Master's Thesis in Biology 2024

Using GWAS data to investigate the biological basis of schizophrenia

Student: Caroline Sophie Jessen
Student number: 201908696

Main supervisor: Jesper Givskov Sørensen, Department of Biology – Genetics, Ecology and Evolution
Supervisor: Doug Speed, Center for Quantitative Genetics and Genomics

ECTS: 60

AARHUS UNIVERSITY
DEPARTMENT OF BIOLOGY

CENTER FOR QUANTITATIVE GENETICS AND GENOMICS

Table of Contents

Abstract

Schizophrenia is a chronic and complex mental disorder that typically emerges in late adolescence or early adulthood. It is characterized by a range of symptoms that affect a person's thoughts, emotions, and behavior. The course of the disease varies significantly among individuals, with some experiencing single episodes, while others have repeated and chronic course of illness with highly impaired functioning. Schizophrenia has a strong genetic component, with heritability estimates ranging from 60% to 80%, but the exact cause of the disorder is still not fully understood. The aim of the study is to investigate the genetic basis of schizophrenia. This study utilizes data from the latest schizophrenia genome wide association study (GWAS) from the Psychiatric Genomics Consortium (PGC) to construct a polygenic risk score (PRS) model. The schizophrenia PRS is calculated for UK Biobank individuals to explore its relationship with the severity of the disorder and treatment response, using prescription medications from primary care linked data as a severity indicator. The study found no significant relationship between schizophrenia PRS and the prescription level measures of severity, however, studies consistently reveal a positive correlation. Furthermore, the study examines the genetic correlations between schizophrenia and other psychiatric disorders, as well as various health-related phenotypes. The results of this study emphasize the role of PRS as a marker of genetic susceptibility to schizophrenia and highlight its potential in predicting the severity of the disorder and treatment resistance. The methods employed in this study show promise in predicting severity of psychiatric disorders, however, larger sample sizes are required to detect significant signals.

Introduction

Schizophrenia is a complex and debilitating mental disorder that despite several significant advances in understanding its clinical aspects, remains poorly understood. In this thesis, I will explore the illness by reviewing the existing literature on schizophrenia, to provide a concise overview of clinical aspects, and the genetic basis through previous findings. I will introduce the methods for identifying genetic associations and genetic features used for prediction of severity and treatment response. The findings from the existing literature will then be used to further inform the experimental part of this thesis, in which I will investigate the prediction of schizophrenia using genetic and clinical data.

GWAS

A genome-wide association study (GWAS) is a method for identifying associations between genetic variations and a specific phenotype. GWAS has been used to study a wide range of traits, such as height and body mass index (Yengo et al., 2018), and diseases like diabetes (Xue et al., 2018), cancer (Sud et al., 2017), and schizophrenia (Trubetskoy et al., 2022). GWAS involves analyzing genetic data from large cohorts of individuals of the same ancestry to explain variation of complex traits, by linking common genetic markers to the phenotype of interest. Single nucleotide polymorphisms (SNPs) are a substitution of a nucleotide at a specific genomic position and are often used as genetic markers in GWAS. The 1000 genome project have found that the typical human genome differs from the reference genome at approximately 4 to 5 million sites (Auton et al., 2015), however most of these variations have a minimal effect on the phenotype. GWAS aims to identify SNPs that are associated with an individual's (susceptibility to a) phenotype. Depending on the trait of interest, different methods can be used for the association analysis. For quantitative traits linear regression is often used, while associations of binary traits are performed using logistic models.

The simplest method for performing a GWAS is using an additive single SNP association analysis, where every SNP is tested individually for association with the phenotype. The genotyping data at each position is encoded as 0, 1, or 2, denoting the number of copies of a given variant. The model assumes that the effects of SNPs are additive, and the genotype and phenotype are assumed to have a linear relationship by the linear regression model.

The additive linear regression model assumes $Y_i = \alpha + \beta X_{i,j} + \varepsilon_i$ where Y_i is the observed phenotypic trait for individual i, $X_{i,j}$ is the individual's genotype at SNP j, α is the intercept, β is the slope, and ε_i is the random error. The coefficients α and β are estimated using least squares regression to find the association between the SNP and the phenotype.

The additive logistic model used for binary traits assumes $logit(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta X_{i,j}$ where p_i is the probability that individual i has the phenotype $Y_i = 1$. $X_{i,j}$ again, represents the individual's phenotype at SNP j. Estimation of the coefficients can be done using maximum likelihood. Both the linear and the logistic models can be expanded to include covariates such as age, sex, and principal components. Including covariates may increase the detection power of the GWAS. Following the estimation of the coefficients, significance of the association can be computed. In human genetics the threshold for genome-wide significance is $p < 5 \times 10^{-8}$. This significance criterion is equivalent to a Bonferroni correction for 1,000,000 independent common SNPs in the human genome.

There are many things that needs to be taken into consideration when performing and analyzing GWAS, including natural variation in the genome among different ancestries, or the signals from associated individual from the same families. Before performing a GWAS the most important quality control (QC) steps include ensuring that both the individual level data and the marker data is fit for use (Marees et al., 2018). Per-individual QC includes removal of individuals with discordant sex information, missing genotype data, divergent ancestry, and duplicated individuals (Anderson et al., 2010). Per-marker QC includes removal of SNPs with high missing data rates or significant deviation from Hardy-Weinberg equilibrium, and the removal of very rare variants with a minor allele frequency (MAF) less than 1% (Anderson et al., 2010). These quality control steps are important to ensure the integrity and reliability of the genotyping data in GWAS. They also address potential sources of bias and errors, reduce false-positive and false-negative associations, and improve the accuracy of the study.

GWAS analyses may have multiple purposes, such as detection of causal variants and enhancing the biological understanding of the genetic architecture of a trait. In recent years the statistical power of GWAS has increased drastically due to large-scale biobanks and collaborative efforts. Furthermore, the increase in diversity in study populations has improved the accuracy and generalizability of GWAS findings (Abdellaoui et al., 2023). GWAS have provided insights into the genetic architecture of complex traits and diseases, and contributed to advancements in disease prevention, treatment, and drug development (Abdellaoui et al., 2023; Visscher et al., 2017). GWAS results have also become a valuable tool for predicting an individual's genetic risk for developing certain diseases or traits.

PRS

Risk prediction using Polygenic Risk Scores (PRS) is a common application of GWAS (Uffelmann et al., 2021). A PRS is an estimate of an individual's genetic predisposition to a certain trait and can be calculated for any individual based on their genome information with the use of a PRS model (Andlauer & Nöthen, 2020; Choi et al., 2020; Jian-Ping Zhang et al., 2019)
. The models are typically based on GWAS results, that assign weights to the individual genetic variants. In a classical PRS model, the genetic risk score for an individual is calculated by summing the number of risk alleles at each variant, weighted by its effect size (log(OR) for binary traits or beta coefficient for quantitative traits). This results in a single score that represents the individual's genetic liability for a disease or continuous trait (Lewis & Vassos, 2020). Single predictor regression or classical PRS is the simplest way to construct a polygenic risk score.

For quantitative traits:

$$PRS = \beta_1 SNP_1 + \beta_2 SNP_2 + \cdots + \beta_n SNP_n$$

For binary traits:

$$PRS = OR_1 SNP_1 + OR_2 SNP_2 + \cdots + OR_n SNP_n$$

Where the SNP_i represents the allele count of SNP i, β_i and OR_i represents the effect size of the SNP as calculated from the GWAS. The classical PRS assumes that all SNPs contribute equally to the expression of a trait, whereas more advanced tools assume different contributions from individual SNPs using a heritability model (Speed et al., 2020). Heritability models specify the amount of heritability contributed to genetic variants across the genome. In models with a different heritability model stronger associated SNPs often have a higher weight than less associated SNPs when estimating a person's liability for a certain trait.

A PRS by itself does not give much information about a person's liability to a specific disease, as it is just an arbitrary number with no fixed scale (Andlauer & Nöthen, 2020). To interpret polygenic scores, they should be compared with a PRS of individuals with known case/control status. A higher-than-average PRS should be an indication of a higher risk or liability of a specific phenotype, while a lower-than-average PRS is indicative of the opposite. For polygenic traits, the PRS alone often explain only a small part of the variation in the phenotype.

There are many ways to calculate a PRS. One can use either summary statistics from previous GWAS, or construct PRS using individual level data if phenotype and genotype data are available (Choi et al., 2020). One of the big challenges when constructing PRS models is which SNPs to

include in the model and how these should be weighted. Due to linkage disequilibrium (LD), many SNPs near the causal loci can achieve a higher importance in the model due to statistically significant associations caused by proximity to the causal loci, rather than being causal themselves. Therefore, thinning or fine mapping of results are essential. The calculation of a PRS require two types of data: GWAS results which provide the estimated effect size of each SNP, and genotype data from a group of individuals independent from the individuals used in the GWAS. These are the individuals for whom the PRS will be estimated based on their genotypes.

Similarly to performing GWAS, quality control measures are necessary to ensure the quality and independence of data and minimize errors before estimating polygenic risk scores (Choi et al., 2020; Lee et al., 2011). Firstly, the trait must be heritable to some extent for the PRS to be relevant and useful. Choi et al. (2020) recommends only performing PRS analyses on phenotypes with an estimated SNP heritability $h^2_{SNP} > 0.05$. It is further recommended that the PRS is calculated for at least 100 individuals (this is the target group), because comparison with other scores are necessary for the interpretation of the score (Choi et al., 2020). The QC prevent errors and ensure that the genome build is consistent across all data. Additionally, duplicates, mismatches between reported and genetic sex, and overlapping samples must be eliminated. Furthermore, the target data must have similar ancestry as the data used to generate GWAS results, to avoid issues with predicting on different ancestries. The recommended QC criteria for PRS are equivalent to that of performing a GWAS (Marees et al., 2018) (Choi et al., 2020).

The accuracy and predictive ability of polygenic risk scores can be evaluated using various metrics, including the area under the receiver operating characteristic curve (AUC) and R^2 values (Choi et al., 2020; Uffelmann et al., 2021). AUC is a statistic used to evaluate the performance of a predictive model, particularly in the context of binary classification problems. It is calculated based on the Receiver Operating Characteristic (ROC) curve, which plots the true positive rate (sensitivity) against the false positive rate (1-specificity) at various classification thresholds (Wray et al., 2010). The AUC represents the probability that a randomly selected diseased individual will have a higher predicted risk score than a randomly selected non-diseased individual. It ranges from 0 to 1, with a higher AUC indicating better predictive performance (Wray et al., 2010). The R^2 is a measure that quantifies the proportion of variance in liability explained by a trait or disease that can be explained by the genetic variants included in the polygenic risk score (Wray et al., 2010) (Choi et al., 2020). A higher R^2 value indicates a greater proportion of genetic variance explained, making it a more accurate predictor of genetic risk. The liability scale is used to model the underlying genetic risk or liability for a particular disease or trait. On the liability scale,

individuals are positioned along a continuum, ranging from low liability (low risk) to high liability (high risk) for the trait or disease (Wray et al., 2010).

The variability of a trait that a PRS can explain is limited by the SNP heritability (h^2_{SNP}) of the trait itself (Wray et al., 2021). It would only be possible to obtain this value if there are no errors in the GWAS data, and all associated SNPs are correctly identified. This is very hard as the effect size estimates contain a lot of error and consequently the predictive power of PRS is usually substantially lower (Choi et al., 2020) (Lee et al., 2011).

Heritability

Heritability is the proportion of phenotypic variation that is due to genetic factors (Lee et al., 2011; Trubetskoy et al., 2022). The total phenotypic variance (V_P) is due to the genotypic variance (V_G), the environmental variance (V_E) and the (potential) interaction between genotype and environment.

$$V_P = V_G + V_E + V_{G\times E}$$

Often it is assumed that the interaction $V_{G\times E}$ is small enough to ignore.

Broad-sense Heritability

Broad-sense heritability (H^2) considers all genetic factors, and is the contribution of genotypic variance to the total phenotypic variance:

$$H^2 = \frac{V_G}{V_P}$$

Heritability values range from 0 to 1. A higher value indicate that the environment has a smaller impact on the phenotypic variance, largely due to genotypic differences among the population. Low heritability values indicate that environmental factors have the largest impact on the phenotype.

The V_G includes all types of genetic variation and does not distinguish between alleles with additive, dominant or epistatic effects.

Narrow-sense Heritability

Narrow-sense heritability (h^2) is the proportion of phenotypic variance that is due only to additive genetic variance (V_A). The total genotypic variance (V_G) is due to both additive genetic variance, dominance variance (V_D) and interactive variance (V_I). Total genoypic variance:

$$V_G = V_A + V_D + V_I$$

Narrow-sense heritability is based on the portion of phenotypic variance that is based on additive variance.

$$h^2 = \frac{V_A}{V_P}$$

When using GWAS results to estimate heritability, a part of the heritability will be missing or cannot be explained by the results. h^2_{GWAS} and h^2_{SNP} are the variation in the trait that can be explained by variation in the significant SNPs identified in a GWAS, and the heritability of all common SNPs, respectively.

Review: Schizophrenia

Schizophrenia is a complex primary psychiatric disorder, which is characterized by fundamental disturbances in thinking, perception, and emotions. It is classified under the International Classification of Diseases (ICD) as a mental disorder with symptoms falling into six domains: positive symptoms, negative symptoms, depressive mood symptoms, manic mood symptoms, psychomotor symptoms, and cognitive symptoms (https://icd.who.int/browse/2024-01/mms/en#405565289). Apart from having a large impact on affected individuals and their families, it is also considered to be one of the most expensive mental disorders (Rössler et al., 2005). According to the World Health Organization (WHO) approximately 24 million people worldwide are affected with schizophrenia and only a minority of those affected fully recover from the illness. Schizophrenia is associated with a reduced life expectancy, primarily due to suicide and an increased morbidity (Rössler et al., 2005). The etiology of schizophrenia is believed to involve multiple genes of small effect and potentially some rare large-effect genes, along with environmental factors (Haller et al., 2014). Although schizophrenia has been intensively researched for a long time, there continues to be a need for more research to improve the understanding of the causes of schizophrenia and the possible avenues of treating this disease.

