schizophrenia, bipolar disorder with psychotic features, other psychotic disorders) and unrelated heathy subjects. I hypothesise that the variance explained will be higher after including both PRSs burden and CNV burden to the analysis compared to models including

only the PRSs. I also performed an exploratory analysis examining whether the classification accuracy of the models was accurate enough (>90% predictive accuracy) to be considered for application to a clinical setting. I also explored whether the inclusion of CNV burden in the models increased the variance explained. I performed another exploratory analysis to investigate whether carrying a schizophrenia associated CNV increases the risk of developing psychosis.

4.3. Methods

The study was performed using two independent samples, the Molecular Psychiatry Laboratory (MPL) dataset from UCL, and the Psychosis Endophenotypes International Consortium (PEIC) family study (Bramon et al, 2014).

Sample collection: *MPL sample*

The initial MPL control sample consisted of 3,549 volunteers, of which 996 were healthy control subjects and 2,553 patients with psychosis. Patients were recruited through mental health services across UK, and the control group consisted of volunteers who saw the study advertised on the internet or in posters and leaflets across general hospitals, universities and community centres. Healthy controls were interviewed with the initial clinical screening questions of the Schedule for Affective Disorders and Schizophrenia or the Structured Clinical Interview (SADS-L) and selected on the basis of not having a past or present personal history of any Research Diagnostic Criteria (RDC; Spitzer, Endicott, & Robins, 1978) defined mental disorder. Heavy drinking and a family history of schizophrenia, alcohol dependence or bipolar disorder were also used as exclusion criteria for controls.

In regards to the patients with psychosis, in order for an ICD-10 diagnosis to be ascertained or ruled out, all cases were interviewed by a psychiatrist or trained researcher using the lifetime version of the Schedule for Affective Disorders and

Schizophrenia-Lifetime Version (SADS-L; Spitzer 1977). Case participants were also rated with the 90-item Operational Criteria Checklist (OPCRIT; McGuffin, 1991).

All cases and controls were of UK or Irish ancestry (Datta et al, 2007). UK National Health Service multi-centre and local research ethics approvals were obtained, and all subjects gave signed informed consent.

Sample collection: *PEIC sample*

The initial PEIC data set consisted of 6,935 participants, of which 3,891 were healthy control subjects and 1,820 patients with psychosis. The samples were collected through seven centres across Australia (Perth) and Europe (Edinburgh, Heidelberg, Holland, London, Munich and Pamplona). Patients with psychosis were recruited through voluntary organisations, advertisements in the local press and from mental health teams from the different sites. Controls were recruited by advertisements in the local press and job centres.

More information regarding the participants collected from each site are presented in table S9. All participants provided written informed consent and the study was approved by the respective ethical committees at each of the participating centres. An overall approval by the ethics committee at the Institute of Psychiatry Psychology and Neuroscience was also granted.

In order for a DSM-IV (American Psychiatric Association, 2000) diagnosis to be ascertained or ruled out, a psychiatrist interviewed all participants using either the SADS-L for DSM Disorders or the Schedules for Clinical Assessment in Neuropsychiatry (Endicott & Spitzer, 1978b; Spitzer, Williams, & Gibbon, 1992; Wing, Babor, Brugha, Cooper, et al., 1990). A total of 77 participants also completed the 90-item OPCRIT checklist (McGuffin, 1991). Participants with a history of neurologic disease or head injury resulting in loss of consciousness were excluded.

DNA extraction and quantification: *MPL sample*

Genomic DNA was obtained from frozen whole blood or saliva samples. DNA was extracted from blood samples using phenol-chloroform and BACC-DNA Extraction kits (Illustra Nucleon Genomic, GE Healthcare, UK). The DNA from saliva samples was extracted according to manufacturer's instructions (Oragene kits, DNA Genotek, Ottowa, Canada). The DNA concentration was quantified by PicoGreen fluorimetry (Invitrogen, Paisley, UK).

DNA extraction and quantification: *PEIC sample*

Genomic DNA was obtained from blood for all participants. The blood samples were sent for genotyping to the Wellcome Trust Sanger Institute, Cambridge, United Kingdom. They were processed in 96-well plate format; each plate carried a positive and a negative control. DNA concentrations were quantified using a PicoGreen assay (Invitrogen, Life Technologies, Grand Island, New York) and an aliquot assayed by agarose gel electrophoresis. A sample passed quality control if the original DNA concentration was at least 50 ng/mL and the DNA was not degraded.

Genotyping Methodology and Quality Control: *MPL sample*

Genotyping was performed on the Illumina PsychArray beadchip at the Broad Institute, MA, US. Quality control were conducted in University College London. Stringent quality control was applied to the genotype information. 62 individuals were excluded on the basis of having degraded or insufficient DNA, or incorrect gender assignments, determined as a mismatch between the reported gender and gender inferred from genetic data. 66 more individuals were excluded for showing poor signal-to-noise ratio in the genotyping assay.

The quality control was conducted at University College London. SNPs with a minor allele frequency <0.5% and SNPs deviating significantly from the Hardy-

Weinberg equilibrium ($p<10^{-6}$) were excluded from the analysis. SNPs from the X or Y chromosomes or mitochondrial DNA were also excluded. Participants with excessive heterozygosity (more than 10 standard deviations above the mean), missing genotype data above 10% and participants with evidence of relatedness were also excluded. In total 212 individuals from the MPL dataset failed quality control.

Genotyping Methodology and Quality Control: *PEIC sample*

To track sample identity, 30 single nucleotide polymorphisms (SNPs) including sex chromosome markers were typed on the Sequenom platform before entry to the whole genome genotyping pipeline. Of the initial 6,935 samples, 347 failed quality control due to degraded or insufficient DNA or incorrect sex classification. The remaining samples were genotyped with the Genome-wide Human SNP Array 6.0 at Affymetrix Services Laboratory (Affymetrix, 2015). Of the samples sent for genotyping, 1,022 showed poor signal-to-noise ratio in the genotyping assay and were excluded from further analysis. Genotype calling was conducted at Wellcome Trust Centre for Human Genetics (University of Oxford) using the CHIAMO algorithm (Marchini, Howie, Myers, McVean, & Donnelly, 2007; Wellcome Trust Case Control Consortium, 2007), modified for use with the Affymetrix 6.0 genotyping array.

The quality control were conducted at King's College London and University College London. Standard quality control procedures were implemented, as described in previous work (Ripke, 2014). SNPs with a study-wide missing data rate over 5%, four or more Mendelian inheritance errors identified with Pedstats (Wigginton & Abecasis, 2005), departure from Hardy-Weinberg equilibrium ($p<10^{-6}$) or minor allele frequency (MAF <0.02) were excluded. SNPs from the X or Y chromosomes or mitochondrial DNA were also excluded from the analysis. Lastly, 9,499 SNPs were removed after visual inspection of the genotyping intensity plots in the program Evoker (Morris, Randall, Maller, & Barrett, 2010) as they were deemed to be poorly genotyped.

Participants were also excluded due to divergent genome-wide heterozygosity (when inbreeding coefficients were F > 0.076 or F < −0.076 as estimated with PLINK (Purcell et al., 2007), or when samples had more than 2% missing data across all SNPs. Lastly, duplicates and monozygotic twins were excluded by removing one of each pair showing identity by descent greater than 95% (PLINK). In total 2,698 samples failed quality control.

Genotype imputation and Quality Control

Genotype imputation was run in parallel in the two datasets. The genotypes that passed the initial SNP quality control were submitted to the Sanger Imputation Server (McCarthy et al., 2016; https://imputation.sanger.ac.uk/). The EAGLE2/PBWT (Durbin, 2014; Loh et al., 2016) pipeline was used for pre-phasing and imputation against the Haplotype Reference Consortium panel (r1.1). This yielded ~39.1 million imputed variants. Since different arrays were used on each dataset, different quality control thresholds were also employed. The resulting genotypes were hard-called using a 0.8 genotype probability threshold and all variants with an INFO score < 0.8 for PEIC and <0.9 for MPL were excluded.

Further quality control of imputed genotypes was performed using PLINK. Imputed SNP exclusion criteria were: missing data rate of over 5%; minor allele frequency < 1%; departure from the Hardy-Weinberg equilibrium ($p < 1e^{-6}$); Mendelian error rate > 10%; and cases vs. controls data missingness significance < $5e^{-6}$. Sample exclusion criteria following imputation were: missing data rate of over 5%, Mendelian error rate > 5% and |inbreeding coefficient| > 0.1. LDAK (Speed, Cai, Johnson, Nejentsev, & Balding, 2017) was used to identify duplicates or twins as pairs of individuals with a kinship coefficient > 0.95 (based on a thinned set of SNPs) and to remove one of each pair. Figures S2 and S3 in appendix 3 present flowcharts of the quality control filtering for SNPs associated with schizophrenia and bipolar disorder in the two datasets and the SNP overlap between each dataset with the leave out version of the disorder specific PGC and the SNP overlap between the MPL and PEIC datasets.

Population Structure Analysis

Principal component analysis (PCA) using EIGENSOFT version 3.0 (Patterson, Price, & Reich, 2006) on a thinned set of SNPs was performed to investigate the genetic structure of the data. Due to the multicentre nature of this study, the first three ancestry principal components (PCs) were included as covariates to control for population stratification.

Kinship matrix

To account for known and cryptic relatedness between individuals, a kinship matrix was generated based on a LD-pruned set of SNPs (102,112 SNPs selected with pruning parameters: $r^2 = 0.2$; window = 1000Kb) using LDAK (Speed et al., 2017) and added as a random effect to the linear mixed model regressions. I set all kinship values below 0.025 to 0 in order to speed up the mixed model regressions.

Polygenic Risk Score Analysis

Polygenic risk scores (PRS) were calculated for both schizophrenia and bipolar disorder using PRSice (http://prsice.info/) (Euesden, Lewis, & O'Reilly, 2015). The PRS scores for each subject were calculated by weighting the number of risk alleles they carried for each SNP by the logarithm of the corresponding odds-ratio (logOR), summing them across all the SNPs. The odds ratios were obtained from the Psychiatric Consortium mega-analysis of genome-wide association studies for schizophrenia (Ripke, Neale, Corvin, Walters, et al., 2014) with a sample of 31,658 cases and 42,022 healthy subjects and for bipolar disorder (Stahl et al., 2019) with a sample of 20,352 cases and 31,358 healthy subjects. As both the MPL and PEIC datasets contributed samples to the Psychiatric Genomics Consortium GWASs, I used summary statistics generated from a PGC subset from which my samples were removed for the PRS calculations. Clumped summary statistics were used for the PRS calculation (independently of whether they were clumped before

running PRSice or clumped by the program). Linkage disequilibrium pruning was employed to identify SNPs in linkage equilibrium with each other. Significance thresholds of $p < 0.01$ and $p < 0.1$ were applied to select the SNPs used for the calculation of the schizophrenia and bipolar PRSs respectively, according to those used in the PGC studies (Ripke et al., 2014; Ruderfer et al., 2018; Stahl et al., 2019). Only the schizophrenia PRS score set at $p < 0.01$ and the bipolar PRS score set at $p < 0.1$, were used in the analyses.

CNV Analysis

The PennCNV algorithm was used to identify CNVs (Wang et al., 2007), and both the log R ratio (LRR) and the B-allele frequency (BAF) were calculated according to the PennCNV-Affy protocol. Data was adjusted for genomic waves and analysis done with standard PennCNV settings.

As a guideline I consulted the papers by Marshall (2017) and Kirov (2016) to set the initial QC cutoff points, but after plotting the BAF-drift, LRR-standard deviation and waviness factor, I noticed that the thresholds were more lenient for MPL dataset and quite sctict for PEIC dataset, resulting in a big loss of sample. Therefore, the cutoff points for each parameter were slightly adjusted for each dataset upon examination of the plots of each parameter.

Quality control analysis was performed as follows: for MPL individuals with BAF-drift > 0.003, LRR-standard deviation > 0.35 or an absolute waviness factor of 0.015 were excluded; for PEIC individuals were excluded if they had BAF-drift > 0.005, LRR-standard deviation > 0.4 or an absolute waviness factor of 0.02. Additionally, individuals with more than 100 CNVs were also excluded from both datasets, since this indicates low quality samples. A total of 2,141 (60.32%) and 3,258 (46.98%) individuals passed sample level QC in MPL and PEIC datasets respectively.

Quality control analysis at CNV-level was conducted in both datasets by merging adjacent calls if the distance between them was less than 25% of the combined

length and by removing CNVs with 10 or less SNPs. CNVs with density of less than one probe per 20 kb or length < 5kb or were also excluded. To include only rare CNVs, I used PLINK to filter out variants with frequency >1%.

After sample-level and CNV-level QC, 1,585 individuals from MPL and 1,494 individuals from PEIC were excluded, leaving us with 5,516 CNVs in 2,141 individuals and 19,292 CNVs in 3,258 individuals for downstream analysis for MPL and PEIC datasets respectively.

Choosing CNV burden measures

The CNV deletion burden, duplication burden and total CNV burden (deletions and duplications combined) were calculated in the following ways: Number of genes affected by CNVs, length of CNVs, CNV intolerance (pli score) and number of genes with low CNV intolerance affected. In both datasets, only the number of genes affected by deletions was significantly different between cases and controls. Taking into consideration that the mega-analysis by Marshall (2016) also indicated the number of genes affected as the most appropriate CNV measure, I decided this was the only CNV burden measure I would include for subsequent analyses in this study. For more information on additional CNV burden measures refer to figures S9 and S10. CNVs were mapped to genes using the “refGene” database, downloaded from UCSC Genome Browser (Kent et al., 2002).

Exploring potential differences in initial samples and the samples included in the analysis

An additional 577 individuals from MPL dataset and 223 individuals from PEIC, were excluded from missing information on age. The final datasets consisted of 1,266 out of the initial 3.549 for the MPL dataset, and of 2,429 out of the initial 6,935 individuals for PEIC dataset. Figures S4 and S5 demonstrate flowcharts of

the process of exclusion due to quality control criteria and age missingness in both datasets.

Statistical Analysis

All data were coded and analyzed using R (R Core Team, 2013). I ran linear regression models adjusted for age and sex to explore the differences in PRS and CNV burdens between the three diagnostic groups (cases with schizophrenia, cases with bipolar disorder, healthy controls).

I used generalized logistic mixed-effects models with the clinical status as outcome and schizophrenia PRS, bipolar PRS and CNV burden (n of genes affected by deletions) as predictors in each model. Age, gender and three ancestry PCs were used as fixed effect covariates. A kinship matrix for each dataset was added as a random effect to account for relatedness between individuals. The p-value threshold of significance for the regression model was set at $p = 0.05$. I also examined the effect of the interaction of CNV burden and the two PRSs by adding it both as a two-way and a three-way interaction to the model. However, when I compared the model with and without the interaction terms, there was no significant difference in the model's classification accuracy. Therefore, I removed the interaction terms from further analyses.

The lme4qtl package (Ziyatdinov et al., 2018) was used for the regression models. The proportion of variance explained by the genetic liability measures was calculated as Nagelkerke's pseudo-R^2 through the comparison of the full regression model against a reduced model with covariates only.

Full model: Clinical group versus controls ~ Schizophrenia PRS + Bipolar PRS + CNV burden (n of genes affected by deletions) + age + gender + pc1 +pc2 +pc3 + (1|kinship matrix)

Reduced model: Clinical group versus controls ~ age + gender + pc1 +pc2 +pc3 + (1|kinship matrix)

The R package pROC (Robin et al., 2011) was used to calculate the area under the receiver operator characteristic (ROC) curve (AUC) in the models by using the predicted case-control status from the full regression models and the real case-control status. I then performed a prognostic accuracy meta-analysis of the areas under the ROC curve using *MedCalc version 16.4.3* (MedCalc Software bv, 2016).

Given that 74% of the cases in MPL dataset and 65% in PEIC dataset had a diagnosis of schizophrenia, to rule out the possibility that my results are driven by this subgroup, I also run the analysis for schizophrenia patients versus controls and bipolar patients versus controls separately as shown in table 4.3.

In order to investigate whether CNV burden was actually contributing to the models' predictive power, I also run the regression models with the schizophrenia and bipolar PRSs but without the CNV burden and used these models as a reference compared to the full model to calculate the variance explained. To explore whether carrying a schizophrenia CNV increased the risk for developing psychosis, I also run logistic mixed-effects models with the clinical status as outcome and schizophrenia PRS, bipolar PRS and CNV carrier status (carrier or non-carrier) as predictors in each model. Age, gender, three ancestry PCs and the kinship matrix were also added to these models.

4.4. Results

Demographics

The MPL sample consisted of 1,049 patients diagnosed with psychosis (771 cases with schizophrenia and 278 cases with bipolar disorder) and 217 healthy controls. In the MPL dataset there was no significant difference with regards to the age of the cases with psychosis and the controls (mean diff = .46, p = .659). There were more males in the patient group compared to the control group ($\chi 2$ = 133.34, p < .001).

PEIC sample consisted of 701 patients with psychosis (531 cases with schizophrenia/schizoaffective or schizophreniform disorder, 70 cases with bipolar disorder and 100 with other psychotic disorder) and 1,728 healthy controls). In PEIC dataset the patient group was significantly younger compared to the control group (mean difference = 12.01, $p < .001$). Age and gender are included as covariates in all analyses. There were more males in the patient group compared to the control group ($\chi 2 = 99.20$, $p < .001$). The participants characteristics of the two datasets are summarized in table 4.1.

I also investigated whether there were significant differences regarding age and gender between the excluded individuals from each dataset and the individuals included in the analyses. Neither in MPL (age: $t = 2.3$, $p = .07$, gender: ($\chi 2 = 30.3$, $p = .06$) nor in PEIC dataset (age: $t = 3.4$, $p = .08$, gender: ($\chi 2 = 0.39$, $p = .53$) were any significant differences.

Table 4.1 Demographic characteristics of the MPL and PEIC samples

		MPL		PEIC	
		Cases	**Controls**	**Cases**	**Controls**
Age, years: mean (SD)		45.63 *(12.40)*	46.09 *(13.36)*	34.43 *(10.38)*	46.44 *(16.42)*
Sex, female: n *(%)*		355 *(33.84%)*	166 *(76.49%)*	204 *(29.10%)*	888 *(51.38%)*
Sub-diagnostic groups n *(%)*	Schizophrenia	771 *(73.49%)*		459 *(65.47%)*	
	Schizoaffective			43 *(6.13%)*	
	Schizophreniform disorder			29 *(4.13%)*	
	Bipolar disorder with psychosis	278 *(20.78%)*		70 *(9.98%)*	
	Brief psychotic disorder			15 *(2.13%)*	
	Delusional disorder			13 *(1.85%)*	
	Psychosis disorder NOS			72 *(10.27%)*	
	Total	1,049	217	701	1,728

SD = Standard deviation; NOS = Not otherwise specified

I explored group differences in bipolar and schizophrenia PRS scores and CNV burden in both datasets. The bipolar polygenic scores differed significantly between the three groups in both MPL (F(2,1263) = 21.81, p < 0,001) and PEIC (F(2,2327) = 51,04, p < 0,001) datasets and in all subgroup comparisons. The schizophrenia polygenic scores also differed significantly across the three groups in both MPL (F(2,1263) = 61.45, p < 0,001) and PEIC datasets (F(2,1263) = 101.5, p < 0,001) and in all the subgroup comparisons in MPL dataset and the comparisons of bipolar cases or schizophrenia cases versus controls in PEIC dataset, but not for the comparison of cases with schizophrenia versus cases with bipolar disorder. No significant differences were found in either dataset when exploring differences in CNV burden, as measured by the number of genes affected by deletions, across the three disgnostic groups. The results of the group comparisons for both datasets are presented in table 4.2.

Table 4.2 Group differences in bipolar PRS, schizophrenia PRS and CNV burden in MPL and PEIC datasets

	Total Sample	Schizophrenia cases – Bipolar cases	Bipolar cases – Controls	Schizophrenia cases – Controls
		MPL dataset		
Bipolar PRS	**F(2,1263) = 21.81**	**-.22**	**.54**	**.32**
	***p* < .001**	***p* = .003**	***p* < .001**	***p* < .001**
Schizophrenia PRS	**F(2,1263) = 61.45**	**.34**	**.50**	**.85**
	***p* < .001**	***p* < .001**	***p* < .001**	***p* < .001**

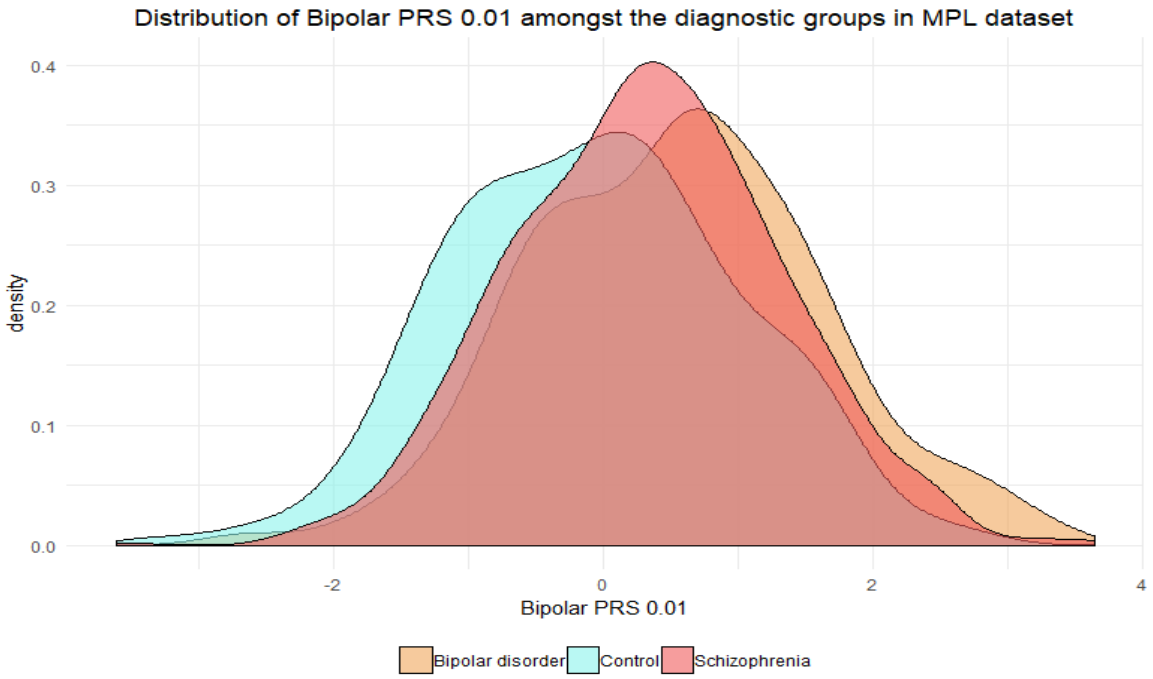
CNV burden	F(2,1263) = 1.01 *p* = .364	.20 *p* = .617	.47 *p* = .321	.67 *p* = .098
		PEIC dataset		
Bipolar PRS	**F(2,2327 = 51.04** ***p* < .001**	**-.53** ***p* < .001**	**.83** ***p* < .001**	**.30** ***p* < .001**
Schizophrenia PRS	**F(2,2327) = 101.5** ***p* < .001**	.03 *p* = .810	**.59** ***p* < .001**	**.62** ***p* < .001**
CNV burden	F(2,2327) = 1.87 *p* = .154	.04 *p* = .974	.65 *p* = .505	.69 *p* = .099

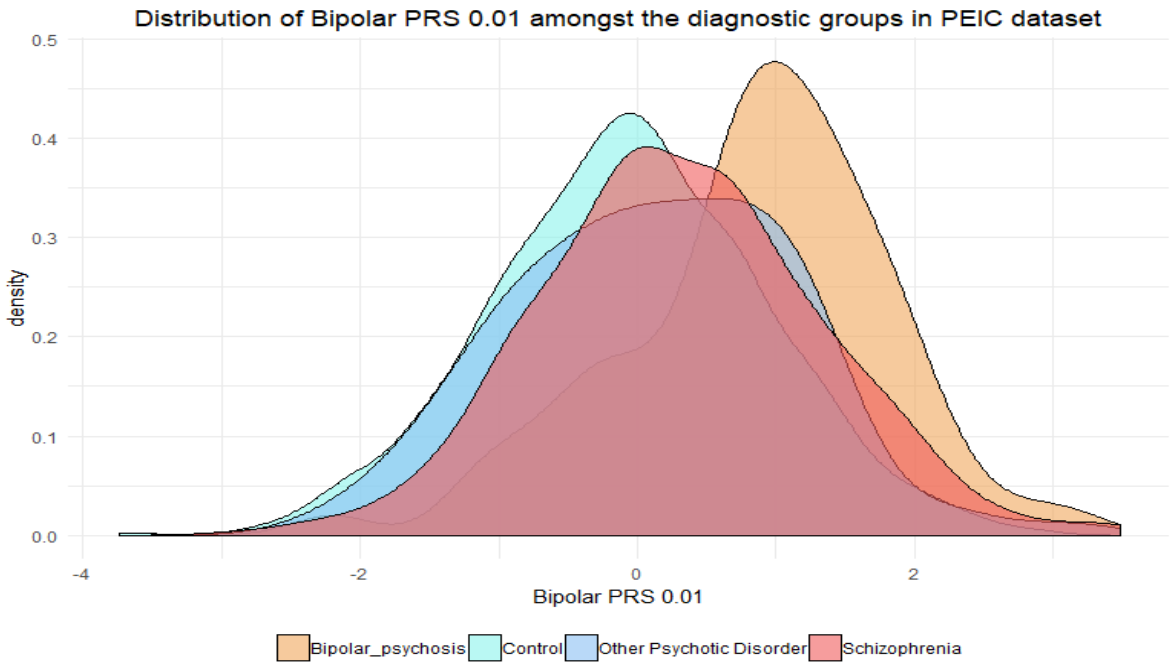
For the 3 group comparisons the mean differences and p values are reported. Models are adjusted by age and sex.

Significant results are highlighted in bold

Figures 4.1 and 4.2 include density plots demonstrating the distributions of bipolar and schizophrenia PRSs across the diagnostic groups in MPL and PEIC datasets. The group differences in PRSs are also depicted in boxplots as shown in figure 4.3. As shown in the figures, the bipolar cases score higher in the bipolar polygenic scores, and the schizophrenia cases score higher in the schizophrenia polygenic scores in both datasets.

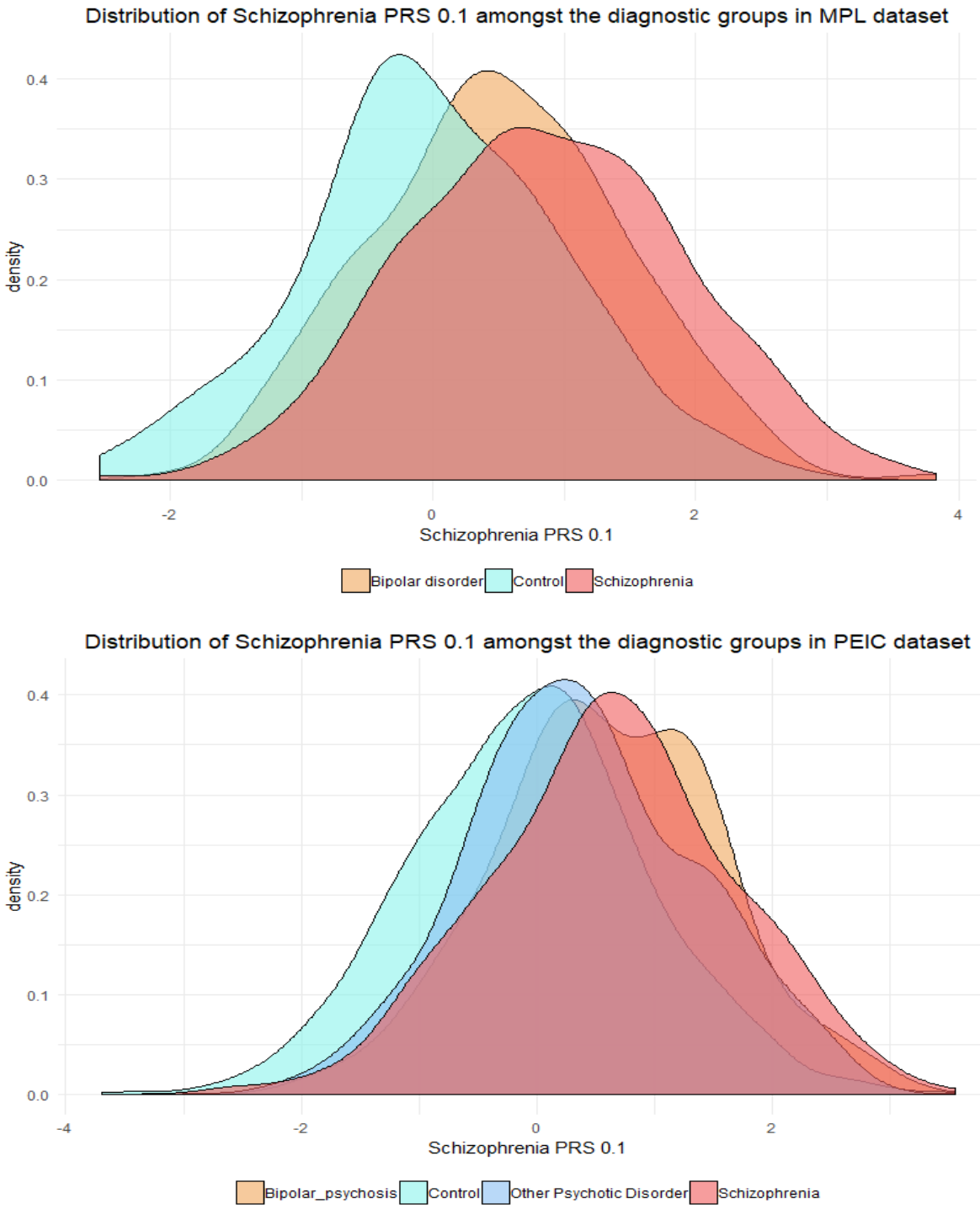
Figure 4.1 Density plots demonstrating the distribution of bipolar polygenic risk score (PRS) amongst the diagnostic groups in MPL and PEIC datasets.





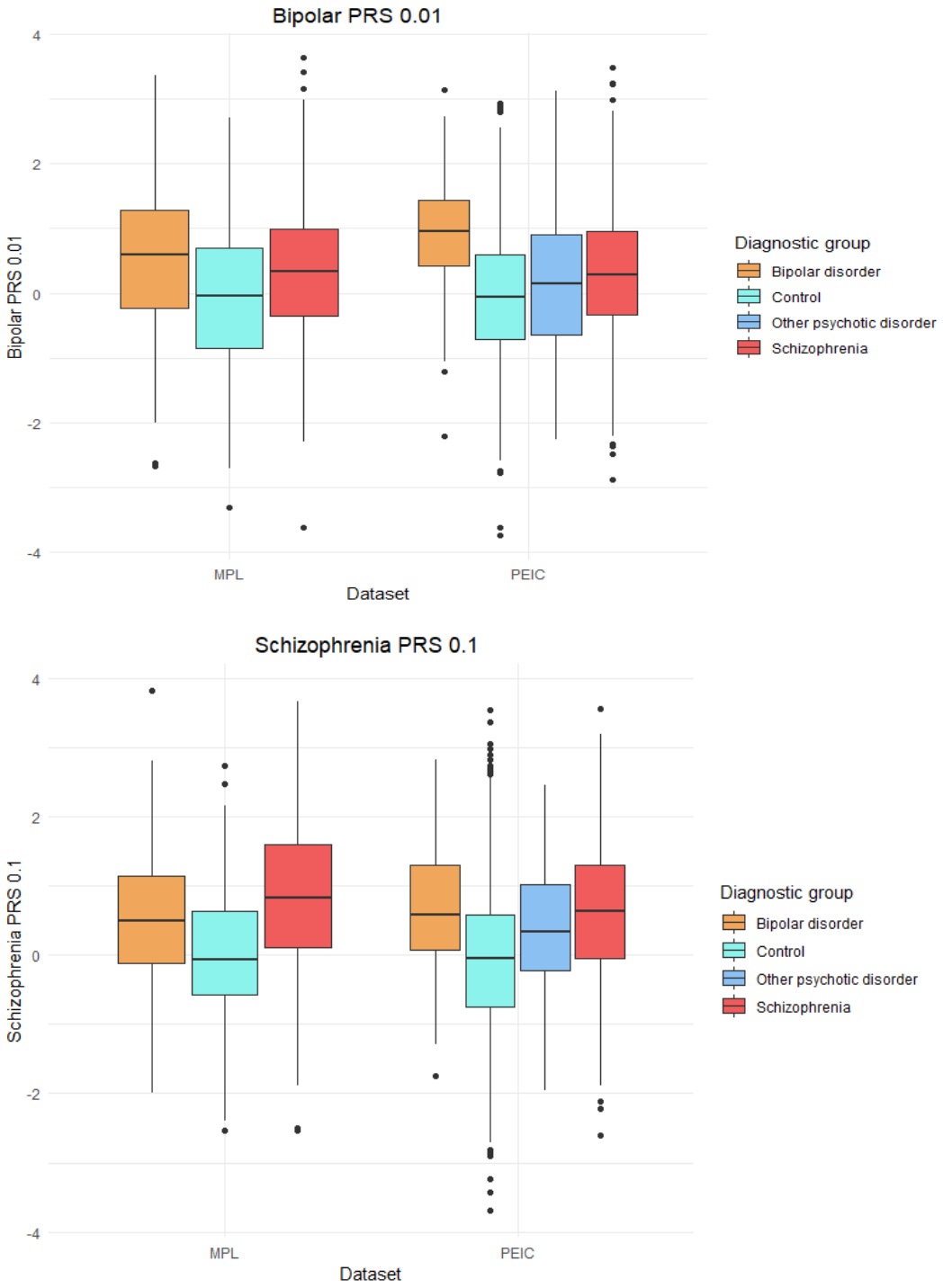
Risk profile scores for bipolar disorder are standardised scores calculated using the p-value threshold of 0.01.

Figure 4.2 Density plots demonstrating the distribution of schizophrenia polygenic risk score (PRS) amongst the diagnostic groups in MPL and PEIC datasets.



Risk profile scores for schizophrenia are standardised scores calculated using the p-value threshold of 0.1.

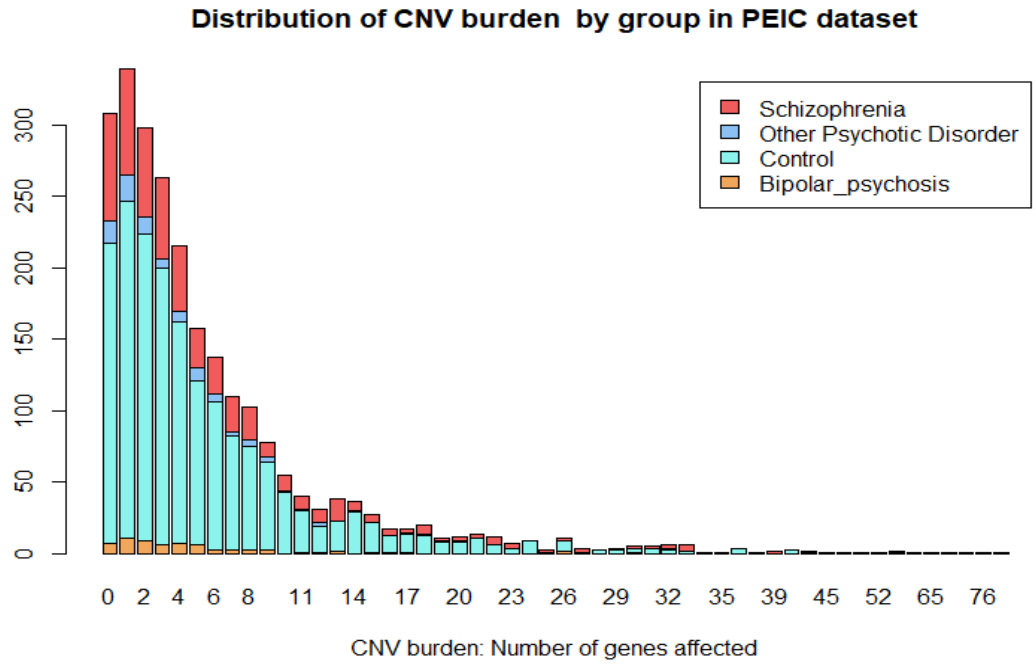
Figure 4.3 Boxplots of bipolar disorder and schizophrenia polygenic risk score (PRSs) between the diagnostic groups in MPL and PEIC datasets.

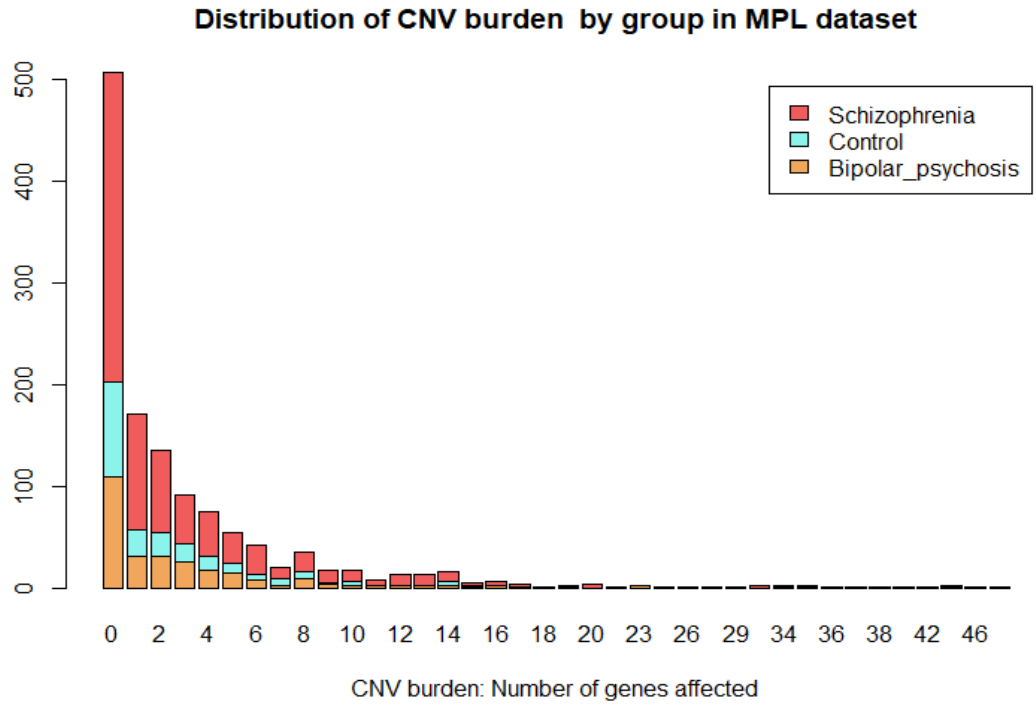


Bipolar disorder risk scores are standardised scores calculated using the p-value threshold 0.01 and schizophrenia risk scores with the p-value threshold of 0.1.

Figure 4.4 demonstrates the distribution of CNV burden (as measured by number of genes affected by deletions) by group in MPL and PEIC datasets. In both datasets the distributions are positively skewed with the majority of individuals not having any genes affected by CNV deletions.

Figure 4.4 Distribution of CNV burden (as measured by number of genes affected by deletions) by group in MPL and PEIC datasets.





Based on their polygenic risk scores, samples in both datasets were allocated to deciles (decile 10% corresponds to the lowest PRS and decile 100% to the highest PRS). Figures 4.5 and 4.6 demonstrate the case control distributions per decile in MPL and PEIC datasets respectively. It is apparent that the higher the PRS score for both schizophrenia and bipolar PRSs the more patients are concentrated on that decile. There is little difference in the middle deciles, as one would expect from a normal distribution.

Similarly, the odds of having broadly defined psychosis increased progressively across PRS deciles. Tables S10-S13 show the case control ratios across PRS deciles and the odds ratios by PRS deciles and for MPL and PEIC datasets.

I performed a meta-analysis of the odds ratios from PEIC and MPL datasets. Compared with individuals in the central deciles (fifth and sixth), those at the tenth and highest decile had an OR for psychosis of 2.29 (95% CI 1.38 – 4.20) for schizophrenia PRS. For the bipolar PRS no difference was found between central and highest deciles (OR = 0.99, 95% CI 0.66 – 1.12) (For more information regarding the meta-analyses of odd ratios refer to table S14).

Figure 4.5 Case and control distribution in the polygenic risk score (PRS) deciles in MPL dataset

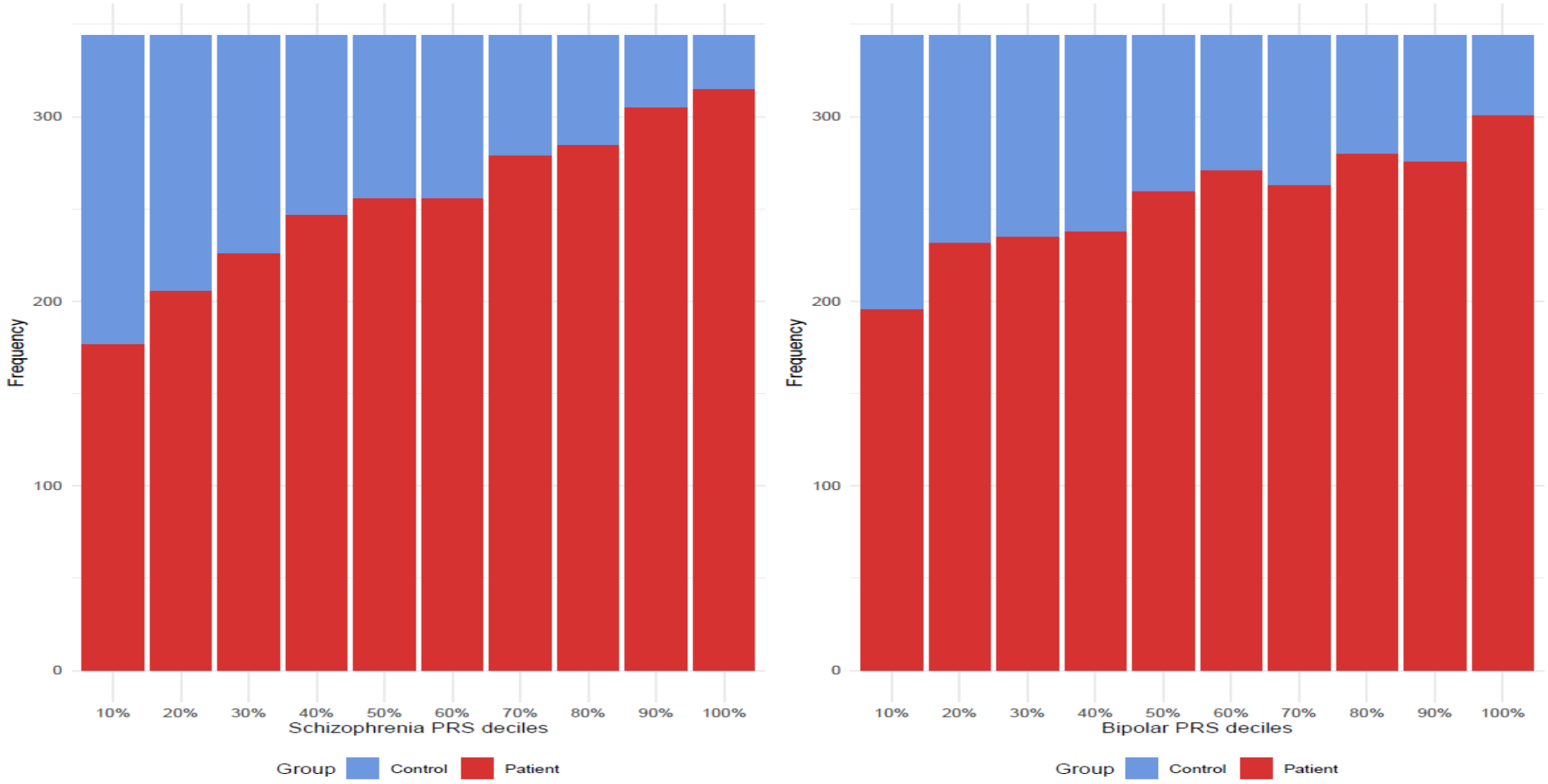
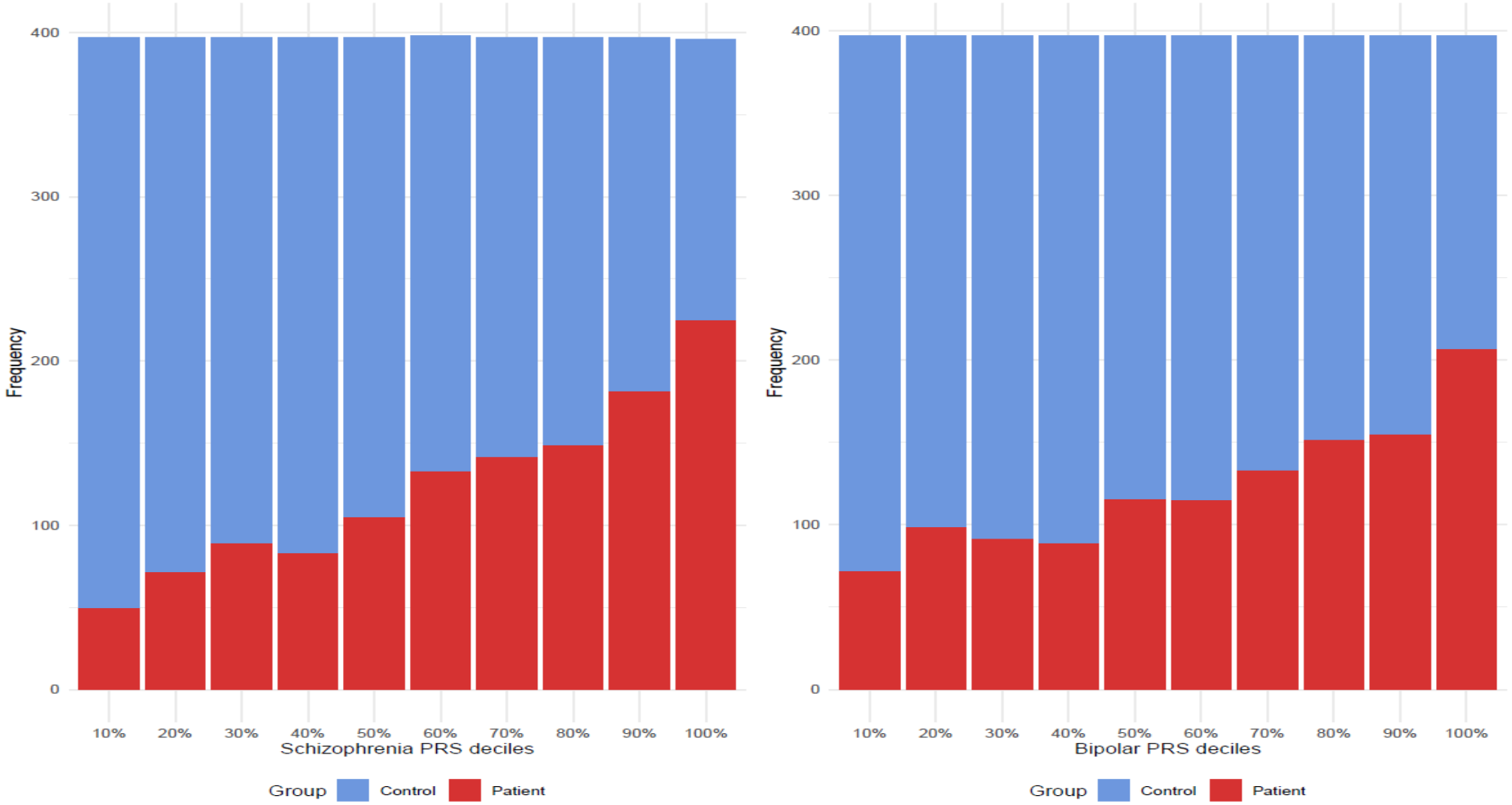


Figure 4.6 Case and control distribution in the polygenic risk score (PRS) deciles in PEIC dataset



Comparisons of cases versus controls in MPL and PEIC datasets separately

As it can be seen in the plots 4.5 and 4.6, there is a significant difference in the case-control ratio between the two datasets across the entire sample and across each decile. The MPL dataset consists of 1,049 cases and the PEIC dataset of 701, whereas the MPL dataset includes only 217 controls and the PEIC dataset includes 1,728. However, as it can be seen in table 4.3 the variance explained by each model and the AUCs are quite similar in both datasets and the unequal ratio has not affected the outcomes.

The variance in disease risk explained by the schizophrenia PRS, the bipolar PRS and the CNV burden was 11.8% (AUC: .81) for MPL and 10.9% (AUC: .81) for PEIC for the comparison of all psychosis cases versus controls, 13.1% (AUC: .73) for MPL and 14.3% (AUC: .82) for PEIC when I compared bipolar patients versus controls, and 13.6% (AUC: .84) for MPL and 10.3% (AUC: .83) for PEIC for the comparison of schizophrenia cases versus controls. The predictive accuracy for all models varied from fair to good. The ROC curves for each dataset are demonstrated on figures 4.7 and 4.8.

Table 4.3 Results of the mixed-effects logistic regression models for the three group comparisons for MPL and PEIC datasets.

MPL dataset	**PEIC dataset**
All psychosis patients versus controls	
N: 1,049 cases; 217 controls	***N: 701 cases; 1,728 controls***
R^2: 11.8% **AUC:** .81	**R^2:** 10.9% **AUC:** .81
Bipolar patients with psychosis versus controls	
N: 278 cases; 217 controls	***N: 70 cases; 1,728 controls***
R^2: 13.1% **AUC:** .73	**R^2:** 14.3% **AUC:** .82
Schizophrenia patients versus controls	
N: 771 cases; 217 controls	***N: 531 cases; 1,728 controls***
R^2: 13.6% **AUC:** .84	**R^2:** 10.3% **AUC:** .83
R^2 shows the variance explained by the full model compared to a model with covariates only AUC shows the model's predictive accuracy: .90 – 1 = *excellent*; .80 – 90 = *good*; .70 - .80 = *fair*; .60 – 70 = *poor*; .50 – 60 = *fail*	

Figure 4.7 ROC curves for the three group comparisons in MPL dataset

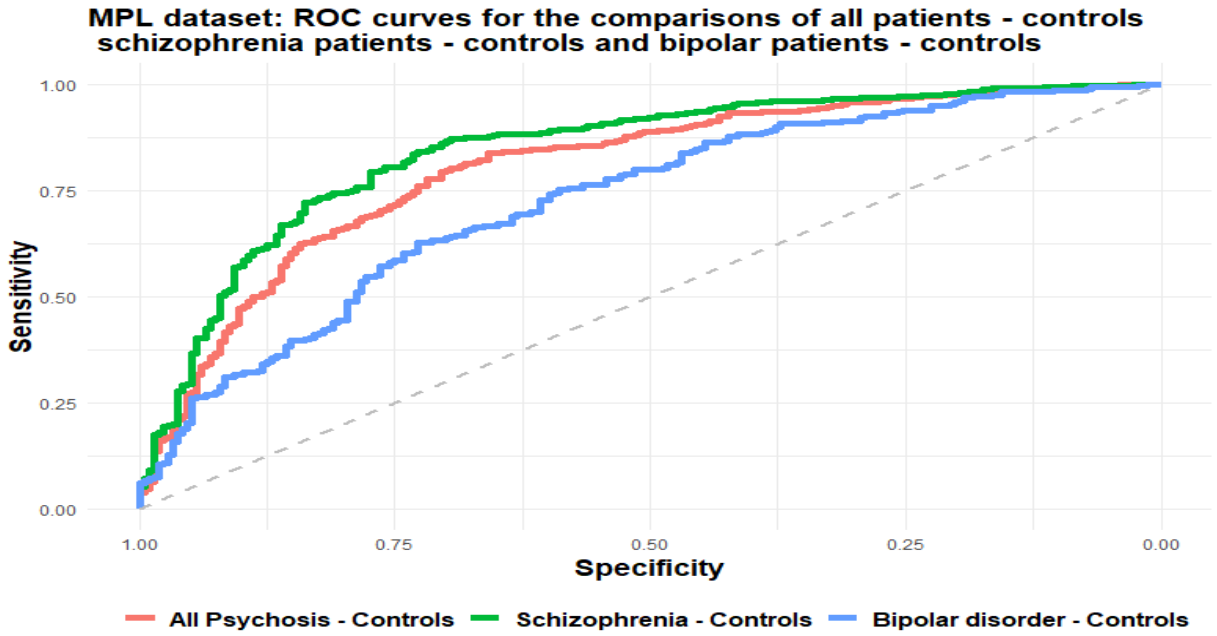
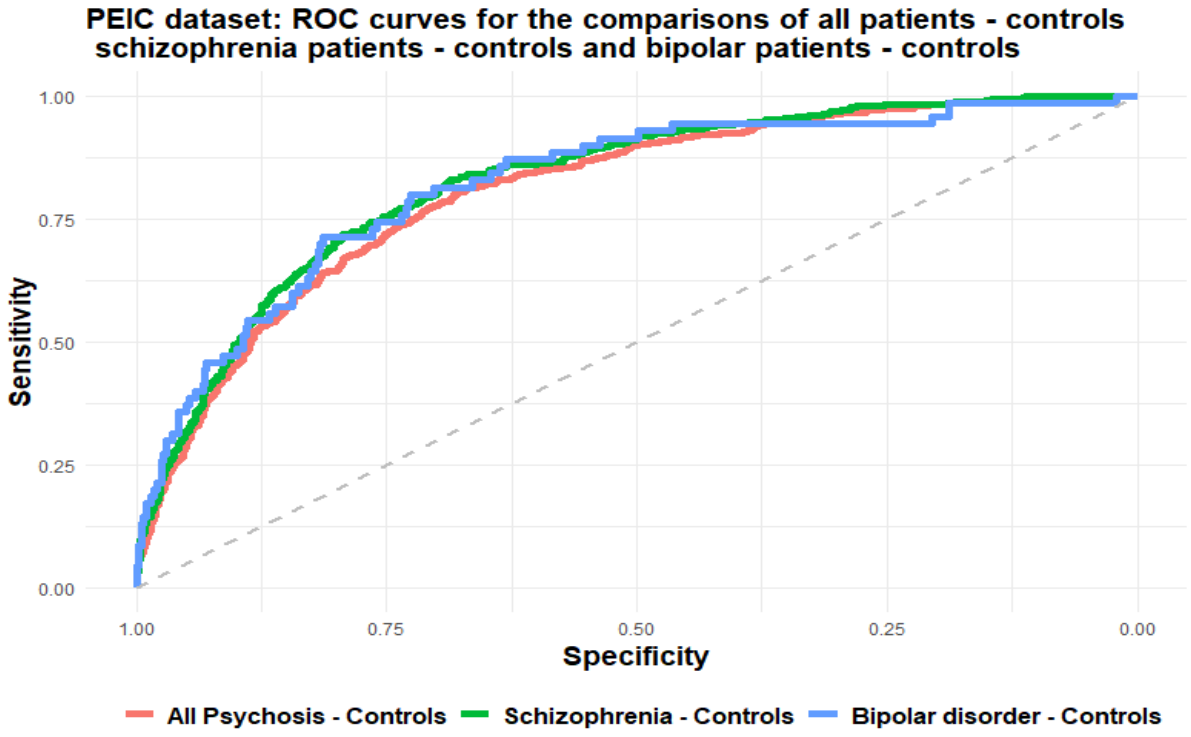


Figure 4.8 ROC curves for the three group comparisons in PEIC dataset



When I ran the regression models with and without the CNV burden, the variance explained was slightly improved in the full model, which included the CNV burden and the statistical difference between the models with and without the CNV burden was significant for the comparison of schizophrenia patients versus controls in both datasets and for all psychosis cases versus controls in MPL dataset (For more information, refer to table S15).

Investigate whether carrying schizophrenia associated CNVs increase psychosis risk

When I explored whether carrying a specific schizophrenia associated CNV (for a list of these CNVs refer to tables S17-S19, and for the distribution of schizophrenia associated CNVs in the diagnostic groups per dataset refer to table S20) had an effect on the risk of developing psychosis, I did not find a significant difference between CNV carriers versus non carriers either in MPL (p = .58) or in PEIC (p = .45) datasets.

Meta-analysis of MPL and PEIC datasets

Our combined sample consisted of 1,750 patients diagnosed with psychosis (1,303 cases with schizophrenia, schizoaffective or schizophreniform disorder, 348 cases with bipolar disorder and 100 with other psychotic disorder) and 1,945 unrelated healthy controls without personal or family history of psychosis (For more information, refer to table S11).

I assessed the heterogeneity between the two datasets which was found to be high only for the comparison of bipolar cases versus controls (Q=7.58, p =.005, I^2=86.82). Therefore, the random effects model is reported for this comparison. For the rest of the comparisons, I report the fixed effects models. In the meta-analyses the classification accuracy of my models measured as the AUCs were 81%, 77% and 83% for the comparisons of all cases vs controls, bipolar cases vs controls

and schizophrenia cases vs controls respectively. The full results from the meta-analyses are presented in table 4.4.

Table 4.4 Results from the meta-analysis of MPL and PEIC datasets for all group comparisons

	Q	AUC	SE	Lower CI	Upper CI	*p* value
All psychosis cases vs controls	.01 (p=.933)	.81	.01	.79	.82	<.001
Bipolar cases vs controls	7.58 (p=.005)	.77	.03	.71	.84	<.001
Schizophrenia cases vs controls	1.03 (p=.308)	.83	.01	.81	.85	<.001

Fixed effects models are reported for the comparisons of all cases vs controls and schizophrenia cases versus controls. Random effects model is reported for the comparison of bipolar cases vs controls.

4.5. Discussion

The aim of the present study was to explore the joint contributions of schizophrenia and bipolar PRSs and CNV burden on psychosis risk. I used two independent, large samples and then combined them in a meta-analysis to increase the statistical power.

I have provided evidence that compared to healthy controls, patients with psychosis have significantly higher PRS for both schizophrenia and bipolar disorder. This is in agreement with several previous studies (Bergen et al., 2019; Calafato et al., 2018; Derks et al., 2012; Tesli et al., 2014; Trotta et al., 2016;

Vassos, Forti, et al., 2017). In the clinical subgroup comparisons of bipolar and schizophrenia patients, both PRSs were able to distinguish the two diagnostic groups in MPL dataset, but only the bipolar PRS could do so in the PEIC dataset. No significant differences on CNV burden were found for any of the comparisons.

To decide which approach to use to measure CNV burden I calculated it in several ways including number of genes affected, length of CNVs, CNV intolerance and number of genes affected with low CNV intolerance. I explored the effects of these measures both for CNV deletions and duplications and the only CNV burden as measured by number of genes affected had significant effect on disease risk and therefore was included in the analyses. The effect was utterly the result of deletions, while duplications had no effect. Other studies have also reported that the number of genes affected, especially by deletions, has been the strongest burden metric in their samples (Marshall et al, 2017; Thygesen et al, 2020) and is considered to be the recommended approach.

When I investigated whether carrying a specific schizophrenia associated CNV had an effect on the risk of developing psychosis, I did not find a significant difference between CNV carriers versus non carriers. However, it should be highlighted that in our study, even after combining the MPL and PEIC datasets I only had 76 carriers of schizophrenia related CNVs. Due to the rarity of these CNVs, very large samples are required to have the statistical power to investigate their effect on disease risk (Sriretnakumar et al., 2019). Another limitation, which applies to research on CNVs in general, is that existing genotyping platforms cannot detect most of the structural variations in DNA (Sudmant et al., 2015). Thus, technological advances using DNA sequencing are essential to be able to detect a wider array of rare genetic variations and small CNVs (Brandler et al., 2016).

There is converging evidence of increased CNV and PRS burden in patients with psychosis (Bergen et al., 2019; Calafato et al., 2018; Tesli et al., 2014; Trotta et al., 2016; Vassos, Forti, et al., 2017, Marshall et al., 2017). Conflicting evidence on the interactive effect of CNV burden and schizophrenia PRS exist in research. Of the two studies conducted so far, one reported lower schizophrenia PRS in carriers of schizophrenia associated CNVs compared to non-carriers (Bergen et al., 2019),

whereas the second study did not find a significant difference in PRS scores, between cases with and without a schizophrenia associated CNV (Taniguchi et al., 2019). It should be noted that in the study by Taniguchi and associates, the authors highlight that their sample size with 1,902 participants is much smaller than in the study by Bergen (2019) with 41,321 partixipants and they might not have sufficient statistical power to detect the interaction between CNVs and PRS. When I examined the interactive effects of PRSs with CNV burden in my samples, I did not find evidence of an interaction either.

To my knowledge, this is the first study exploring the joint contributions of CNV burden and schizophrenia and bipolar PRS burdens to the risk of developing broadly defined psychosis, including patients with schizophrenia spectrum disorder, bipolar disorder with psychotic features and other psychoses. I found that the variance explained ranged from 10.9% up to 11.8% depending on the outcome disorder. The AUCs ranged from fair to good in all comparisons. In the meta-analyses of classification accuracy of my models, the areas under the ROC curve were 81%, 77% and 83% for the comparisons of all psychosis cases versus controls, bipolar disorder cases versus controls and schizophrenia cases vs controls respectively. A recent longitudinal study comparing high risk individuals who developed psychosis over a two year period with healthy subjects, reported that schizophrenia PRS could explain 12% of variance with a predictive accuracy of 70% in a European sample and 3% of variance with 62% predictive accuracy in a non European sample (Perkins et al., 2020). Another study by So and Sham (2017) investigating the predictive power of polygenic risk scores for several clinical phenotypes reported a predictive accuracy of 82% for schizophrenia and 68% for bipolar disorder. One limitation that should be addressed is that the uneven proportion of patients and controls in my sample, could have worsened the ROC performance.

Evidence for a considerable genetic overlap between schizophrenia and bipolar disorder with psychotic features has been provided by several GWAS (Ruderfer et al., 2018; Sklar, Ripke, Scott, Andreassen, Cichon, et al., 2011; Smoller, Kendler, Craddock, Lee, Neale, Nurnberger, Ripke, Santangelo, et al., 2013, Ruderfer et

al., 2014). My study adds evidence for the shared genetic architecture of psychotic disorders supporting a continuum model for their aetiology.

However, one limitation is that for the bipolar groups in my study, I solely know they exhibited psychotic symptoms, but it is not clear whether they had manic or depressive psychosis. Markota et al. (2018) provided evidence that schizophrenia PRS was significantly higher in bipolar cases with manic psychosis (p=.007) than in cases with depressive psychosis (p=.045), concluding that the manic psychosis subgroup is the one more genetically similar to schizophrenia. A further limitation is that in this study DNA was obtained both from saliva and blood samples, which could have potentially impacted the quality of DNA samples.

One of the well-established limitations of polygenic risk scores for schizophrenia and bipolar disorder is that they are obtained from European populations only, and when applied to other or admixed populations, they do not perform well. For this study, all the participants were of European ancestry, so accuracy was not affected. However, I acknowledge that including diverse populations in GWAS is crucial so as to reduce biases and address health and research inequalities (De La Vega & Bustamante, 2018). Other studies have also reported that the PRS for schizophrenia varied significantly between different ancestral groups and concluded that PRS contains a strong ancestry component, thus any associations of polygenic contribution as measured by PRS should be interpreted with caution (Curtis, 2018). Nonetheless, several studies have reported PRS to be highly predictive of both schizophrenia and bipolar disorder genetic risk and being able to distinguish between schizophrenia patients, bipolar disorder patients and healthy controls (Ohi et al., 2020; Calafato et al., 2018; Markota et al., 2018; Ranlund et al., 2018; Ripke et al., 2014).

Despite the promising results of PRS, it has been reported to explain only up to 7% of the variability in schizophrenia risk (Ripke et al., 2014) which has been repeatedly described as insufficient to be implemented in a clinical context (Dudbridge, 2013; Vassos et al., 2017, Wray et al., 2013). Nonetheless, the ability of solely one variable explaining 7% of the variation in disease risk should not be undervalued. It constitutes a valuable research tool.

Several environmental and demographic factors have also been consistently associated with increased risk of developing psychosis. The Maudsley Environmental Risk Score for Psychosis (ERS; Vassos et al., 2019) is the first scale to incorporate environmental and demographic risk factors for psychotic disorders, and authors report it could explain approximately 7% of the variability in disease risk. Another recently developed scale, the Psychosis Polyrisk Score (PPS; Oliver, Radua, Reichenberg, Uher, & Fusar-Poli, 2019), also investigated environmental risk factors for psychosis and identified eight factors to be highly correlated with psychosis risk. Another tool combined demographic information regarding age, sex and ethnicity to calculate an estimate of individualized risk for developing psychosis in 33,820 individuals (Paolo Fusar-Poli et al., 2017). The tool was externally validated in a sample of 54,716 individuals and performed quite well on identifying cases at high risk of developing psychosis over a 6-year period, showing promising results.

The development of such a tool for psychosis integrating demographic, genetic and environmental risk components could substantially increase the explained variability in psychosis risk and optimize the accuracy detecting individuals at high risk (Gillett, Vassos, & Lewis, 2019). Guloksuz et al. (2019) investigated the joint interactions of schizophrenia PRS and several well-established environmental risk exposures on psychosis liability. They reported positive additive interactions of PRS with cannabis use and early life adversities including emotional abuse, sexual abuse, neglect and bullying.

Individualized risk scores based on demographic and/or environmental factors are already being extensively used in clinical practice in other fields of medicine to predict high risk for developing cardiovascular disorders (QRISK; Hippisley-Cox et al., 2007), diabetes (AUSDRISK; Chen et al., 2010) and stroke (CHA2DS2-VASc score; Tanaka et al., 2015). At present, polygenic as well as environmental and combined risk scores for psychosis are not accurate enough for clinical practice but are extremely useful for research. The development of such a tool for psychosis integrating demographic, genetic and environmental risk components could substantially increase the explained variability in psychosis risk and optimize the

predictive accuracy of detecting individuals at clinical high risk (Gillett, Vassos, & Lewis, 2019).

Concluding remarks

Furthering our understanding of the genetic mechanisms increasing the risk of developing psychosis could help to develop new biologically informed treatments. Future research on well-established biomarkers and endophenotypes of psychosis, as well as CNV and PRS burdens, and their combination with environmental risk components could lead to the development of a screening tool accurately measuring psychosis risk. Such a tool could be employed by clinicians and could potentially lead to earlier detection and treatment of psychotic disorders, which ultimately leads to better outcome and prognosis.