Clinical aspects of schizophrenia

Schizophrenia is characterized by a wide range of symptoms. These include positive symptoms (such as hallucinations and delusions), negative symptoms (like social withdrawal and lack of

initiative) (Mäkinen et al., 2008), and cognitive dysfunction. The course of the disease varies significantly among individuals, with some experiencing single episodes, others having repeated episodes, and some facing a chronic course with impaired functioning (Albus, 2012). The age of onset for schizophrenia differ between men and women, with males tending to have an onset in early to mid 20s, and women in mid to late 20s. Earlier onset of the disease has been linked to more severe course of illness (Carbon & Correll, 2014). Based on the heterogeneity of symptoms and course of illness it is likely that schizophrenia is not a singular disease entity (Tandon et al., 2009; Trastulla et al., 2023). Previous studies have found significant variation in the incidence and course of illness, with urbanity, migration, and being male leading to a higher risk of developing the disease (Tandon et al., 2008). The lifetime prevalence of schizophrenia is <1% (Saha et al., 2005).

Genetic basis of schizophrenia

A meta-analysis of 12 twin studies of schizophrenia conducted by Sullivan et al. (2003) found the heritability of schizophrenia to be 81% with a 95% confidence interval of 73%-90%. A study of a Swedish cohort estimated the heritability for schizophrenia at 64.3% (Lichtenstein et al., 2009), but it is generally accepted that approx. 80% of the liability for the illness is attributable to the contribution of genetic factors. However, while many genetic variants have been found to be associated with schizophrenia, no major gene or locus explaining a substantial part of the heritability has yet been identified (Tandon et al., 2008).

Through meta-analyses Trubetskoy et al. (2022) identified 313 independent SNPs at 287 distinct genomic loci associated with the disease that that exceeded genome-wide significance. The SNP-based heritability (h^2_{SNP}) for European individuals was estimated to 0.24 (Trubetskoy et al., 2022). Furthermore, family history of psychotic disorders is a strong predictor of schizophrenia, with a positive family history being associated with approx. 10-fold risk of schizophrenia (Lu et al., 2018). The point prevalence of schizophrenia is 4.5 per 1000 people, and no significant difference in prevalence has been found between males and females (Abel et al., 2010; Tandon et al., 2008). A GWAS for male and female individuals had a genetic correlation statistically indistinguishable from 1, indicating that the common variant genetic liability to schizophrenia is essentially identical in males and females (Trubetskoy et al., 2022). However, there is a difference in disease risk between the two sexes with men having a higher incidence at a young age, and earlier onset compared to women. The clinical presentation and symptom profile also differ, with depressive symptoms being more likely for women and men being more likely to experience negative symptoms (Abel et al., 2010).

Psychiatric comorbidities

Schizophrenia is associated with a range of comorbidities, including cardiovascular disease, obesity, diabetes, smoking, anxiety disorders, depressive syndromes, intellectual disability, and personality disorders (Tandon et al., 2009). The presence of comorbidities in patients with schizophrenia is generally associated with more severe psychopathology and poorer outcomes (Buckley et al., 2008). Insomnia is also common comorbidity in schizophrenia (Cohrs, 2008; Robertson et al., 2019). Comorbid depression occurs in 50% of patients, and approximately 47% of patients have a lifetime diagnosis of comorbid substance abuse (Buckley et al., 2008). A study by Lichtenstein et al. (2009) found schizophrenia and bipolar disorder to be associated. The comorbidity between schizophrenia and bipolar disorder is mainly due to additive genetic effects common to both disorders (Lichtenstein et al., 2009). Panic disorder, posttraumatic stress disorder (Hübel et al.), and obsessive-compulsive disorder (Hübel et al.) are also prevalent among patients with schizophrenia (Buckley et al., 2008).

Environmental risk factors

While schizophrenia is highly heritable (Sullivan et al., 2003), multiple environmental factors have also been found to be associated with the disease (Tandon et al., 2008). Cannabis use has been linked to poor outcomes in schizophrenia patients and has the potential to accelerate psychosis in individuals with preexisting liability (Henquet et al., 2005; Johnson et al., 2021). Smoking and schizophrenia has also been found to share a common genetic component (Hartz et al., 2018). Early life factors include prenatal infection, malnutrition, and complications, as well as a history of winter birth (Martínez-Ortega et al., 2011; Tandon et al., 2008). Vitamin D deficiency has also been linked to schizophrenia, which may also be related to developmental factors (McGrath et al., 2010).

Neurobiology of schizophrenia

Schizophrenia is associated with structural abnormalities in the brain, including reductions in brain volume, gray matter, white matter, and increased ventricular volume (Keshavan et al., 2008). The anatomical characteristics of the brain are believed to have a shared genetic foundation with the disorder (Lee et al., 2016). Structural irregularities within certain cerebral regions could be the foundation for hallucinations and abnormalities in the superior temporal gyrus and gray matter may contribute to an increased predisposition to psychotic episodes (van Tol et al., 2014). Specific defects in areas governing verbal memory and speech articulation may increase the risk of auditory hallucinations. Additionally, atypical morphometry in white matter of the postcentral gyrus and parietal lobule could enhance the neural communication that underpins the perception of voices (van Tol et al., 2014). It is believed that presynaptic anomalies in the dopamine system and

glutamate abnormalities are implicated in the pathophysiology of schizophrenia (Howes & Murray, 2014; Keshavan et al., 2008). Elevated dopamine synthesis, increased dopamine release, and increased baseline synaptic dopamine levels is associated with the development of more severe psychotic symptoms (Howes & Murray, 2014).

Irregularities in the development of brain lateralization and deviations in cerebral dominance may play a significant role in the onset of schizophrenia (Keshavan et al., 2008). The prevalence of atypical handedness (left-handedness, non-right-handedness, and mixed-handedness) is significantly higher in schizophrenia patients compared to the general population (Dragovic & Hammond, 2005), and may support that these factors could be linked to genes increasing the vulnerability to the condition. Brain abnormalities in schizophrenia are inconsistent but occur in areas that normally show sexual dimorphism. This suggests that the same factors driving sex differences in normal brain development are also involved in sex differences in schizophrenia (Abel et al., 2010).

Treatment

The most common treatment option for schizophrenia is antipsychotic medication, however several other treatment options exist such as therapy, psychosocial treatment and education (Patel et al., 2014). The goal of treatment is to relieve symptoms and increase quality of life for affected individuals. Antipsychotic medication generally have a good effect on positive symptoms, but negative symptoms are harder to treat (Galderisi et al., 2021). Negative symptoms affect approximately 40% of individuals with schizophrenia (Correll & Schooler, 2020). Treatment-resistant schizophrenia (TRS) is a term used to describe individuals who do not show a response to antipsychotics (Howes et al., 2017). Approximately 15% of patients do not experience a reduction of symptoms with antipsychotic medication and can be classified as treatment resistant. A study of 408 patients with TRS uncovered several associations and clinical markers related to TRS (Correll et al., 2019). TRS exhibits a dimensional gradient from non-TRS to more severe forms, characterized by greater illness severity and chronicity markers. Furthermore, TRS displays heterogeneity, with some patients resistant from disease onset and others developing resistance over time (Correll et al., 2019).

Treatment response have proven hard to predict with several studies having attempted to identify potential predictors of antipsychotic response. A study by Zhang et al. (2019) investigated polygenic risk scores as a predictor of treatment response and found that higher PRS significantly predicted a poorer response to treatment. This indicates that PRS may have potential utility as prognostic biomarker (Jian-Ping Zhang et al., 2019). Another study found male sex, poor

premorbid adjustment, early onset, and lower educational level associated with reduced odds for remission (Carbon & Correll, 2014). Additionally, negative symptoms have been confirmed as major predictors of poor outcome, as there is a close association between these symptoms and real-life functioning (Carbon & Correll, 2014).

First-generation and second-generation antipsychotics

Two generations of antipsychotics are available for treatment of schizophrenia: first-generation (typical) and second-generation (atypical). Second-generation antipsychotics has advantages over the first-generation preparations in that they produce fewer side effects and are generally better tolerated by patients (Hartling et al., 2012). They are therefore recommended as the first choice for treatment, especially for negative symptoms (Galderisi et al., 2021). Benzo- diazepines and -thiazepines are classes of second-generation compounds that produce sedative and antidepressant effects, and have been found to be effective in treating anxiety disorders, epilepsy, muscle spasms, and insomnia (Robertson et al., 2019; Zamani & Doustkhah, 2022). Several studies have used Clozapine as a measure for treatment resistance (Pardiñas et al., 2022; Ruderfer et al., 2016). Clozapine is a second-generation drug that has been estimated to be approximately 30% effective in controlling schizophrenic episodes in TRS cases (Patel et al., 2014). However, due severe side effects of clozapine such as agranulocytosis and seizures, it is not recommended as a first-line treatment (Patel et al., 2014).

Project description

As previously stated, schizophrenia is a complex psychiatric disorder influenced by genetic and environmental factors. The following parts of this study aims to investigate the relationship between genetic markers and clinical outcomes, such as diagnosis and treatment response, through genomic analysis and healthcare statistics, for the identification of potential subtypes of schizophrenia. This project will use data from the latest genome-wide association study of Schizophrenia (PGC3) by the Psychiatric Genomics Consortium, and from the UK Biobank. PRS model constructed based on the PGC3 data will be applied to individuals in the UK Biobank cohort, to produce a measure of the individual's genetic predisposition to schizophrenia. I will explore whether the PRS values correlate with clinical outcomes, and further evaluate the ability of PRS to predict severity and treatment success.

The primary hypothesis is that there is a positive correlation between the schizophrenia PRS and the severity of the disorder. Specifically, I expect that patients with higher PRS, which suggest a greater genetic predisposition to the disorder, will require a longer duration of treatment. This assumes that a higher PRS is indicative of more severe manifestations of the disorder.

Additionally, it is hypothesized that treatment-resistant cases are an expression of more severe symptoms, which should be quantifiable by a higher PRS. The use of clozapine, typically reserved for treatment-resistant cases, and multiple different drug prescription are expected to be associated with a higher PRS under the assumption that a high number of distinct prescriptions may serve as a proxy for treatment resistance. I expect to observe genetic correlations between schizophrenia and phenotypes that are linked to a more severe course of illness, such as earlier age of onset or environmental factors like smoking. These hypotheses will be tested using GWAS and PRS to investigate the genetic basis of schizophrenia and its relationship to other psychiatric disorders, such as bipolar and major depression disorder.

Methods

Data

Psychiatric Genomics Consortium

The Psychiatric Genomics Consortium (PGC) is an international consortium of researchers investigating the genetic basis of multiple psychiatric disorders, including schizophrenia, major depressive episode, and bipolar disorder. Since 2007 the aim has been to identify loci associated with these disorders through genome wide association studies, to improve our understanding of the biology underlying psychiatric illness to improve diagnosis, treatment, and quality of life of patients suffering from mental disorders (Sullivan et al., 2018). Summary statistics results from GWASs performed by the PGC are available for download from their website: https://pgc.unc.edu/for-researchers/download-results/.

Schizophrenia summary statistics data

For the construction of the polygenic risk score I used the summary statistics from the most recent and largest study to date on schizophrenia (PGC3) published by the PGC (Trubetskoy et al., 2022). The primary data of the PGC3 study included 7,585,078 SNPs with MAF ≥ 1% from 175,799 individuals. Of those, 74,776 individuals were diagnosed with schizophrenia (cases) and 101,023 were controls. 74.3% of the individuals were of European descent. To control for ancestry, only the data set of individuals of European descent were used for the PRS construction. The European subset (PGC3_SCZ_wave3.european.autosome.public.v3.vcf.tsv.gz) contained 52,017 cases and 75,889 controls, and data was based on 76 different cohorts The genome reference ensemble used is the Genome Reference Consortium Human Build 37 (GRCh37).

Major Depressive Disorder summary statistics data

The major depressive disorder summary statistics for genetic correlation was sourced from PGC paper on major depressive disorder. Genome-wide association meta-analysis of 9.6 million imputed SNPs from seven cohorts of European ancestry. The study included a total of 135,458 cases with major depression and 344,901 controls of European ancestry. The results of the analysis identified 44 loci associated with major depression (Wray et al., 2018). Summary statistics data was available for the full study, but as one of the seven cohorts were the UK Biobank I downloaded the option without the UK Biobank. The summary statistics included 45,591 cases and 97,947 controls. Data was downloaded from

https://figshare.com/articles/dataset/MDD2_MDD2018_GWAS_sumstats_w_o_UKBB/21655784].

Bipolar Disorder summary statistics data

The summary statistics for bipolar disorder was sourced from the most recent PGC paper about bipolar disorder (Mullins et al., 2021). Results excluded the UK Biobank individuals and come from a Meta-analysis of 41,917 BIP cases, 371,549 controls from 57 datasets. Quality control performed by the PGC included filtering SNPs from all included datasets using a MAF ≥ 1%. Data was downloaded from https://figshare.com/articles/dataset/bip2021_noUKBB/22564402.