Chapter 5. General discussion

In this thesis I have performed a scoping review on an assortment of clinical phenotypes, including schizophrenia and bipolar disorder and associated CNVs, have investigated the relationship between CNV size and effect size for the schizophrenia phenotype and have performed the first meta-analysis of the literature of 16p11.2 distal deletion on schizophrenia. Besides, I have investigated the interrelationships between cognitive, electrophysiological and brain structure endophenotypes of psychotic disorders. Finally, I have also examined the joint contributions of CNV and PRS burden on the risk of developing broadly and narrowly defined psychosis in two large datasets.

In the present chapter, I discuss the main findings of my thesis, their clinical implications, as well the strengths and limitations of my research.

5.1. Main findings and future directions

In the first chapter I conducted the first scoping review investigating the relationships between several clinical phenotypes including schizophrenia, bipolar disorder, intellectual disability, ASD, ADHD and cognitive functioning, and CNVs. With the data identified by the review, I populated CNV catalog, a repository created by myself and my supervisors to investigate associations of CNVs with clinical phenotypes. I performed a meta-analysis of 5 studies investigating the effect of 16p11.2 distal deletions on schizophrenia risk. I found that carriers have a higher risk of developing schizophrenia compared to non carriers with an odds ratio of 2.41.

Additionally, with the current data from CNVcatalog, I investigated the relationship between variant size and effect size for schizophrenia associated CNV loci and reported a positive significant association for total CNVs and CNV deletions but not for duplications. Thus, larger CNVs resulted in greater risk of developing schizophrenia, confirming our hypothesis. This could be due to larger CNVs being

more likely to affect more genes and result in protein changes/loss of function with important biological impact. The number of genes affected, especially by deletions has been reported to be the CNV burden measure with the highest impact on psychosis risk (Marshall, 2016).

In the second chapter I investigated the interrelationships between several established endophenotypes of psychosis. I did not find an association between the amplitude and latency of the P300 ERP, supporting the notion that amplitude is an index of attention and working memory, whereas the latency measures processing speed (Ford, 2014; Näätänen, 1990). However, this contradicts previous studies (Hall et al., 2006, 2014; Polich et al., 1992, 1997), that have reported significant negative associations between the two measures. Therefore, additional research is required to clarify their relationship.

P300 amplitude was positively associated with digit span and block design. The P300 amplitude is an attention-driven, context-updating mechanism, which subsequently feeds into memory stores and is thus expected to be associated with cognitive tasks that require attention and working memory, such as the digit span and block design tasks. All the cognitive endophenotypes (digit span, block design, and the Rey Auditory Verbal Learning Test) were associated with each other.

Furthermore, I examined whether the relationships between endophenotype pairs were consistent in the three participant groups (patients with psychosis, their unaffected relatives and healthy controls). The relationships were consistent for all endophenotype pairs, differing for some of the cognitive pairings only in the strengths of the relationships, with the endophenotype correlation being stronger in the controls.

I have also found that not only patients with psychosis, but also to a lesser extent their unaffected relatives, exhibit reduced amplitudes and prolonged latencies of the P300 when compared to healthy controls. This is in agreement with previous research on that field (Bestelmeyer et al., 2009; Díez et al., 2013; Price et al., 2006; Schulze et al., 2008; Thaker, 2008) and endorses the theory of a continuum of psychosis across the population (Allardyce, Suppes, & van Os, 2007; DeRosse &

Karlsgodt, 2015; Esterberg & Compton, 2009; Ian, Jenner, & Cannon, 2010). In addition, I have replicated previous findings supporting that several cognitive measures (digit span and block design) are true endophenotypes of psychosis. We, however, found no significant differences in lateral ventricular volume or performance in verbal memory when comparing unaffected relatives of the patients with healthy subjects, indicating they might actually be biomarkers of disease progression, rather than endophenotypes for psychotic disorders (Lenzenweger, 2013).

Future studies should focus on genetic analyses of psychosis endophenotypes. Using polygenic risk score methods, they can investigate how much of the common genetic risk is shared between psychotic disorders and the endophenotypes of interest (Hart et al., 2014). Apart from examining only the genetic associations between these endophenotypes and psychotic disorders, future research could also focus on investigating whether SNPs and CNVs that have been associated with psychosis overlap with SNPs and CNVs associated with, for example, neurocognitive performance (Greenwood, Shutes-David, & Tsuang, 2019).

Moreover, in the third chapter of my thesis I have provided evidence that the polygenic risk scores (PRS) for both schizophrenia and bipolar disorder are able to accurately distinguish between cases with broadly defined psychosis and healthy individuals in an independent sample. With the classification accuracies being 81%, 83% and 77% for the comparisons of psychosis vs controls, schizophrenia vs controls and bipolar disorder vs controls respectively, I believe that investigating the effects of PRS in psychosis is a promising area of research. Apart from distinguishing between cases and controls, the PRS could further inform research on biomarkers and endophenotypes of psychosis. It is anticipated that the prevalence of abnormalities in an endophenotype for psychosis would be higher amongst individuals with a higher polygenic score (Fullerton & Nurnberger, 2019). PRS could also be used to investigate whether phenotypes associated with psychotic disorders have a genetic basis (Hong Lee et al., 2013). If there is a genetic overlap between these phenotypes and psychosis, then these phenotypes

would also correlate with the PRS. Conversely, phenotypes solely due to psychosocial factors or environmental exposures would not be correlated to PRS.

Like, the disease itself, most endophenotypes for psychosis are thought to be both under genetic as well as environmental influence. Future research could also aim at integrating genetic risk scores with environmental risk factors that have been linked to psychosis. The recently developed Environmental Risk Score for Psychosis (ERS; Vassos et al., 2019) incorporates six key environmental risk factors with consistent evidence that they increase schizophrenia risk including belonging to an ethnic minority, urbanicity, paternal age, obstetric complications, cannabis use and childhood adversities. The authors report the ERS could explain 7% of the variability in disease risk. Therefore, developing predictive models comprising both genetic and environmental influences could increase the explained variability in disease risk and result in more accurate predictions (Gillett et al., 2019).

Although the PRS and the ERS are not accurate enough for any clinical use, as the training datasets to generate them become larger, there is evidence that the classification accuracy tends to improve (Wray, 2014; Calafato, 2018). By integrating environmental risk factors, as well as gene by gene, and gene by environment interactions accuracy may improve further and there is hope that the PRS could potentially be used in clinical settings for risk reduction counsel (Calafato, 2018). Such a screening tool could eventually lead to earlier intervention for psychosis, reducing the existing delays in access to treatment. The use of polygenic risk scores could in the long run help to develop further biologically informed treatments. A recent study investigated the effect of schizophrenia PRS on response to antipsychotics in patients with first episode psychosis, and have reported that individuals with low PRS responded better to treatment (Zhang et al., 2019).

Regarding the contribution of CNV burden on psychotic disorders, it has been repeatedly reported that CNV burden is increased in schizophrenia and bipolar disorder cases when compared to healthy controls (Malhotra et al., 2011; Marshall et al., 2017; Stone, O'Donovan, Gurling, Kirov, Blackwood, Corvin, Craddock,

Ardlie, et al., 2008; Walsh et al., 2008; Xu, Roos, Levy, Van Rensburg, et al., 2008). In my sample, I also found that it was contributing significantly on the prediction of case control status. However, in my study CNV burden could only explain a very small percentage (0.1%) of the variation in disease risk.

When I explored whether carrying a schizophrenia associated CNV, had a significant effect on the risk of developing psychosis, I did not find a significant difference between CNV carriers and non-carriers. However, it should be highlighted that in my sample I only had 76 carriers of schizophrenia related CNVs, and I might have been underpowered to explore their effect. Not surprisingly, most of the schizophrenia CNVs I found in my sample, have also been associated with other neurodevelopmental disorders, supporting the notion of a neurodevelopmental continuum model (Davis et al., 2016; Owen & O'Donovan, 2017).

Regarding the joint contributions of CNV burden and PRS, only two studies so far have explored their combined effect and interaction in schizophrenia risk (Bergen et al., 2019; Taniguchi et al., 2019). Bergen et al (2019) reported that schizophrenia patients who were carrying schizophrenia associated CNVs had lower schizophrenia PRS compared to non-carriers but higher PRSs than healthy controls. The authors concluded that the PRS was diminished in proportion to the effect size of the CNV, therefore the higher the CNV's odds ratio, the lower the PRS needed to be in order to become ill. The second study by Taniguchi et al. (2019) reported no significant difference in PRS scores in cases with and without a schizophrenia associated CNV. To my knowledge, my study was the first one to explore the joint effect of SNPs and CNVs not only on schizophrenia risk but also on the liability for bipolar disorder and for psychosis broadly defined and my findings were consistent with the study by Taniguchi (2019). However, both in Taniguchi's study and ours the sample size is much smaller than in the study by Bergen (2019) and I may lack the statistical power to detect the interactive effects of PRSs and CNVs.

5.2. Strengths and limitations

There are limitations to this thesis, and despite the fact that they have already been discussed in each chapter, some are relevant to the overall thesis and will be highlighted here.

Firstly, the sample I used for the second and third chapters consists of patients with broadly defined psychosis including schizophrenia, schizoaffective disorder, schizophreniform disorder, other psychoses and bipolar disorder with psychotic features. Despite strong evidence for overlapping aetiology, symptomatology and risk factors (Bramon & Sham, 2001; Laursen, Agerbo, & Pedersen, 2009; Lee, Yang, Goddard, Visscher, & Wray, 2012; Pearlson, 2015; Smoller, Kendler, Craddock, Lee, Neale, Nurnberger, Ripke, Santangelo, et al., 2013), one should be cautious grouping them all together under the umbrella term of psychosis.

There are several factors that are quite distinct between different psychotic illnesses and using a broadly defined psychosis group could potentially add noise to the phenotype definition. In my work, I have addressed this issue by carrying out additional analyses with each clinical subgroup analysed separately. Additionally, since there is not an established biological test to classify patients into diagnostic categories, diagnoses are entirely based on clinical observation and self-reported symptoms. Taking into consideration the commonalities of the clinical presentation of psychotic disorders, grouping them together under a broader clinical phenotype could be seen as an advantage. Also, this way we acquire larger samples to study, which consequently results in greater statistical power.

A further limitation is the confounding effects of antipsychotic medication on endophenotype performance. The majority of patients included in the second and third chapters were taking antipsychotic medication, which has been repeatedly reported to affect brain structure and functioning, confounding MRI performance and resulting to a slowing of the EEG signal (Goozée, Handley, Kempton, & Dazzan, 2014; Huhtaniska, Jääskeläinen, et al., 2017; Hyun, Myung, & Ung, 2011; Roiz-Santiañez, Suarez-Pinilla, & Crespo-Facorro, 2015). A study with 84 patients with broadly defined psychosis provided evidence that atypical antipsychotics

(clozapine, aripiprazole, olanzapine and risperidone) caused EEG modifications, which were greater for clozapine (Dias Alves, Micoulaud-Franchi, Simon, & Vion-Dury, 2018). A recent systematic review of 14 papers with 665 patients also suggested that antipsychotics, especially clozapine induced EEG slowing (Jackson & Seneviratne, 2019). Other studies however, have even provided evidence that antipsychotics could help normalise EEG changes associated with psychotic disorders (Su, Cai, Shi, & Wang, 2012; Zhou, Zhu, & Chen, 2013). Furthermore, there is extensive research on the confounding effects of antipsychotic medication in cognitive endophenotype performance (MacKenzie et al., 2018; Harvey, 2006; Woodward et al., 2005). Antipsychotic medication has been reported to improve multiple areas of cognitive functioning including selective attention, delayed recall, verbal fluency and verbal and short-term memory (Goozee et al., 2016; Bervoets et al., 2012; Johnsen et al., 2013). Unfortunately, for approximately half of my samples I did not have more detailed information in regard to the medication the patients were receiving and the dosage in order to account for this effect in my analyses.

Nonetheless, it is quite challenging to determine which EEG, cognitive or brain morphometry abnormalities could be attributed to the actual illness and which could result from the medication. This is why studying unaffected relatives of individuals with psychosis is of high importance. The unaffected relatives are genetically similar to the patients and have an increased risk of developing psychosis, but do not exhibit the associated symptomatology and are not receiving antipsychotic medication. However, when studying unaffected relatives, one should consider that apart from the shared genetics, relatives may also have several environmental, social and psychological common factors with the patients. Shared environment has also be reported to influence brain activity and function, including EEG performance (Rasetti & Weinberger, 2011).

A further probable limitation is that since data were collected across several sites there is high heterogeneity of methods between study sites; differences in cognitive test versions and variation on the EEG and MRI protocols employed, which introduces greater variability into the phenotypic data. To account for that I used

standardised measures. I also adjusted all models for centre and for key methodological variables. Nevertheless, one advantage of multi-centre studies is the collection of large samples, resulting in higher statistical power. Therefore, I believe that the benefits of a larger sample outweight the limitations of having to adjust for variations in the methodology to collect the data across multiple centres. Another strength of my study is the use of EEG data. EEG accurately measures neural function at extremely high time resolution and has the highest temporal resolution of all the imaging techniques. Besides, it is a non-invasive and safe technique, and the equipment required is fairly inexpensive and simple to operate.

In regards to PRS burden, it should be highlighted that since the majority of GWAS studies are conducted on populations of European ancestry, their utility in non-European populations is limited. A study by Curtis (2018) explored the distribution of schizophrenia PRS in different ancenstry groups. He reported that polygenic scores have a strong ancenstry component and concluded that results should be interpreted with caution and principal components should always be employed. In our study, all participants were of European ancestry and I also used three principal components to account for population stratification.

Despite the impact of ancestry on polygenic scores, several studies have reported PRS to be highly predictive of psychosis risk if applied to a similar population (Calafato et al., 2018; Ranlund et al., 2018, Toulopoulou et al., 2019). Despite PRS's promising results, it has been reported to explain only up to 7% of the variability in disease risk (Stephan Ripke, Neale, Corvin, Walter, et al., 2014), which has been deemed insufficient to be implemented in a clinical context (Dudbridge, 2013; Vassos, Di Forti, et al., 2017). Nevertheless, the ability of only one variable being able to explain 7% of the variation in psychosis risk should not be underestimated.

One limitation of research on CNVs is that current genotyping platforms do not have the power to detect most of the structural variations in DNA (Sudmant et al., 2015). Technological advances are required to be able to detect a broader array of rare genetic variations and smaller CNVs (Brandler et al., 2016). It should also be emphasized that my study was limited by sample size due to having only 76

CNV carriers. As a result of the rarity of individual CNVs, large samples are required so as to have the statistical power of detecting significant effects of CNVs on disease risk (Sriretnakumar et al., 2019).

Ethical considerations of how genetic information on PRS and CNV burden could be used have also been raised. The fact that the predictive ability of PRS is "race-restricted" (Regalado, 2018) has been argued to be the main ethical matter of implementing PRS into a clinical setting (Palk et al., 2019). Several studies highlight the underrepresentation of non-European populations, particularly populations of African ancestry, in current psychiatric genetics research (Martin et al., 2018; Dalvie et al., 2015; Campbell & Tishkoff, 2008). This would ultimately result to non-European populations being left behind with regard to genetic research and consequent treatment advances. Current studies cannot be generalised to wider populations. A diverse data collection would result in larger, generalizable samples and advance progress.

Moreover, there is a multitude of concerns that being identified with either having high schizophrenia/bipolar PRS score or carrying a schizophrenia associated CNV could result in an aggravation on health disparities and increased discrimination and stigmatisation of mental health patients (Palk et al., 2019). Communicating the results sometimes involves oversimplification or even exaggeration of the findings in an attempt to capture attention (Caufield & Condit, 2012). Merely disseminating non-actionable genetic information to the public without a clinician explaining the results and elucidating what they could mean can easily lead to their misinterpretation and subsequently result to an exacerbation of the inequitable assumptions regarding mental illness.

Several genetic services, including 23andMe and MyHeritage DNA, provide individuals with an easy and relatively affordable way to get a genetic test and learn about their individual disease susceptibilities for a multitude of conditions. Providing information on clinically actionable genetic mutations is undoubtedly of immense importance. However, some of these services provide polygenic risk scores for several traits and illnesses including breast cancer, type 2 diabetes and heart disease. Polygenic scores have been repeatedly reported as significant

predictors for liability in several physical ilnnesses including breast cancer (Mavaddat et al., 2019; Palmer, 2020) and cardiovascular disease (Dikilitas et al. 2019), but they are not ready yet for clinical use. These genetic services claim that PRS is a comprehensive approach for assessing the genetic risk for complex polygenic conditions, but without explaining that its clinical implementation is precluded due to several limitations, including the poor risk prediction for individuals of non-European ancestry.

5.3. General conclusions

In conclusion, I have investigated endophenotypes across several domains that are associated with psychotic disorders as well as genetic variants increasing the risk of their manifestation. The findings have contributed to research in mental health by demonstrating additional evidence that:

i) 16p11.2 distal deletions significantly increase the risk for schizophrenia.

ii) CNV size is associated with schizophrenia risk for CNV deletions but not for duplications, with larger CNVs conferring greater risk.

iii) Impairments in cognitive functioning are promising psychosis endophenotypes, since they are also found, albeit to a lesser extent, in the unaffected relatives of the patients.

iv) P300 amplitude and latency, as well as deficits in block design and digit span are potential psychosis endophenotypes.

v) CNV burden, as measured by number of genes affected and PRSs for both schizophrenia and bipolar disorder could explain approximately 11% of variance in disease risk.

vi) The predictive accuracy of my models incorporating CNV burden and schizophrenia and bipolar PRSs is not yet high enough for the models to be considered for clinical use.

Apart from providing evidence on the aforementioned points, I have also been involved in the development of CNV catalog, a repository incorporating data from CNV association studies that facilitates a multitude of meta-analytical procedures and graphical illustrations.

The existing classification systems distinguish between psychiatric diagnoses, such as schizophrenia and bipolar disorder, largely on the basis of symptom clusters. Further research on the neurobiological mechanisms and the genetic architecture of psychiatric disorders could help to identify biologically defined subgroups and to provide additional evidence to improve our current classification systems reflecting neurobiological distinctions.

A better understanding of the genetic mechanisms underlying psychosis will help to develop new biologically informed treatments. Future research could focus on how several biomarkers of psychosis, CNV burden and polygenic risk scores could be used to develop a screening tool to measure psychosis risk. This tool, if it were sufficiently accurate, could be used by clinicians and would lead to earlier detection and treatment of psychosis, which lead to better outcome/prognosis. Furthermore, investigating well-established endophenotypes of psychosis is important in psychiatric research by shedding light to the mechanisms by which genetic risk factors increase the risk of developing psychosis.

Findings from this thesis add to a growing body on literature on genetic influences in psychoses and contribute to knowledge that could hopefully in the future improve the lives of people affected by psychotic disorders.

Bibliography

Abel, K. M., Wicks, S., Susser, E. S., Dalman, C., Pedersen, M. G., Mortensen, P. B., & Webb, R. T. (2010). Birth weight, schizophrenia, and adult mental disorder: is risk confined to the smallest babies?. *Archives of general psychiatry*, *67*(9), 923-930.

Affymetrix. (2015). Affymetrix.

Allardyce, J., Suppes, T., & van Os, J. (2007). Dimensions and the psychosis phenotype. *International Journal of Methods in Psychiatric Research*, *16*(1), 34–40.

Almasy, L., & Blangero, J. (2001). Endophenotypes as quantitative risk factors for psychiatric disease: rationale and study design. *American Journal of Medical Genetics*, *105*(1), 42–44.

American Psychiatric Association. (2000). *Diagnostic and Statistical manual for mental disorders (DSM IV TR)*. Washington: American Psychiatric Association

American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders. *BMC Med*, *17*, 133-137.

Andreasen, N. C., Flaum, M., & Arndt, S. (1992). The Comprehensive Assessment of symptoms and History (CASH). An Instrument for Assessing Diagnosis and Psychopathology. *Archives of General Psychiatry*, *49*(8), 615.

Andreassen, O. A., Thompson, W. K., Schork, A. J., Ripke, S., Mattingsdal, M., Kelsoe, J. R., … 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).

Antoniou, A. C., & Chenevix-Trench, G. (2010). Common genetic variants and cancer risk in Mendelian cancer syndromes. *Current Opinion in Genetics & Development*, *20*(3), 299– 307.

Antonova, E., Sharma, T., Morris, R., & Kumari, V. (2004). The relationship between brain structure and neurocognition in schizophrenia: a selective review. *Schizophrenia Research*, *70*(2–3), 117–145.

Arnfred, S. M. (2006). Exploration of auditory P50 gating in schizophrenia by way of difference waves. *Behavioral and Brain Functions*, *2*(1), 6.

Arseneault, L., Cannon, M., Fisher, H. L., Polanczyk, G., Moffitt, T. E., & Caspi, A. (2011). Childhood trauma and children's emerging psychotic symptoms: a genetically sensitive longitudinal cohort study. *American Journal of Psychiatry*, *168*(1), 65-72.

Ayres, A. M., Busatto, G. F., Menezes, P. R., Schaufelberger, M. S., Coutinho, L., Murray, R. M., Scazufca, M. (2007). Cognitive deficits in first-episode psychosis: a population-based study in São Paulo, Brazil. *Schizophrenia Research*, *90*(1–3), 338–343.

Baddeley, A. (1992). Working memory. *Science (New York, N.Y.)*, *255*(5044), 556–559.

Bagshaw, A. T. M., Horwood, L. J., Liu, Y., Fergusson, D. M., Sullivan, P. F., & Kennedy, M. A. (2013). No Effect of Genome-Wide Copy Number Variation on Measures of Intelligence in a New Zealand Birth Cohort. *PLoS ONE*, *8*(1), 1–6.

Bailey, E. (2015). *shinyBS: Twitter Bootstrap Components for Shiny*.

Balan, S., Iwayama, Y., Toyota, T., Toyoshima, M., Maekawa, M., & Yoshikawa, T. (2014). 22q11.2 deletion carriers and schizophrenia-associated novel variants. *British Journal of Psychiatry*, *204*(5), 398–399.

Barta, P. E., Dhingra, L., Royall, R., & Schwartz, E. (1997). Improving sterological estimates for the volume of structures identified in three-dimensional arrays of spatial data. *Journal of Neuroscience Methods*, *75*(2), 111–118.

Bashore, T. R., Wylie, S. A., Ridderinkhof, K. R., & Martinerie, J. M. (2014). Response-specific slowing in older age revealed through differential stimulus and

response effects on P300 latency and reaction time. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, *21*(6), 633–673.

Bassett, A. S., Chow, E. W., & Weksberg, R. (2000). Chromosomal abnormalities and schizophrenia. *American Journal of Medical Genetics*, *97*(1), 45–51.

Bassett, A. S., Scherer, S. W., & Brzustowicz, L. M. (2010). Copy number variations in schizophrenia: critical review and new perspectives on concepts of genetics and disease. *The American Journal of Psychiatry*, *167*(8), 899–914.

Bergen, S. E., O'dushlaine, C. T., Ripke, S., Lee, P. H., Ruderfer, D. M., Akterin, S., ... & Ösby, U. (2012). Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. *Molecular psychiatry*, *17*(9), 880-886.

Bergen, S. E., Ploner, A., Howrigan, D., O'Donovan, M. C., Smoller, J. W., Sullivan, P. F., … Kendler, K. S. (2019). Joint Contributions of Rare Copy Number Variants and Common SNPs to Risk for Schizophrenia. *American Journal of Psychiatry*, *176*(1), 29–35.

Bervoets, C., Morrens, M., Vansteelandt, K., Kok, F., de Patoul, A., Halkin, V., ... & Sabbe, B. (2012). Effect of aripiprazole on verbal memory and fluency in schizophrenic patients. *CNS drugs*, *26*(11), 975-982.

Bestelmeyer, P. E. G., Phillips, L. H., Crombie, C., Benson, P., & Clair, D. S. (2009). The P300 as a possible endophenotype for schizophrenia and bipolar disorder: Evidence from twin and patient studies. *Psychiatry Research*, *169*(3), 212–219.

Birkett, P., Sigmundsson, T., Sharma, T., Toulopoulou, T., Griffiths, T. D., Reveley, A., & Murray, R. (2008). Executive function and genetic predisposition to schizophrenia—the Maudsley family study. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *147B*(3), 285–293.

Blacker, D., Bertram, L., Saunders, A. J., Moscarillo, T. J., Albert, M. S., Wiener, H., Tanzi, R. E. (2003). Results of a high-resolution genome screen of 437 Alzheimer's Disease families. *Human Molecular Genetics*, *12*(1), 23–32.

Blackwood, D. H., St Clair, D. M., Muir, W. J., & Duffy, J. C. (1991). Auditory P300 and eye tracking dysfunction in schizophrenic pedigrees. *Archives of General Psychiatry*, *48*(10), 899–909.

Blakey, R., Ranlund, S., Zartaloudi, E., Cahn, W., Calafato, S., Colizzi, M., Bramon, E. (2018). Associations between psychosis endophenotypes across brain functional, structural, and cognitive domains. *Psychological Medicine*, *48*(08), 1325–1340.

Bochukova, E. G., Huang, N. I., Keogh, J., Henning, E., Purmann, C., Blaszczyk, K., ... & Hurles, M. E. (2010). Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature*, *463*(7281), 666-670.

Bodatsch, M., Brockhaus-Dumke, A., Klosterkötter, J., & Ruhrmann, S. (2015). Forecasting Psychosis by Event-Related Potentials—Systematic Review and Specific Meta-Analysis. *Biological Psychiatry*, *77*(11), 951–958.

Bogren, M., Mattisson, C., Isberg, P.-E., & Nettelbladt, P. (2009). How common are psychotic and bipolar disorders? A 50-year follow-up of the Lundby population. *Nordic Journal of Psychiatry*, *63*(4), 336–346.

Boos, H. B. M., Aleman, A., Cahn, W., Hulshoff Pol, H., & Kahn, R. S. (2007). Brain volumes in relatives of patients with schizophrenia: A meta-analysis. *Archives of General Psychiatry*, *64*(3), 297–304.

Bora, E., Lin, A., Wood, S. J., Yung, A. R., McGorry, P. D., & Pantelis, C. (2014). Cognitive deficits in youth with familial and clinical high risk to psychosis: a systematic review and meta-analysis. *Acta Psychiatrica Scandinavica*, *130*(1), 1–15.

Bora, E., & Murray, R. M. (2014). Meta-analysis of cognitive deficits in ultra-high risk to psychosis and first-episode psychosis: do the cognitive deficits

progress over, or after, the onset of psychosis? *Schizophrenia Bulletin*, *40*(4), 744–755.

Bora, E., & Pantelis, C. (2015). Meta-analysis of Cognitive Impairment in First-episode Bipolar Disorder: Comparison with First-episode Schizophrenia and Healthy Controls. *Schizophrenia Bulletin*, *41*(5), 1095-1104.

Bora, E., Yucel, M., & Pantelis, C. (2009). Cognitive functioning in schizophrenia, schizoaffective disorder and affective psychoses: meta-analytic study. *The British Journal of Psychiatry : The Journal of Mental Science*, *195*(6), 475–482.

Bora, E., Yücel, M., & Pantelis, C. (2010). Cognitive impairment in affective psychoses: a meta-analysis. *Schizophrenia Bulletin*, *36*(1), 112–125.

Bousman, C. A., Yung, A. R., Pantelis, C., Ellis, J. A., Chavez, R. A., Nelson, B., … Foley, D. L. (2013). Effects of NRG1 and DAOA genetic variation on transition to psychosis in individuals at ultra-high risk for psychosis. *Translational Psychiatry*, *3*(4), 251.

Boydell, J., Onwumere, J., Dutta, R., Bhavsar, V., Hill, N., Morgan, C., … Fearon, P. (2014). Caregiving in first-episode psychosis: social characteristics associated with perceived 'burden' and associations with compulsory treatment. *Early Intervention in Psychiatry*, *8*(2), 122–129.

Braff, D. L. (2015). The importance of endophenotypes in schizophrenia research. *Schizophrenia Research*, *163*(1–3), 1–8.

Braff, D. L., Freedman, R., Schork, N. J., & Gottesman, I. I. (2006). Deconstructing Schizophrenia: An Overview of the Use of Endophenotypes in Order to Understand a Complex Disorder. *Schizophrenia Bulletin*, *33*(1), 21–32.

Bramon, E., & Sham, P. C. (2001). The common genetic liability between schizophrenia and bipolar disorder: a review. *Current Psychiatry Reports*, *3*(4), 332–337.

Bramon, E., Rabe-Hesketh, S., Sham, P., Murray, R. M., & Frangou, S. (2004). Meta-analysis of the P300 and P50 waveforms in schizophrenia. *Schizophrenia research*, *70*(2-3), 315-329.

Bramon, E., McDonald, C., Croft, R. J., Landau, S., Filbey, F., Gruzelier, J. H., … Murray, R. M. (2005). Is the P300 wave an endophenotype for schizophrenia? A meta-analysis and a family study. *NeuroImage*, *27*(4), 960–968.

Brandler, W. M., Antaki, D., Gujral, M., Noor, A., Rosanio, G., Chapman, T. R., … Sebat, J. (2016). Frequency and Complexity of de Novo Structural Mutation in Autism. *American Journal of Human Genetics*, *98*(4), 667–679.

Breteler, M. M., van Amerongen, N. M., van Swieten, J. C., Claus, J. J., Grobbee, D. E., van Gijn, J., … van Harskamp, F. (1994). Cognitive correlates of ventricular enlargement and cerebral white matter lesions on magnetic resonance imaging. The Rotterdam Study. *Stroke; a Journal of Cerebral Circulation*, *25*(6), 1109–1115.

Brewer, M. B., Campbell, D. T., & Crano, W. D. (1970). Testing a Single-Factor Model as an Alternative to the Misuse of Partial Correlations in Hypothesis-Testing Research. *Sociometry*, *33*(1), 1.

Brillinger, D. R. (2001). Does Anyone Know When the Correlation Coefficient Is Useful? A Study of the Times of Extreme River Flows. *Technometrics*, *43*(January 2015), 266–273.

Brown, J. A. (2011). Talking about Life after Early Psychosis: The Impact on Occupational Performance. *Canadian Journal of Occupational Therapy*, *78*(3), 156–163.

Burbach, J. P. H., & van der Zwaag, B. (2009). Contact in the genetics of autism and schizophrenia. *Trends in Neurosciences*, *32*(2), 69–72.

Burnside, R. D., Pasion, R., Mikhail, F. M., Carroll, A. J., Robin, N. H., Youngs, E. L., ... & Potluri, V. R. (2011). Microdeletion/microduplication of proximal

15q11. 2 between BP1 and BP2: a susceptibility region for neurological dysfunction including developmental and language delay. *Human genetics*, *130*(4), 517-528.

Cahn, W., Rais, M., Stigter, F. P., van Haren, N. E. M., Caspers, E., Hulshoff Pol, H. E., … Kahn, R. S. (2009). Psychosis and brain volume changes during the first five years of schizophrenia. *European Neuropsychopharmacology : The Journal of the European College of Neuropsychopharmacology*, *19*(2), 147–151.

Calafato, M. S., Thgesen, J. H., Ranlund, S., Zartaloudi, E., Cahn, W., Crespo-Facorro, B., … Bramon, E. (2018). Use of schizophrenia and bipolar disorder polygenic risk scores to identify psychotic disorders. *British Journal of Psychiatry*, *213*(3), 535–541.

Campbell, M. C., & Tishkoff, S. A. (2008). African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu. Rev. Genomics Hum. Genet.*, *9*, 403-433.

Cannon, T. D., & Keller, M. C. (2006). Endophenotypes in the genetic analyses of mental disorders. *Annual Review of Clinical Psychology*, *2*, 267–290.

Cantor-Graae, E., & Selten, J. P. (2005). Schizophrenia and migration: a meta-analysis and review. *American Journal of Psychiatry*, *162*(1), 12-24.

Cardno, A. G. (2014). Genetics and psychosis. *Advances in Psychiatric Treatment*, *20*(1), 69– 70.

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.

Cardno, A. G., Marshall, E. J., Coid, B., Macdonald, A. M., Ribchester, T. R., Davies, N. J., … Murray, R. M. (1999). Heritability estimates for psychotic disorders: The Maudsley Twin psychosis series. *Archives of General Psychiatry*, *56*(2), 162–168.

Cardno, A. G., Marshall, E. J., Coid, B., Macdonald, A. M., Ribchester, T. R., Davies, N. J., … Murray, R. M. (1999). Heritability Estimates for Psychotic Disorders. *Archives of General Psychiatry*, *56*(2), 162.

Cariaga-Martinez, A., Saiz-Ruiz, J., & Alelú-Paz, R. (2016). From linkage studies to epigenetics: what we know and what we need to know in the neurobiology of schizophrenia. *Frontiers in neuroscience*, *10*, 202.

Caseras, X., Tansey, K. E., Foley, S., & Linden, D. (2015). Association between genetic risk scoring for schizophrenia and bipolar disorder with regional subcortical volumes. *Translational Psychiatry*, *5*(12), e692-e692.

Caulfield, T., & Condit, C. (2012). Science and the sources of hype. *Public Health Genomics*, *15*(3-4), 209-217.

Chang, H., Li, L., Li, M., & Xiao, X. (2017). Rare and common variants at 16p11. 2 are associated with schizophrenia. *Schizophrenia research*, *184*, 105-108.

Charney, A. W., Stahl, E. A., Green, E. K., Chen, C.-Y., Moran, J. L., Chambert, K., … Sklar, P. (2019). Contribution of Rare Copy Number Variants to Bipolar Disorder Risk Is Limited to Schizoaffective Cases. *Biological Psychiatry*, *86*(2), 110–119.

Chen, J., Calhoun, V. D., Perrone-Bizzozero, N. I., Pearlson, G. D., Sui, J., Du, Y., & Liu, J. (2016). A pilot study on commonality and specificity of copy number variants in schizophrenia and bipolar disorder. *Translational Psychiatry*, *6*(5), 824.

Chen, K. C., Lee, I. H., Yang, Y. K., Landau, S., Chang, W. H., Chen, P. S., … Bramon, E. (2013). P300 waveform and dopamine transporter availability: a controlled EEG and SPECT study in medication-naive patients with schizophrenia and a meta-analysis. *Psychological Medicine*, *44*(10), 2151-2162.

Chen, L., Magliano, D. J., Balkau, B., Colagiuri, S., Zimmet, P. Z., Tonkin, A. M., … Shaw, J. E. (2010). AUSDRISK: An Australian Type 2 Diabetes Risk

Assessment Tool based on demographic, lifestyle and simple anthropometric measures. *Medical Journal of Australia*, *192*(4), 197–202.

Chen, X., Dunham, C., Kendler, S., Wang, X., O'Neill, F. A., Walsh, D., & Kendler, K. S. (2004). Regulator of G-protein signaling 4 (RGS4) gene is associated with schizophrenia in Irish high density families. *American Journal of Medical Genetics*, *129*(1), 23–26.

Cheng, C.-H., Chan, P.-Y. S., Liu, C.-Y., & Hsu, S.-C. (2016). Auditory sensory gating in patients with bipolar disorders: A meta-analysis. *Journal of Affective Disorders*, *203*, 199– 203.

Cho, Y., Ryu, S., Huh, I., Cho, E. Y., Oh, H., Lee, Y. S., ... & Hong, K. S. (2015). Effects of genetic variations in NRG1 on cognitive domains in patients with schizophrenia and healthy individuals. *Psychiatric genetics*, *25*(4), 147-154.

Chow, E. W. C., Ho, A., Wei, C., Voormolen, E. H. J., Crawley, A. P., & Bassett, A. S. (2011). Association of schizophrenia in 22q11.2 deletion syndrome and gray matter volumetric deficits in the superior temporal gyrus. *The American Journal of Psychiatry*, *168*(5), 522– 529.

Chu, C. S., Chow, P. C. K., Cohen-Woods, S., Gaysina, D., Tang, K. Y., & McGuffin, P. (2017). The DAOA gene is associated with schizophrenia in the Taiwanese population. *Psychiatry research*, *252*, 201-207.

Clifton, N. E., Pocklington, A. J., Scholz, B., Rees, E., Walters, J. T. R., Kirov, G., & Hall, J. (2017). Schizophrenia copy number variants and associative learning. *Molecular psychiatry*, *22*(2), 178-182.

Collier, A. K., Wolf, D. H., Valdez, J. N., Turetsky, B. I., Elliott, M. A., Gur, R. E., & Gur, R. C. (2014). Comparison of auditory and visual oddball fMRI in schizophrenia. *Schizophrenia research*, *158*(1-3), 183-188.

Collip, D., Habets, P., Marcelis, M., Gronenschild, E., Lataster, T., Lardinois, M., … Myin-Germeys, I. (2013). Hippocampal volume as marker of daily life stress sensitivity in psychosis. *Psychological Medicine*, *43*(07), 1377–1387.

Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., … Hurles, M. E. (2010). Origins and functional impact of copy number variation in the human genome. *Nature*, *464*(7289), 704–712.

Conroy, M. A., & Polich, J. (2007). Normative Variation of P3a and P3b from a Large Sample. *Journal of Psychophysiology*, *21*(1), 22–32.

Consortium, W. T. C. C., & P. E. I. C. (2014). A Genome-Wide Association Analysis of a Broad Psychosis Phenotype Identifies Three Loci. *Biological Psychiatry*, *75*(5), 386–397.

Cooper, G. M., Coe, B. P., Girirajan, S., Rosenfeld, J. A., Vu, T. H., Baker, C., … Eichler, E. E. (2011). A copy number variation morbidity map of developmental delay. *Nature Genetics*, *43*(9), 838–846.

Cotton, S. M., McCann, T. V., Gleeson, J. F., Crisp, K., Murphy, B. P., & Lubman, D. I. (2013). Coping strategies in carers of young people with a first episode of psychosis. *Schizophrenia Research*, *146*(1–3), 118–124.

Cox, D. M., & Butler, M. G. (2015). The 15q11. 2 BP1–BP2 microdeletion syndrome: A review. *International journal of molecular sciences*, *16*(2), 4068-4082.

Craddock, N. (2013). Genome-wide association studies: what a psychiatrist needs to know. *Advances in Psychiatric Treatment*, *19*(2), 82–88.

Crespo-Facorro, B., Roiz-Santiáñez, R., Pelayo-Terán, J. M., Rodríguez-Sánchez, J. M., Pérez-Iglesias, R., González-Blanch, C., … Vázquez-Barquero, J. L. (2007). Reduced thalamic volume in first-episode non-affective psychosis: correlations with clinical variables, symptomatology and cognitive functioning. *NeuroImage*, *35*(4), 1613–1623.

Crespo-Facorro, B., Roiz-Santiáñez, R., Pérez-Iglesias, R., Tordesillas-Gutiérrez, D., Mata, I., Rodríguez-Sánchez, J. M., … Vázquez-Barquero, J. L. (2009). Specific brain structural abnormalities in first-episode schizophrenia. A comparative study with patients with schizophreniform disorder, non-schizophrenic

non-affective psychoses and healthy volunteers. *Schizophrenia Research*, *115*(2–3), 191–201.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. In *Interjournal*. Complex Systems.

Curtis, D. (2018). Polygenic risk score for schizophrenia is more strongly associated with ancestry than with schizophrenia. *Psychiatric Genetics*, *28*(5), 85–89.

Cutsuridis, V., Kumari, V., & Ettinger, U. (2014). Antisaccade performance in schizophrenia: a neural model of decision making in the superior colliculus. *Frontiers in Neuroscience*, *8*, 13.

Dalvie, S., Koen, N., Duncan, L., Abbo, C., Akena, D., Atwoli, L., ... & Mall, S. (2015). Large scale genetic research on neuropsychiatric disorders in African populations is needed. *EBioMedicine*, *2*(10), 1259-1261.

Dahoun, T., Trossbach, S. V, Brandon, N. J., Korth, C., & Howes, O. D. (2017). The impact of Disrupted-in-Schizophrenia 1 (DISC1) on the dopaminergic system: a systematic review. *Translational Psychiatry*, *7*(1), e1015.

Dapunt, J., Kluge, U., & Heinz, A. (2017). Risk of psychosis in refugees: a literature review. *Translational Psychiatry*, *7*(6), e1149.

Davies, G., Welham, J., Chant, D., Torrey, E. F., & McGrath, J. (2003). A Systematic Review and Meta-analysis of Northern Hemisphere Season of Birth Studies in Schizophrenia. *Schizophrenia Bulletin*, *29*(3), 587–593.

Davis, J., Eyre, H., Jacka, F. N., Dodd, S., Dean, O., McEwen, S., … Berk, M. (2016). A review of vulnerability and risks for schizophrenia: Beyond the two hit hypothesis. *Neuroscience and Biobehavioral Reviews*, 65, 185–194.

Dawn Teare, M., & Barrett, J. H. (2005). Genetic linkage studies. *The Lancet*, *366*(9490), 1036– 1044.

De Hert, M., Correll, C. U., Bobes, J., Cetkovich-Bakmas, M., Cohen, D., Asai, I., … Leucht, S. (2011). Physical illness in patients with severe mental disorders. I. Prevalence, impact of medications and disparities in health care. *World Psychiatry : Official Journal of the World Psychiatric Association (WPA)*, *10*(1), 52–77.

De La Vega, F. M., & Bustamante, C. D. (2018). Polygenic risk scores: A biased prediction? *Genome Medicine*, *10*(1).

De Peri, L., Crescini, A., Deste, G., Fusar-Poli, P., Sacchetti, E., & Vita, A. (2012). Brain Structural Abnormalities at the Onset of Schizophrenia and Bipolar Disorder: A Meta-analysis of Controlled Magnetic Resonance Imaging Studies. *Current Pharmaceutical Design*, *18*(4), 486–494.

Derks, E. M., Vorstman, J. A. S., Ripke, S., Kahn, R. S., Ophoff, R. A., & Ophoff, R. A. (2012). Investigation of the Genetic Association between Quantitative Measures of Psychosis and Schizophrenia: A Polygenic Risk Score Analysis. *PLoS ONE*, *7*(6), e37852.

DeRosse, P., & Karlsgodt, K. H. (2015). Examining the Psychosis Continuum. *Current Behavioral Neuroscience Reports*, *2*(2), 80–89.

Dias Alves, M., Micoulaud-Franchi, J. A., Simon, N., & Vion-Dury, J. (2018). Electroencephalogram Modifications Associated With Atypical Strict Antipsychotic Monotherapies. *Journal of Clinical Psychopharmacology*, *38*(6), 555–562.

Dickinson, D., Iannone, V. N., & Gold, J. M. (2002). Factor structure of the Wechsler Adult Intelligence Scale–III in schizophrenia. *Assessment*, *9*(2), 171–180.

Dickinson, D., Ragland, J. D., Calkins, M. E., Gold, J. M., & Gur, R. C. (2006). A comparison of cognitive structure in schizophrenia patients and healthy controls using confirmatory factor analysis. *Schizophrenia Research*, *85*(0), 20–29.

Díez, Á., Suazo, V., Casado, P., Martín-Loeches, M., Molina, V., Díez, A., … Molina, V. (2013). Spatial distribution and cognitive correlates of gamma noise power in schizophrenia. *Psychological Medicine*, *43*(6), 1175–1185.

Dikilitas, O., Schaid, D. J., Kosel, M. L., Chute, C. G., Jarvik, G. P., Denny, J. C., ... & Wiesner, G. L. (2019). Predictive Value of Polygenic Risk Scores for Coronary Heart Disease in Three Racial/Ethnic Groups. *Circulation*, *140*(Suppl_1), A11611-A11611.

Ding, L., Styblo, M., Drobná, Z., & Hegde, A. N. (2016). Expression of the longest RGS4 splice variant in the prefrontal cortex is associated with single nucleotide polymorphisms in schizophrenia patients. *Frontiers in psychiatry*, *7*, 26.

Dong, C., Nabizadeh, N., Caunca, M., Cheung, Y. K., Rundek, T., Elkind, M. S. V, … Wright, C. B. (2015). Cognitive correlates of white matter lesion load and brain atrophy: the Northern Manhattan Study. *Neurology*, *85*(5), 441–449.

Dong, S., Reder, L. M., Yao, Y., Liu, Y., & Chen, F. (2015). Individual differences in working memory capacity are reflected in different ERP and EEG patterns to task difficulty. *Brain Research*, *1616*, 146–156.

Dorph-Petersen, K.-A., Pierri, J. N., Perel, J. M., Sun, Z., Sampson, A. R., & Lewis, D. A. (2005). The influence of chronic exposure to antipsychotic medications on brain size before and after tissue fixation: a comparison of haloperidol and olanzapine in macaque monkeys. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, *30*(9), 1649–1661.

Dragulescu, A. A., & Arendt, C. (2018). *xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files*.

Dudbridge, F. (2013). Power and Predictive Accuracy of Polygenic Risk Scores. *PLoS Genetics*, *9*(3).

Durbin, R. (2014). Efficient haplotype matching and storage using the positional Burrows- Wheeler transform (PBWT). *Bioinformatics*, *30*(9), 1266–1272.

Dutt, A., McDonald, C., Dempster, E., Prata, D., Shaikh, M., Williams, I., … Bramon, E. (2009). The effect of COMT, BDNF, 5-HTT, NRG1 and DTNBP1 genes on hippocampal and lateral ventricular volume in psychosis. *Psychological Medicine*, *39*(11), 1783–1797.

Earls, H. A., Curran, T., & Mittal, V. (2016). A Meta-analytic Review of Auditory Event-Related Potential Components as Endophenotypes for Schizophrenia: Perspectives From First-Degree Relatives. *Schizophrenia Bulletin*, *42*(6), 1504–1516.

Eaton, W. W., Mortensen, P. B., & Frydenberg, M. (2000). Obstetric factors, urbanization and psychosis. *Schizophrenia Research*, *43*(2–3), 117–123.

Ek, M., Wicks, S., Svensson, A. C., Idring, S., & Dalman, C. (2015). Advancing Paternal Age and Schizophrenia: The Impact of Delayed Fatherhood. *Schizophrenia Bulletin*, *41*(3), 708–714.

Ekman, M., Granstrom, O., Omerov, S., Jacob, J., & Landen, M. (2013). The societal cost of schizophrenia in Sweden. *The Journal of Mental Health Policy and Economics*, *16*(1), 13– 25.

Endicott, J., & Spitzer, R. L. (1978a). A diagnostic interview. The schedule for affective disorders and schizophrenia. *Archives of General Psychiatry*, *35*(7), 837–844.

Endicott, J., & Spitzer, R. L. (1978b). A Diagnostic Interview. *Archives of General Psychiatry*, *35*(7), 837–844.

Erickson, M. A., Ruffle, A., & Gold, J. M. (2016). A meta-analysis of mismatch negativity in schizophrenia: from clinical risk to disease specificity and progression. *Biological psychiatry*, *79*(12), 980-987.

Escamilla, M., Hare, E., Dassori, A. M., Peralta, J. M., Ontiveros, A., Nicolini, H., ... & Almasy, L. (2009). A schizophrenia gene locus on chromosome 17q21 in a new set of families of Mexican and central american ancestry: evidence from the

NIMH Genetics of schizophrenia in latino populations study. *American Journal of Psychiatry*, *166*(4), 442-449.

Esterberg, M. L., & Compton, M. T. (2009). The psychosis continuum and categorical versus dimensional diagnostic approaches. *Current Psychiatry Reports*, *11*(3), 179–184.

Ethridge, L. E., Hamm, J. P., Pearlson, G. D., Tamminga, C. A., Sweeney, J. A., Keshavan, M. S., & Clementz, B. A. (2015). Event-Related Potential and Time-Frequency Endophenotypes for Schizophrenia and Psychotic Bipolar Disorder. *Biological Psychiatry*, *77*(2), 127–136.

Euesden, J., Lewis, C. M., & O'Reilly, P. F. (2015). PRSice: Polygenic Risk Score software. *Bioinformatics*, *31*(9), 1466–1468.

Facal, F., & Costas, J. (2019). Evidence of association of the DISC1 interactome gene set with schizophrenia from GWAS. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *95*, 109729.

Fannon, D., Tennakoon, L., Sumich, A., O'Ceallaigh, S., Doku, V., Chitnis, X., … Sharma, T. (2000). Third ventricle enlargement and developmental delay in first-episode psychosis: preliminary findings. *The British Journal of Psychiatry*, *177*(4), 354–359.

Fanous, A. H. (2010). Genetic Studies of Schizophrenia and Bipolar Disorder. *FOCUS*, *8*(3), 323–338.

Farrell, M. S., Werge, T., Sklar, P., Owen, M. J., Ophoff, R. A., O'Donovan, M. C., ... & Sullivan, P. F. (2015). Evaluating historical candidate genes for schizophrenia. *Molecular psychiatry*, *20*(5), 555.

Fatouros-Bergman, H., Cervenka, S., Flyckt, L., Edman, G., & Farde, L. (2014). Meta-analysis of cognitive performance in drug-naïve patients with schizophrenia. *Schizophrenia Research*, *158*(1–3), 156–162.

Fazel, S., Gulati, G., Linsell, L., Geddes, J. R., & Grann, M. (2009). Schizophrenia and violence: systematic review and meta-analysis. *PLoS Medicine*, *6*(8), e1000120.

Fernandez, T. V., Sanders, S. J., Yurkiewicz, I. R., Ercan-Sencicek, A. G., Kim, Y. S., Fishman, D. O., ... & Bilguvar, K. (2012). Rare copy number variants in tourette syndrome disrupt genes in histaminergic pathways and overlap with autism. *Biological psychiatry*, *71*(5), 392-402.

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., … Dale, A. M. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, *33*(3), 341–355.

Fjell, A. M., & Walhovd, K. B. (2001). P300 and neuropsychological tests as measures of aging: scalp topography and cognitive changes. *Brain Topography*, *14*(1), 25–40.

Flanagan, J. M., Popendikyte, V., Pozdniakovaite, N., Sobolev, M., Assadzadeh, A., Schumacher, A., ... & Petronis, A. (2006). Intra-and interindividual epigenetic variation in human germ cells. *The American Journal of Human Genetics*, *79*(1), 67-84.

Flint, J., & Munafò, M. R. (2007). The endophenotype concept in psychiatric genetics. *Psychological Medicine*, *37*(2), 163–180.

Flomen, R. H., Collier, D. A., Osborne, S., Munro, J., Breen, G., St Clair, D., & Makoff, A. J. (2006). Association study of CHRFAM7A copy number and 2bp deletion polymorphisms with schizophrenia and bipolar affective disorder. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *141B*(6), 571–575.

Forbes, N. F., Carrick, L. A., McIntosh, A. M., & Lawrie, S. M. (2009). Working memory in schizophrenia: a meta-analysis. *Psychological Medicine*, *39*(06), 889–905.

Ford, J. M. (2014). Decomposing P300 to identify its genetic basis. *Psychophysiology*, *51*(12), 1325–1326.

Forsyth, J. K., Nachun, D., Gandal, M. J., Geschwind, D. H., Anderson, A. E., Coppola, G., & Bearden, C. E. (2019). Synaptic and Gene Regulatory Mechanisms in Schizophrenia, Autism, and 22q11.2 CNV Mediated Risk for Neuropsychiatric Disorders. *BioRxiv*, 555490.

Fraguas, D., Díaz-Caneja, C. M., Pina-Camacho, L., Janssen, J., & Arango, C. (2016). Progressive brain changes in children and adolescents with early-onset psychosis: A meta-analysis of longitudinal MRI studies. *Schizophrenia Research*, *173*(3), 132–139.

Francks, C., Maegawa, S., Laurén, J., Abrahams, B. S., Velayos-Baeza, A., Medland, S. E., ... & Timmusk, T. (2007). LRRTM1 on chromosome 2p12 is a maternally suppressed gene that is associated paternally with handedness and schizophrenia. *Molecular psychiatry*, *12*(12), 1129-1139.

Frangou, S., Sharma, T., Sigmudsson, T., Barta, P., Pearlson, G., & Murray, R. M. (1997). The Maudsley Family Study. 4. Normal planum temporale asymmetry in familial schizophrenia. A volumetric MRI study. *The British Journal of Psychiatry*, *170*(4), 328–333.

Fu, C. H. Y., Steiner, H., & Costafreda, S. G. (2013). Predictive neural biomarkers of clinical response in depression: A meta-analysis of functional and structural neuroimaging studies of pharmacological and psychological therapies. *Neurobiology of Disease*, *52*, 75–83.

Fuggetta, G., Bennett, M. A., Duke, P. A., & Young, A. M. J. (2014). Quantitative electroencephalography as a biomarker for proneness toward developing psychosis. *Schizophrenia Research*, *153*(1–3), 68–77.

Fullerton, J. M., & Nurnberger, J. I. (2019). Polygenic risk scores in psychiatry: Will they be useful for clinicians? *F1000Research*, *8*, 1293.

Fusar-Poli, P., Deste, G., Smieskova, R., Barlati, S., Yung, A. R., Howes, O., … Borgwardt, S. (2012a). Cognitive functioning in prodromal psychosis: a meta-analysis. *Archives of General Psychiatry*, *69*(6), 562–571.

Fusar-Poli, P., Deste, G., Smieskova, R., Barlati, S., Yung, A. R., Howes, O., … Borgwardt, S. (2012). Cognitive Functioning in Prodromal Psychosis. *Archives of General Psychiatry*, *69*(6), 1–6.

Fusar-Poli, P., Rutigliano, G., Stahl, D., Davies, C., Bonoldi, I., Reilly, T., & McGuire, P. (2017). Development and validation of a clinically based risk calculator for the transdiagnostic prediction of psychosis. *JAMA Psychiatry*, *74*(5), 493–500.

Fusar-Poli, P., Smieskova, R., Kempton, M. J., Ho, B. C., Andreasen, N. C., & Borgwardt, S. (2013). Progressive brain changes in schizophrenia related to antipsychotic treatment? A meta-analysis of longitudinal MRI studies. *Neuroscience & Biobehavioral Reviews*, *37*(8), 1680–1691.

Garver, D. L., & L., D. (1987). Methodological Issues Facing the Interpretation of High-risk Studies: Biological Heterogeneity. *Schizophrenia Bulletin*, *13*(3), 525–529.

Georgieva, L., Rees, E., Moran, J. L., Chambert, K. D., Milanova, V., Craddock, N., ... & O'Donovan, M. C. (2014). De novo CNVs in bipolar affective disorder and schizophrenia. *Human molecular genetics*, *23*(24), 6677-6683.

Geschwind, D. H. (2011). Genetics of autism spectrum disorders. *Trends in Cognitive Sciences*, *15*(9), 409–416.

Geschwind, D. H., & Flint, J. (2015). Genetics and genomics of psychiatric disease. *Science*, *349*(6255), 1489–1494.

Giaroli, G., Bass, N., Strydom, A., Rantell, K., & McQuillin, A. (2014). Does rare matter? Copy number variants at 16p11.2 and the risk of psychosis: A systematic review of literature and meta-analysis. *Schizophrenia Research*, *159*(2–3), 340–346.

Gillett, A. C., Vassos, E., & Lewis, C. M. (2019). Transforming Summary Statistics from Logistic Regression to the Liability Scale: Application to Genetic and Environmental Risk Scores. *Human Heredity*, *83*(4), 210–224.

Girirajan, S., Rosenfeld, J. A., Coe, B. P., Parikh, S., Friedman, N., Goldstein, A., … Eichler, E. E. (2012). Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *The New England Journal of Medicine*, *367*(14), 1321–1331.

Gladsjo, J. A., McAdams, L. A., Palmer, B. W., Moore, D. J., Jeste, D. V., & Heaton, R. K.(2004). A six-factor model of cognition in schizophrenia and related psychotic disorders: relationships with clinical symptoms and functional capacity. *Schizophrenia Bulletin*, *30*(4), 739.

Glahn, D. C., Bearden, C. E., Cakir, S., Barrett, J. A., Najt, P., Serap Monkul, E., … Soares, J. C. (2006). Differential working memory impairment in bipolar disorder and schizophrenia: effects of lifetime history of psychosis. *Bipolar Disorders*, *8*(2), 117–123.

Glahn, D. C., Knowles, E. E. M., McKay, D. R., Sprooten, E., Raventós, H., Blangero, J., … Almasy, L. (2014). Arguments for the sake of endophenotypes: examining common misconceptions about the use of endophenotypes in psychiatric genetics. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics : The Official Publication of the International Society of Psychiatric Genetics*, *165B*(2), 122–130.

Glessner, J. T., Wang, K., Cai, G., Korvatska, O., Kim, C. E., Wood, S., ... & Imielinski, M. (2009). Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*, *459*(7246), 569-573.

Glessner, J. T., Reilly, M. P., Kim, C. E., Takahashi, N., Albano, A., Hou, C., ... & Imielinski, M. (2010). Strong synaptic transmission impact by copy number variations in schizophrenia. *Proceedings of the National Academy of Sciences*, *107*(23), 10584-10589.

Goes, F. S., & Sawa, A. (2017). Psychosis Beyond the 22q11.2 Deletion: Do Additional Genetic Factors Play a Role? *American Journal of Psychiatry*, *174*(11), 1027–1029.

Goes, F. S., Zandi, P. P., Miao, K., McMahon, F. J., Steele, J., Willour, V. L., … Potash, J. B. (2007). Mood-Incongruent Psychotic Features in Bipolar Disorder: Familial Aggregation and Suggestive Linkage to 2p11-q14 and 13q21-33. *American Journal of Psychiatry*, *164*(2), 236–247.

Gogtay, N., Sporn, A., Clasen, L. S., Greenstein, D., Giedd, J. N., Lenane, M., … Rapoport, J. L. (2003). Structural Brain MRI Abnormalities in Healthy Siblings of Patients With Childhood-Onset Schizophrenia. *American Journal of Psychiatry*, *160*(3), 569–571.

Goldstein, J. M., Buka, S. L., Seidman, L. J., & Tsuang, M. T. (2010). Specificity of Familial Transmission of Schizophrenia Psychosis Spectrum and Affective Psychoses in the New England Family Study's High-Risk Design. *Archives of General Psychiatry*, *67*(5), 458–467.

Gómez-de-Regil, L., Kwapil, T. R., & Barrantes-Vidal, N. (2014). Predictors of expressed emotion, burden and quality of life in relatives of Mexican patients with psychosis. *Journal of Psychiatric and Mental Health Nursing*, *21*(2), 170–179.

González-Blanch, C., Crespo-Facorro, B., Álvarez-Jiménez, M., Rodríguez-Sánchez, J. M., Pelayo-Terán, J. M., Pérez-Iglesias, R., & Vázquez-Barquero, J. L. (2007). Cognitive dimensions in first-episode schizophrenia spectrum disorders. *Journal of Psychiatric Research*, *41*(11), 968–977.

Goodin, D. S., Squires, K. C., Henderson, B. H., & Starr, A. (1978). Age-related variations in evoked potentials to auditory stimuli in normal human subjects. *Electroencephalography and Clinical Neurophysiology*, *44*(4), 447–458.

Gooding, D. C., Gjini, K., Burroughs, S. A., & Boutros, N. N. (2013). The association between psychosis proneness and sensory gating in cocaine-

dependent patients and healthy controls. *Psychiatry Research*, *210*(3), 1092–1100.

Goozée, R., Handley, R., Kempton, M. J., & Dazzan, P. (2014). A systematic review and meta-analysis of the effects of antipsychotic medications on regional cerebral blood flow (rCBF) in schizophrenia: Association with response to treatment. *Neuroscience and Biobehavioral Reviews*, 43, 118–136.

Goozee, R., Reinders, A. A., Handley, R., Marques, T., Taylor, H., O'Daly, O., ... & Dazzan, P. (2016). Effects of aripiprazole and haloperidol on neural activation during the n-back in healthy individuals: a functional MRI study. *Schizophrenia research*, *173*(3), 174-181.

Gordovez, F. J. A., & McMahon, F. J. (2020). The genetics of bipolar disorder. *Molecular Psychiatry*, 1-16.

Gottesman, I. I. (1991). *Schizophrenia genesis: The origins of madness.* WH Freeman/Times Books/Henry Holt & Co.

Gottesman, I. I., & Gould, T. D. (2003). The Endophenotype Concept in Psychiatry: Etymology and Strategic Intentions. *Amercian Journal of Psychiatry*, *160*(April), 636–645.

Gottesman, I. I., & Shields, J. (1967). A polygenic theory of schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*, *58*(1), 199–205.

Gottesman, I. I., & Shields, J. (1973). Genetic theorizing and schizophrenia. *The British Journal of Psychiatry : The Journal of Mental Science*, *122*(566), 15–30.

Grant, B. F., Stinson, F. S., Hasin, D. S., Dawson, D. A., Chou, S. P., Ruan, W. J., & Huang, B. (2005). Prevalence, correlates, and comorbidity of bipolar I disorder and axis I and II disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *The Journal of Clinical Psychiatry*, *66*(10), 1205–1215.

Gratten, J., Wray, N. R., Keller, M. C., & Visscher, P. M. (2014). Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nature Neuroscience*, *17*(6), 782–790.

Green, E. K., Rees, E., Walters, J. T. R., Smith, K.-G., Forty, L., Grozeva, D., … Kirov, G. (2016). Copy number variation in bipolar disorder. *Molecular Psychiatry*, *21*(1), 89–93.

Greenwood, T. A., Swerdlow, N. R., Gur, R. E., Cadenhead, K. S., Calkins, M. E., Dobie, D. J., Braff, D. L. (2013). Genome-Wide Linkage Analyses of 12 Endophenotypes for Schizophrenia From the Consortium on the Genetics of Schizophrenia. *American Journal of Psychiatry*, *170*(5), 521–532.

Greenwood, T., Shutes-David, A., & Tsuang, D. W. (2019). Endophenotypes in Schizophrenia: Digging Deeper to Identify Genetic Mechanisms. *Journal of Psychiatry and Brain Science*, *4*(2).

Grozeva, D., Kirov, G., Ivanov, D., Jones, I. R., Jones, L., Green, E. K., ... & McGuffin, P. (2010). Rare copy number variants: a point of rarity in genetic risk for bipolar disorder and schizophrenia. *Archives of general psychiatry*, *67*(4), 318-327.

Grozeva, D., Conrad, D. F., Barnes, C. P., Hurles, M., Owen, M. J., O'Donovan, M. C., … Kirov, G. (2011). Independent estimation of the frequency of rare CNVs in the UK population confirms their role in schizophrenia. *Schizophrenia Research 135*(1-3), 1-7.

Guloksuz, S., Pries, L. K., Delespaul, P., Kenis, G., Luykx, J. J., Lin, B. D., ... & Yalınçetin, B. (2019). Examining the independent and joint effects of molecular genetic liability and environmental exposures in schizophrenia: results from the EUGEI study. *World Psychiatry*, *18*(2), 173-182.

Gur, R. E., Calkins, M. E., Gur, R. C., Horan, W. P., Nuechterlein, K. H., Seidman, L. J., & Stone, W. S. (2007). The Consortium on the Genetics of

Schizophrenia: neurocognitive endophenotypes. *Schizophrenia Bulletin*, *33*(1), 49–68.

Gur, R. E. (2015). Neurocognitive performance in family-based and case-control studies of schizophrenia. *Schizophrenia Research*, *163*(1), 17–23.

Gureje, O., Herrman, H., Harvey, C., Morgan, V., & Jablensky, A. (2017). The Australian National Survey of Psychotic Disorders : profile of psychosocial disability and its risk factors. *Psychological Medicine*, *32*, 639–647.

Guyatt, A. L., Stergiakouli, E., Martin, J., Walters, J., O'Donovan, M., Owen, M., ... & Zammit, S. (2018). Association of copy number variation across the genome with neuropsychiatric traits in the general population. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *177*(5), 489-502.

Habets, P., Marcelis, M., Gronenschild, E., Drukker, M., & Van Os, J. (2011). Reduced cortical thickness as an outcome of differential sensitivity to environmental risks in schizophrenia. *Biol Psychiatry*, *69*(5), 487–494.

Haijma, S. V., Van Haren, N., Cahn, W., Koolschijn, P. C. M. P., Hulshoff Pol, H. E., & Kahn, R. S. (2013a). Brain Volumes in Schizophrenia: A Meta-Analysis in Over 18000 Subjects. *Schizophrenia Bulletin*, *39*(5), 1129–1138.

Hall, M.-H., Levy, D. L., Salisbury, D. F., Haddad, S., Gallagher, P., Lohan, M., … Smoller, J. W. (2014). Neurophysiologic effect of GWAS derived schizophrenia and bipolar risk variants. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics*, *165B*(1), 9–18.

Hall, M. H., Schulze, K., Bramon, E., Murray, R. M., Sham, P., & Rijsdijk, F. (2006). Genetic overlap between P300, P50, and duration mismatch negativity. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, *141*(4), 336–343.

Hall, M. H., Chen, C. Y., Cohen, B. M., Spencer, K. M., Levy, D. L., Öngür, D., & Smoller, J. W. (2015). Genomewide association analyses of electrophysiological endophenotypes for schizophrenia and psychotic bipolar

disorders: a preliminary report. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *168*(3), 151-161.

Hall, M. H., Schulze, K., Rijsdijk, F., Picchioni, M., Ettinger, U., Bramon, E., … Sham, P. (2006). Heritability and Reliability of P300, P50 and Duration Mismatch Negativity. *Behavior Genetics*, *36*(6), 845–857.

Hall, M. H., & Smoller, J. W. (2010). A new role for endophenotypes in the GWAS era: functional characterization of risk variants. *Harvard Review Psychiatry*, *18*(1), 67–74.

Harrison, P. J. (2015). Recent genetic findings in schizophrenia and their therapeutic relevance. *Journal of Psychopharmacology (Oxford, England)*, *29*(2), 85–96.

Harrison, P. J., & Owen, M. J. (2003). Genes for schizophrenia? Recent findings and their pathophysiological implications. *The Lancet*, *361*(9355), 417–419.

Hart, A. B., Gamazon, E. R., Engelhardt, B. E., Sklar, P., Kähler, A. K., Hultman, C. M., … Palmer, A. A. (2014). Genetic variation associated with euphorigenic effects of d-amphetamine is associated with diminished risk for schizophrenia and attention deficit hyperactivity disorder. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(16), 5968–5973.

Hartberg, C. B., Sundet, K., Rimol, L. M., Haukvik, U. K., Lange, E. H., Nesvåg, R., … Agartz, I. (2011). Subcortical brain volumes relate to neurocognition in schizophrenia and bipolar disorder and healthy controls. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, *35*(4), 1122–1130.

Harvey, P. D. (2006). Cognitive and functional effects of atypical antipsychotic medications. *The Journal of clinical psychiatry*, *67*(10), e13.

Heinrichs, R. W., & Zakzanis, K. K. (1998). Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. *Neuropsychology*, *12*(3), 426–445.

Hermens, D. F., Ward, P. B., Hodge, M. A. R., Kaur, M., Naismith, S. L., & Hickie, I. B. (2010). Impaired MMN/P3a complex in first-episode psychosis: Cognitive and psychosocial associations. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *34*(6), 822–829.