UK Biobank

A large part of the data for this project was sourced from the UK Biobank. The UK Biobank is a biomedical database and research resource providing a wide range of genetic and health related information about approximately 500,000 participants. The aim of the data base is to investigate the factors contributing to diseases in middle- and old age, as well as the effects of lifestyle, environmental, and genetic variables on human health, morbidity, and mortality (Ollier et al., 2005; Sudlow et al., 2015). Recruitment of participants took place across 22 assessment centers across the United Kingdom to ensure variation in the socioeconomic position, ethnicity, and urban-rural mix of candidates (Ollier et al., 2005). At the time of enrollment, the participants ranged in age from 40 to 69. The UK Biobank has gathered, and continues to collect, a wide range of phenotypic and health-related data. Genome-wide genotyping data have been collected from every individual along with the participant's demographics, health and lifestyle indicators, biological measurements, and biomarkers obtained from urine and blood samples (Bycroft et al., 2018; Elliott et al., 2008; Ollier et al., 2005). Data was collected through verbal interviews, online surveys, direct evaluations, and electronic health records (EHR). New examinations and long-term follow-up, ensures that the collection of biomedical data keeps growing, and new data is regularly uploaded to the database. Imaging data of the brain, cardiac and abdominal regions were added in 2014 and primary care linked data recorded by healthcare professionals was made available in 2019.

Genotype data

The individual-level genetic data from the UK Biobank was restricted to only include data from white British unrelated individuals. The available data was post-ancestral filtering and quality control, which involved the removal of SNPs with low frequency of call rate. The genotype file

contained information about 392,214 unique individuals. Furthermore 13 covariates for the genetic analyses of these individuals were available for this data set. The covariates included age (data field 21022), sex (31), Townsend deprivation index (189) and ten principal components. Further information about the generation and methods for quality control of these data sets are described in (Zhang et al., 2021).

Primary care data

Primary care data is available for approximately 45% of the cohort (230,000 UK Biobank participants), with the newest entries being as recent as 2017 depending on the data provider. Primary care data within the UK Biobank refers to data recorded by health care professionals working at setting where individuals first seek advice or treatment for health concerns e.g., general practitioners (Wierenga et al.) or nurses at a local general practice. The primary care data contains three different datasets. GP prescription records (Data-Field 42039) contains prescribed medications including prescription date, drug code, and, where available, drug name and quantity. GP clinical event records (42040) with coded clinical events such as consultations, laboratory tests, diagnoses, and procedures, and GP registration records (42038) with a variety of administrative codes (e.g. referrals to specialist hospital clinics).

I used the GP prescription records (GP_scripts.txt). The GP_scripts data in the UK Biobank includes information about prescriptions issued to participants. The data is available for participants from England, Scotland, and Wales. Since data was obtained from three different registration records the completeness and accuracy may vary. The variables included in the data is shown in table 1. The UK Biobank purposefully carried out limited amounts of data cleaning prior to release, and the amount of information about each individual varies considerably within the datasets.

Table 1. *Data available in the UK Biobank primary care linked prescription file (gp_scripts.txt).*

Prescriptions dataset from UK Biobank: gp_scripts.txt	
Column name	**Description**
eid	Unique 7 digit participant identifyer
data_provider	1=England (Vison), 2=Scotland, 3=England (TPP), 4=Wales
issue_date	Issue date of the prescription
read_2	Read v2
bnf_code	BNF code
dmd_code	DM+D code
drug_name	Drug name
quantity	Quantity issued

UK Biobank Brain Imaging

In 2014 the UK Biobank initiated the collection of imaging data, aiming at collection data from 100,000 of its participants (Littlejohns et al., 2020; Miller et al., 2016). The brain imaging data used for this project was sourced from the Oxford Brain Imaging Genetics Server – BIG40. BIG40 contains GWAS results from approx. 4000 UK Biobank brain-imaging derived phenotypes from 40,000 UK Biobank participants (Smith et al., 2021). The full data set was reduced to 33,224 individuals after quality control (Alfaro-Almagro et al., 2018).. I downloaded results from the full GWAS results (discovery and replication combined), as well as the sex specific discovery sample consisting of 22,138 individuals of where 11,624 were genetic females (Smith et al., 2021). Linear association tests were performed on 17,103,079 genetic variants and thee GWAS results are available for download from https://open.win.ox.ac.uk/ukbiobank/big40/.

1000 Genome reference panel

The 1000 Genomes Project is an international research consortium with the aim of establishing a detailed catalogue of common human genetic variation with a frequency of at least 1%. Data from the 1000 genomes project is freely available to the worldwide scientific community though public databases (https://www.internationalgenome.org). 1000 Genomes Project combined genomic data from 2,504 individuals using the human genome reference GRCh37 build. Individuals from five different continental regions, including Europe, were sampled to create the 1000 genomes reference panel of human genotypes and major classes of structural variants in diverse populations (Auton et al., 2015). Analyses of GWAS summary statistics require a reference panel of similar ancestry to the studied population. This reference panel is used for genotype imputation to estimate correlations between predictors when accounting for LD. The 1000 Genomes Project was used as reference for the genetic correlation. The panel used contained genotype information from 404 non-Finnish European individuals to match the ancestry of the PGC summary statistics. The file was already available on the GenomeDK cluster. Details on how the file was created can be found here https://dougspeed.com/reference-panel/. Due to the small sample size of the reference genome the minor allele frequency (MAF) was increased from 0.005 to 0.01.

Data preparation

Prescription data

The raw gp_scripts.txt file contained information from 222,000 individuals. This file was subsetted to only include individuals in my genotype file, resulting in a new file with 44,910,331 lines (including the header) and information about 177,412 individuals, averaging approximately 253 prescriptions per participant. The genotype file itself contained information about 392,214 individuals. Through literature search (Tandon et al., 2010) and online databases (sundhed.dk), 71 different antipsychotics used in treatment of schizophrenia was identified (Table 2).

Table 2. List of first- and second-generation antipsychotics medication prescribed for schizophrenia.

First-generation antipsychotics (N=53)	
Acetophenazine Benperidol Blonanserin Bromperidol Butaperazine Chlorproethazine Chlorpromazine Chlorprothixene Clocapramine Clopenthixol Clotiapine Cyamemazine Dixyrazine Droperidol Fluanisone Flupenthixol Flupentixol Fluphenazine Fluspirilene Haloperidol Levomepromazine Loxapine Melperone Molindone Moperone Mosapramine Nemonapride	Oxypertine Penfluridol Perazine Periciazine Perphenazine Phenothiazine Pimozide Pipamperone Piperacetazine Pipoptiazine Prochlorperazine Promazine Propericiazine Sulforidazine Sulpiride Sultopride Thiopropazate Thioproperazine Thiothixene Thioridazine Tiapride Timiperone Trifluoperazine Trifluperidol Triflupromazine Zuclopenthixol
Second-generation antipsychotics (N=18)	
Benzo (diaze- or thiaze-) pines • Asenapine • Clozapine • Olanzapine • Quetiapine • Zotepine Amisulpride Aripiprazole Brexpiprazole	Cariprazine Iloperidone Lumateperone Lurasidone Paliperidone Perospirone Pimavanserin Risperidone Sertindole Ziprasidone

Using the list of first-, and second-generation antipsychotics from table and the BNF code corresponding to antipsychotics the rows of the data was filtered based on two conditions: A BNF

code of “0402” or “04.02” or a drug name containing any of the antipsychotic drug names (combining both first and second-generation drugs). 26,907 entries matched the BNF criteria and using the list of specific antipsychotic drugs 23,868 entries were identified. This resulted in a new file containing 306,413 lines, representing 30,723 unique individuals with antipsychotic prescriptions.

Due to inconsistencies in the drug_name column two new columns were generated. One contained only the drug and the second contained dosage information. The rest of the information was filtered out.

Additional Data Sets

A separate data sets for individuals with first-generation, second-generation, and benzo (diaze- or thiaze-) pine prescriptions was ceated. The breakdown is as follows:

- First-gen prescriptions: 106,297 lines (26612 unique individuals)
- Second-gen prescriptions: 77,207 lines (1525 unique individuals)
- Benzo prescriptions: 52,332 lines (1115 unique individuals)

I also identified 619 individuals who have been prescribed both first-generation and second-generation antipsychotics, and 10 individuals who had been prescribed Clozapine which is often used in treatment resistant cases. For each of the 30,723 unique individuals the total number of prescriptions, and the number of unique drugs prescribed was calculated.

Additional Patient-Level Metrics

To test the hypothesis that different treatment parameters may correlate with PRS and schizophrenia severity an additional dataset exclusively for schizophrenia patients was created. I identified 704 individuals with a schizophrenia diagnosis in the genotype file. Among these, only 283 individuals had prescription data available. The 283 schizophrenias-diagnosed patients collectively had 3,353 unique antipsychotic prescriptions but only 218 individuals had received antipsychotic prescriptions. Based on this information a binary dataset (n=283), with 218 cases (individuals with antipsychotic prescriptions) and 65 controls (individuals without antipsychotic prescriptions) was created. For each patient the total number of antipsychotics prescriptions, the number of distinct drugs, and the duration of treatment was calculated. The duration of treatment was determined based on the earliest and latest prescriptions of antipsychotics. Furthermore, it was determined whether the individual had a clozapine prescription or not.

The relationship between PRS and these additional patient metrics was examined. I performed a Spearman rank correlation to analyze the relationship between PRS and duration (days) of antipsychotic prescriptions. For the two binary datasets comparing patients with antipsychotic

prescriptions to patients without, and whether the patient had a clozapine prescription, a T-test was performed to test the difference in PRS means between the groups. A linear model for the number of total and distinct prescriptions was fitted to the data.

Phenotype creation

To create phenotypes relevant to the psychiatric disorders, I utilized the International Classification of Diseases (ICD-10) codes available for individuals in the UK Biobank (field 41270). Specifically, I focused on the following ICD-10 codes:

- Schizophrenia: F20
- Depression: F32
- Bipolar Disorder: F31

These codes were used to identify individuals with a diagnosis of these psychiatric disorders. For each individual, I searched for the presence of the respective ICD-10 code. If the code was present, a value of 1 was assigned; otherwise, a value of 0 was assigned to create a phenotype for each of the three psychiatric disorders.

Prescription phenotypes

Seven phenotype files in the format "ID ID Pheno", containing identification number in the first two columns, and phenotype values in the third, were created. The phenotypes are as follows:

- Antipsychotics prescription or not (Binary):
    - Case (1): Individuals with antipsychotic prescriptions (30,723 individuals).
    - Control (0): Individuals without antipsychotic prescriptions.
- First-Generation Antipsychotics (Binary Phenotype):
    - Case (1): Individuals with first-generation antipsychotic prescriptions (26,612 individuals).
    - Control (0): Individuals without first-generation antipsychotic prescriptions.
- Second-Generation Antipsychotics (Binary Phenotype):
    - Case (1): Individuals with second-generation antipsychotic prescriptions (1,525 individuals).
    - Control (0): Individuals without second-generation antipsychotic prescriptions.
- Benzodiazepines (Binary Phenotype):
    - Case (1): Individuals with benzodiazepine prescriptions (1,115).
    - Control (0): Individuals without benzodiazepine prescriptions.

- Individuals Prescribed Both First and Second-Generation Antipsychotics (Binary Phenotype):
    - Case (1): Individuals who have been prescribed both first-generation and second-generation antipsychotics (619 individuals).
    - Control (0): Individuals who have not been prescribed both generations of antipsychotics.
- Clozapine Phenotype (Binary Phenotype):
    - Case (1): Individuals with a clozapine prescription (10 individuals).
    - Control (0): Individuals without a clozapine prescription.
- Number of Prescriptions per Individual (Continuous Phenotype). Out of the 30,723 individuals with antipsychotic prescriptions, 14,782 have more than one prescription.

Predictors for schizophrenia

From the review of schizophrenia seven additional traits, which were also obtainable from the UK Biobank data, and previously found to be associated with schizophrenia were identified. The phenotypes are listed in table 3. The continuous phenotype for standing height (Data field 50) was used a non-psychiatric control.

Table 3. Description of the alternative severity phenotype for schizophrenia.

Phenotype	Coding	Description	UK Biobank code	Citation	Cases
First psychotic episode Code 2046	Continuous	Age at first psychotic episode. Range: 0 (As long as patient can remember/always) – 76 years old.	Code 2046	(Zhan et al., 2023)	4860
Handedness Code 1707	Binary	Handedness. 0 Right-handed or 1 left- or mixed handed	Code 1707	(Dragovic & Hammond, 2005)	43.899
Vitamin D deficiency Code 30890	Continuous	Vitamin D deficiency.	Code 30890	(McGrath et al., 2010)	357.488
Winter birth Code 52	Binary	Based on month of birth. 1 winter birth (dec, jan, feb) 0 (remaining months)	Code 52	(Martínez-Ortega et al., 2011)	95.286
Insomnia Code 1200	Continuous	1 never/rarely 2 sometimes 3 usually	Code 1200	(Cohrs, 2008; Robertson et al., 2019)	392.214
Ever smoked tobacco Code 20160	Binary	1 yes 0 no	Code 20160	(Hartz et al., 2018)	235.401
Ever used cannabis Code 20453	Continuous	0 No 1 Yes, 1-2 times 2 Yes, 3-10 times 3 Yes, 11-100 times 4 Yes, more than 100 times	Code 20453	(Henquet et al., 2005; Johnson et al., 2021)	19.799

Brain image phenotypes

The summary statistics from the GWAS of six brain imaging phenotypes that have previously been found associated with schizophrenia were downloaded from the online data server BIG40 (https://open.win.ox.ac.uk/ukbiobank/big40/). The six phenotypes were:

- Volume of grey matter
- Volume of white matter
- Total brain volume (grey+white matter)
- Volume of ventricular cerebrospinal fluid
- Volume of hippocampus
- Volume of grey matter in Superior Temporal Gyrus

Previously it has been found that a reduction in all parts, excluding ventricular volume where an increase in volume, is associated with schizophrenia (Keshavan et al., 2008). These summary statistics were downloaded for males and females separately, with the intention of investigating

whether differences in brain structure between men and women might be able to explain the heterogeneity in disease expression related to sex.