Heslenfeld, D. J. (2003). Visual Mismatch Negativity. In J. Polich (Ed.), *Detection of change: event-related potential and fMRI findings.* (pp. 41–59). Norwell, MA:: Kluwer Academic Publishers.

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.

Hilker, R., Helenius, D., Fagerlund, B., Skytthe, A., Christensen, K., Werge, T. M., … Glenthøj, B. (2018). Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide Danish Twin Register. *Biological Psychiatry*, *83*(6), 492–498.

Hipp, D. R. (2000). *SQLite*.

Hippisley-Cox, J., Coupland, C., Vinogradova, Y., Robson, J., May, M., & Brindle, P. (2007). Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: Prospective open cohort study. *British Medical Journal*, *335*(7611), 136–141.

Hippolyte, L., Maillard, A. M., Rodriguez-Herreros, B., Pain, A., Martin-Brevet, S., Ferrari, C., ... & Reigo, A. (2016). The number of genomic copies at the 16p11. 2 locus modulates language, verbal memory, and inhibition. *Biological psychiatry*, *80*(2), 129-139.

Ho, B.-C., Andreasen, N. C., Ziebell, S., Pierson, R., & Magnotta, V. (2011). Long-term antipsychotic treatment and brain volumes: a longitudinal study of first-episode schizophrenia. *Archives of General Psychiatry*, *68*(2), 128–137.

Hodgkinson, C. A., Goldman, D., Jaeger, J., Persaud, S., Kane, J. M., Lipsky, R. H., & Malhotra, A. K. (2004). Disrupted in schizophrenia 1 (DISC1): association with schizophrenia, schizoaffective disorder, and bipolar disorder. *American Journal of Human Genetics*, *75*(5), 862–872.

Hoeffding, L. K., Trabjerg, B. B., Olsen, L., Mazin, W., Sparsø, T., Vangkilde, A., ... & Werge, T. (2017). Risk of psychiatric disorders among individuals with the 22q11. 2 deletion or duplication: a Danish nationwide, register-based study. *JAMA psychiatry*, *74*(3), 282-290.

Horan, W. P., Braff, D. L., Nuechterlein, K. H., Sugar, C. A., Cadenhead, K. S., Calkins, M. E., … Green, M. F. (2008). Verbal working memory impairments in individuals with schizophrenia and their first-degree relatives: Findings from the Consortium on the Genetics of Schizophrenia. *Schizophrenia Research*, *103*(1–3), 218–228.

Huang, A. Y., Yu, D., Davis, L. K., Sul, J. H., Tsetsos, F., Ramensky, V., ... & McGrath, L. M. (2017). Rare copy number variants in NRXN1 and CNTN6 increase risk for Tourette syndrome. *Neuron*, *94*(6), 1101-1111.

Hubbard, L., Tansey, K. E., Rai, D., Jones, P., Ripke, S., Chambert, K. D., ... & Zammit, S. (2016). Evidence of common genetic overlap between schizophrenia and cognition. *Schizophrenia bulletin*, *42*(3), 832-842

Huguet, G., Schramm, C., Douard, E., Jiang, L., Labbe, A., Tihy, F., ... & Poline, J. B. (2018). Measuring and estimating the effect sizes of copy number variants on general intelligence in community-based samples. *JAMA psychiatry*, *75*(5), 447-457.

Huhtaniska, S., Jääskeläinen, E., Hirvonen, N., Remes, J., Murray, G. K., Veijola, J., … Miettunen, J. (2017). Long-term antipsychotic use and brain changes

in schizophrenia - a systematic review and meta-analysis. *Human Psychopharmacology: Clinical and Experimental*, *32*(2), e2574.

Hulshoff Pol, H. E., Schnack, H. G., Bertens, M. G., van Haren, N. E., van der Tweel, I., Staal, W. G., ... & Kahn, R. S. (2002). Volume changes in gray matter in patients with schizophrenia. *American Journal of Psychiatry*, *159*(2), 244-250.

Staal, W. G., … Kahn, R. S. (2002). Volume changes in gray matter in patients with schizophrenia. *The American Journal of Psychiatry*, *159*(2), 244–250.

Hyun, J., Myung, J. B., & Ung, G. K. (2011). Effects of psychotropic drugs on quantitative EEG among patients with schizophrenia-spectrum disorders. *Clinical Psychopharmacology and Neuroscience*, *9*, 78–85.

Ian, K., Jenner, J. A., & Cannon, M. (2010). Psychotic symptoms in the general population - an evolutionary perspective. *The British Journal of Psychiatry : The Journal of Mental Science*, *197*(3), 167–169.

Ivleva, E. I., Moates, A. F., Hamm, J. P., Bernstein, I. H., O'Neill, H. B., Cole, D., … Tamminga, C. A. (2014). Smooth Pursuit Eye Movement, Prepulse Inhibition, and Auditory Paired Stimuli Processing Endophenotypes Across the Schizophrenia-Bipolar Disorder Psychosis Dimension. *Schizophrenia Bulletin*, *40*(3), 642–652.

Ivleva, E. I., Morris, D. W., Osuji, J., Moates, A. F., Carmody, T. J., Thaker, G. K., … Tamminga, C. A. (2012). Cognitive endophenotypes of psychosis within dimension and diagnosis. *Psychiatry Research*, *196*(1), 38–44.

Jackson, A., & Seneviratne, U. (2019, June 1). EEG changes in patients on antipsychotic therapy: A systematic review. *Epilepsy and Behavior*, *95*, 1–9.

Jasper, H. (1958). Report to the committee on methods of clinical examination in electroencephalography. *Electroencephalography and Clinical Neurophysiology*, *10*, 371– 375.

Jia, P., Wang, L., Fanous, A. H., Chen, X., Kendler, K. S., Zhao, Z., & International Schizophrenia Consortium. (2012). A bias-reducing pathway enrichment analysis of genome-wide association data confirmed association of the MHC region with schizophrenia. *Journal of medical genetics*, *49*(2), 96-103.

Johansson, V., Kuja-Halkola, R., Cannon, T. D., Hultman, C. M., & Hedman, A. M. (2019). A population-based heritability estimate of bipolar disorder – In a Swedish twin sample. *Psychiatry Research*, *278*, 180–187.

Johns, L. C., & van Os, J. (2001). The continuity of psychotic experiences in the general population. *Clinical Psychology Review*, *21*(8), 1125–1141.

Johnsen, E., Jørgensen, H. A., Kroken, R. A., & Løberg, E. M. (2013). Neurocognitive effectiveness of quetiapine, olanzapine, risperidone, and ziprasidone: a pragmatic, randomized trial. *European psychiatry*, *28*(3), 174-184.

Johnstone, E. C., Ebmeier, K. P., Miller, P., Owens, D. G. C., & Lawrie, S. M. (2005). Predicting schizophrenia: findings from the Edinburgh High-Risk Study. *The British Journal of Psychiatry*, *186*(1), 18–25.

Jongsma, H. E., Turner, C., Kirkbride, J. B., & Jones, P. B. (2019). International incidence of psychotic disorders, 2002–17: a systematic review and meta-analysis. *The Lancet Public Health*, *4*(5), 229–244.

Kaminsky, E. B., Kaul, V., Paschall, J., Church, D. M., Bunke, B., Kunig, D., … Martin, C. L. (2011). An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, *13*(9), 777–784.

Karayiorgou, M., Morris, M. A., Morrow, B., Shprintzen, R. J., Goldberg, R., Borrow, J., … Lasseter, V. K. (1995). Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(17), 7612–7616.

Karlsson, R., Graae, L., Lekman, M., Wang, D., Favis, R., Axelsson, T., … Paddock, S. (2012). MAGI1 Copy Number Variation in Bipolar Affective Disorder and Schizophrenia. *BPS*, *71*, 922–930.

Kaur, M., Battisti, R. a., Ward, P. B., Ahmed, A., Hickie, I. B., & Hermens, D. F. (2011). MMN/P3a deficits in first episode psychosis: Comparing schizophrenia-spectrum and affective-spectrum subgroups. *Schizophrenia Research*, *130*(1–3), 203–209.

Keilp, J. G., Sweeney, J. A., Jacobsen, P., Solomon, C., St.Louis, L., Deck, M., … Mann, J. J. (1988). Cognitive impairment in schizophrenia: Specific relations to ventricular size and negative symptomatology. *Biological Psychiatry*, *24*(1), 47–55.

Kempton, M. J., Stahl, D., Williams, S. C. R., & DeLisi, L. E. (2010a). Progressive lateral ventricular enlargement in schizophrenia: A meta-analysis of longitudinal MRI studies. *Schizophrenia Research*, *120*(1–3), 54–62.

Kempton, M. J., Stahl, D., Williams, S. C. R., & DeLisi, L. E. (2010b). Progressive lateral ventricular enlargement in schizophrenia: A meta-analysis of longitudinal MRI studies. *Schizophrenia Research*, *120*(1–3), 54–62.

Kendler, K., & Neale, M. (2010). Endophenotype: a conceptual analysis. *Molecular Psychiatry*, *15*(8), 789–797.

Kety, S. S., Wender, P. H., Jacobsen, B., Ingraham, L. J., Jansson, 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.

Kim, D., Kim, J., Koo, T., Yun, H., & Won, S. (2015). Shared and distinct neurocognitive endophenotypes of schizophrenia and psychotic bipolar disorder. *Clinical Psychopharmacology and Neuroscience*, *13*(1), 94–102.

Kim, M.-S., Kang, S. S., Youn, T., Kang, D. H., Kim, J. J., & Kwon, J. S. (2003). Neuropsychological correlates of P300 abnormalities in patients with

schizophrenia and obsessive–compulsive disorder. *Psychiatry Research: Neuroimaging*, *123*(2), 109–123.

Kim, M., Seo, H.-J., Yun, H., Jung, Y.-E., Park, J. H., Lee, C.-I., … Bahk, W.-M. (2015). The Relationship between Cognitive Decline and Psychopathology in Patients with Schizophrenia and Bipolar Disorder. *Clinical Psychopharmacology and Neuroscience*, *13*(1), 103–108.

Kim, M., Lee, T. H., Kim, J. H., Hong, H., Lee, T. Y., Lee, Y., ... & Kwon, J. S. (2018). Decomposing P300 into correlates of genetic risk and current symptoms in schizophrenia: an inter-trial variability analysis. *Schizophrenia research*, *192*, 232-239.

Kim, Y., Zerwas, S., Trace, S. E., & Sullivan, P. F. (2011). Schizophrenia genetics: where next? *Schizophrenia Bulletin*, *37*(3), 456–463.

Kirkbride, J. B. (2017). Migration and psychosis: our smoking lung? *World Psychiatry : Official Journal of the World Psychiatric Association (WPA)*, *16*(2), 119–120.

Kirkbride, J. B., Boydell, J., Ploubidis, G. B., Morgan, C., Dazzan, P., McKenzie, K., … Jones, P. B. (2008). Testing the association between the incidence of schizophrenia and social capital in an urban area. *Psychological Medicine*, *38*(08), 1083–1094.

Kirkbride, J. B., Errazuriz, A., Croudace, T. J., Morgan, C., Jackson, D., Boydell, J., ... & Jones, P. B. (2012). Incidence of schizophrenia and other psychoses in England, 1950–2009: a systematic review and meta-analyses. *PloS one*, *7*(3).

Kirov, G., Grozeva, D., Norton, N., Ivanov, D., Mantripragada, K. K., Holmans, P., … O'Donovan, M. C. (2009). Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Human Molecular Genetics*, *18*(8), 1497–1503.

Kirov, G. (2010). The role of copy number variation in schizophrenia. *Expert review of neurotherapeutics*, *10*(1), 25-32.

Kirov, G., Pocklington, A. J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., ... & Grozeva, D. (2012). De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Molecular psychiatry*, *17*(2), 142-153.

Kirov, G., Rees, E., Walters, J. T. R., Escott-Price, V., Georgieva, L., Richards, A. L., … Owen, M. J. (2014). The Penetrance of Copy Number Variations for Schizophrenia and Developmental Delay. *Biological Psychiatry*, *75*(5), 378–385.

Konopaske, G. T., Dorph-Petersen, K. A., Pierri, J. N., Wu, Q., Sampson, A. R., & Lewis, D. A. (2007). Effect of chronic exposure to antipsychotic medication on cell numbers in the parietal cortex of macaque monkeys. *Neuropsychopharmacology*, *32*(6), 1216–1223.

Korver, N., Quee, P. J., Boos, H. B. M., Simons, C. J. P., & de Haan, L. (2012). Genetic Risk and Outcome of Psychosis (GROUP), a multi site longitudinal cohort study focused on gene-environment interaction: objectives, sample characteristics, recruitment and assessment methods. *International Journal of Methods in Psychiatric Research*, *21*(3), 205–221.

Kovalchik, S. (2017). *RISmed: Download Content from NCBI Databases*.

Krigolson, O. E. (2018). Event-related brain potentials and the study of reward processing: Methodological considerations. *International Journal of Psychophysiology*, *132*, 175-183.

Kujala, A., & Naatanen, R. (2003). Auditory environment and change detection as indexed by the mismarch gegativity (MMN). In J. Polich (Ed.), *Detection of Change: Event-Related Potential and fMRI Findings*. Boston: Springer US.

Kumra, S., Giedd, J. N., Vaituzis, A. C., Jacobsen, L. K., McKenna, K., Bedwell, J., … Rapoport, L. (2014). Childhood-onset psychotic disorders: magnetic resonance imaging of volumetric differences in brain structure. *American Journal of Psychiatry 157*(9), 1467-1474).

Lal, D., Ruppert, A.-K., Trucks, H., Schulz, H., de Kovel, C. G., Kasteleijn-Nolst Trenité, D., … Sander, T. (2015). Burden analysis of rare microdeletions suggests a strong impact of neurodevelopmental genes in genetic generalised epilepsies. *PLoS Genetics*, *11*(5), e1005226.

Lam, M. T., & McDiarmid, M. (2016). Increasing number of databases searched in systematic reviews and meta-analyses between 1994 and 2014. *Journal of the Medical Library Association: JMLA*, *104*(4), 284.

Large, M. M., & Nielssen, O. (2011). Violence in first-episode psychosis: A systematic review and meta-analysis. *Schizophrenia Research*, *125*(2–3), 209–220.

Laursen, T. M., Agerbo, E., & Pedersen, C. B. (2009). Bipolar disorder, schizoaffective disorder, and schizophrenia overlap: A new comorbidity index. *Journal of Clinical Psychiatry*, *70*(10), 1432–1438.

Lee, J., Green, M. F., Calkins, M. E., Greenwood, T. A., Gur, R. E. … Braff, D. L. (2015). Verbal working memory in schizophrenia from the Consortium on the Genetics of Schizophrenia (COGS) Study: The moderating role of smoking status and antipsychotic medications. *Schizophrenia Research*, *163*(1), 24–31.

Lee, S. H., Ripke, S., Neale, B. M., Faraone, S. V., Purcell, S. M., Perlis, R. H., … Wray, N. R. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics*, *45*(9), 984–994.

Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M., & Wray, N. R. (2012). Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*, *28*(19), 2540– 2542.

Lee, T. Y., Hong, S. Bin, Shin, N. Y., & Kwon, J. S. (2015). Social cognitive functioning in prodromal psychosis: A meta-analysis. *Schizophrenia Research*, *164*(1–3), 28–34.

Leeson, V. C., Sharma, P., Harrison, M., Ron, M. A., Barnes, T. R., & Joyce, E. M. (2011). IQ trajectory, cognitive reserve, and clinical outcome following a first episode of psychosis: a 3-year longitudinal study. *Schizophrenia bulletin*, *37*(4), 768-777.

Lencz, T., Guha, S., Liu, C., Rosenfeld, J., Mukherjee, S., DeRosse, P., ... & Darvasi, A. (2013). Genome-wide association study implicates NDST3 in schizophrenia and bipolar disorder. *Nature communications*, *4*(1), 1-10.

Lenzenweger, M. F. (2013). Endophenotype, Intermediate phenotype, biomarkers: Definitions, concept comparisons, clarifications. *Depression and Anxiety*, *30*(3), 185–189.

Leppa, V. M., Kravitz, S. N., Martin, C. L., Andrieux, J., Le Caignec, C., Martin-Coignard, D., ... & Geschwind, D. H. (2016). Rare inherited and de novo CNVs reveal complex contributions to ASD risk in multiplex families. *The American Journal of Human Genetics*, *99*(3), 540-554.

Lerer, B., Segman, R. H., Hamdan, A., Kanyas, K., Karni, O., Kohn, Y., ... & Macciardi, F. (2003). Genome scan of Arab Israeli families maps a schizophrenia susceptibility gene to chromosome 6q23 and supports a locus at chromosome 10q24. *Molecular psychiatry*, *8*(5), 488-498.

Leucht, S., Arbter, D., Engel, R. R., Kissling, W., & Davis, J. M. (2009). How effective are second-generation antipsychotic drugs? A meta-analysis of placebo-controlled trials. *Molecular Psychiatry*, *14*(4), 429–447.

Leucht, S., Tardy, M., Komossa, K., Heres, S., Kissling, W., Salanti, G., & Davis, J. M. (2012). Antipsychotic drugs versus placebo for relapse prevention in schizophrenia: a systematic review and meta-analysis. *The Lancet*, *379*(9831), 2063–2071.

Levinson, D. F., Duan, J., Oh, S., Wang, K., Sanders, A. R., Shi, J., … Gejman, P. V. (2011). Copy Number Variants in Schizophrenia: Confirmation of Five Previous Findings and New Evidence for 3q29 Microdeletions and VIPR2 Duplications. *American Journal of Psychiatry*, *168*(3), 302–316.

Li, Z., Chen, J., Xu, Y., Yi, Q., Ji, W., Wang, P., … Shi, Y. (2016). Genome-wide Analysis of the Role of Copy Number Variation in Schizophrenia Risk in Chinese. *Biological Psychiatry*, *80*(4), 331–337.

Lichtenstein, P., Yip, B. H., Björk, C., Pawitan, Y., Cannon, T. D., Sullivan, P. F., & Hultman, C. M. (2009a). Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *The Lancet*, *373*(9659), 234–239.

Lieberman, J. A., Stroup, T. S., McEvoy, J. P., Swartz, M. S., Rosenheck, R. A., Perkins, D. O., … Hsiao, J. K. (2005). Effectiveness of Antipsychotic Drugs in Patients with Chronic Schizophrenia. *New England Journal of Medicine*, *353*(12), 1209–1223.

Light, G. A., Swerdlow, N. R., Thomas, M. L., Calkins, M. E., Green, M. F., Greenwood, T. A., … Turetsky, B. I. (2015). Validation of mismatch negativity and P3a for use in multi-site studies of schizophrenia: Characterization of demographic, clinical, cognitive, and functional correlates in COGS-2. *Schizophrenia Research*, *163*(1–3), 63–72.

Liu, J., Ulloa, A., Perrone-Bizzozero, N., Yeo, R., Chen, J., & Calhoun, V. D. (2012). A pilot study on collective effects of 22q13. 31 deletions on gray matter concentration in schizophrenia. *PloS one*, *7*(12).

Loh, P.-R., Danecek, P., Palamara, P. F., Fuchsberger, C., Reshef, Y. A., Finucane, H. K., … Abecasis, G. R. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. *Nature Genetics*, *48*(11), 1443.

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., … Parkinson, H. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research*, *45*(D1), D896–D901.

MacKenzie, N. E., Kowalchuk, C., Agarwal, S. M., Costa-Dookhan, K. A., Caravaggio, F., Gerretsen, P., ... & Graff-Guerrero, A. (2018). Antipsychotics, metabolic adverse effects, and cognitive function in schizophrenia. *Frontiers in Psychiatry*, *9*, 622.

MacDonald, J. R., Ziman, R., Yuen, R. K. C., Feuk, L., & Scherer, S. W. (2014). The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Research*, *42*(Database issue), D986-92.

McDonald-McGinn, D. M., Sullivan, K. E., Marino, B., Philip, N., Swillen, A., Vorstman, J. A., ... & Scambler, P. J. (2015). 22q11. 2 deletion syndrome. *Nature reviews Disease primers*, *1*(1), 1-19.

Malhotra, D., McCarthy, S., Michaelson, J. J., Vacic, V., Burdick, K. E., Yoon, S., … Sebat, J. (2011a). High frequencies of de novo cnvs in bipolar disorder and schizophrenia. *Neuron*, *72*(6), 951–963.

Malhotra, D., McCarthy, S., Michaelson, J. J., Vacic, V., Burdick, K. E., Yoon, S., … Sebat, J. (2011b). High Frequencies of De Novo CNVs in Bipolar Disorder and Schizophrenia. *Neuron*, *72*(6), 951–963.

March, D., Hatch, S. L., Morgan, C., Kirkbride, J. B., Bresnahan, M., Fearon, P., & Susser, E. (2008). Psychosis and Place. *Epidemiologic Reviews*, *30*(1), 84–100.

Marchini, J., Howie, B., Myers, S., McVean, G., & Donnelly, P. (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature Genetics*, *39*(7), 906–913.

Markota, M., Coombes, B. J., Larrabee, B. R., McElroy, S. L., Bond, D. J., Veldic, M., ... & Kung, S. (2018). Association of schizophrenia polygenic risk score with manic and depressive psychosis in bipolar disorder. *Translational psychiatry*, *8*(1), 1-7.

Marshall, C. R., Howrigan, D. P., Merico, D., Thiruvahindrapuram, B., Wu, W., Greer, D. S., … CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium. (2017). Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nature Genetics*, *49*(1), 27–35.

Martin, A. R., Teferra, S., Möller, M., Hoal, E. G., & Daly, M. J. (2018). The critical needs and challenges for genetic architecture studies in Africa. *Current opinion in genetics & development*, *53*, 113-120.

Massuda, R., Bücker, J., Czepielewski, L. S., Narvaez, J. C., Pedrini, M., Santos, B. T., … Gama, C. S. (2013). Verbal memory impairment in healthy siblings of patients with schizophrenia. *Schizophrenia Research*, *150*(2–3), 580–582.

Mata, I., Perez-Iglesias, R., Roiz-Santiañez, R., Tordesillas-Gutierrez, D., Gonzalez-Mandly, A., Vazquez-Barquero, J. L., & Crespo-Facorro, B. (2009). A Neuregulin 1 Variant Is Associated with Increased Lateral Ventricle Volume in Patients with First-Episode Schizophrenia. *Biological Psychiatry*, *65*(6), 535–540.

Matheson, S. L., Shepherd, A. M., & Carr, V. J. (2017). *How much do we know about schizophrenia and how well do we know it? Evidence from the Schizophrenia Library*. *Psychological medicine*, *44*(16), 3387-3405.

Mavaddat, N., Michailidou, K., Dennis, J., Lush, M., Fachal, L., Lee, A., ... & Yang, X. (2019). Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *The American Journal of Human Genetics*, *104*(1), 21-34.

McCarroll, S. A., Hadnott, T. N., Perry, G. H., Sabeti, P. C., Zody, M. C., Barrett, J. C., … The International HapMap Consortium, T. I. H. (2006). Common deletion polymorphisms in the human genome. *Nature Genetics*, *38*(1), 86–92.

McCarroll, S. A., Kuruvilla, F. G., Korn, J. M., Cawley, S., Nemesh, J., Wysoker, A., … Altshuler, D. (2008). Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nature Genetics*, *40*(10), 1166–1174.

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., … Sharp, K. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*, *48*(10), 1279.

McCarthy, S. E., Makarov, V., Kirov, G., Addington, A. M., McClellan, J., Yoon, S., … Sebat, J. (2009). Microduplications of 16p11.2 are associated with schizophrenia. *Nature Genetics*, *41*(11), 1223–1227.

McDonald, C., Grech, A., Toulopoulou, T., Schulze, K., Chapple, B., Sham, P., … Murray, R. M. (2002). Brain volumes in familial and non-familial schizophrenic probands and their unaffected relatives. *American Journal of Medical Genetics*, *114*(6), 616–625.

McDonald, C., Marshall, N., Sham, P. C., Bullmore, E. T., Schulze, K., Chapple, B., … Murray, R. M. (2006). Regional brain morphometry in patients with schizophrenia or bipolar disorder and their unaffected relatives. *The American Journal of Psychiatry*, *163*(3), 478– 487.

McGuffin, P., Owen, M. J. & Gottesman, I. I. (Eds). (2004). *Psychiatric genetics and genomics*. Oxford University Press.

McIntosh, A. M., Harrison, L. K., Forrester, K., Lawrie, S. M., & Johnstone, E. C. (2005). Neuropsychological impairments in people with schizophrenia or bipolar disorder and their unaffected relatives. *The British Journal of Psychiatry : The Journal of Mental Science*, *186*, 378–385.

McIntosh, A. M., Job, D. E., Moorhead, T. W. J., Harrison, L. K., Forrester, K., Lawrie, S. M., & Johnstone, E. C. (2004). Voxel-based morphometry of patients

with schizophrenia or bipolar disorder and their unaffected relatives. *Biological Psychiatry*, *56*(8), 544–552.

McIntosh, A. M., Job, D. E., Moorhead, T. W. J., Harrison, L. K., Lawrie, S. M., & Johnstone, E. C. (2005). White matter density in patients with schizophrenia, bipolar disorder and their unaffected relatives. *Biological Psychiatry*, *58*(3), 254–257.

McIntosh, A. M., Gow, A., Luciano, M., Davies, G., Liewald, D. C., Harris, S. E., ... & Deary, I. J. (2013). Polygenic risk for schizophrenia is associated with cognitive change between childhood and old age. *Biological psychiatry*, *73*(10), 938-943.

Mena, A., Ruiz-Salas, J. C., Puentes, A., Dorado, I., Ruiz-Veguilla, M., & De la Casa, L. G. (2016). Reduced prepulse inhibition as a biomarker of schizophrenia. *Frontiers in behavioral neuroscience*, *10*, 202.

Mendlewicz, J., & Rainer, J. D. (1977). Adoption study supporting genetic transmission in manic–depressive illness. *Nature*, *268*(5618), 327–329.

Merikangas, K. R., Akiskal, H. S., Angst, J., Greenberg, P. E., Hirschfeld, R. M. A., Petukhova, M., & Kessler, R. C. (2007). Lifetime and 12-Month Prevalence of Bipolar Spectrum Disorder in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *64*(5), 543.

Moberget, T., Doan, N. T., Alnæs, D., Kaufmann, T., Córdova-Palomera, A., Lagerberg, T. V, … Westlye, L. T. (2017). Cerebellar volume and cerebellocerebral structural covariance in schizophrenia: a multisite mega-analysis of 983 patients and 1349 healthy controls. *Molecular Psychiatry, 23*(6), 1512.

Mokhtari, R., & Lachman, H. M. (2016). The major histocompatibility complex (MHC) in schizophrenia: a review. *Journal of clinical & cellular immunology*, *7*(6).

Morales-Muñoz, I., Martínez-Gras, I., Ponce, G., de la Cruz, J., Lora, D., Rodríguez-Jiménez, R., … Rubio, G. (2017). Psychological symptomatology and

impaired prepulse inhibition of the startle reflex are associated with cannabis-induced psychosis. *Journal of Psychopharmacology*, *31*(8), 1035–1045.

Moreno-De-Luca, D., Mulle, J. G., Kaminsky, E. B., Sanders, S. J., Myers, S. M., Adam, M. P., … Ledbetter, D. H. (2010). Deletion 17q12 Is a Recurrent Copy Number Variant that Confers High Risk of Autism and Schizophrenia. *The American Journal of Human Genetics*, *87*(5), 618–630.

Morgan, C., & Gayer-Anderson, C. (2016). Childhood adversities and psychosis: evidence, challenges, implications. *World Psychiatry : Official Journal of the World Psychiatric Association (WPA)*, *15*(2), 93–102.

Morgan, C., Kirkbride, J., Hutchinson, G., Craig, T., Morgan, K., Dazzan, P., … Fearon, P. (2008). Cumulative social disadvantage, ethnicity and first-episode psychosis: a case-control study. *Psychological Medicine*, *38*(12), 1701.

Morris, J. A., Randall, J. C., Maller, J. B., & Barrett, J. C. (2010). Evoker: a visualization tool for genotype intensity data. *Bioinformatics (Oxford, England)*, *26*(14), 1786–1787.

Morrow, E. M. (2010). Genomic Copy Number Variation in Disorders of Cognitive Development. *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(11), 1091–1104. Muller, K., Wickham, H., James, D. A., & Falcon, S. (2019). *RSQLite: "SQLite" Interface for R*.

Munn, Z., Peters, M. D., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC medical research methodology*, *18*(1), 143.

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, *13*(02), 201–233.

Näätänen, R., Kujala, T., Escera, C., Baldeweg, T., Kreegipuu, K., Carlson, S., & Ponton, C. (2012). The mismatch negativity (MMN)–a unique window to

disturbed central auditory processing in ageing and different clinical conditions. *Clinical Neurophysiology*, *123*(3), 424-458.

Narayanan, B., O'Neil, K., Berwise, C., Stevens, M. C., Calhoun, V. D., Clementz, B. A., … Pearlson, G. D. (2014). Resting State Electroencephalogram Oscillatory Abnormalities in Schizophrenia and Psychotic Bipolar Patients and Their Relatives from the Bipolar and Schizophrenia Network on Intermediate Phenotypes Study. *Biological Psychiatry*, *76*(6), 456–465.

Neil, A. L., Carr, V. J., Mihalopoulos, C., Mackinnon, A., & Morgan, V. A. (2014). Costs of psychosis in 2010: Findings from the second Australian National Survey of Psychosis. *Australian & New Zealand Journal of Psychiatry*, *48*(2), 169–182.

Ng, M. Y. M., Levinson, D. F., Faraone, S. V, Suarez, B. K., DeLisi, L. E., Arinami, T., … Lewis, C. M. (2009). Meta-analysis of 32 genome-wide linkage studies of schizophrenia. *Molecular Psychiatry*, *14*(8), 774–785.

Notaras, M. J., Vivian, B., Wilson, C., & van den Buuse, M. (2017). Interaction of reelin and stress on immobility in the forced swim test but not dopamine-mediated locomotor hyperactivity or prepulse inhibition disruption: Relevance to psychotic and mood disorders. *Schizophrenia Research*.

Nowakowska, B. (2017). Clinical interpretation of copy number variants in the human genome. *Journal of Applied Genetics*, *58*(4), 449–457.

O'Donovan, M. C., Craddock, N., Norton, N., Williams, H., Peirce, T., Moskvina, V., … Owen, M. J. (2008). Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nature Genetics*, *40*(9), 1053–1055.

O'Donovan, M. C., & Owen, M. J. (2016). The implications of the shared genetics of psychiatric disorders. *Nature medicine*, *22*(11), 1214.

Oh, H., Abe, J., Negi, N., & DeVylder, J. (2015). Immigration and psychotic experiences in the United States: Another example of the epidemiological paradox?. *Psychiatry research*, *229*(3), 784-790.

Ohi, K., Nishizawa, D., Shimada, T., Kataoka, Y., Hasegawa, J., Shioiri, T., ... & Ikeda, K. (2020). Polygenetic Risk Scores for Major Psychiatric Disorders Among Schizophrenia Patients, Their First-Degree Relatives and Healthy Subjects. *International Journal of Neuropsychopharmacology*.

Olabi, B., Ellison-Wright, I., McIntosh, A. M., Wood, S. J., Bullmore, E., & Lawrie, S. M. (2011). Are There Progressive Brain Changes in Schizophrenia? A Meta-Analysis of Structural Magnetic Resonance Imaging Studies. *Biological Psychiatry*, *70*(1), 88–96.

Oliver, D., Radua, J., Reichenberg, A., Uher, R., & Fusar-Poli, P. (2019). Psychosis Polyrisk score (PPS) for the detection of individuals at-risk and the prediction of their outcomes. *Frontiers in Psychiatry*, *10*, 174.

Owen, M. J., & O'Donovan, M. C. (2017). Schizophrenia and the neurodevelopmental continuum: evidence from genomics. *World Psychiatry*, *16*(3), 227–235.

Owens, E. M., Bachman, P., Glahn, D. C., & Bearden, C. E. (2016). Electrophysiological Endophenotypes for Schizophrenia. *Harvard Review of Psychiatry*, *24*(2), 129–147.

Palk, A. C., Dalvie, S., De Vries, J., Martin, A. R., & Stein, D. J. (2019, February 27). Potential use of clinical polygenic risk scores in psychiatry - Ethical implications and communicating high polygenic risk. *Philosophy, Ethics, and Humanities in Medicine*, *14*(1), 4.

Palmer, J. R. (2019). Polygenic risk scores for breast cancer risk prediction: Lessons learned and future opportunities. *JNCI: Journal of the National Cancer Institute*.

Pardiñas, A. F., Holmans, P., Pocklington, A. J., Escott-Price, V., Ripke, S., Carrera, N., … Walters, J. T. R. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nature Genetics*, *50*(3), 381–389.

Park, H., Kim, J.I., Ju, Y. S., Gokcumen, O., Mills, R. E., Kim, S., … Seo, J.S. (2010). Discovery of common Asian copy number variants using integrated high-resolution array CGH and massively parallel DNA sequencing. *Nature Genetics*, *42*(5), 400–405.

Park, S., & Gooding, D. C. (2014). Working memory impairment as an endophenotypic marker of a schizophrenia diathesis. *Schizophrenia Research: Cognition*, *1*(3), 127–136.

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, *2*(12), 190.

Pearlson, G. D. (2015). Etiologic, Phenomenologic, and Endophenotypic Overlap of Schizophrenia and Bipolar Disorder. *Annual Review of Clinical Psychology*, *11*(1), 251– 281.

Pedersen, C. B., & Mortensen, P. B. (2001). Evidence of a dose-response relationship between urbanicity during upbringing and schizophrenia risk. *Archives of General Psychiatry*, *58*(11), 1039–1046.

Perälä, J., Saarni, S. I., Ostamo, A., Pirkola, S., Haukka, J., Härkänen, T., … Suvisaari, J. (2008). Geographic variation and sociodemographic characteristics of psychotic disorders in Finland. *Schizophrenia Research*, *106*(2–3), 337–347.

Perälä, J., Suvisaari, J., Saarni, S. I., Kuoppasalmi, K., Isometsä, E., Pirkola, S., … Lönnqvist, J. (2007). Lifetime Prevalence of Psychotic and Bipolar I Disorders in a General Population. *Archives of General Psychiatry*, *64*(1), 19.

Perkins, D. O., Olde Loohuis, L., Barbee, J., Ford, J., Jeffries, C. D., Addington, J., ... & Mathalon, D. H. (2020). Polygenic risk score contribution to

psychosis prediction in a target population of persons at clinical high risk. *American Journal of Psychiatry*, *177*(2), 155-163.

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *British Medical Journal*, *316*(7139), 1236–1238.

Persson, E. (2016). *Programmatistic Conversion of PDF Tables*

Peters, M. D., Godfrey, C. M., Khalil, H., McInerney, P., Parker, D., & Soares, C. B. (2015). Guidance for conducting systematic scoping reviews. *International journal of evidence-based healthcare*, *13*(3), 141-146.

Pierson, A., Jouvent, R., Quintin, P., Perez-Diaz, F., & Leboyer, M. (2000). Information processing deficits in relatives of manic depressive patients. *Psychological Medicine*, *30*(3), 545–555.

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson Jr, R., ... & Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*(2), 127-152.

Pilowsky, L. S., Kerwin, R. W., & Murray, R. M. (1993). Schizophrenia: A Neurodevelopmental Perspective. *Neuropsychopharmacology*, *9*(1), 83–91.

Polich, J. (1992). On the correlation between P300 amplitude and latency. *Bulletin of the Psychonomic Society*, *30*(1), 5–8.

Polich, J. (1997). Meta-analysis of P300 normative aging studies. *Psychophysiology*, *33*(4), 334–353.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, *118*(10), 2128–2148.

Polich, J. (2011). Neurophychology of P300. In E. S. Kappenman & S. J. Luck (Eds.), *The Oxford Handbook of Event-Related Potential Components* (pp. 89–294).

Polich, J., Alexander, J. E., Bauer, L. O., Kuperman, S., Morzorati, S., O'Connor, S. J., … Begleiter, H. (1997). P300 topography of amplitude/latency correlations. *Brain Topography*, *9*(4), 275–282.

Polich, J., Howard, L., & Starr, A. (1983). P300 latency correlates with digit span. *Psychophysiology*, *20*(6), 665–669.

Polich, J., Howard, L., & Starr, A. (1985). Effects of Age on the P300 Component of the Event-related Potential From Auditory Stimuli: Peak Definition, Variation, and Measurement. *Journal of Gerontology*, *40*(6), 721–726.

Potash, J. B., Chiu, Y.-F., MacKinnon, D. F., Miller, E. B., Simpson, S. G., McMahon, F. J., … DePaulo, J. R. (2003). Familial aggregation of psychotic symptoms in a replication set of 69 bipolar disorder pedigrees. *American Journal of Medical Genetics*, *116B*(1), 90–97.

Prata, D. P., Costa-Neves, B., Cosme, G., & Vassos, E. (2019). Unravelling the genetic basis of schizophrenia and bipolar disorder with GWAS: A systematic review. *Journal of psychiatric research*, *114*, 178-207.

Price, G. W., Michie, P. T., Johnston, J., Innes-Brown, H., Kent, A., Clissa, P., & Jablensky, A. V. (2006). A Multivariate Electrophysiological Endophenotype, from a Unitary Cohort, Shows Greater Research Utility than Any Single Feature in the Western Australian Family Study of Schizophrenia. *Biological Psychiatry*, *60*(1), 1–10.

Priebe, L., Degenhardt, F., Strohmaier, J., Breuer, R., Herms, S., Witt, S. H., … Cichon, S. (2013). Copy number variants in German patients with schizophrenia. *PloS One*, *8*(7), e64035.

Psychiatric GWAS Consortium Bipolar Disorder Working Group. (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genetics*, *43*(10), 977–983.

Psychosis Endophenotypes International Consortium, P. E. I., Wellcome Trust Case-Control Consortium 2, W. T. C.-C. C., Bramon, E., Pirinen, M., Strange,

A., Lin, K., … Spencer, C. C. A. (2014). A genome-wide association analysis of a broad psychosis phenotype identifies three loci for further investigation. *Biological Psychiatry*, *75*(5), 386–397.

Purcell, S. M., Moran, J. L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., … Sklar, P. (2014). A polygenic burden of rare disruptive mutations in schizophrenia. *Nature, 506*(7487), 185.

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., & Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, *460*(7256), 748–752.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575.

Qiu, F., Xu, Y., Li, K., Li, Z., Liu, Y., DuanMu, H., … Li, X. (2012). CNVD: Text mining-based copy number variation in disease database. *Human Mutation*, *33*(11), e2375–e2381.

Qi, X., Wen, Y., Li, P., Liang, C., Cheng, B., Ma, M., ... & Zhang, F. (2020). An integrative analysis of genome-wide association study and regulatory SNP annotation datasets identified candidate genes for bipolar disorder. *International journal of bipolar disorders*, *8*(1), 1-7.

R Core Team. (2013). *R: A Language and Environment for Statistical Computing*.

Ramchurn, A., de Fockert, J. W., Mason, L., Darling, S., & Bunce, D. (2014). Intraindividual reaction time variability affects P300 amplitude rather than latency. *Frontiers in Human Neuroscience*, *8*, 557.

Ramos-Quiroga, J. A., Sánchez-Mora, C., Casas, M., Garcia-Martínez, I., Bosch, R., Nogueira, M., ... & Bayés, M. (2014). Genome-wide copy number

variation analysis in adult attention-deficit and hyperactivity disorder. *Journal of psychiatric research*, *49*, 60-67.

Ranlund, S., Adams, R. A., Díez, Á., Constante, M., Dutt, A., Hall, M.-H., … Bramon, E. (2016). Impaired prefrontal synaptic gain in people with psychosis and their relatives during the mismatch negativity. *Human Brain Mapping*, *37*(1), 351–365.

Ranlund, S., Calafato, S., Thygesen, J. H., Lin, K., Cahn, W., Crespo-Facorro, B., … Bramon, E. (2018). A polygenic risk score analysis of psychosis endophenotypes across brain functional, structural, and cognitive domains. *American Journal of Medical Genetics, Part B Neuropsychiatric Genetics*, *177*(1), 21–34.

Ranlund, S., Nottage, J., Shaikh, M., Dutt, A., Constante, M., Walshe, M., … Bramon, E. (2014). Resting EEG in psychosis and at-risk populations - A possible endophenotype? *Schizophrenia Research*, *153*(1–3), 96–102.

Rasetti, R., & Weinberger, D. R. (2011, June). Intermediate phenotypes in psychiatric disorders. *Current Opinion in Genetics and Development*, 21, 340–348.

Rasic, D., Hajek, T., Alda, M., & Uher, R. (2014). Risk of mental illness in offspring of parents with schizophrenia, bipolar disorder, and major depressive disorder: A meta-analysis of family high-risk studies. *Schizophrenia Bulletin*, *40*(1), 28–38.

Rees, E., Kendall, K., Pardiñas, A. F., Legge, S. E., Pocklington, A., Escott-Price, V., … Kirov, G. (2016). Analysis of Intellectual Disability Copy Number Variants for Association With Schizophrenia. *JAMA Psychiatry*, *73*(9), 963–969.

Rees, E., Walters, J. T. R., Georgieva, L., Isles, A. R., Chambert, K. D., Richards, A. L., … Kirov, G. (2014). Analysis of copy number variations at 15 schizophrenia-associated loci. *British Journal of Psychiatry*, *204*(02), 108–114.

Reichenberg, A., Caspi, A., Harrington, H., Houts, R., Keefe, R. S. E., Murray, R. M., … Moffitt, T. E. (2010). Static and dynamic cognitive deficits in

childhood preceding adult schizophrenia: a 30-year study. *The American Journal of Psychiatry*, *167*(2), 160–169.

Reichenberg, A., & Mollon, J. (2016, December 1). Challenges and opportunities in studies of cognition in the prodrome to psychosis: No detail is too small. *JAMA Psychiatry*, Vol. 73, pp. 1249–1250.

Regalado, A. (2018). *White-people-only DNA tests show how unequal science has become.* MIT Technology Review. https://www.technologyreview.com/s/ 612322/white-people-only-dna-tests-show-how-unequal-science-has-become/. Accessed 14 Nov 2018.

Rey, A. (1964). *L'Examen clinique en psychologie*. Paris: Presses Universitaires de France.

Rijsdijk, F. V., Gottesman, I. I., McGuffin, P., & Cardno, A. G. (2011). Heritability estimates for psychotic symptom dimensions in twins with psychotic disorders. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *156*(1), 89–98.

Ripke, S., Neale, B. M., Corvin, A., Walter, J. T. R., & Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, *511*(7510), 421–427.

Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J. L., Kähler, A. K., Akterin, S., … Sullivan, P. F. (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genetics*, *45*(10), 1150–1159.

Ripke, S., Sanders, A. R., Kendler, K. S., Levinson, D. F., Sklar, P., Holmans, P. A., … Gejman, P. V. (2011). Genome-wide association study identifies five new schizophrenia loci. *Nature Genetics*, *43*(10), 969–976.

Ripke, S., Walters, J. T., O'Donovan, M. C., & Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2020). Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *MedRxiv*.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*(1), 77.

Roiz-Santiañez, R., Suarez-Pinilla, P., & Crespo-Facorro, B. (2015). Brain Structural Effects of Antipsychotic Treatment in Schizophrenia: A Systematic Review. *Current Neuropharmacology*, *13*(4), 422–434.

Rolstad, S., Pålsson, E., Ekman, C. J., Eriksson, E., Sellgren, C., & Landén, M. (2015). Polymorphisms of dopamine pathway genes NRG 1 and LMX 1A are associated with cognitive performance in bipolar disorder. *Bipolar disorders*, *17*(8), 859-868.

Rothman, K. J. (1990). No adjustments are needed for multiple comparisons. *Epidemiology (Cambridge, Mass.)*, *1*(1), 43–46.

Rudelfer, D. M. (2013). Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Molecular Psychiatry*, *19*(9), 1017–1024.

Ruderfer, D. M., Fanous, A. H., Ripke, S., McQuillin, A., Amdur, R. L., Schizophrenia Working Group of the Psychiatric Genomics Consortium, … Kendler. (2014). Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Molecular Psychiatry*, *19*(9), 1017–1024.

Ruderfer, D. M., Ripke, S., McQuillin, A., Boocock, J., Stahl, E. A., Pavlides, J. M. W., … Kendler, K. S. (2018). Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell*, *173*(7), 1705–1715.

Rujescu, D., Ingason, A., Cichon, S., Pietiläinen, O. P. H., Barnes, M. R., Toulopoulou, T., … Collier, D. A. (2009). Disruption of the neurexin 1 gene is associated with schizophrenia. *Human Molecular Genetics*, *18*(5), 988–996.

Salvador-Oliván, J. A., Marco-Cuenca, G., & Arquero-Avilés, R. (2019). Errors in search strategies used in systematic reviews and their effects on

information retrieval. *Journal of the Medical Library Association: JMLA*, *107*(2), 210.

Sánchez-Morla, E. M., García-Jiménez, M. A., Barabash, A., Martínez-Vizcaíno, V., Mena, J., Cabranes-Díaz, J. A., … Santos, J. L. (2008). P50 sensory gating deficit is a common marker of vulnerability to bipolar disorder and schizophrenia. *Acta Psychiatrica Scandinavica*, *117*(4), 313–318.

Saperstein, A. M., Fuller, R. L., Avila, M. T., Adami, H., McMahon, R. P., Thaker, G. K., & Gold, J. M. (2006). Spatial Working Memory as a Cognitive Endophenotype of Schizophrenia: Assessing Risk for Pathophysiological Dysfunction. *Schizophrenia Bulletin*, *32*(3), 498–506.

Savitz, D. A., & Olshan, A. F. (1995). Multiple comparisons and related issues in the interpretation of epidemiologic data. *American Journal of Epidemiology*, *142*(9), 904–908.

Schalkwyk, L. C., Meaburn, E. L., Smith, R., Dempster, E. L., Jeffries, A. R., Davies, M. N., ... & Mill, J. (2010). Allelic skewing of DNA methylation is widespread across the genome. *The American Journal of Human Genetics*, *86*(2), 196-212.

Schnack, H. G., Hulshoff Pol, H. E., Baaré, W. F. C., Viergever, M. A., & Kahn, R. S. (2001). Automatic Segmentation of the Ventricular System from MR Images of the Human Brain. *NeuroImage*, *14*(1), 95–104.

Schnack, H. G., Hulshoff Pol HE, Baaré, W. F., Staal, W. G., Viergever, M. A., & Kahn, R. S. (2001). Automated separation of gray and white matter from MR images of the human brain. *NeuroImage*, *13*(1), 230–237.

Schulze, K. K., Hall, M. H. M.-H., Mcdonald, C., Marshall, N., Walshe, M., Murray, R. M., & Bramon, E. (2008). Auditory P300 in patients with bipolar disorder and their unaffected relatives. *Bipolar Disorders*, *10*(3), 377–386.

Schulze, K., MacCabe, J. H., Rabe-Hesketh, S., Crawford, T., Marshall, N., Zanelli, J., … McDonald, C. (2006). The relationship between eye movement and

brain structural abnormalities in patients with schizophrenia and their unaffected relatives. *Journal of Psychiatric Research*, *40*(7), 589–598.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., ... & Leotta, A. (2007). Strong association of de novo copy number mutations with autism. *Science*, *316*(5823), 445-449.

Seidman, L. J., Hellemann, G., Nuechterlein, K. H., Greenwood, T. A., Braff, D. L., Cadenhead, K. S., … Green, M. F. (2015). Factor structure and heritability of endophenotypes in schizophrenia: Findings from the Consortium on the Genetics of Schizophrenia (COGS-1). *Schizophrenia Research*, *163*(1–3), 73–79.

Seidman, L. J., Shapiro, D. I., Stone, W. S., Woodberry, K. A., Ronzio, A., Cornblatt, B. A., … Woods, S. W. (2016). Association of neurocognition with transition to psychosis: Baseline functioning in the second phase of the north American prodrome longitudinal study. *JAMA Psychiatry*, *73*(12), 1239–1248.

Semlitsch, H. V., Anderer, P., Schuster, P., & Presslich, O. (1986). A Solution for Reliable and Valid Reduction of Ocular Artifacts, Applied to the P300 ERP. *Psychophysiology*, *23*(6), 695–703.

Shaikh, M., Hall, M.-H., Schulze, K., Dutt, A., Li, K., Williams, I., … Bramon, E. (2013). Effect of DISC1 on the P300 waveform in psychosis. *Schizophrenia Bulletin*, *39*(1), 161–167.

Sharma, T., Lancaster, E., Lee, D., Lewis, S., Sigmundsson, T., Takei, N., … Murray, R. (1998). Brain changes in schizophrenia. Volumetric MRI study of families multiply affected with schizophrenia--the Maudsley Family Study 5. *The British Journal of Psychiatry: The Journal of Mental Science*, *173*, 132–138.

Sheffield, J. M., Gold, J. M., Strauss, M. E., Carter, C. S., MacDonald, A. W., Ragland, J. D., … Barch, D. M. (2014). Common and specific cognitive deficits in schizophrenia: relationships to function. *Cognitive, Affective & Behavioral Neuroscience*, *14*(1), 161–174.

Shelley, A. M., Ward, P. B., Catts, S. V., Michie, P. T., Andrews, S., & McConaghy, N. (1991). Mismatch negativity: an index of a preattentive processing deficit in schizophrenia. *Biological psychiatry*, *30*(10), 1059-1062.

Shenton, M. E., Dickey, C. C., Frumin, M., & McCarley, R. W. (2001). A review of MRI findings in schizophrenia. *Schizophrenia Research*, *49*(1–2), 1–52.

Shi, J., Levinson, D. F., Duan, J., Sanders, A. R., Zheng, Y., Pe'er, I., … Gejman, P. V. (2009). Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature*, *460*(7256), 753–757.

Simpson, N. H., Ceroni, F., Reader, R. H., Covill, L. E., Knight, J. C., Hennessy, E. R., ... & Fisher, S. E. (2015). Genome-wide analysis identifies a role for common copy number variants in specific language impairment. *European Journal of Human Genetics*, *23*(10), 1370-1377.

Sitskoorn, M. M., Aleman, A., Ebisch, S. J. H., Appels, M. C. M., & Kahn, R. S. (2004). Cognitive deficits in relatives of patients with schizophrenia: a meta-analysis. *Schizophrenia Research*, *71*(2–3), 285–295.

Sklar, P., Ripke, S., Scott, L. J., Andreassen, O. A., Cichon, S., Craddock, N., … Purcell, S. M. (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genetics*, *43*(10), 977–985.

Smoller, J. W., & Finn, C. T. (2003). Family, twin, and adoption studies of bipolar disorder. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, *123C*(1), 48– 58.

Smoller, J. W., Kendler, K. S., Craddock, N., Lee, P. H., Neale, B. M., Nurnberger, J. I., … et al. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *The Lancet*, *381*(9875), 1371–1379.

Snitz, B. E., Macdonald, A. W., & Carter, C. S. (2006). Cognitive deficits in unaffected first-degree relatives of schizophrenia patients: a meta-analytic review of putative endophenotypes. *Schizophrenia Bulletin*, *32*(1), 179–194.

So, H. C., & Sham, P. C. (2017). Exploring the predictive power of polygenic scores derived from genome-wide association studies: a study of 10 complex traits. *Bioinformatics*, *33*(6), 886-892.

Souza, V. B. N., Muir, W. J., Walker, M. T., Glabus, M. F., Roxborough, H. M., Sharp, C. W., … Blackwood, D. H. R. (1995). Auditory P300 event-related potentials and neuropsychological performance in schizophrenia and bipolar affective disorder. *Biological Psychiatry*, *37*(5), 300–310.

Speed, D., Cai, N., Johnson, M. R., Nejentsev, S., & Balding, D. J. (2017). Reevaluation of SNP heritability in complex human traits. *Nature Genetics*, *49*(7), 986–992.

Spitzer, R. L., Endicott, J., & Robins, E. (1978). Research Diagnostic Criteria. *Archives of General Psychiatry*, *35*(6), 773.

Spitzer, R. L., Williams, J. B., & Gibbon, M. (1992). The Structured Clinical Interview for DSM-III-R (SCID). *Archives of General Psychiatry*, *49*(8), 624–629.

Spitzer, R. L., Williams, J. W., Gibbon, M., & First, M. B. (1992). The Structured Clinical Interview for DSM-III-r (SCID): I: History, rationale, and description. *Archives of General Psychiatry*, *49*(8), 624–629.

Sriretnakumar, V., Zai, C. C., Wasim, S., Barsanti-Innes, B., Kennedy, J. L., & So, J. (2019). Copy number variant syndromes are frequent in schizophrenia: progressing towards a CNV-schizophrenia model. *BioRxiv*, 596213.

St Clair, D. (2009). Copy number variation and schizophrenia. *Schizophrenia bulletin*, *35*(1), 9-12.

Stahl, E. A., Breen, G., Forstner, A. J., McQuillin, A., Ripke, S., Trubetskoy, V., … Bipolar Disorder Working Group of the Psychiatric Genomics Consortium.

(2019). Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature Genetics*, *51*(5), 793–803.

Stankiewicz, P., & Lupski, J. R. (2010). Structural Variation in the Human Genome and its Role in Disease. *Annual Review of Medicine*, *61*(1), 437–455.

Steel, R. M., Whalley, H. C., Miller, P., Best, J. J. K., Johnstone, E. C., & Lawrie, S. M. (2002). Structural MRI of the brain in presumed carriers of genes for schizophrenia, their affected and unaffected siblings. *J. Neurol. Neurosurg. Psychiatry*, *72*(4), 455–458.

Steen, R. G., Mull, C., McClure, R., Hamer, R. M., & Lieberman, J. A. (2006). Brain volume in first-episode schizophrenia: systematic review and meta-analysis of magnetic resonance imaging studies. *The British Journal of Psychiatry , 188*(6), 510–518.

Stefansson, H., Meyer-Lindenberg, A., Steinberg, S., Magnusdottir, B., Morgen, K., Arnarsdottir, S., … Stefansson, K. (2014a). CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature*, *505*(7483), 361–366.

Stefansson, H., Ophoff, R. A., Steinberg, S., Andreassen, O. A., Cichon, S., Rujescu, D., … Myin-Germeys, I. (2009). Common variants conferring risk of schizophrenia. *Nature*, *460*(7256), 744–747.

Stefansson, H., Rujescu, D., Cichon, S., Pietiläinen, O. P. H., Ingason, A., Steinberg, S., … Myin-Germeys, I. (2008a). Large recurrent microdeletions associated with schizophrenia. *Nature*, *455*(7210), 232–236.

Stefansson, H., Sigurdsson, E., Steinthorsdottir, V., Bjornsdottir, S., Sigmundsson, T., Ghosh, S., … Stefansson, K. (2002). Neuregulin 1 and susceptibility to schizophrenia. *American Journal of Human Genetics*, *71*(4), 877–892.

Steinberg, S., Jong, S. De, Mattheisen, M., Costas, J., Demontis, D., Jamain, S., … Agartz, I. (2014). Common variant at 16p11. 2 conferring risk of psychosis. *Molecular Psychiatry*, *19*(1), 108–114.

Stone, J. L., O'Donovan, M. C., Gurling, H., Kirov, G. K., Blackwood, D. H. R., Corvin, A., … Ardlie, K. (2008). Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*, *455*(7210), 237–241.

Stone, J. L., O'Donovan, M. C., Gurling, H., Kirov, G. K., Blackwood, D. H. R., Corvin, A., … Sklar, P. (2008). Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*, *455*(7210), 237–241.

Stone, W. S., Giuliano, A. J., Tsuang, M. T., Braff, D. L., Cadenhead, K. S., Calkins, M. E., … Seidman, L. J. (2011). Group and site differences on the California Verbal Learning Test in persons with schizophrenia and their first-degree relatives: Findings from the Consortium on the Genetics of Schizophrenia (COGS). *Schizophrenia Research*, *128*(1–3), 102–110.

Stone, W. S., Mesholam-Gately, R. I., Braff, D. L., Calkins, M. E., Freedman, R., Green, M. F., Seidman, L. J. (2015). California Verbal Learning Test-II performance in schizophrenia as a function of ascertainment strategy: Comparing the first and second phases of the Consortium on the Genetics of Schizophrenia (COGS). *Schizophrenia Research*, *163*(1), 32–37.

Stone J.L., O'Donovan M.C., Gurling H., Kirov G.K., Blackwood D.H.R., Corvin A. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*, *455*(7210), 237.

Stranger, B. E., Stahl, E. A., & Raj, T. (2011). Progress and Promise of Genome-Wide Association Studies for Human Complex Trait Genetics. *Genetics*, *187*(2), 367–383.

Strasser, H. C., Lilyestrom, J., Ashby, E. R., Honeycutt, N. A., Schretlen, D. J., Pulver, A. E., … Pearlson, G. D. (2005). Hippocampal and ventricular volumes in psychotic and nonpsychotic bipolar patients compared with schizophrenia patients and community control subjects: a pilot study. *Biological Psychiatry*, *57*(6), 633–639.

Su, L., Cai, Y., Shi, S., & Wang, L. (2012). P300 latency and amplitude that occur in patients with schizophrenia during treatment with antipsychotic medication. *Shanghai Archives of Psychiatry*, *24*(4), 200–207.

Sudmant, P. H., Rausch, T., Gardner, E. J., Handsaker, R. E., Abyzov, A., Huddleston, J., … Korbel, J. O. (2015). An integrated map of structural variation in 2,504 human genomes. *Nature*, *526*(7571), 75–81.

Sullivan, P. F., Daly, M. J., & O'Donovan, M. (2012a). Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nature Reviews Genetics*, *13*(8), 537– 551.

Sullivan, P. F., Kendler, K. S., & Neale, M. C. (2003). Schizophrenia as a Complex Trait: evidence from a meta-analysis of twin studies. *Archives of General Psychiatry*, *60*(12), 1187.

Suvisaari, J. M., Taxell-Lassas, V., Pankakoski, M., Haukka, J. K., Lönnqvist, J. K., & Häkkinen, L. T. (2013). Obstetric complications as risk factors for schizophrenia spectrum psychoses in offspring of mothers with psychotic disorder. *Schizophrenia Bulletin*, *39*(5), 1056–1066.

Swaminathan, G. J., Bragin, E., Chatzimichali, E. A., Corpas, M., Bevan, A. P., Wright, C. F., ... & Firth, H. V. (2012). DECIPHER: web-based, community resource for clinical interpretation of rare variants in developmental disorders. *Human molecular genetics*, *21*(1), 37-44.

Szatkiewicz, J., Fromer, M., Nonneman, R., Ancalade, N., Johnson, J., Stahl, E., … Ruderfer, D. (2019). Characterization of single gene copy number variants in schizophrenia. *BioRxiv*, 550863.

Szatkiewicz, J. P., O'Dushlaine, C., Chen, G., Chambert, K., Moran, J. L., Neale, B. M., … Sullivan, P. F. (2014). Copy number variation in schizophrenia in Sweden. *Molecular Psychiatry*, *19*(7), 762–773.

Tanaka, K., Yamada, T., Torii, T., Furuta, K., Matsumoto, S., Yoshimura, T., … Murai, H. (2015). Pre-admission CHADS2, CHA2DS2-VASc, and R2CHADS2

scores on severity and functional outcome in acute ischemic stroke with atrial fibrillation. *Journal of Stroke and Cerebrovascular Diseases*, *24*(7), 1629–1635.

Tandon, R., Keshavan, M., & Narsallah, H. (2008). Schizophrenia, “Just the Facts” What we know in 2008. 2. Epidemiology and etiology. *Schizophrenia Research*, *102*(1–3), 1–18.

Taniguchi, S., Ninomiya, K., Kushima, I., Saito, T., Shimasaki, A., Sakusabe, T., ... & Ikeda, M. (2020). Polygenic risk scores in schizophrenia with clinically significant copy number variants. *Psychiatry and clinical neurosciences*, *74*(1), 35-39.

Tansey, K. E., Rees, E., Linden, D. E., Ripke, S., Chambert, K. D., Moran, J. L., ... & Owen, M. J. (2016). Common alleles contribute to schizophrenia in CNV carriers. *Molecular psychiatry*, *21*(8), 1085-1089.

Taylor, M., & Perera, U. (2015). NICE CG178 Psychosis and Schizophrenia in Adults: Treatment and Management - an evidence-based guideline? *The British Journal of Psychiatry*, *206*(5), 357–359.

Tesli, M., Espeseth, T., Bettella, F., Mattingsdal, M., Aas, M., Melle, I., … Andreassen, O. A. (2014). Polygenic risk score and the psychosis continuum model. *Acta Psychiatrica Scandinavica*, *130*(4), 311–317.

Thaker, G. K. (2008). Neurophysiological endophenotypes across bipolar and schizophrenia psychosis. *Schizophrenia Bulletin*, *34*(4), 760–773.

The international Schizophrenia Consortium. (2008). Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*, *455*, 237–241.

Tienari, P., Wynne, L. C., Läksy, K., Moring, J., Nieminen, P., Sorri, A., … Wahlberg, K.-E. (2003). Genetic Boundaries of the Schizophrenia Spectrum: Evidence From the Finnish Adoptive Family Study of Schizophrenia. *American Journal of Psychiatry*, *160*(9), 1587– 1594.

Tiihonen, J., Lönnqvist, J., Wahlbeck, K., Klaukka, T., Niskanen, L., Tanskanen, A., & Haukka, (2009). 11-year follow-up of mortality in patients with schizophrenia: a population-based cohort study (FIN11 study). *The Lancet*, *374*(9690), 620–627.

Tiihonen, J., Suokas, J. T., Suvisaari, J. M., Haukka, J., & Korhonen, P. (2012). Polypharmacy With Antipsychotics, Antidepressants, or Benzodiazepines and Mortality in Schizophrenia. *Archives of General Psychiatry*, *69*(5), 476.

Tochigi, M., Nishida, A., Shimodera, S., Okazaki, Y., & Sasaki, T. (2013). Season of birth effect on psychotic-like experiences in Japanese adolescents. *European Child & Adolescent Psychiatry*, *22*(2), 89–93.

Toomey, R., Faraone, S. V., Seidman, L. J., Kremen, W. S., Pepple, J. R., & Tsuang, M. T. (1998). Association of neuropsychological vulnerability markers in relatives of schizophrenic patients. *Schizophrenia Research*, *31*(2–3), 89–98.

Torrey, E. F., Buka, S., Cannon, T. D., Goldstein, J. M., Seidman, L. J., Liu, T., … Yolken, R. H. (2009). Paternal age as a risk factor for schizophrenia: How important is it? *Schizophrenia Research*, *114*(1–3), 1–5.

Toulopoulou, T., Goldberg, T. E., Mesa, I. R., Picchioni, M., Rijsdijk, F., Stahl, D., … Murray, R. M. (2010). Impaired intellect and memory: a missing link between genetic risk and schizophrenia? *Archives of General Psychiatry*, *67*(9), 905–913.

Trotta, A., Iyegbe, C., Di Forti, M., Sham, P. C., Campbell, D. D., Cherny, S. S., … Fisher, H. L. (2016). Interplay between Schizophrenia Polygenic Risk Score and Childhood Adversity in First-Presentation Psychotic Disorder: A Pilot Study. *PloS One*, *11*(9), e0163319.

Turetsky, B. I., Cannon, T. D., & Gur, R. E. (2000). P300 subcomponent abnormalities in schizophrenia: III. Deficits in unaffected siblings of schizophrenic probands. *Biological Psychiatry*, *47*(5), 380–390.

Turetsky, B. I., Dress, E. M., Braff, D. L., Calkins, M. E., Green, M. F., Greenwood, T. A., … Light, G. (2015). The utility of P300 as a schizophrenia endophenotype and predictive biomarker: Clinical and socio-demographic modulators in COGS-2. *Schizophrenia Research*, *163*(1–3), 53–62.

Üçok, A., Direk, N., Koyuncu, A., Keskin-Ergen, Y., Yüksel, Ç., Güler, J., … Devrim-Üçok, M. (2013). Cognitive deficits in clinical and familial high risk groups for psychosis are common as in first episode schizophrenia. *Schizophrenia Research*, *151*(1–3), 265–269.

Umesh, S., Nizamie, S. H., Goyal, N., Tikka, S., & Bose, S. (2016). Social anhedonia and gamma band abnormalities as a composite/multivariate endophenotype for schizophrenia: a dense array EEG study. *Early Intervention in Psychiatry*, *12*(3), 362-371

Vacic, V., McCarthy, S., Malhotra, D., Murray, F., Chou, H. H., Peoples, A., ... & Iakoucheva, L. M. (2011). Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. *Nature*, *471*(7339), 499-503.

Van der Loo, M. (2014). *The stringdist package for approximate string matching*.

Van Dinteren, R., Arns, M., Jongsma, M. L. A., & Kessels, R. P. C. (2014). P300 development across the lifespan: a systematic review and meta-analysis. *PLoS ONE*, *9*(2), e87347.

Van Erp, T. G. M., Hibar, D. P., Rasmussen, J. M., Glahn, D. C., Pearlson, G. D., Andreassen, O. A., … Turner, J. A. (2016). Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. *Molecular Psychiatry*, *21*(4), 547–553.

Van Haren, N. E., Cahn, W., Hulshoff Pol, H. E., & Kahn, R. S. (2013). Confounders of excessive brain volume loss in schizophrenia. *Neuroscience and Biobehavioral Reviews*, *37*(10 Pt 1), 2418–2423.

Van Os, J., Hanssen, M., Bijl, R. V., Vollebergh, W., FJM, F., J, S., … ME, S. (2001). Prevalence of Psychotic Disorder and Community Level of Psychotic Symptoms. *Archives of General Psychiatry*, *58*(7), 663.

Van Tricht, M. J., Nieman, D. H., Koelman, J. T., Mensink, A. J., Bour, L. J., Van der Meer, J. N., ... & De Haan, L. (2015). Sensory gating in subjects at ultra high risk for developing a psychosis before and after a first psychotic episode. *The world journal of biological psychiatry*, *16*(1), 12-21.

Van Winkel, R., Esquivel, G., Kenis, G., Wichers, M., Collip, D., Peerbooms, O., ... & Van Os, J. (2010). Genome-Wide Findings in Schizophrenia and the Role of Gene–Environment Interplay. *CNS neuroscience & therapeutics*, *16*(5).

Varese, F., Smeets, F., Drukker, M., Lieverse, R., Lataster, T., Viechtbauer, W., … Bentall, R. (2012). Childhood Adversities Increase the Risk of Psychosis: A Meta-analysis of Patient-Control, Prospective- and Cross-sectional Cohort Studies. *Schizophrenia Bulletin*, *38*(4), 661–671.