Preprocessing the raw PGC3 schizophrenia summary statistics data

The raw summary statistics data contained data from 7,659,767 variants and was structured into 14 columns, representing the following information: chromosome (CHROM), rsID (ID), base pair position (POS), reference allele (A1), alternate allele (A2), frequency of A1 in cases (FCAS) and controls (FCON), imputation information score (IMPINFO), beta coefficient (BETA), standard error (SE), p-value (PVAL), and effective sample sizes for cases (NCAS), controls (NCON), and total (NEFF). The reference genome used was the Genome Reference Consortium Human Build 37 (GRCh37). Data preprocessing involved a combination of R scripting and Unix shell commands (bash and awk) to ensure data integrity and reformat the summary statistics for compatibility with MegaPRS in LDAK. Information about the correct format for summary statistics in LDAK can be found here: https://dougspeed.com/summary-statistics/. The new summary statistics file was constructed with the extraction of eight columns from the raw file as seen in table 4.

Table 4. PGC3 schizophrenia summary statistics after preprocessing for LDAK format.

Predictor	A1	A2	N	Direction	P
Name of the predictor in "Chr:BP" format	Reference allele	Alternate allele	Sample size NCAS+NCON	Indication of effect for the reference allele	p-value

The allele frequency (AF) was calculated using the formula:

$$AF = \frac{F_{cases} \cdot N_{cases} + F_{controls} \cdot N_{controls}}{N_{cases} + N_{controls}}$$

AF was employed for minor allele frequency (MAF) quality control with a threshold of MAF > 0.01. All predictors passed the MAF quality control, which was anticipated due to prior quality checks by the Psychiatric Genomics Consortium (PGC). The summary statistics were filtered to retain only predictors present within the genotype data, thereby reducing computational demands. Duplicate predictors identified in the PGC data file were removed.

Schizophrenia PRS based on PGC data

The PRS model for schizophrenia was constructed based on the most recent PGC3 Schizophrenia paper summary statistics using MegaPRS in LDAK. Based on the constructed model the PRS for 392,214 cleaned QC UK Biobank individuals was calculated. When calculating the PRS, a

reference panel is required for accounting for regions of high LD. Here I used the UK biobank data in combination with a list of high-LD regions downloaded from https://dougspeed.com/high-ld-regions/. The accuracy of the PRS model was estimated using jackknifing, which reports the $R^2$and the standard deviation. The variance explained by the PRSs (R^2) was converted to the liability scale using a schizophrenia prevalence of 1%.

GWAS prescription and other phenotypes

A GWAS was performed on all described phenotypes to identify genetic associations related to antipsychotic prescriptions using the 392,214 UK Biobank individuals. Logistic regression with a score test was used for the binary phenotypes and single SNP linear regression on the number of prescriptions. For all GWASs the previously described covariates were included in the model. All GWASs were conducted using LDAK (version 5.2). The results of the prescription GWAS analyses were uploaded to FUMA for gene annotation (Watanabe et al., 2017). Results were also used to calculate the heritability of the individual phenotypes, and for genetic correlation analyses with schizophrenia, bipolar disorder, major depression, and height.

Genetic correlation

Genetic correlation was calculated based on a heritability model. The Human Default Model (Berrandou et al., 2023) which assumes that more common SNPs have a higher SNP heritability h^2_{SNP} that less common SNPs was used. The heritability model is used to calculate taggings that specifies the expected heritability tagged by each predictor. Calculation of taggings also require a reference panel. Here, I used the 1000 genome for 404 non-Finnish Europeans. Taggings contain one row for each predictor, and the corresponding value express the relative expected heritability of the predictor.

Software tools

LDAK

LDAK is a tool for genomic analyses including GWAS, PRS and estimating SNP heritability. The software is available from https://dougspeed.com. LDAK version 5.2 was used.

FUMA

FUMA (Functional Mapping and Annotation of GWAS) is an online mapping and annotation tool for GWAS results. FUMA maps significant SNPs from the uploaded summary statistics to genes using the SNP2GENE function. Further biological interpretation is made possible using the

GENE2FUNC function which annotates the function and expression patterns of the identified genes (Watanabe et al., 2017). FUMA is available at https://fuma.ctglab.nl.

FUMA was used to annotate the results of the six prescription phenotypes. Two phenotypes, prescription of first-generation antipsychotics and antipsychotic prescription, were further investigated using FUMA's GENE2FUNC. The GWAS Catalog within FUMA (MacArthur et al., 2017) was used to annotate SNPs to traits. GWAS Catalog is a database of previously reported associations between SNPs and traits. The GENE2FUNC also links to several external databases. I have used GeneCard (https://www.genecards.org) which is a human gene database with information about all annotated and predicted human genes.

Results

The study population from UK Biobank consisted of 392,214 individuals, of which 212,094 (54.1%) are females and 180,120 (45.9%) are males. In total 704 individuals had been diagnosed with schizophrenia (F20 from the ICD-10 chapter). Prevalence of schizophrenia in the study population was ~0.18%. There was a difference in the disease prevalence among males and females, with males having a prevalence of ~0.25% with 445 cases and females ~0.12% with 259 cases. Within the study population, 1319 and 21,559 individuals had a diagnosis of bipolar disorder (ICD-10 code F31) and depression (ICD-10 code F32), respectively. A total of 127 individuals had overlapping diagnosis between schizophrenia and bipolar disorder, and 270 individuals had been diagnosed with both schizophrenia and depression.

SNP heritability of Schizophrenia

The polygenic risk score model was constructed using MegaPRS in LDAK. The best model was a bayesr (mixture of gaussians) with a correlation of 0.4, explaining 16% of the variance in the data. The PRS model was then used to estimate scores for the UK Biobank individuals. A significant higher schizophrenia PRS was found for schizophrenia individuals compared to non-diagnosed individuals ($P < 0.05$; one-tailed $p = 2.2 \times 10^{-16}$) (fig. 1a). The PRS for schizophrenia explains approx. 9% of the variance in lability to the disease, with a liability squared correlation of 0.087 (SD = 0.01). The area under the curve (AUC) was 0.71 (SD = 0.01). Figure 1b shows the prevalence of schizophrenia across strata of schizophrenia PRS. It shows a steep increase in the schizophrenia prevalence in the upper tail.

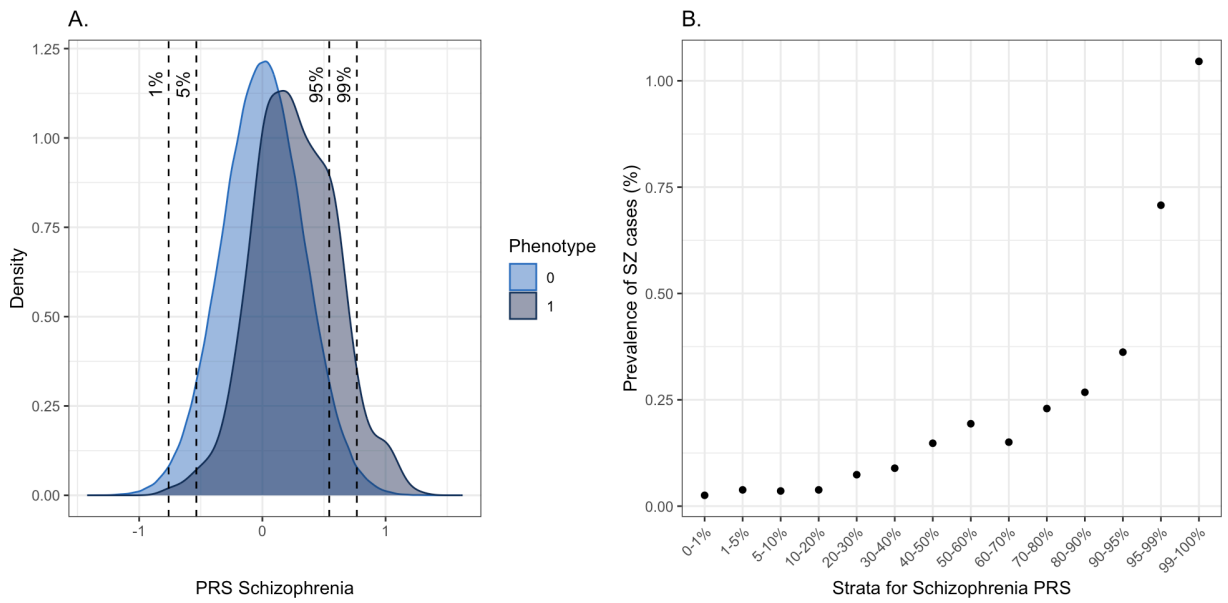


Figure 1. Distribution of Polygenic Risk Scores (PRS) for Schizophrenia. *A. Density Plot: The density plot illustrates the distribution of PRS for individuals with (phenotype 1) and without schizophrenia (phenotype 0). The l, 5, 95, and 99% percentiles are marked.* B. *Prevalence of Schizophrenia Cases: The scatter plot illustrates the prevalence of schizophrenia for different schizophrenia PRS strata. The y-axis represents the prevalence of schizophrenia cases as a percentage, and the x-axis represents the different strata of schizophrenia PRS.*

The heritability of schizophrenia is defined as the proportion of variance in liability explained by genetic factors. As previously mentioned, the lifetime risk of schizophrenia is <1% (Saha et al., 2005), and estimated heritability (h^2) 80% (Sullivan et al., 2003). SNP heritability (h^2_{SNP}), the estimated variance in liability associated with common SNPs, was estimated from the schizophrenia summary statistics data from PGC3 to 30.6% (SD = 0.35%). This is considerably higher than the 24% (SE = 7%) estimated by the PGC3 for the European cohort when restricting to genome-wide significant SNPs (Trubetskoy et al., 2022). The AUC estimate of the PGC3 dataset for all ancestries was estimated to 0.74 (Trubetskoy et al., 2022), while it in this study it was estimated to 0.71 (SD = 0.01). Across all cohorts the PGC3 obtained a liability scale variance explained estimate of 7.3%, while the estimated variance in liability explained in this study is 8.8% (SD = 1%). The conversion to the liability scale was done using a schizophrenia prevalence estimate of 1%. The ascertainment in the data was 40%.

Genetic correlation of psychiatric disorders

The h^2_{SNP} of bipolar disorder has been estimated to explain between 17-23% of the phenotypic variance of the disorder (Stahl et al., 2019), and for major depressive episode h^2_{SNP} has been

estimated to be 8.7% (Wray et al., 2018). In this study, the estimated h^2_{SNP} for bipolar disorder was 38.8% (SD = .7%), while for major depressive disorder was 3.7% (SD = 0.15%). The
The prevalence of major depressive disorder was assumed to be 0.15, and the ascertainment rate was 0.30 (Wray et al., 2018). For bipolar disorder, the prevalence was assumed to be 0.01, and the ascertainment rate was 0.10 (Mullins et al., 2021). Height was used as a non-psychiatric control and have previously been estimated to have a SNP heritability of 45% (Yang, Benyamin, et al., 2010), while I estimated it to be 72.1% (SD = 0.4%) using individuals from the UK Biobank.

Using the SumHer function in LDAK, I estimated the genetic correlations between schizophrenia and two psychiatric disorders: bipolar disorder and major depressive disorder, with height serving as a control variable (fig. 2). The analysis revealed a significant positive genetic correlation between schizophrenia and bipolar disorder, with a correlation coefficient approaching 0.7. Similarly, major depressive disorder exhibited a notable positive genetic correlation with schizophrenia, evidenced by a correlation coefficient near 0.4. In contrast, height, used as a non-psychiatric control trait, showed no significant genetic correlation with schizophrenia. The correlation coefficient for height was approximately zero, confirming its validity as a control in this genetic correlation study.

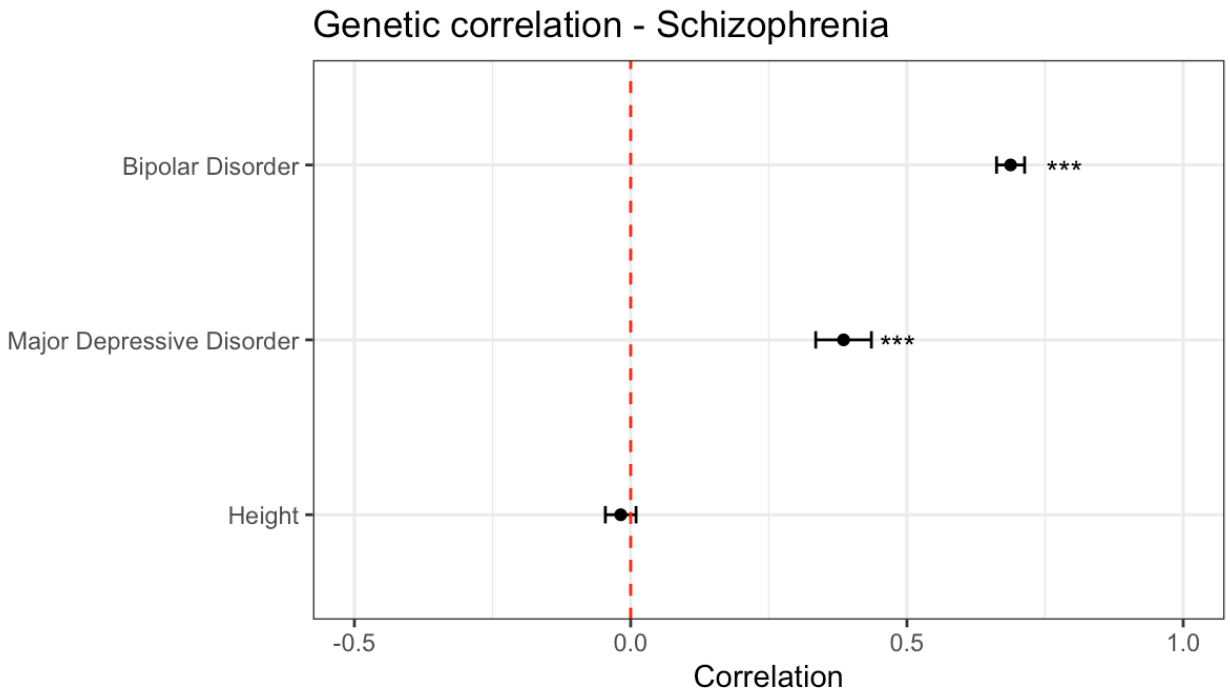


Figure 2. Genetic Correlation Between Schizophrenia and bipolar disorder, major depressive disorder, and Height: *LDAK SumHer was used to estimate the genetic correlations between schizophrenia and two related psychiatric diseases; bipolar disorder (BP) and major depressive disorder (MDD). Height served as a control. The x-axis represents the correlation values, ranging from -0.5 to 1.0. The error bars denote the 95% confidence interval of the correlation estimates. Asterisks mark the significance of the correlation corresponding to a two-tailed z-score: * p≤0.05, ** p≤0.01, *** p≤0.001.*