Vassos, E., Collier, D. A., Holden, S., Patch, C., Rujescu, D., St Clair, D., & Lewis, C. M. (2010). Penetrance for copy number variants associated with schizophrenia. *Human molecular genetics*, *19*(17), 3477-3481.

Vassos, E., Pedersen, C. B., Murray, R. M., Collier, D. A., & Lewis, C. M. (2012). Meta-analysis of the association of urbanicity with schizophrenia. *Schizophrenia bulletin*, *38*(6), 1118-1123.

Vassos, E., Di Forti, M., Coleman, J., Iyegbe, C., Prata, D., Euesden, J., … Breen, G. (2017). An Examination of Polygenic Score Risk Prediction in Individuals With First-Episode Psychosis. *Biological Psychiatry*, *81*(6), 470–477.

Vassos, E., Sham, P., Kempton, M., Trotta, A., Stilo, S. A., Gayer-Anderson, C., … Morgan, C. (2019). The Maudsley environmental risk score for psychosis. *Psychological Medicine*, 1– 8.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48.

Vita, A., De Peri, L., Silenzi, C., & Dieci, M. (2006). Brain morphology in first-episode schizophrenia: a meta-analysis of quantitative magnetic resonance imaging studies. *Schizophrenia research*, *82*(1), 75-88.

Voisey, J., Swagell, C. D., Hughes, I. P., Connor, J. P., Lawford, B. R., Young, R. M., & Morris, C. P. (2010). A polymorphism in the dysbindin gene (DTNBP1) associated with multiple psychiatric disorders including schizophrenia. *Behavioral and Brain Functions*, *6*(1), 41.

Vrijenhoek, T., Buizer-Voskamp, J. E., van der Stelt, I., Strengman, E., Genetic Risk and Outcome in Psychosis (GROUP) Consortium, C., Sabatti, C., … Veltman, J. A. (2008). Recurrent CNVs disrupt three candidate genes in schizophrenia patients. *American Journal of Human Genetics*, *83*(4), 504–510.

Wahlbeck, K., Forsén, T., Osmond, C., Barker, D. J. P., Eriksson, J. G., E, M., … RM, M. (2001). Association of Schizophrenia With Low Maternal Body Mass Index, Small Size at Birth, and Thinness During Childhood. *Archives of General Psychiatry*, *58*(1), 48.

Walhovd, K. B., & Fjell, A. M. (2003). The relationship between P3 and neuropsychological function in an adult life span sample. *Biological Psychology*, *62*(1), 65–87.

Walsh, T., McClellan, J. M., McCarthy, S. E., Addington, A. M., Pierce, S. B., Cooper, G. M., … Sebat, J. (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*, *320*(5875), 539–543.

Walsh, T., McClellan, J. M., McCarthy, S. E., Addington, A. M., Pierce, S. B., Cooper, G. M., … Sebat, J. (2008). Rare Structural Variants Disrupt Multiple Genes in Neurodevelopmental Pathways in Schizophrenia. *Science*, *320*(5875), 539–543.

Walters, J. T. R., Corvin, A., Owen, M. J., Williams, H., Dragovic, M., Quinn, E. M., … Article, O. (2010). Psychosis Susceptibility Gene ZNF804A and Cognitive Performance in Schizophrenia. *Archives of General Psychiatry*, *67*(7), 692–700.

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. F. A., … Bucan, M. (2007). PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Research*, *17*(11), 1665–1674.

Waters, F., Price, G., Dragović, M., & Jablensky, A. (2009). Electrophysiological brain activity and antisaccade performance in schizophrenia patients with first-rank (passivity) symptoms. *Psychiatry Research*, *170*(2–3), 140–149.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale - Revised Manual*. New York: Psychological Corporation.

Weickert, C. S., Straub, R. E., McClintock, B. W., Matsumoto, M., Hashimoto, R., Hyde, T. M., … Kleinman, J. E. (2004). Human Dysbindin (DTNBP1) Gene Expression inNormal Brain and in Schizophrenic Prefrontal Cortex and Midbrain. *Archives of General Psychiatry*, *61*(6), 544.

Weisbrod, M., Hill, H., Niethammer, R., & Sauer, H. (1999). Genetic influence on auditory information processing in schizophrenia: P300 in monozygotic twins. *Biological Psychiatry*, *46*(5), 721–725.

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.

Wellcome Trust Case Control Consortium, Craddock, N., Hurles, M. E., Cardin, N., Pearson, R. D., Plagnol, V., … Donnelly, P. (2010). Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature*, *464*(7289), 713–720.

Wender, P. H., Rosenthal, D., Kety, S. S., Schulsinger, F., & Welner, J. (1974). Crossfostering. A research strategy for clarifying the role of genetic and experiential factors in the etiology of schizophrenia. *Archives of General Psychiatry*, *30*(1), 121–128.

Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., … QUADAS-2 Group. (2011). QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Annals of Internal Medicine*, *155*(8), 529.

Wicham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Wickham, H., & Murray, R. M. (1997). Can biological markers identify endophenotypes predisposing to schizophrenia? *International Review of Psychiatry*, *9*(4), 355–364.

Wigginton, J. E., & Abecasis, G. R. (2005). PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics (Oxford, England)*, *21*(16), 3445–3447.

Wiles, N. J., Zammit, S., Bebbington, P., Singleton, N., Meltzer, H., & Lewis, G. (2006). Self-reported psychotic symptoms in the general population: results from the longitudinal study of the British National Psychiatric Morbidity Survey. *The British Journal of Psychiatry: The Journal of Mental Science*, *188*(6), 519–526.

Williams, N. M., Zaharieva, I., Martin, A., Langley, K., Mantripragada, K., Fossdal, R., ... & Gustafsson, O. (2010). Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. *The Lancet*, *376*(9750), 1401-1408.

Williams, H. J., Norton, N., Dwyer, S., Moskvina, V., Nikolov, I., Carroll, L., … O'Donovan, M. C. (2011). Fine mapping of ZNF804A and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder. *Molecular Psychiatry*, *16*(4), 429–441.

Williams, N. M., Preece, A., Spurlock, G., Norton, N., Williams, H. J., McCreadie, R. G., … O'Donovan, M. C. (2004). Support for RGS4 as a susceptibility gene for schizophrenia. *Biological Psychiatry*, *55*(2), 192–195.

Williams, N. M., Zaharieva, I., Martin, A., Langley, K., Mantripragada, K., Fossdal, R., … Thapar, A. (2010). Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. *Lancet (London, England)*, *376*(9750), 1401–1408.

Wing, J. K., Babor, T., Brugha, T., Burke, J., Cooper, J. E., Giel, R., … Sartorius, N. (1990). SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Archives of General Psychiatry*, *47*(6), 589–593.

Wing, J. K., Babor, T., Brugha, T., Cooper, J. E., Giel, R., & Sartorius, N. (1990). SCAN: Schedules fonr Clinical Assessment in Neuropsychiatry. *Archives of General Psychiatry*, *47*(6), 589–593.

Winterer, G., Coppola, R., Sc, D., Goldberg, T. E., Ph, D., Egan, M. F., … Weinberger, D. R. (2003). P300 and genetic risk for schizophrenia. *Archives of General Psychiatry*, *60*(11), 1158–1167.

Wittorf, A., Klingberg, S., & Wiedemann, G. (2004). Secondary verbal memory: a potential endophenotype of schizophrenia. *Journal of Psychiatric Research*, *38*(6), 601–612.

Wobrock, T., Gruber, O., Schneider-Axmann, T., Wölwer, W., Gaebel, W., Riesbeck, M., … Falkai, P. (2009). Internal capsule size associated with outcome in first-episode schizophrenia. *European Archives of Psychiatry and Clinical Neuroscience*, *259*(5), 278– 283.

Wolf, D. H., Satterthwaite, T. D., Calkins, M. E., Ruparel, K., Elliott, M. A., Hopson, R. D., … Gur, R. E. (2015). Functional Neuroimaging Abnormalities in Youth With Psychosis Spectrum Symptoms. *JAMA Psychiatry*, *72*(5), 456.

Wood, S. J., Pantelis, C., Proffitt, T., Phillips, L. J., Stuart, G. W., Buchanan, J. A., … McGorry, P. D. (2003). Spatial working memory ability is a marker of risk-for-psychosis. *Psychological Medicine*, *33*(7), 1239-1247.

Woodward, N. D., Purdon, S. E., Meltzer, H. Y., & Zald, D. H. (2005). A meta-analysis of neuropsychological change to clozapine, olanzapine, quetiapine, and risperidone in schizophrenia. *International Journal of Neuropsychopharmacology*, *8*(3), 457-472.

World Health Organisation. (1993). *International Classification of Diseases*. World Health Organization.

Wray, N. R., Lee, S. H., Mehta, D., Vinkhuyzen, A. A. E., Dudbridge, F., & Middeldorp, C. M. (2014). Research Review: Polygenic methods and their application to psychiatric traits. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *55*(10), 1068–1087.

Wray, N. R., Yang, J., Hayes, B. J., Price, A. L., Goddard, M. E., & Visscher, P. M. (2013). Pitfalls of predicting complex traits from SNPs. *Nature Reviews Genetics*, *14*(7), 507-515.

Wright, I. C., Rabe-Hesketh, S., Woodruff, P. W. R., David, A. S., Murray, R. M., & Bullmore, E. T. (2000). Meta-analysis of regional brain volumes in schizophrenia. *The American Journal of Psychiatry*, *157*(1), 16–25.

Xu, B., Roos, J. L., Levy, S., Van Rensburg, E. J., Gogos, J. A., & Karayiorgou, M. (2008). Strong association of de novo copy number mutations with sporadic schizophrenia. *Nature Genetics*, *40*(7), 880–885.

Yuan, J., Hu, J., Li, Z., Zhang, F., Zhou, D., & Jin, C. (2017). A replication study of schizophrenia-related rare copy number variations in a Han Southern Chinese population. *Hereditas*, *154*(1), 2.

Zarrei, M., MacDonald, J. R., Merico, D., & Scherer, S. W. (2015). A copy number variation map of the human genome. *Nature Reviews Genetics*, *16*(3), 172–183.

Zhang, C., Lu, W., Wang, Z., Ni, J., Zhang, J., Tang, W., & Fang, Y. (2016). A comprehensive analysis of NDST3 for schizophrenia and bipolar disorder in Han Chinese. *Translational psychiatry*, *6*(1), 701-701.

Zhang, J. P., Robinson, D., Yu, J., Gallego, J., Fleischhacker, W. W., Kahn, R. S., … Lencz, T. (2019). Schizophrenia polygenic risk score as a predictor of antipsychotic efficacy in first-episode psychosis. *American Journal of Psychiatry*, *176*(1), 21–28.

Zhao, Q., Li, T., Zhao, X., Huang, K., Wang, T., Li, Z., ... & Feng, G. (2013). Rare CNVs and tag SNPs at 15q11. 2 are associated with schizophrenia in the Han Chinese population. *Schizophrenia bulletin*, *39*(3), 712-719.

Zhou, Z., Zhu, H., & Chen, L. (2013). Effect of AripiprazoLe on Mismatch Negativity (MMN) in Schizophrenia. *PLoS ONE*, *8*(1).

Ziats, M. N., Goin-Kochel, R. P., Berry, L. N., Ali, M., Ge, J., Guffey, D., ... & Penney, L. S. (2016). The complex behavioral phenotype of 15q13. 3 microdeletion syndrome. *Genetics in Medicine*, *18*(11), 1111-1118.

Ziyatdinov, A., Vázquez-Santiago, M., Brunel, H., Martinez-Perez, A., Aschard, H., & Soria, J. M. (2018). lme4qtl: linear mixed models with flexible covariance structure for genetic studies of related individuals. *BMC Bioinformatics*, *19*(1), 68.

Appendices

Appendix 1. Supplementary material for chapter 2

Table S1 Columns included in the excel sheet labelled “CNVs_formated”

Column name	Description
Marker †	Marker name usually the loci where the CNV is located followed by del or.dup specifying if the CNV is a deletion or a duplication (i.e. 15q11.2.del)
alternative_name	If any other name(s) are given in the paper for this CNV they should be included here. Multiple names are accepted given as; separated (i.e. nrxn1.del; 2p16.del).
build1 †	Genomic builds that the paper coordinates align to: hg16, hg17, hg18, hg19, hg38
Chr †	The chromosomal position of the CNV: 1-22, x or y
start_bp1 † **A**	The start position of the CNV in base pairs
stop_bp1 † **A**	The stop position of the CNV in base pairs
position_note	If the position is not directly stated in the table I state here where in the paper the position was obtained from
gene_name	Name of the suggested causative gene(s) in the region. Multiple gene-names can be given separated by; (i.e..DLG1; PAK2)
gene_note	Suggested mechanism or notes about the gene(s) from the paper. Multiple notes can be given separated by; and will be matched with multiple gene names (i.e. DLG1 note; PAK2 note)

freq_pheno1 **B**	Phenotype name for which n_carriers_pheno and sample_size_pheno are given. I refer to existing phenotypes in the database and use similar spelling
n_carriers_pheno1 **B**	Number of CNV carriers with phenotype
sample_size_pheno1 **B**	Total sample size with phenotype1 (including CNV carriers)
assoc_pheno1 **C**	Phenotype name for which association is reported. I refer to existing phenotypes in the database and use similar spelling
quantitative_n1 **C**	If assoc_pheno is quantitative the number of carriers tested is indicated here, otherwise if the phenotype is case control this is either left blank or with NA
p_value1 **C**	P-value given for association with assoc_pheno
effect_size1 **C**	Odds-ratio or effect size given for assoc_pheno
ci_95_low1 **C**	Lower 95 confidence interval given for association with phenotype
ci_95_high1 **C**	Higher 95 confidence interval given for association with phenotype
assoc_note1 **C**	Notes specifying the specific association model used (i.e. Cochran-Mantel-Haenszel), if p-values have been adjusted and the format of the effect size if not given as odds ratio.

† *Essential/obligatory these columns must always be included*

A, B, C *Multiple entries allowed when marked by ascending numbers (letter indicate groups), where all marked group columns are expected for all sets of entries. (i.e. build1, start.bp1, stop.bp1, build2, start.bp2, stop.bp2 etc.)*

Table S2 Columns included in the excel sheet labelled “References”

Column name	Description
reference †	Short reference in the form of LastnameYear (i.e.Thygesen2016)
table_name †	Table name in the form of table1 or S-table1 if it is a supplement table
title †	Full title of the paper
year †	Publication year
genotyping_method†	Genotyping method employed: snp-assay, array-cgh etc
type †	Type of publication: population, multi-sample, family-based, case-report etc
pmid†	PMID as found on Pubmed. If PMID is not available DOI
sample	Name of the sample used as stated in the article. Multiple names can be given as; separated (i.e. CLOZUK; CLOZUK1)
age_mean	Mean age of sample. If the sample is case control multiple mean ages can be given as: phenotype1(43.5); phenotype2(55.5). If mean age is not given a range can also be specified as: (40-55).
age_sd	Standard deviation of sample age. If the sample is case control multiple standard deviations be given as: phenotype1(2.5); phenotype2(3.5)
male_female_ratio	Sample male to female ratio. If the sample is case control multiple male to female ratios can be given as: phenotype1(0.6); phenotype2(0.4)
ethnicity †	Ethnicity of the samples examined. Multiple is an accepted answer if unknown please state unknown

ref_note	Can be left blank if no notable info is to be included
related_reference	Reference (short references as described above) of table/paper already included in database where sample is also used or in other way related.
related_samples_pmid	If related sample paper is not included in the database give PMID of papers using the same samples to help identify how samples and papers are linked. Multiple PMID are accepted given as; separated (e.g 22424231; 19675094). If the PMID is unknown or unavailable the title of the related article is stated in the next column.
related_samples_title	Title of articles of related samples should only be given if PMID is unknown or cannot be given. Multiple titles can be given as; separated (i.e. title1; title2)
relation_type	One of the following four relation types: subset of sample included here, same sample used here, this sample is a smaller subset of a larger sample used in the related article(s), other

---

† *Essential/obligatory these columna must always be included*

A, B, C *Multiple entries allowed when marked by ascending numbers (letter indicate groups), where all marked group columns are expected for all sets of entries. (i.e. build1, start.bp1, stop.bp1, build2, start.bp2, stop.bp2 etc.)*

Table S3 Table presenting the CNVs included in the analyses of variant size against the effect size

Marker	CNV type	CNV size (kb)	Odds ratio	Reference
2q11.2	deletion	935107	9.3	Rees, 2016
2q13	deletion	618609	3.6	Rees, 2016
2q13	duplication	618609	1.7	Rees, 2016
3p11.2	deletion	264019	2.4	Rees, 2016
3q28-29	deletion	266284	4.1	Rees, 2016
3q29	deletion	1634659	18	Rees, 2016
WBS	duplication	1397977	5.2	Rees, 2016
TAR	deletion	412862	1.2	Rees, 2016
15q11.2	deletion	289217	1.8	Rees, 2016
15q13.3	deletion	1382131	4.6	Rees, 2016
TAR	duplication	412862	1.9	Rees, 2016
16p13.11	deletion	782034	1.1	Rees, 2016
16p13.11	duplication	782034	1.7	Rees, 2016
16p12.1	deletion	481754	3.3	Rees, 2016
16p11.2	deletion	223587	1.7	Rees, 2016
16p11.2	duplication	223587	1.2	Rees, 2016
16p11.2	deletion	549933	0.61	Rees, 2016
16p11.2	duplication	549933	11	Rees, 2016
1q21.1	deletion	866457	6.81	Rees, 2016
17q11.2	deletion	1157584	0.2	Rees, 2016
17q12	duplication	1401528	2.2	Rees, 2016
22q11.2	duplication	2429394	0.2	Rees, 2016
1q21.1	duplication	866457	2.3	Rees, 2016
22q11.2	deletion	1733519	0	Rees, 2016
NRXN1	deletion	1114031	4.5	Rees, 2016
2p15-16.1	duplication	169284	0.37	Rees, 2016
17p11.2	deletion	1270000	1.89	Green, 2016
VIPR2	duplication	120000	0.36	Green, 2016

17p11.2	deletion	1270000	3.62	Rees, 2014 (a)
VIPR2	duplication	120000	1.54	Rees, 2014 (a)
3q29	deletion	1610000	57.65	Rees, 2014 (a)
WBS	duplication	1400000	11.35	Rees, 2014 (a)
PWS/AS	duplication	3610000	13.2	Rees, 2014 (a)
15q11.2	deletion	290000	2.15	Rees, 2014 (a)
15q13.3	deletion	1350000	7.52	Rees, 2014 (a)
16p13.11	duplication	790000	2.3	Rees, 2014 (a)
16p11.2	deletion	230000	3.39	Rees, 2014 (a)
16p11.2	duplication	560000	11.52	Rees, 2014 (a)
1q21.1	deletion	820000	8.35	Rees, 2014 (a)
17q12	deletion	1390000	6.64	Rees, 2014 (a)
1q21.1	duplication	820000	3.45	Rees, 2014 (a)
NRXN1	deletion	1110000	9.01	Rees, 2014 (a)
Xq28	duplication	425000	8.9	Marshall, 2016
13q12.11	duplication	26180	0.36	Marshall, 2016
Xq28	duplication	5243	0.35	Marshall, 2016
8q22.2(VPS13B)	deletion	864314	14.5	Marshall, 2016
22q11.2	deletion	2350000	67.7	Marshall, 2016
22q11.2	duplication	2350000	0.15	Marshall, 2016
16p12.2-p11.2	duplication	800000	0.63	Szatkiewicz, 2014
VIPR2	duplication	100000	0.25	Szatkiewicz, 2014
3q29	deletion	1600000	16.32	Szatkiewicz, 2014
WBS	duplication	1400000	6.27	Szatkiewicz, 2014
16p13.11	deletion	900000	0.94	Szatkiewicz, 2014
16p13.11	duplication	900000	2.51	Szatkiewicz, 2014
16p11.2	duplication	700000	6.28	Szatkiewicz, 2014
1q21.1	deletion	3000000	6.27	Szatkiewicz, 2014
17q12	duplication	1400000	6.27	Szatkiewicz, 2014
22q11.2	deletion	3100000	16.32	Szatkiewicz, 2014
22q11.2	duplication	3100000	0.18	Szatkiewicz, 2014
1q21.1	duplication	3000000	2.51	Szatkiewicz, 2014

3q29	duplication	1600000	3.76	Szatkiewicz, 2014
NRXN1	deletion	1100000	6.27	Szatkiewicz, 2014
15q11q13	deletion	2600000	4.39	Szatkiewicz, 2014
15q11q13	duplication	2600000	0.63	Szatkiewicz, 2014
16p12.2-p11.2	deletion	800000	2.51	Szatkiewicz, 2014
4q35.1	duplication	90000	4.12	Rees, 2014 (b)
1p36	duplication	260000	8.66	Rees, 2014 (b)
6p24.2	duplication	150000	4.12	Rees, 2014 (b)
15q21.3	duplication	170000	1.71	Rees, 2014 (b)
4q35.2	deletion	60000	2.97	Rees, 2014 (b)
16p12.1	deletion	480000	2.72	Rees, 2014 (b)
2q37.3	duplication	120000	1.43	Rees, 2014 (b)
4q25	duplication	150000	8.66	Rees, 2014 (b)
5q33.1	deletion	140000	11.14	Rees, 2014 (b)
9p24.2	deletion	100000	6.19	Rees, 2014 (b)
3q29	deletion	1400000	49.5	Grozeva, 2011
15q11.2	deletion	600000	2.2	Grozeva, 2011
15q13.3	deletion	1600000	8.3	Grozeva, 2011
16p13.11	duplication	1200000	2.1	Grozeva, 2011
1q21.1	deletion	1400000	9.2	Grozeva, 2011
17q12	deletion	1500000	18.4	Grozeva, 2011
17p12	deletion	1400000	5.9	Grozeva, 2011
17p11.2	deletion	1400000	0.92	Li, 2016
1p36	duplication	1400000	6.73	Li, 2016
7q36.3	duplication	800000	5.5	Li, 2016
PWS/AS	deletion	1100000	1.48	Li, 2016
16p13.11	duplication	1900000	1.26	Li, 2016
16p11.2	deletion	300000	1.83	Li, 2016
16p11.2	duplication	2000000	7.96	Li, 2016
1q21.1	deletion	3000000	8.57	Li, 2016
22q11.2	deletion	3200000	11.01	Li, 2016
22q11.2	duplication	3200000	0.3	Li, 2016

1q21.1	duplication	4900000	1.37	Li, 2016
15q11q13	deletion	1500000	1.83	Li, 2016
7q36.3	duplication	120000	1	Priebe, 2013
15q11.2	deletion	470000	0.66	Priebe, 2013
16p13.11	deletion	810000	0.99	Priebe, 2013
16p13.11	duplication	810000	1.99	Priebe, 2013
2p16.3	duplication	1600000	1	Priebe, 2013
16p11.2	duplication	640000	1	Priebe, 2013
1q21.1	deletion	1350000	2.98	Priebe, 2013
22q11.2	deletion	2800000	1.99	Priebe, 2013
NRXN1	deletion	1600000	1.99	Priebe, 2013

Figure S1 Database schema for CNVcatalog.

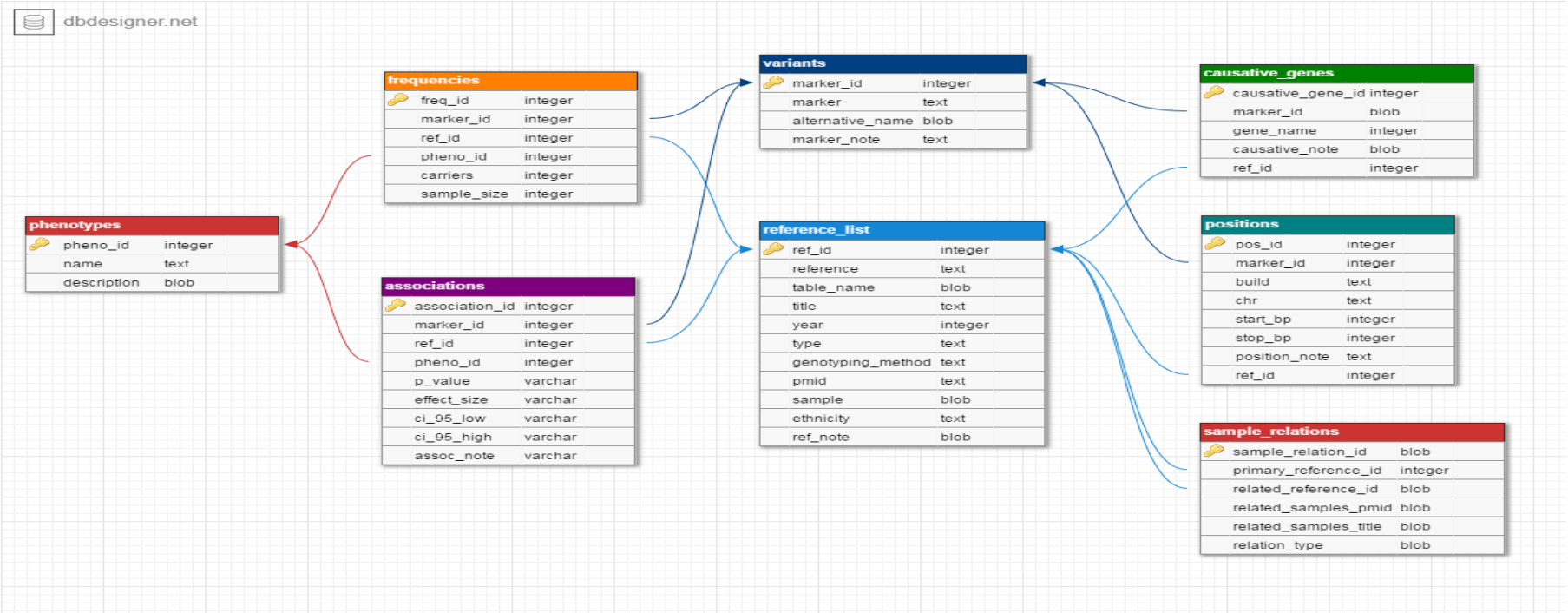


Figure demonstrating the eight linked tables used by CNV catalog to store the data. They describe: the phenotypes of interest, their frequencies, the CNV name, the genomic position, the associated genes, association results, inter-study relations and a reference list.

Appendix 2. Supplementary material for chapter 3

Table S4 Study sites and sample sizes.

Affiliation	City	Country	Number of participants				Endophenotypes contributed
			Total	C	R	P	
The University of Western Australia	Perth	Australia	893	224	260	409	P300, RAVLT
Heidelberg University	Heidelberg	Germany	78	23	19	36	P300, LVV
Ludwig-Maximilians, University of Munich	Munich	Germany	2185	2185	-	-	Block Design, Digit Span
GROUP consortium: University of Amsterdam, University of Groningen, Maastricht University, University of Utrecht	Amsterdam, Groningen, Maastricht, Utrecht	Holland	2993	1484	722	787	Block Design, RAVLT, LVV
Fundacion Argibide, Pamplona	Pamplona	Spain	69	-	-	69	Digit Span, RAVLT
Universidad de Cantabria, Santander	Santander	Spain	630	359	-	271	LVV, Digit Span, RAVLT
University of Edinburgh	Edinburgh	United Kingdom	160	87	-	73	LVV, Block Design, Digit Span

Institute of Psychiatry, King's College London	London	United Kingdom	1746	693	486	567	P300, LVV, Block Design, Digit Span, RAVLT

C = controls; R = relatives, P = patients; LVV = lateral ventricular volume; RAVLT = Ray Auditory Verbal Learning Task

Table S5 Table describing the MRI acquisition and processing methods used on each site.

Site	MRI methods
Holland (Maastricht)	Scanner used: 3 T Siemens (Erlangen, Germany). Acquisition sequence: Either a modified driven equilibrium Fourier transform (MDEFT), or a magnetization prepared rapid acquisition gradient echo (MPRAGE). Acquisition protocol either; i) Flip angle = 15°, TR = 7.92 ms, TE = 2.4 ms, or ii) Flip angle = 9°, TR = 2250 ms, TE = 2.6 ms. Images were analysed using Freesurfer. Automatic labelling of each MRI voxel was carried out based on probabilistic information derived from training on a manually labelled dataset (Fischl et al., 2002). For full details see (Collip et al., 2013; Habets et al., 2011).
Perth	Scanner used: 1 T Siemens Magnetom (Erlangen, Germany). Acquisition sequence: Magnetisation prepared rapid acquisition gradient echo (MPRAGE). Acquisition protocol: Flip angle = 12°, repetition time (TR) = 10 ms, echo time (TE) = 4 ms.
Germany (Heidelberg)	Scanner used: 1.5 T (Tesla) Phillips. Acquisition sequence: Magnetisation prepared rapid acquisition gradient echo (MPRAGE). Acquisition protocol:

	Flip angle = 15°, TR = 11.4 ms, TE = 4.4 ms. Images were analysed using a region of interest tool in the software Analyze, and lateral ventricular volume was defined according to borders described in the literature (Shenton et al., 2001). For full details see (Wobrock et al., 2009).
Munich	Scanner used: 3 T Siemens (Erlangen, Germany). Acquisition sequence: Spoiled gradient-recalled acquisition in the steady state (GRASS) (SPGR). Acquisition protocol either: i) Flip angle = 15°, TR = 7.92 ms, TE = 2.4 ms, or ii) Flip angle = 9°, TR = 2250 ms, TE = 2.6 ms.
Holland (Utrecht)	Scanner used: 1.5 T Philips NT. Acquisition sequence: Fast field echo (FFE). Acquisition protocol: Flip angle = 30°, TR = 30 ms, TE = 4.6 ms. Images were analysed using a Histogram method validated previously by the research group (Schnack, Hulshoff Pol, Baaré, Viergever, & Kahn, 2001). For full details see (Hulshoff Pol et al., 2002; Schnack, Hulshoff Pol HE, et al., 2001).
Spain (Santander and Pamplona)	Scanner used: 1.5 T General Electric Signa System (GE Medical Systems, Milwaukee, WI). Acquisition sequence: Spoiled gradient-recalled acquisition in

	the steady state (GRASS) (SPGR). Acquisition protocol: Flip angle = 45°, TR = 24 ms, TE = 5 ms. Images were analysed using the software BRAINS2, including automatic measurements of brain areas. For full details see (Crespo-Facorro et al., 2009; Mata et al., 2009)
United Kingdom (Edinburgh)	Scanner used: 1 T Siemens Magnetom (Erlangen, Germany). Acquisition sequence: Magnetisation prepared rapid acquisition gradient echo (MPRAGE). Acquisition protocol: Flip angle = 12°, repetition time (TR) = 10 ms, echo time (TE) = 4 ms. Images were analysed using a regions of interest analysis using the semi-automated programme Analyze, and lateral ventricular volume was defined by the autotrace and included frontal, occipital and temporal horns. For full details see (McIntosh, Harrison, et al., 2005; McIntosh et al., 2004; McIntosh, Job, et al., 2005).
United Kingdom (London)	Scanner used: 1.5 T General Electric (USA) Signa System. Acquisition sequence: Spoiled gradient recall (SPGR) echo. One of the following acquisition protocols was used: Flip angle = 35°, TR = 35 ms, TE = 5 ms; Flip angle = 20°, TR = 14.7 ms, TE =

	3.7 ms; Flip angle = 20°, TR = 9.8 ms, TE = 2.3 ms; or Flip angle = 20°, TR = 13.1 ms, TE = 5.8 ms. Images were analysed using MEASURE, an image analysis program that uses stereologically unbiased estimation of volume. Lateral ventricular volume included the body, frontal, occipital and temporal horns, and choroid plexus where visible. For full details see (Dutt et al., 2009; Frangou et al., 1997; McDonald et al., 2002, 2006; K. Schulze et al., 2006).

Table S6 Family sizes.

Number of family members participating	Number of families	% of families	Number of individuals	% of total sample
1	5545	84.00%	5545	63.34%
2	456	6.91%	912	10.42%
3	306	4.64%	918	10.49%
4	214	3.24%	856	9.78%
5	49	0.74%	245	2.80%
6	17	0.26%	102	1.17%
7	10	0.15%	70	0.80%
8	2	0.03%	16	0.18%
9	1	0.02%	9	0.11%
11	1	0.02%	11	0.13%

Table S7 Group interactions on associations between endophenotypes.

Endophenotype relationship	**Controls** Standardised increase in association (95% CI)	**Relatives** Est. difference from controls (95% CI)	**Patients** Est. difference from controls (95% CI)	**Overall test of interaction effect**
Digit Span x Block Design N=2754	0.31 (0.27 to 0.34) p < 0.001	0.18 (0.02 to 0.35) p = 0.028	0.28 (0.19 to 0.38) p < 0.001	p < 0.001
RAVLT del x Block Design N=2137	0.21 (0.15 to 0.26) p < 0.001	-0.04 (-0.14 to 0.05) p = 0.390	0.19 (0.09 to 0.29) p < 0.001	p < 0.001
RAVLT imm x Block Design	0.24 (0.18 to 0.29) p < 0.001	-0.02 (-0.14 to 0.06) p = 0.427	0.12 (0.02 to 0.23) p = 0.018	p = 0.010

Regressions on standardised scores including interactions terms between group (patient, relative, controls) and predictor, adjusted for covariates (age, gender and study site), using robust standard errors to account for correlations within families. Shown for controls are the regression coefficients for the associations between the two cognitive tasks, and shown for relatives and patients are the changes in slope from that of controls. RAVLT del = Rey Auditory Verbal Learning Task delayed recall; CI = Confidence Interval.

Table S8 Comparison between full models[1] (in the chapter, including age, sex and group) and models excluding age and sex[2].

This table shows that despite imbalances in demographic variables across the clinical groups, the full and reduced models are stable and there is no collinearity between clinical group and demographic variables.

	Total Sample	**Patients – Controls**	**Patients – Relatives**	**Relatives – Controls**
Endophenotype:	Global p-value*	Mean difference (95% CI)	Mean difference (95% CI)	Mean difference (95% CI)
P300 amplitude[1]	< 0.001	-0.50 (-0.71 to -0.29) p < 0.001	-0.16 (-0.32 to -0.01) p = 0.061	-0.34 (-0.54 to -0.14) p = 0.001
P300 amplitude[2]	< 0.001	-0.57 (-0.79 to -0.36) p < 0.001	-0.14 (-0.30 to -0.02) p = 0.091	-0.44 (-0.63 to -0.25) p < 0.001
P300 latency[1]	< 0.001	0.47 (0.33 to 0.61) p < 0.001	0.03 (-0.14 to 0.19) p = 0.749	0.44 (0.29 to 0.60) p < 0.001

P300 latency[2]	< 0.001	0.43 (0.29 to 0.58) p < 0.001	-0.17 (-0.34 to 0.02) p = 0.030	0.61 (0.46 to 0.75) p < 0.001
Lateral Ventricular Volume[1]	= 0.145	0.20 (0.08 to 0.32)	0.09 (-0.06 to 0.23)	0.11 (-0.04 to 0.25)
Lateral Ventricular Volume[2]	= 0.056	0.27 (0.16 to 0.37)	0.06 (-0.08 to 0.20)	0.11 (-0.04 to 0.25)
Digit Span[1]	< 0.001	-0.72 (-0.88 to -0.55) p < 0.001	-0.14 (-0.32 to 0.05) p = 0.141	-0.58 (-0.77 to -0.39) p < 0.001
Digit Span[2]	< 0.001	-0.72 (-0.88 to -0.55) p < 0.001	-0.04 (-0.22 to 0.13) p = 0.627	-0.67 (-0.86 to -0.49) p < 0.001
Block Design[1]	< 0.001	-0.91 (-1.07 to -0.75) p < 0.001	-0.08 (-0.21 to 0.04) p = 0.190	-0.83 (-0.97 to -0.69) p < 0.001

Block Design[2]	< 0.001	-0.88 (-1.03 to -0.73) p < 0.001	0.22 (0.11 to 0.34) p < 0.001	-1.11 (-1.24 to -0.98) p < 0.001
RAVLT immediate recall[1]	< 0.001	-1.32 (-2.29 to -0.37) p = 0.007	-1.24 (-2.22 to -0.27) p = 0.012	-0.08 (-0.24 to 0.07) p = 0.286
RAVLT immediate recall[2]	< 0.001	-1.40 (-2.14 to -0.66) p < 0.001	-1.21 (-1.98 to -0.46) p = 0.002	-0.18 (-0.36 to -0.01) p = 0.041
RAVLT delayed recall[1]	< 0.001	-0.98 (-2.21 to 0.25) p =0.118	-0.94 (-2.18 to 0.30) p =0.136	-0.03 (-0.20 to 0.13) p = 0.669
RAVLT delayed recall[2]	< 0.001	-1.07 (-2.05 to -0.09) p =0.033	-0.96 (-1.95 to 0.04) p =0.059	-0.11 (-0.29 to 0.65) p = 0.221

All the regression models are conducted on standardised scores for each endophenotype. All models are adjusted for study site and use robust standard errors to account for correlations within families.

[1] Full models (reported in chapter 3) include clinical group, age, sex, study site and where significant a group by study site interaction term.

[2] Reduced models include the same variables as above except for age and sex.

* P-value for the overall test of a group effect. Note that p-values were not produced for the models that include lateral ventricular volume since I used bootstrapping, which is a percentile based method; therefore, I looked at the bias-corrected confidence intervals to check for significance.

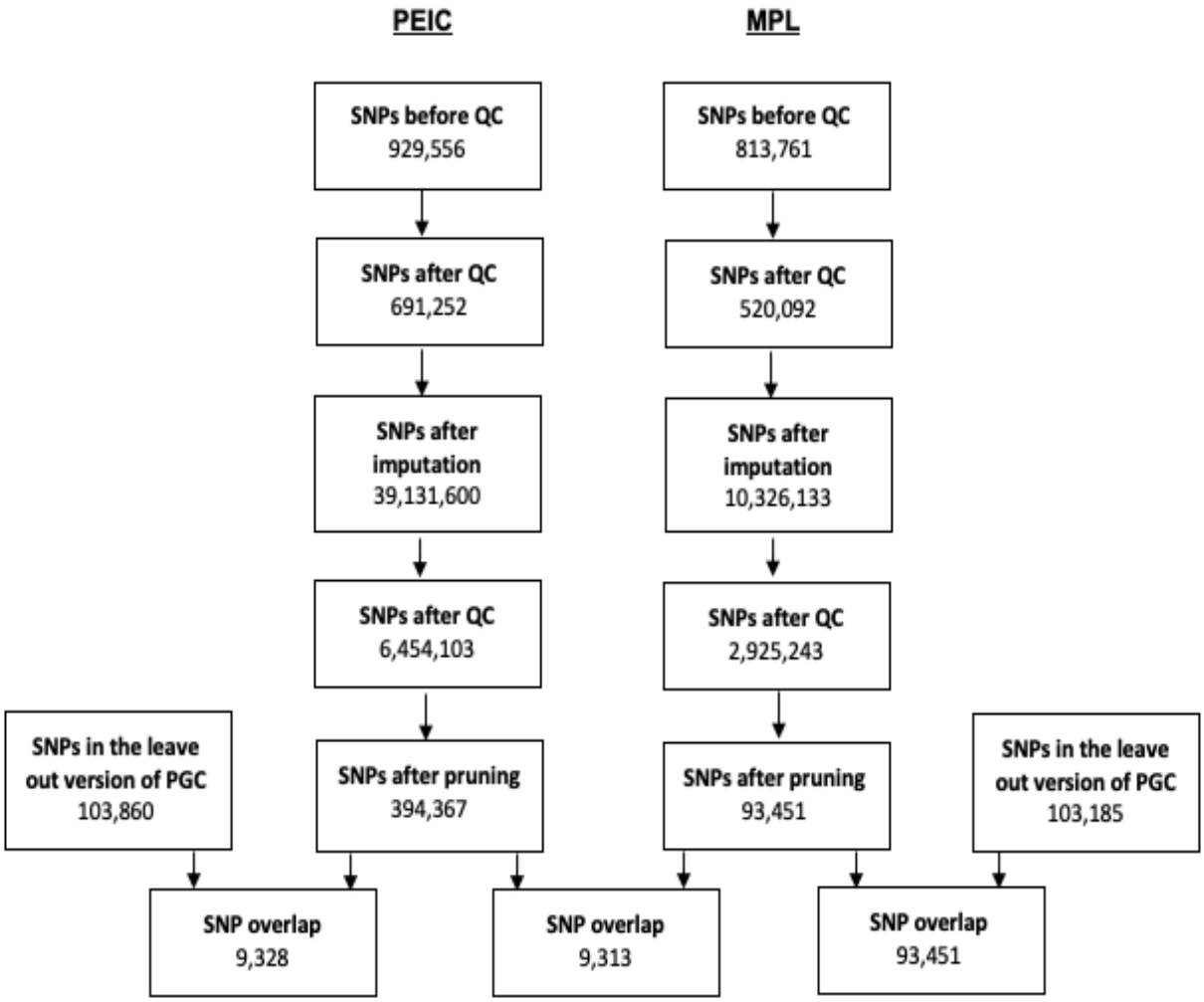
RAVLT = Rey Auditory Verbal Learning Task; CI = Confidence Interval.

Appendix 3. Supplementary material for Chapter 4

Table S9 Samples per research centre in PEIC dataset

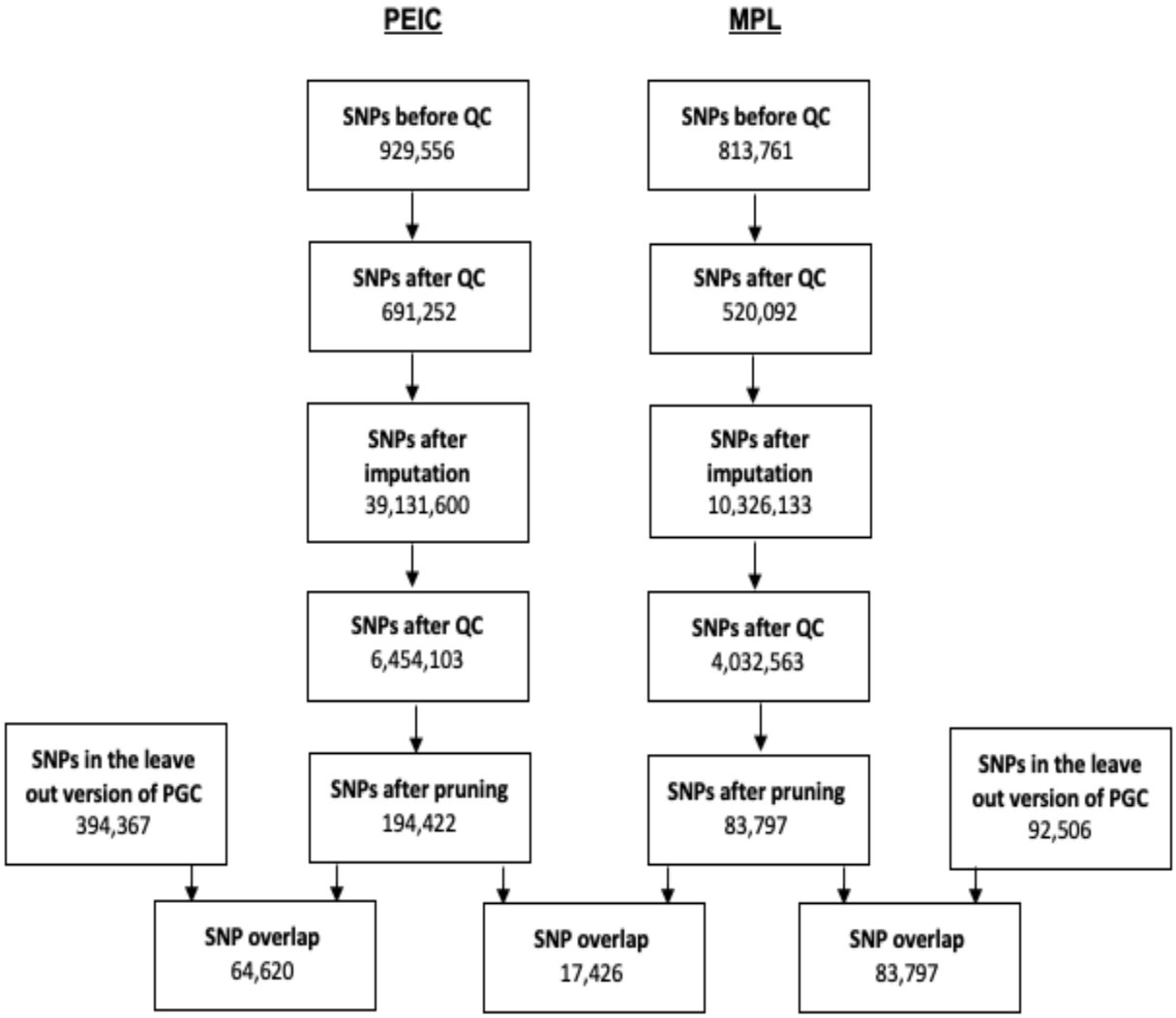
Site		Controls	Cases	Total
London	Institute of Psychiatry – King's College London	228 *(13.19%)*	170 *(24.25%)*	398
Edinburgh	University of Edinburgh	16 *(0.92%)*	29 *(4.13%)*	45
Holland	GROUP Consortium (Universities of Amsterdam, Groningen, Maastricht and Utrecht)	584 *(33.79%)*	245 *(34.95%)*	829
Perth	University of Western Australia	107 *(6.19%)*	206 *(29.38%)*	313
Munich	Ludwig-Maximilians-Universität München	779 *(45.08%)*	0	779
Pamplona	Universidad de Cantabria	0	32 *(4.56%)*	32
Heidelberg	Heidelberg University	14 *(0.81%)*	19 *(2.71%)*	33
Total		1,728	701	2,429

Figure S2. Flowchart illustrating the quality control filtering of schizophrenia SNPs and the SNP overlap between PEIC and MPL datasets.



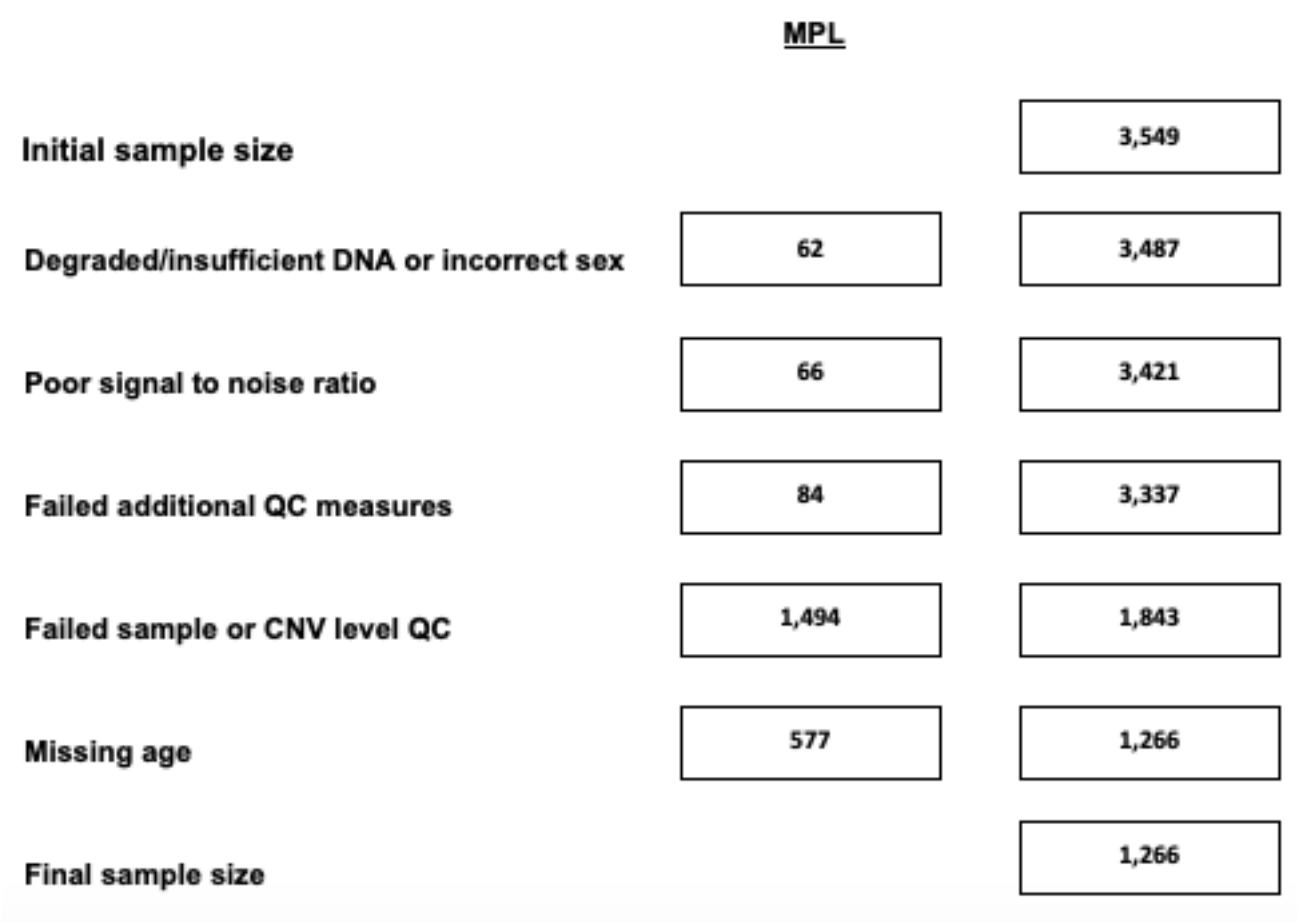
Flowchart demonstrating the quality control filtering of the SNPs in the two datasets (MPL and PEIC) along with the SNP overlap of the two datasets with the leave out verison of the schizozphrenia Psychiatric Genomic Consortium (PGC) data and the overlap between the two datasets

Figure S3. Flowchart illustrating the quality control filtering of bipolar disorder SNPs and the SNP overlap between PEIC and MPL datasets.



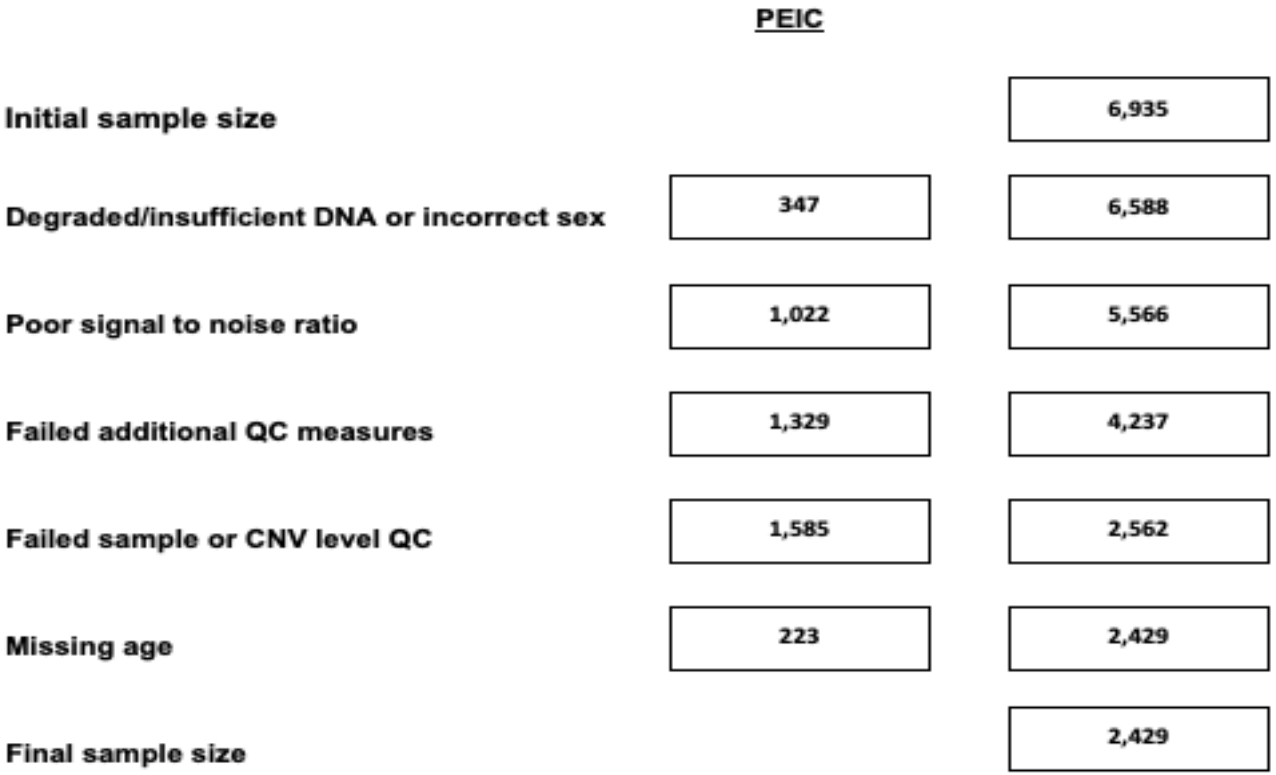
Flowchart demonstrating the quality control filtering of the SNPs in the two datasets (MPL and PEIC) along with the SNP overlap of the two datasets with the leave out verison of the bipolar Psychiatric Genomic Consortium (PGC) data and the overlap between the two datasets

Figure S4. Flowchart illustrating the justifications of exclusion of certain participants from the MPL dataset.



Flowchart demonstrating the reasons for excluding certain individuals from the dataset, the number of participants excluded on each step (on the left column) and the number of remaining participants after each step (on the right column) for the MPL dataset.

Figure S5. Flowchart illustrating the justifications of exclusion of certain participants from the PEIC dataset.



Flowchart demonstrating the reasons for excluding certain individuals from the dataset, the number of participants excluded on each step (on the left column) and the number of remaining participants after each step (on the right column) for the PEIC dataset.

Table S10 Case to control ratios across the polygenic risk score deciles for PEIC dataset

	Schizophrenia				Bipolar Disorder			
Decile	**Controls**	**%**	**Cases**	**%**	**Controls**	**%**	**Cases**	**%**
1	341	87.21	50	12.79	325	81.86	72	18.14
2	325	81.86	72	18.14	298	75.06	99	24.94
3	308	77.58	89	22.42	305	76.83	92	23.17
4	314	79.09	83	20.91	308	77.58	89	22.42
5	292	73.55	105	26.45	281	70.60	117	29.40
6	265	66.58	133	33.42	282	71.21	114	28.79
7	255	64.23	142	35.77	264	66.50	133	33.50
8	248	62.47	149	37.53	245	61.71	152	38.29
9	215	54.16	182	45.84	242	60.80	156	39.20
10	171	43.18	225	56.82	190	47.86	207	52.14

The threshold used to calculate polygenic risk score was PT = 0.01 for bipolar disorder and 0.1 for schizophrenia. Based on their RPS samples were divided into deciles (decile 1 = lowest PRS, 10 = highest PRS). The table reports the case to control ratios for broadly defined psychosis across the deciles.

Table S11 Polygenic risk scores (PRS) deciles and odd ratios (OR) of broadly defined psychotic disorder for PEIC dataset

	Schizophrenia risk score			Bipolar Disorder risk score		
Decile	**OR**	**LCI**	**UCI**	**OR**	**LCI**	**UCI**
1	0.34	0.24	0.47	0.54	0.40	0.73
2	0.52	0.38	0.70	0.81	0.61	1.06
3	0.68	0.51	0.89	0.74	0.55	0.97
4	0.62	0.46	0.82	0.70	0.53	0.93
5 and 6	1.00			1.00		
7	1.30	1.01	1.68	1.23	0.95	1.59
8	1.41	1.09	1.81	1.51	1.17	1.95
9	1.98	1.54	2.54	1.56	1.21	2.01
10	3.08	2.40	3.95	2.65	2.07	3.41

Data are odds ratios (OR) and 95% confidence interval (CI). OR were calculated using the central deciles (5^{th} and 6^{th}) as reference group. LCI = lower confidence interval. UCI = upper confidence interval.

Table S12 Case to control ratios across the polygenic risk score deciles for MPL dataset

	Schizophrenia PRS				Bipolar Disorder PRS			
Decile	**Controls**	**%**	**Cases**	**%**	**Controls**	**%**	**Cases**	**%**
1	167	48.55	177	51.45	148	43.02	196	59.98
2	138	40.12	206	59.88	112	32.56	232	67.44
3	118	34.30	226	65.70	109	31.69	235	68.32
4	97	28.20	247	71.80	106	30.81	238	69.19
5	88	25.58	256	74.42	84	24.42	260	75.58
6	88	25.58	256	74.42	73	21.22	271	78.78
7	65	18.90	279	81.10	81	23.55	263	76.45
8	59	17.15	285	82.85	64	18.60	280	81.40
9	39	11.34	305	88.66	68	19.77	276	80.23
10	29	8.43	315	91.57	43	12.50	301	87.50

The threshold used to calculate polygenic risk score was PT = 0.01 for bipolar disorder and 0.1 for schizophrenia. Based on their RPS samples were divided into deciles (decile 1 = lowest PRS, 10 = highest PRS). The table reports the case to control ratios for broadly defined psychosis across the deciles.

Table S13 Polygenic risk scores (PRS) deciles and odd ratios (OR) of broadly defined psychotic disorder for MPL dataset

	Schizophrenia risk score			Bipolar Disorder risk score		
Decile	**OR**	**LCI**	**UCI**	**OR**	**LCI**	**UCI**
1	0.36	0.28	0.48	0.39	0.30	0.52
2	0.51	0.39	0.68	0.61	0.46	0.82
3	0.66	0.50	0.87	0.64	0.48	0.85
4	0.88	0.66	1.17	0.66	0.50	0.89
5 and 6	1.00			1.00		
7	1.47	1.07	2.04	0.96	0.71	1.30
8	1.66	1.19	2.31	1.29	0.94	1.80
9	2.68	1.86	3.95	1.20	0.87	1.66
10	3.71	2.49	5.74	2.06	1.44	3.01

Data are odds ratios (OR) and 95% confidence interval (CI). OR were calculated using the central deciles (5^{th} and 6^{th}) as reference group. LCI = lower confidence interval. UCI = upper confidence interval.

Table S14 Meta analyses of odd ratios (OR) for schizophrenia and bipolar risk scores combining MPL and PEIC datasets

	Schizophrenia risk score			Bipolar Disorder risk score		
Decile	**OR**	**LCI**	**UCI**	**OR**	**LCI**	**UCI**
1	-1.03	-1.24	-0.82	-0.78	-1.1	-0.47
2	-0.46	-0.87	-0.05	-0.35	-0.62	-0.07
3	-0.35	-0.58	-0.19	-0.38	-0.58	-0.18
4	0.98	-1.2	3.16	-0.38	-0.59	-0.18
5 and 6	1.00			1.00		
7	0.97	-0.16	2.09	0.1	-0.14	0.34
8	0.92	0.12	1.73	0.35	0.15	0.55
9	1.17	0.89	1.46	0.34	0.07	0.6
10	2.29	1.38	4.2	0.99	0.66	1.12

Table S15 Comparison of the full regression models to the models without the CNV burden for MPL and PEIC datasets

	MPL dataset	**PEIC dataset**
	Full model vs model without CNV burden	**Full model vs model without CNV burden**
All Patients vs Controls	**R^2=** .007% **F=** 5.17 ***p***= .022	**R^2=** .001% **F=** 1.91 ***p***= .166
Schizophrenia cases vs Controls	**R^2=** .008% **F=** 5.11 ***p***= .023	**R^2=** .001% **F=** 2.71 ***p***= .009
Bipolar cases vs Controls	**R^2=** .002% **F=** 0.62 ***p***= .431	**R^2=** .002% **F=** 0.19 ***p***= .664

Table S16 Demographics of the combined sample of MPL and PEIC datasets

		Cases	Controls
Age, years: mean (sd)		41.70 *(*13.06*)*	46.06 *(*16.20*)*
Sex, female: n *(%)*		559 *(31.90%)*	1,054 *(54.19%)*
Sub-diagnostic groups n *(%)*	Schizophrenia	1,230 *(70.28%)*	
	Schizoaffective	43 *(2.45%)*	
	Schizophreniform disorder	29 *(1.65%)*	
	Bipolar disorder with psychosis	348 *(19.88%)*	
	Brief psychotic disorder	15 *(0.85%)*	
	Delusional disorder	13 *(0.74%)*	
	Psychosis disorder NOS	72 *(4.11%)*	
	Total	1,750	1,945

Table S17 Schizophrenia associated CNV loci identified in the MPL dataset.

Locus	Chromosome	Start Position (Hg18)	Stop Position (Hg18)	Size (Mb)	Gene of Interest	No. found	Odds Ratio	Frequency in controls	Reference
1q21.1.del	chr1	144800000	146326000	1.5		2	3.8-8.1	0.02-0.07	[1,2,3]
1q21.1.dup	chr1	144800611	146326568	1.5		2	4.2-5.2	0.021	[1,2,3]
15q11.2.del	chr15	20301000	20824174	0.5	CYFIP1	19	1.8-2.1	0.25-0.27	[1,2,3]
15q13.3.I.del	chr15	28723577	30303141	1.6	CHRNA7	2	4.7-15.6	0.009	[1,2,3]
15q13.3.II.del	chr15	29806023	30407419	0.6		2	4.7-14.9	0.019	[2,3]
16p12.1.del	chr16	21854731	22331199	0.5		7	1.8		[3]
16p13.11.dup	chr16	14897345	16199484	1.3	NTAN1, NDE1	1	2-2.2	0.13	[2,3]

16p13.11.del	chr16	15032942	16199484	1.1	2	1.7-1.9	0.039	[2,3]

The loci comprise all schizophrenia associated loci from Marshall et al. 2017[1], Kirov et al. 2014[2] and Stefansson et al. 2014[3], excluding protective loci 22q11.21.dup, 7q11.21.del 7q11.21.dup, 13q12.11.dup, Xq28.dup. No. found indicate number of carriers found in the MPL sample.

Table S18 Schizophrenia associated CNV loci identified in the PEIC dataset.

Locus	Chromosome	Start Position(Hg18)	Stop Position(Hg18)	Size (Mb)	Gene of Interest	No. found	Odds Ratio	Frequency in controls	Reference
1q21.1.del	chr1	144800000	146326000	1.5		3	3.8-8.1	0.02-0.07	[1,2,3]
2p25.3.dup	chr2	1733000	2204000	0.5	MYT1L	2	15.7		[3]
2p16.del	chr2	49900000	51500000	1.6	NRXN1	3	10.7-14.4	0.014	[1,2,3]
7q36.3.dup	chr7	158448321	158810016	0.4	VIPR2	2	3.2-3.5	0.029	[1,3]
15q11.2.del	chr15	20301000	20824174	0.5	CYFIP1	8	1.8-2.1	0.25-0.27	[1,2,3]
15q11.2-13.1.dup	chr15	20322358	26208861	5.9		1	5.1		[2,3]
15q13.3.l.del	chr15	28723577	30303141	1.6	CHRNA7	1	4.7-15.6	0.009	[1,2,3]

16p13.11.dup	chr16	14897345	16199484	1.3	NTAN1, NDE1	7	2-2.2	0.13	[2,3]
16p12.1.del	chr16	21854731	22331199	0.5		1	1.8		[3]
16p11.2.del	chr16	29502984	30100062	0.6		2	0.5-0.9	0.04	[2,3]
16p11.2.dup	chr16	29531748	30105652	0.6		4	8-9.4	0.03	[1,2,3]
22q11.21.large.del	chr22	17285281	19818855	2.5		1	67.7	0.04	[1,2]
22q11.21.del	chr22	19063495	19795780	0.7		1	Inf		[3]

The loci comprise all schizophrenia associated loci from Marshall et al. 2017[1], Kirov et al. 2014[2] and Stefansson et al. 2014[3], excluding protective loci 22q11.21.dup, 7q11.21.del 7q11.21.dup, 13q12.11.dup, Xq28.dup. No. found indicate number of carriers found in the PEIC sample.

Table S19 Schizophrenia associated CNV loci identified in both MPL and PEIC datasets.

Locus	Chromosome	Start Position(Hg18)	Stop Position(Hg18)	Size (Mb)	Gene of Interest	No. found	Odds Ratio	Frequency in controls	Reference
1q21.1.del	chr1	144800000	146326000	1.5		5	3.8-8.1	0.02-0.07	[1,2,3]
1q21.1.dup	chr1	144800611	146326568	1.5		2	4.2-5.2	0.021	[1,2,3]
2p25.3.dup	chr2	1733000	2204000	0.5	MYT1L	2	15.7		[3]
2p16.del	chr2	49900000	51500000	1.6	NRXN1	3	10.7-14.4	0.014	[1,2,3]
7q36.3.dup	chr7	158448321	158810016	0.4	VIPR2	2	3.2-3.5	0.029	[1,3]
15q11.2.del	chr15	20301000	20824174	0.5	CYFIP1	27	1.8-2.1	0.25-0.27	[1,2,3]
15q11.2-13.1.dup	chr15	20322358	26208861	5.9		1	5.1		[2,3]

15q13.3.I.del	chr15	28723577	30303141	1.6	CHRNA7	3	4.7-15.6	0.009	[1,2,3]
15q13.3.II.del	chr15	29806023	30407419	0.6		2	4.7-14.9	0.019	[2,3]
16p13.11.dup	chr16	14897345	16199484	1.3	NTAN1, NDE1	8	2-2.2	0.13	[2,3]
16p13.11.del	chr16	15032942	16199484	1.1		2	1.7-1.9	0.039	[2,3]
16p12.1.del	chr16	21854731	22331199	0.5		8	1.8		[3]
16p11.2.del	chr16	29502984	30100062	0.6		2	0.5-0.9	0.04	[2,3]
16p11.2.dup	chr16	29531748	30105652	0.6		4	8-9.4	0.03	[1,2,3]
22q11.21.large.del	chr22	17285281	19818855	2.5		1	67.7	0.04	[1,2]
22q11.21.del	chr22	19063495	19795780	0.7		1	Inf		[3]

The loci comprise all schizophrenia associated loci from Marshall et al. 2017[1], Kirov et al. 2014[2] and Stefansson et al. 2014[3], excluding protective loci 22q11.21.dup, 7q11.21.del 7q11.21.dup, 13q12.11.dup, Xq28.dup. No. found indicate number of carriers found in both PEIC and MPL datasets.

Table S20 Distribution of schizophrenia associated CNV carriers in sub-diagnosis groups per dataset**.**

	Group	**MPL**	**PEIC**
Schizophrenia associated CNVs	Controls	10 *(29.41%)*	18 *(50%)*
	Schizophrenia/Schizophreniform	19 *(55.88%)*	14 (*38.88%)*
	Bipolar disorder	5 *(14.70%)*	0
	Other psychotic disorder	0	4 *(11.11%)*
	Total	34	36