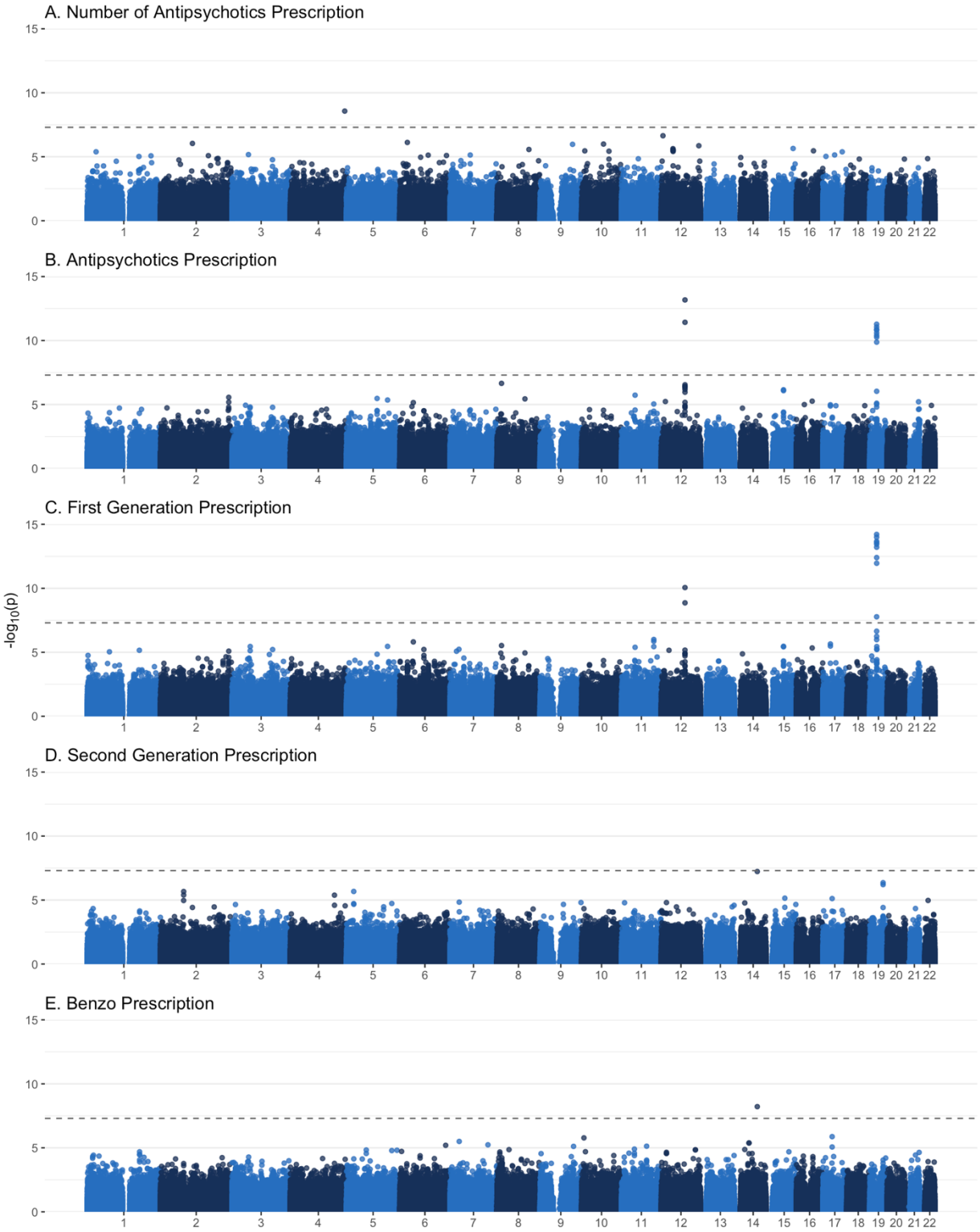
Prescription data

GWAS results

The GWAS results of the six prescription phenotypes are presented as Manhattan plots in figure 3. Except for second-generation antipsychotic prescription (fig. 3d) all other phenotypes found at least one significant SNP passing the genome-wide significance level.

For the number of antipsychotics prescribed the GWAS identified one genomic risk loci on chromosome 4 (fig. 3a). Results for the binary phenotypes antipsychotic prescription or not, and first-generation antipsychotic prescription were very similar (fig. 3b and c). Both analyses identified two genomic risk loci located on chromosome 12 and 19. The loci are mapped to two genes: OTOGL and ZNF91. The GWAS of the benzo prescriptions identify a single genomic risk loci (fig. 3e) which FUMA mapped to two potential genes, one intronic and one untranslated. The

results of the clozapine GWAS identified 77 genomic risk loci (fig. 3f) that FUMA mapped to 73 different genes.



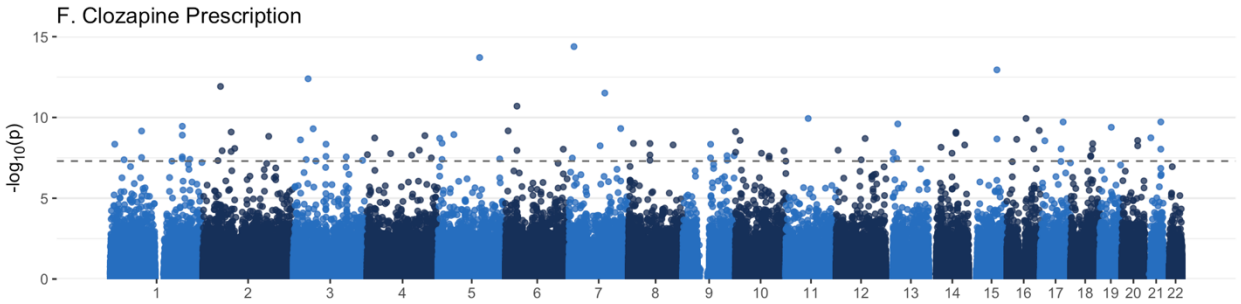


Figure 3. Manhattan plot showing the genome-wide association study (GWAS) results for the six prescription phenotypes*. Each point represents a single nucleotide polymorphism (SNP). The -log10 p-values from the Wald test (A.) and the score tests (B.-F.) are plotted against their genomic positions. The red line represents the genome-wide significance threshold ($p < 5\times10^{-8}$). Chromosomes are labeled on the x-axis.*

PRS correlation with the prescription data

Five patient-level metrics were calculated and investigated for correlation with the patient's PRS for schizophrenia. The results showed that several non-significant but positive correlations suggesting a relationship between the PRS and disease severity (fig. 4).

Antipsychotic prescription

Within the available data 283 individuals were identified who had both a schizophrenia diagnosis and primary care linked. Out of this cohort, 218 patients were documented to have received at least one antipsychotic medication according to the registry. I compared the mean PRS of patients with a prescription and patients without a prescription using a T-test. The analysis revealed that the mean PRS for patients with a prescription was 0.29, while for those without a prescription it was 0.25. The T-test indicated that the difference in means was not statistically significant ($p = 0.43$).

Total number of prescriptions

The total number of prescriptions was plotted against the PRS for schizophrenia in the sample of 218 individuals with both a diagnosis and prescription data (fig. 4c). The association between the polygenic risk score and the number of total prescriptions was investigated using a linear model. The findings indicated that the number of prescriptions had a non-significant effect on the PRS, ($p = 0.82$).

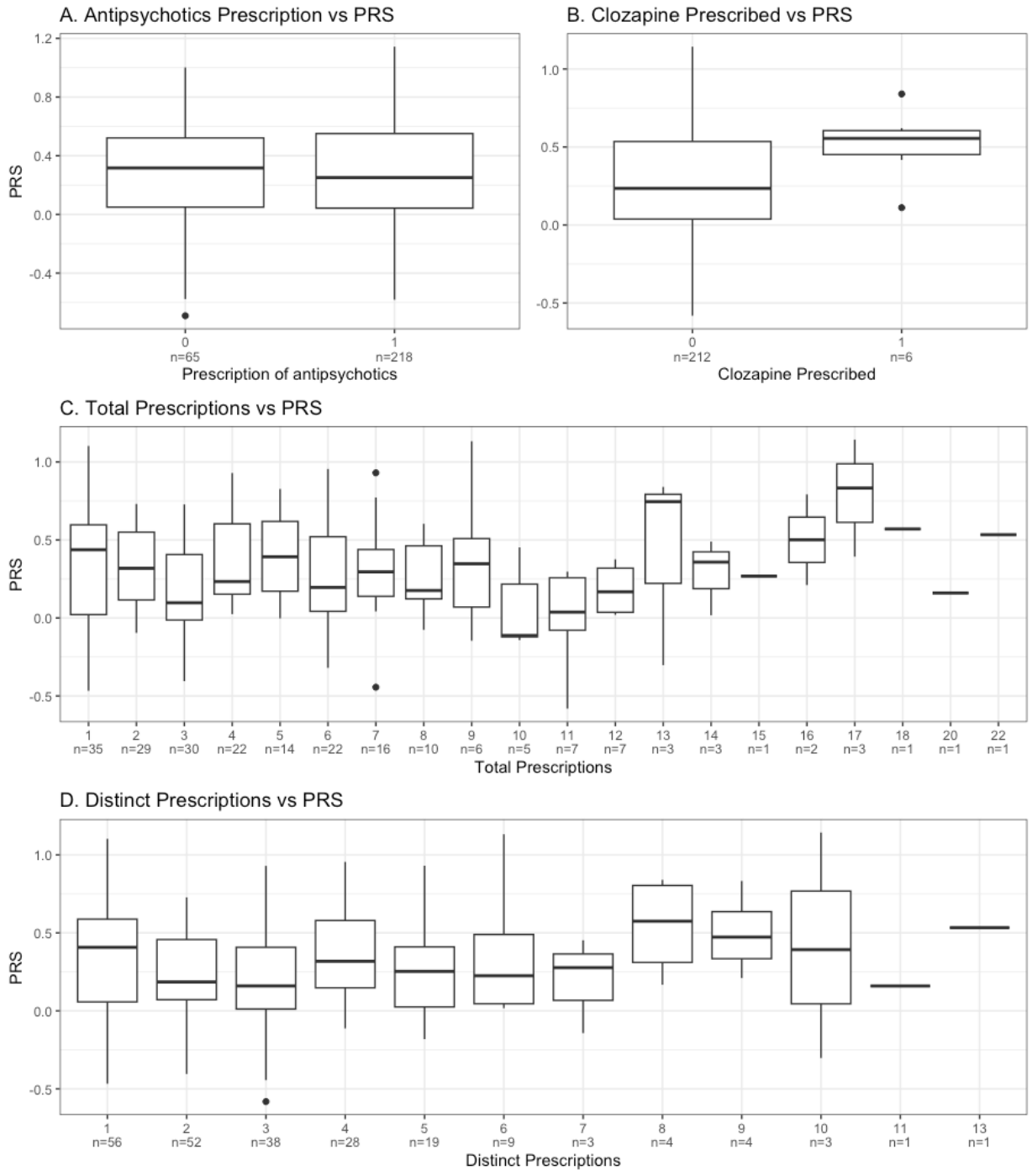


Figure 4. Association of Polygenic Risk Scores with Antipsychotic Prescription Patterns. *A. Antipsychotics Prescription vs PRS. Boxplot comparing the PRS of schizophrenia diagnosed individuals to antipsychotic prescription status. Of 283 patients, 65 did not have an antipsychotic prescription (0), while 218 had at least one antipsychotic prescription (1). B. Clozapine Prescribed vs PRS: Comparison between the PRS of individuals prescribed clozapine (1, n=6) versus individuals not prescribed clozapine (0, n=212). C. Total Prescriptions vs PRS: Boxplot of the relationship between total prescription counts and PRS. D. Distinct Prescriptions vs PRS: The number of distinct antipsychotic drugs plotted against the schizophrenia PRS of the individuals.*

Number of distinct prescriptions

I examined the relationship between the number of distinct prescriptions and the PRS for schizophrenia among the 218 patients. No clear trend was visible in the data (fig. 4d). A linear model was constructed to investigate whether PRS could be predicted by the number of distinct antipsychotic medications the patient had been prescribed. The correlation coefficient was found to be 0.07, and the model's findings suggested no significant correlation ($p = 0.32$).

Clozapine prescription

Of the 218 patients with accessible with available prescription records 6 individuals had been prescribed with clozapine (fig. 4b). The PRS values between the group of patients with a clozapine prescription and the patients without was investigated using a T-test. The group not prescribed clozapine had a mean PRS of 0.29, while the clozapine group had a mean PRS of 0.52. The T-test results indicated that the mean PRS values of the two groups did not differ significantly ($p = 0.07$).

Duration of antipsychotic prescription

The relationship between patients' PRS and the duration (in days) of antipsychotic medication prescriptions and the PRS for schizophrenia was examined for the 218 individuals with available data Using a linear model (fig. 5). The model shows a slight upward trend indicating a positive correlation between PRS and the duration of antipsychotic treatment period. The correlation coefficient estimated using a Spearman's rank correlation (ρ=0.08) suggesting a slight positive trend, implying that higher PRS Schizophrenia scores may be associated with longer durations. However, results were non-significant ($p = 0.24$).

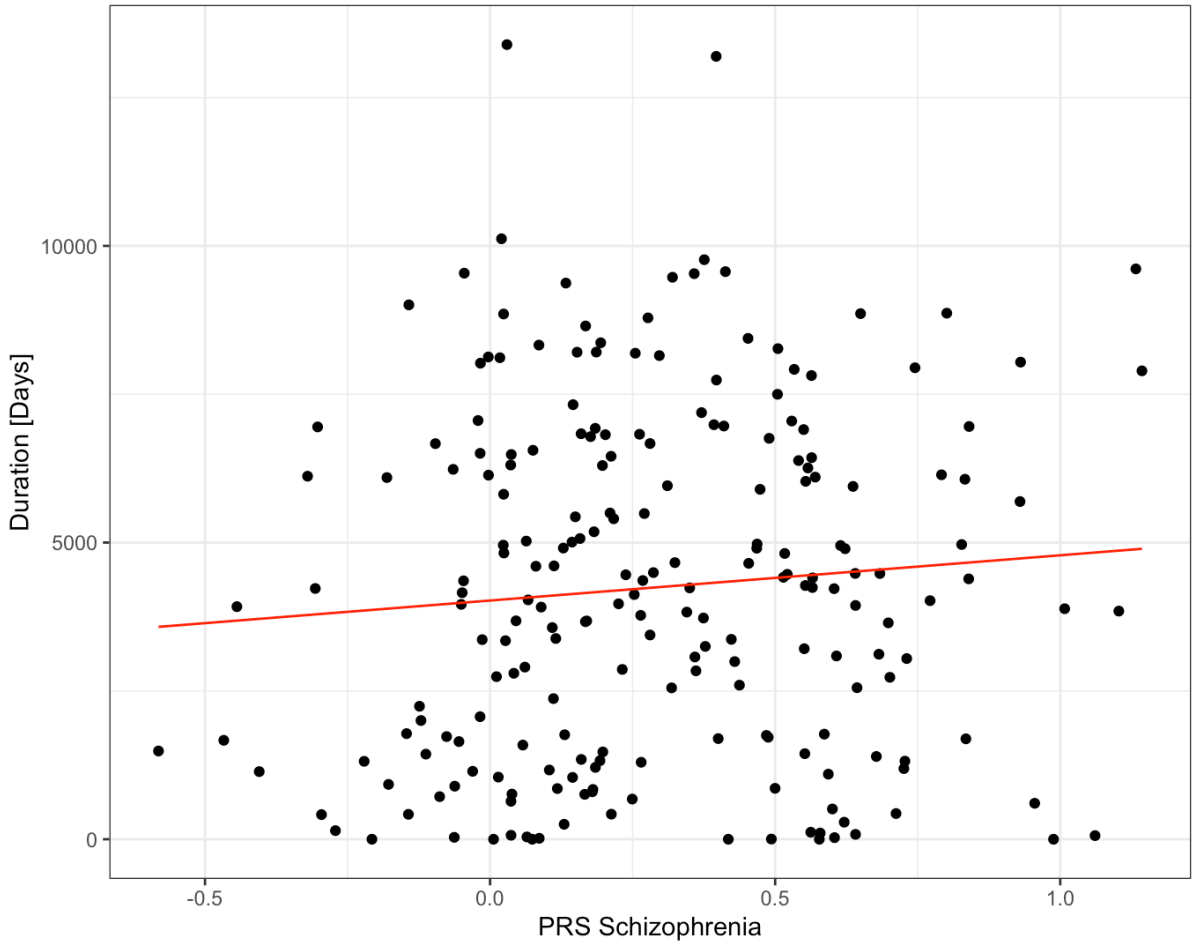


Figure 5. Association Between PRS Schizophrenia and Duration of medical treatment*. Scatterplot with linear regression lines of PRS predicting duration of antipsychotic treatment (days).*

Genetic correlation

Schizophrenia vs. Prescription

The data suggests a complex relationship between different classes of medication and psychiatric traits, with varying degrees of positive correlations (fig. 6). Antipsychotics demonstrated a weak positive correlation with schizophrenia and bipolar disorder, and a moderate and highly significant correlation with Major Depressive Disorder. However, they showed a weak negative correlation with height. The number of antipsychotics prescribed exhibited a moderate positive correlation with all psychiatric traits, but no correlation with height. First-generation medications did not show any correlation with schizophrenia and bipolar disorder but had a moderate correlation (approximately 0.4) with major depressive disorder. Interestingly, these medications showed a negative correlation with height. Second-generation and benzodiazepines medications displayed a strong positive correlation with schizophrenia and bipolar disorder, however with high inaccuracy.

They also showed a strong positive correlation with major depressive disorder, and no correlation with height. The large 95%-confidence interval indicating a high inaccuracy. Clozapine demonstrated a positive correlation, which was only significant for bipolar disorder. None of the medications showed any strong correlation with height.

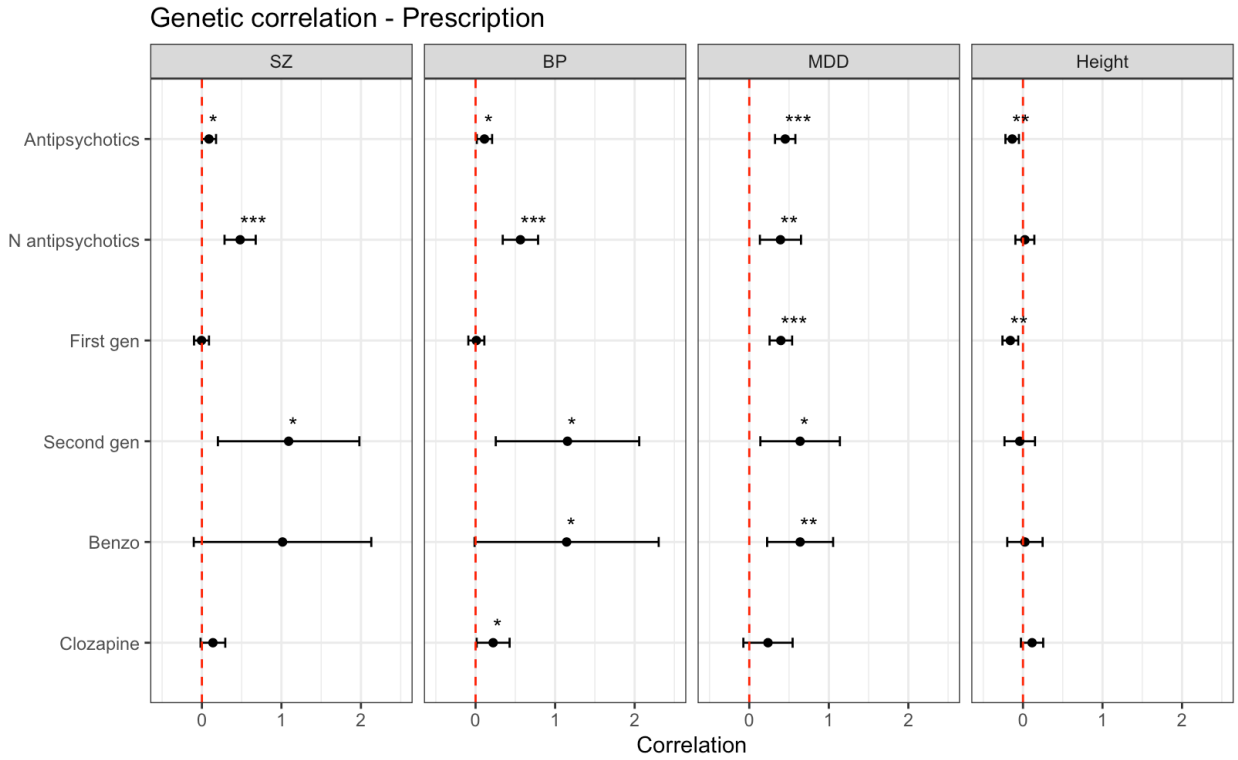
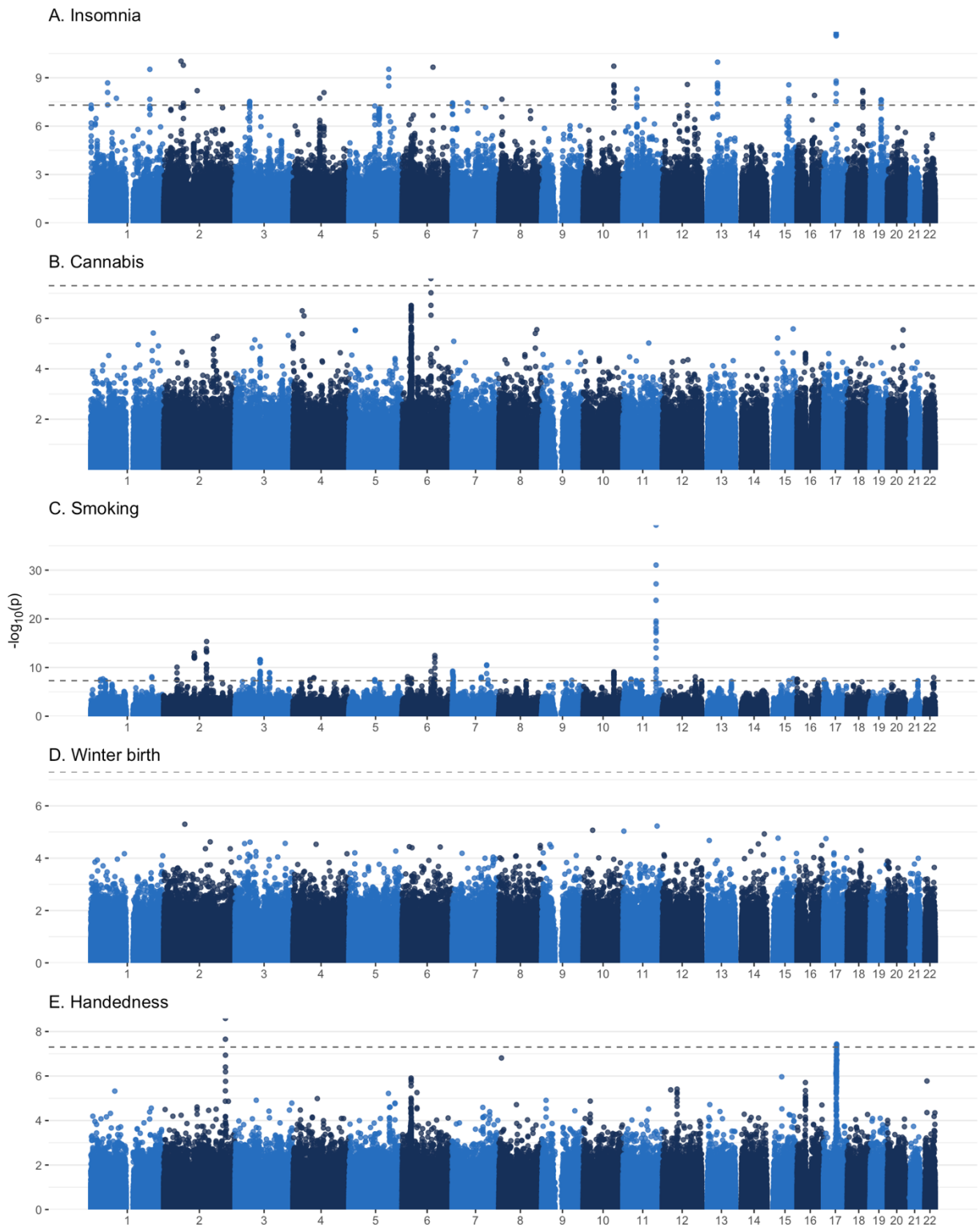


Figure 6. Genetic Correlation Between Antipsychotic Prescription Medications and Psychiatric Disorders: *The subgraphs represent correlations with Schizophrenia (SZ), Bipolar Disorder (BP), Major Depressive Disorder (MDD), and Height (as a non-psychiatric control). Each prescription class (Antipsychotics or not, Number of antipsychotics, First generation antipsychotics, Second generation antipsychotics, Benzo, and Clozapine prescription) is evaluated for all four traits. The error bars represent 95%-confidence intervals, and asterisks denote statistical significance of two-tailed z-score: * p≤0.05, ** p≤0.01, *** p≤0.001.*

Schizophrenia vs. Other phenotypes

The genome-wide association studies (GWAS) aimed to identify genetic variants associated with previously schizophrenia associated traits. The results are presented in Manhattan plots, depicting the significance of each association across the human chromosomes (fig. 7). For insomnia (fig. 7a) 24 genomic risk loci were identified, which could be mapped to 109 different protein-coding genes using FUMA. Only one genomic risk locus was identified for cannabis use (fig. 7b), and no genes could be mapped to this locus. A total of 36 genomic risk loci were identified, with 97 genes mapped to these loci for smoking (fig. 7c). For winter birth (fig. 7d) and Psychosis Age (fig. 7g) the GWAS did not identify any significant hits. Two genomic risk loci on chromosomes 2 and 17 were identified for handedness (fig. 7e), which could be mapped to 12 protein-coding genes.

Vitamin D (fig. 7f) showed a some significantly associated loci on chromosomes 1, 4, 11, and 19. A total of 54 genomic risk loci were identified for this phenotype, which could be mapped to 321 different genes. The largest number of genomic risk loci were identified for height (fig. 7h), with a total of 834. These loci could be mapped to 4509 genes according to FUMA.


A. Insomnia
B. Cannabis
C. Smoking
D. Winter birth
E. Handedness
-log10(p)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22

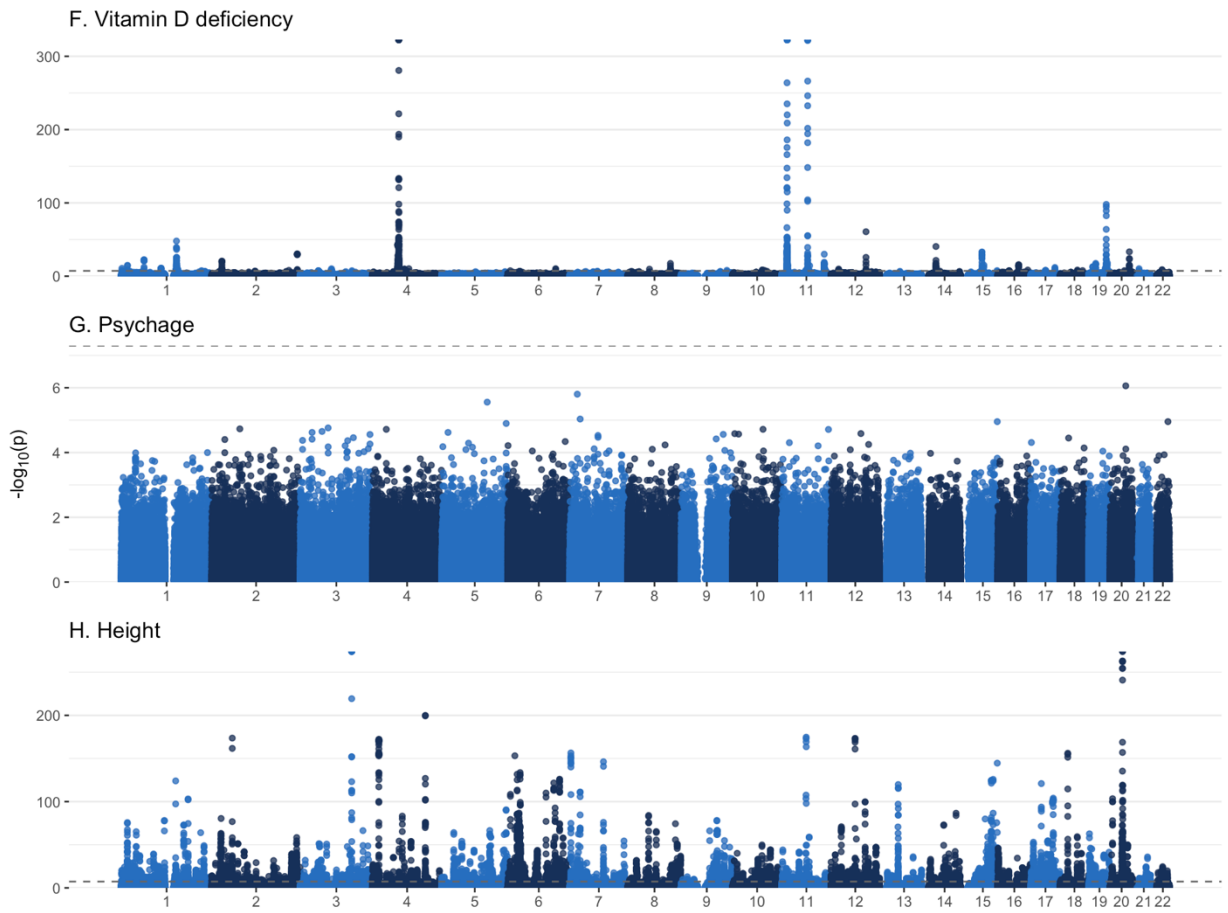


Figure 7. Manhattan plot showing the genome-wide association study (GWAS) results for the eight phenotypes*. Each point represents a single nucleotide polymorphism (SNP). The -log10 p-values from the Wald test (A.) and the score tests (B.-F.) are plotted against their genomic positions. The red line represents the genome-wide significance threshold ($p < 5\times10^{-8}$). Chromosomes are labeled on the x-axis.*

The genetic correlations between various health-related phenotypes and psychiatric disorders were examined using genetic correlation. The variables included vitamin D deficiency, age at first psychosis, winter birth, insomnia, handedness, smoking, and cannabis use. These variables have previously been associated with schizophrenia. Height was included as a non-psychiatric control (fig. 8). Vitamin D deficiency showed a weak negative correlation with psychiatric diseases. This correlation was statistically significant for schizophrenia and major depressive disorder but not for bipolar disorder. Psychosis age showed a negative correlation with all phenotypes, including height. However, error estimates are unreported due to small sampling size and the correlation is not significant. Winter birth did not show any significant genetic correlation. Insomnia showed a significant moderate positive correlation with major depressive disorder and a negative correlation with height. However, it did not show any significant correlation with bipolar disorder and schizophrenia.

Handedness was significantly positively correlated with both schizophrenia and bipolar disorder but did not show any correlation for major depressive disorder or height. Smoking and cannabis use showed a positive correlation with all the psychiatric disorders considered and height, however, the cannabis correlation with major depressive disorder was non-significant. Height was not correlated with schizophrenia and bipolar disorder. However, it showed a significant correlation with major depressive disorder.

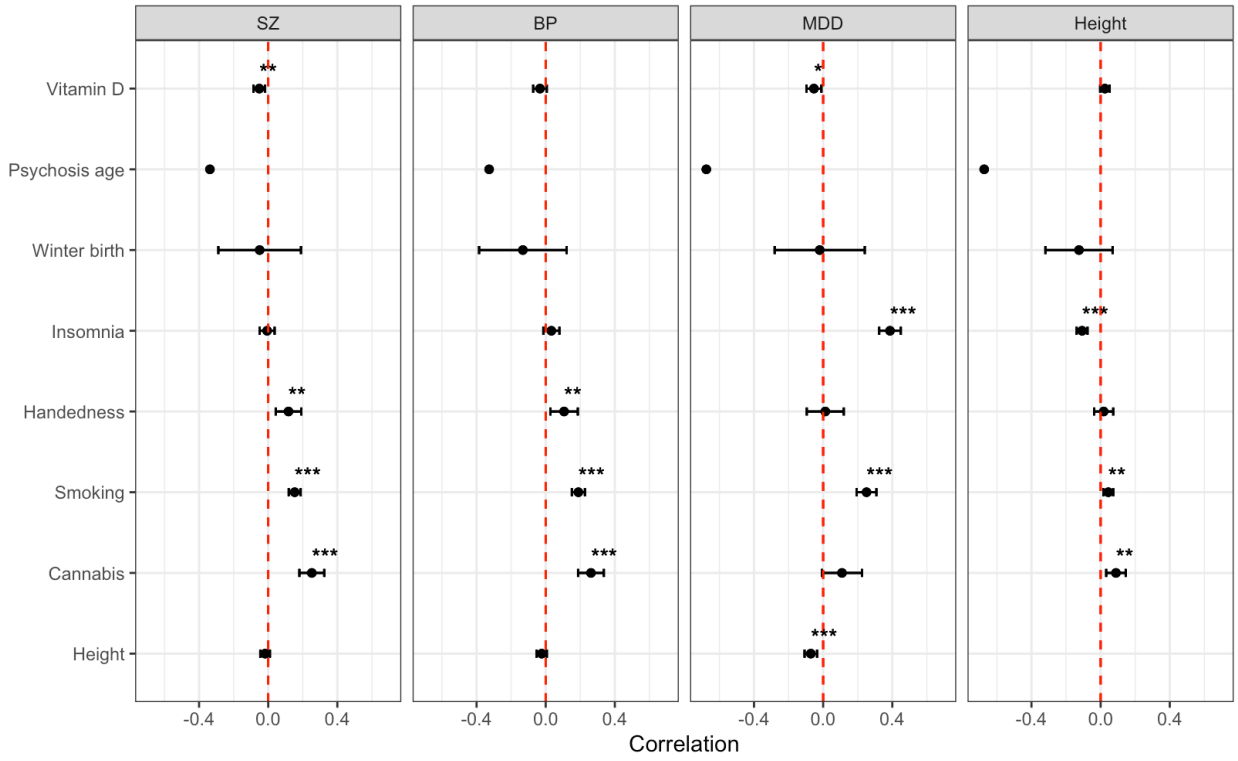


Figure 8. Genetic Correlation Between Health-related Phenotypes and Psychiatric Disorders: *The health-related variables (e.g., Vitamin D, Psychosis age, Winter birth, Insomnia, Handedness, Smoking, Cannabis use) have all previously been found to be associated with schizophrenia. The subgraphs represent correlations with Schizophrenia (SZ), Bipolar Disorder (BP), Major Depressive Disorder (MDD), and Height (as a non-psychiatric control). The error bars represent 95%-confidence intervals, and asterisks denote statistical significance of two-tailed z-score: * p≤0.05, ** p≤0.01, *** p≤0.001.*

Discussion

The study constructed a PRS model for schizophrenia using the latest schizophrenia summary statistics from the Psychiatric Genomics Consortium (PGC) and calculated the schizophrenia PRS for 392,214 individuals from the UK Biobank. The relationship between schizophrenia PRS and antipsychotic prescriptions in individuals with a schizophrenia diagnosis was analyzed. To my knowledge, this is the first study to investigate how schizophrenia PRS is associated with the severity of schizophrenia when using primary care prescription medications as a severity measure of the illness. Other studies have previously investigated the association between medication-use and the associated disease in UK Biobank (Wu et al., 2019), however the GWASs were conducted using self-reported medication-use and not prescriptions provided by healthcare professionals.

Interpretation of results

The primary aim of this thesis was to investigate the biological basis and genetic factors associated with schizophrenia, with primary hypothesis being the existence of a positive correlation between the PRS for schizophrenia and the severity of the disease.

This study found that the PRS for schizophrenia was significantly higher in cases compared to controls, with PRS explaining approximately 9% of the variance in liability to the disease, confirming the hypohesis. Furthermore, the higher prevalence of cases in the upper tail of the PRS distribution presents a clear trend suggesting a correlation between PRS and risk of diagnosis. These results successfully replicate past findings suggesting that polygenic scores can explain a part of the development of the disease (Trubetskoy et al., 2022). The higher estimate of SNP heritability (h^2_{SNP}) in this study compared to the PGC schizophrenia study by Trubetskoy et el. (2022) which uses the same data, may be explained due the use of an improved heritability model (Speed et al., 2017; Speed et al., 2020). Similarly, this may also explain the increase in estimated h^2_{SNP} for bipolar disorder compared to previous studies.

Genetic correlation between psychiatric diseases

Due to previous studies showing association between schizophrenia and other psychiatric disorders, and because of overlap of the disorders within the cases, the genetic correlation between schizophrenia and bipolar and major depressive disorder was investigated. Results of this study found a strong genetic correlation between schizophrenia and bipolar disorder, as well as a moderate genetic correlation between major depressive disorder and schizophrenia. These results are expected due to the cooccurrence of the disorders within the UK Biobank sample. Furthermore, these results are consistent with previous studies, including those by Bulik-Sullivan et al. (2015), which reported similar genetic correlations between schizophrenia and bipolar disorder and major

depressive disorder using three different methods (Bulik-Sullivan et al., 2015). Lichtenstein et al. (2009) estimated the genetic correlation between schizophrenia and bipolar disorder at to 0.60 which is successfully replicated. The genetic correlation between schizophrenia and major depressive disorder is lower, indicating more genetic differences between the two disorders, than between schizophrenia and bipolar disorder.

Prescription as severity measures

To investigate the hypothesis of an association between schizophrenia PRS and the severity of schizophrenia, I explored whether a higher PRS had a positive correlation with five patient-level metrics acting as measures of severity. Results showed a positive, however non-significant, correlation between PRS and these patient-level metrics. The comparison of PRS between schizophrenia patients with and without antipsychotic medication prescription did not show any significant difference in PRS between the two groups. Under the assumption that patients without the need for antipsychotic medication have a less severe course of illness it would be expected that the PRS of these individuals would be lower compared to patients requiring medications. However, results showed very similar and no significant difference between means of the two groups, indicating that the genetic liability for schizophrenia, as measured by the PRS, may not be associated with the severity of the course of illness measured as the need for antipsychotic medication. Several limitations of the available data may complicate the detection of a pattern if a true relationship exists. The medication data only include antipsychotics prescribed by primary health care providers, meaning that prescriptions from hospitalization may be missing. Furthermore, non-adhering patients may not be accurately represented in the data, which could leave out a group of patients with potentially different PRS profiles. These patients may have a higher genetic risk as non-adherence is often associated with more severe disease presentation, cognitive impairment, and comorbidities such as substance abuse and alcohol use (Ascher-Svanum et al., 2006; Haddad et al., 2014).

Harrow & Jobe (2013) reported that most patients continuously on antipsychotics for prolonged periods experienced frequent psychotic symptoms of at least moderate severity and a decrease in function. They further conclude that evidence from longitudinal studies suggest that not all schizophrenia patients require ongoing antipsychotic treatment over an extended time, with several instances of schizophrenia patients with positive outcomes without continuous medication (Harrow & Jobe, 2013). The total number of prescriptions and duration of treatment may therefore have potential, to some extent, for serving as a proxy for severity or course of illness. The expectations related to this was that a greater genetic predisposition to schizophrenia would result

in a longer treatment period, however the results of this study did not find a significant relationship between PRS and these measures. If such a relationship exists, the inability to detect it is most likely due to the nature of constructed variables. When analyzing the association between the total number of antipsychotic and the duration of treatment no other variables were considered. It is plausible that the model could be improved by adding the patients age or diagnosis date as a covariable as it is reasonable to expect longer treatment, and thereby higher number of antipsychotic prescriptions, in older individuals or patients who have had the diagnosis for a long time.

Treatment resistance

In this study, two different metrics were used as measures for treatment resistance, firstly clozapine prescription and secondly high number of distinct antipsychotic prescriptions. The assumption that clozapine prescription can be used as a proxy for treatment resistance has been used in several previous studies (Facal & Costas, 2023; Pardiñas et al., 2022; Wimberley et al., 2017). However, I have not been able to identify studies specifically using the number of distinct antipsychotic drug prescriptions as a proxy for treatment resistance. It is reasonable to assume that a higher number of distinct antipsychotic prescriptions may indicate more complex treatment regimens and potentially greater treatment resistance, as the need for multiple drugs with the same therapeutic target suggests the illness is challenging to treat effectively. This provides an alternative metric beyond just clozapine prescription for treatment resistance. Results of this study provided no statistically significant evidence supporting an association between higher schizophrenia PRS and treatment resistance. Previous results have been mixed, with some studies finding an association while others have not. A study by Zhang et al. (2019) found schizophrenia PRS to be significantly predictive of antipsychotic drug efficacy, with patients with a higher PRS for schizophrenia tending to have poorer treatment response compared to patients with lower PRS (Jian-Ping Zhang et al., 2019). Another study, defining treatment resistance as use of clozapine or failure of two sequentially prescribed antipsychotics, also found a significant association between schizophrenia PRS and treatment resistance (Facal & Costas, 2023). However, the meta-analysis conducted in the study revealed a very small effect of PRS on treatment resistance, and therefore deemed the results unfit for clinical use. I line with the results of this current study, Wimberly et al. (2017) reported a non-significant association between the PRS for schizophrenia and treatment resistance defined as clozapine prescription or hospitalization due to schizophrenia during treatment with antipsychotics. Although I did not find statistically significant evidence for an association between the polygenic risk score for schizophrenia and treatment resistance the possibility of the existence of a true relationship cannot be ruled out. The limited sample size of the study may have hindered the capacity to detect an association. A total of six individuals had a clozapine prescription and

few individuals had a high number of distinct antipsychotic prescriptions which greatly reduced statistical power of the study. Despite these limitations, our findings provide a basis for future research, especially clozapine prescription show great promise due to the large difference in mean PRS between the two groups.

GWAS of prescription data

To further investigate the nature of the antipsychotic-prescription derived severity measures, a GWAS was performed for six constructed prescription phenotypes. The aim was to identify genetic variants associated with antipsychotic medication use, with the intention of advancing the understanding of potential biological factors contributing to need for the specific drug. The GWAS successfully identified SNPs associated with the prescription metrics for five of the six phenotypes. Most interesting were the results of the phenotypes of first-generation antipsychotic prescription and antipsychotic prescription. Both analyses identified the same two genomic risk loci located on chromosome 12 and 19, which were mapped to the genes OTOGL and ZNF91 using FUMA. Both genes have been found associated with the development of the inner ear and vertigo. Furthermore, GeneCards (https://www.genecards.org/cgi-bin/carddisp.pl?gene=ZNF91) mentions ZNF91, a zinc finger gene, as a gene associated with "*neurodevelopmental disorder with cataracts, poor growth, and dysmorphic facies*", which suggest that the GWAS has been able to recapitulate results of psychiatric diseases for which this (first-generation) antipsychotic medication is prescribed. Furthermore, zinc finger genes have been identified as potential susceptibility genes for both schizophrenia and bipolar disorder (Squassina et al., 2019; Sun et al., 2015). None of the studies specifically mentions ZNF91, however both studies investigate the role of several other zinc finger in relation to psychiatric disorders and brain functions. Additionally, Squassina et al. (2019) investigates the role of zinc finger genes in relation to psychotropic medication response and finds that zinc finger genes may be implicated in medication response to not only antipsychotics but also mood stabilizers and antidepressants. Significant hits for the other phenotypes were not as interesting. The high number of significantly associated SNPs with clozapine prescription is likely due to the very low number of samples (n=6), which makes gives the GWAS a low detection power.

Genetic correlation between prescription and psychiatric disorders

The genetic correlation between the antipsychotic-prescription derived phenotypes and the psychiatric disorders were calculated to investigate a potential shared genetic architecture of severity, measured as antipsychotic medication use, and psychiatric disorders. Overall, similar trends were observed in schizophrenia and bipolar disorder, suggesting a stronger association and potentially overlapping genetics between the two disorders compared to that of schizophrenia and

major depressive disorder. However, there was also a clear suggestion of overlap with major depressive disorder, which is to be expected from the initial genetic correlation analysis between the disorders. The genetic correlation of the prescription measures with schizophrenia but also with bipolar disorder and depression, makes it challenging to determine specificity for schizophrenia. This is likely because some medications used for schizophrenia are also used for bipolar disorder and depression (Rybakowski, 2023), complicating the interpretation of the results. Furthermore, the overlap in medication usage across different conditions suggests that some genetic findings might be linked to general psychiatric treatments rather than specific to schizophrenia. Particularly interesting are the phenotypes of antipsychotic and first-generation antipsychotic prescription, which showed a clear significant hit in the GWAS (fig. 3b and c). Antipsychotic prescription is only weakly correlated, and first-generation is not at all correlated with schizophrenia and bipolar disorder. However, both phenotypes are significantly correlated with major depression suggesting that this might be a hit for depression and not for schizophrenia. Despite not being able to clearly determine whether the genetic hits are specific to schizophrenia, the method is able to detect shared genetic architecture between these measures and psychiatric disorders. These findings highlight the complex nature of the underlying genetic basis shared across these three psychiatric disorders. Ultimately, to ensure a "pure" schizophrenia hit, individuals with overlapping psychiatric diagnoses should be removed from the data set, however this would significantly reduce the sample size and consequently also the power to detect associations.

Other severity phenotypes

To investigate the hypothesis that genetic correlations exist between schizophrenia and phenotypes linked to more severe courses of illness, eight different traits were selected which have previously been found to be associated with severity of schizophrenia. A GWAS was performed on all phenotypes. Two of the eight GWAS did not identify any significant associations. This was expected for winter birth as it is not a genetic trait, and therefore a GWAS should not identify any risk loci. Age at first psychotic episode on the other hand is a genetic trait. Early onset of schizophrenia has been linked to a higher severity of clinical symptoms, and in a recent review by Zhan et al (2023) authors summarized past GWAS results. Only three GWAS have been performed on age at onset of schizophrenia, however the top genetic loci identified within each study did not overlap with results from the other studies (Zhan et al., 2023). The current study is also not able to replicate previous results, as no variants were found to be significantly associated with the age at first psychotic episode. This could be due to the small sample size of the phenotype (table 3).
As expected, based on the GWAS results, winter birth and age at first psychotic episode show little potential as a severity measure for schizophrenia and other psychiatric disorders. The study did

find a significant genetic correlation between schizophrenia and four of the severity-linked phenotypes, suggesting common genetics between these traits. The correlation with vitamin D deficiency was negative, indication a weak negative relationship. Handedness, smoking and cannabis use was positive. The similarity between the genetic correlations for schizophrenia and bipolar disorder was striking, with the two having almost identical genetic correlation with the phenotypes. These results again highlight the shared genetic architecture between the two disorders. Major depression disorder differed from the other two psychiatric diseases, by being significantly correlated with insomnia, highlighting insomnia as a potential measure for major depressive disorder.

Possible sources of bias in UK Biobank

Previous studies have reported a similar prevalence of schizophrenia between men and women (Abel et al., 2010; Saha et al., 2005), however within the UK Biobank cohort the prevalence of schizophrenia was twice as high for males as for females. In general, the 0.18% prevalence of schizophrenia within the UK Biobank cohort was much lower than the generally accepted ~1% in the general population (Saha et al., 2005), suggesting an underrepresentation of affected individuals which makes the identification of potential genetic associations more challenging. This disparity in prevalence between the UK Biobank cohort and the general population likely reflects participation biases of the UK Biobank. Even though recruitment of participants for the UK Biobank took place across the entire United Kingdom to ensure as much variation as possible in terms of a wide range of variables (Ollier et al., 2005), the cohort is still not fully representative of the UK population due to a lower participation rate among certain demographic groups, such as individuals with lower socioeconomic status (Fry et al., 2017). Fry et al. (2017) suggested that there is evidence suggesting that UK Biobank is suffering under a “healthy volunteer” selection bias. These biases may have a substantial effect on the results of genetic studies (Schoeler et al., 2023), and can potentially limit the detection power. The underrepresentation of individuals with schizophrenia in the UK Biobank cohort highlights the need for more inclusive and representative sampling in genetic studies to better understand the genetic underpinnings of this complex disorder.

At the time of recruiting, participants of the UK Biobank range in age from 40-69 to study the development and progression of diseases in middle to older age groups (Ollier et al., 2005). Schizophrenia, however, is typically diagnosed earlier in life, with the peak onset in the 20s, which could limit the study's ability to capture the full scope of schizophrenia in the population due to multiple factors. One being that there might be less chance of people with severe psychiatric disorders participating in the study, as they may have difficulty engaging with the comprehensive

assessment process required by the UK Biobank. Therefore, being less likely to capture individuals with more severe course of illness who may have significant cognitive and functional impairments that hinder their ability to participate in such a study.

Limitations

Case/control ratio

Several studies have investigated the optimal case/control ratio in (Akobeng, 2016; Turgut & Koca, 2024; Yang, Wray, et al., 2010). Smaller sample size for case/control studies may limit the detection power of the study, while larger sample sizes improve the statistical power and precision of the study findings. However, increasing sample size alone is not enough, the effective sample size also depends on the case/control ratio. The case/control ratio is also an important factor to consider, as it can impact the overall statistical power and efficiency of the study The existing literature suggests that an optimal case-control ratio, often around 1:1, can maximize the statistical power and efficiency of such studies, allowing for more reliable and meaningful conclusions to be drawn from the data (Cai et al., 2023; Turgut & Koca, 2024; Yang, Wray, et al., 2010).

Prescription data

I encountered several challenges related to making the prescription data fit for analyses investigating the use of prescription medication as a measure of severity of schizophrenia. I found two previous studies that illustrates the usefulness and develops tool for handling the primary care linked UK Biobank data (including GP_scripts) (Darke et al., 2021; Stroganov et al., 2022). Both studies succeeded in mapping and linking the data, but also highlight limitations and emphasize the need for data cleaning, curation, and standardization to make the UK Biobank dataset useful for research purposes. Unfortunately, the methods developed did not match the aim of my analyzes so I was unable to use their pipelines for preparing the data.

The GP prescriptions data is not available for all individuals, and contain several inconsistencies in reporting. It was found that the formatting within each column varied substantially. The BNF codes are a British medication reference containing information about the drug, chemical, dosage, and side effects of British prescription drugs. The first two characters of the BNF code informs about the target of the drug, and the following two specifies the section. The fourth chapter (04) refers to central nervous system, and the second section (02) to drugs used in psychoses and related disorders. All antipsychotic medication should therefore be registred with a BNF code with 0402 as the first four characters (https://openprescribing.net/bnf/). Several entries in the dataset were missing BNF codes, and the format of BNF codes varied. Most of the codes didn't reference a

specific drug, but only contained the information about the drug class. Furthermore, there was substantial inconsistencies in drug names and dosage, and several unreported entries.

Conclusion

This was the first study to examine the association between PRS for schizophrenia and the severity of the disorder, using prescription medications from primary care as a measure of severity. Schizophrenia PRS was positively correlated with all prescription level measures of severity, however, these correlations were non-significant. Due to the high genetic correlation with major depressive disorder and bipolar disorder, schizophrenia PRS show potential as a tool for assessing genetic predisposition to other psychiatric traits. In conclusion, the results of the study support the role of PRS as an indicator of genetic predisposition to schizophrenia and highlights its potential utility in predicting the severity of the disorder and treatment resistance. Future research, with larger sample sizes may allow for a more in-depth exploration of the genetic mechanisms underlying schizophrenia and its relationship with other psychiatric disorders. While the methods of this study might have potential predicting severity of psychiatric, larger sample sizes are needed to detect signal and may allow for a more in-depth exploration of genetic mechanisms underlying schizophrenia.

Data availability

The data used for this project were obtained from UK Biobank and from the Psychiatric Genomics Consortium (PGC).

Code availability

The code used for generating and analyzing data for this thesis can be found in GitHub repository: https://github.com/CarolineSophie/Master-s-thesis

Acknowledgement

I would like to thank Doug Speed and Jesper Givskov Sørensen for supervising the project.
I want to thank Jasper Hof for constructive discussions and helpful suggestions for statistical analyses and my code. I appreciate the immense support and encouragement (especially during the last month of writing), and for listening and taking the time to help me when I have needed it.

Leyí Zhang for many great lunches and discussions, and for efficiently preparing data for LDAK format. All other office companions during my thesis process - Takiy, Dora, and Patrick – for being helpful and for sharing advice and experiences. Finally, family and friends for moral support.

References

Abdellaoui, A., Yengo, L., Verweij, K. J. H., & Visscher, P. M. (2023). 15 years of GWAS discovery: Realizing the promise. *The American Journal of Human Genetics*, *110*(2), 179-194. https://doi.org/https://doi.org/10.1016/j.ajhg.2022.12.011

Abel, K. M., Drake, R., & Goldstein, J. M. (2010). Sex differences in schizophrenia. *International Review of Psychiatry*, *22*(5), 417-428. https://doi.org/10.3109/09540261.2010.515205

Akobeng, A. K. (2016). Understanding type I and type II errors, statistical power and sample size. *Acta Paediatrica*, *105*(6), 605-609. https://doi.org/https://doi.org/10.1111/apa.13384

Albus, M. (2012). Clinical Courses of Schizophrenia. *Pharmacopsychiatry*, *45*(S 01), S31-S35. https://doi.org/10.1055/s-0032-1308968

Alfaro-Almagro, F., Jenkinson, M., Bangerter, N. K., Andersson, J. L. R., Griffanti, L., Douaud, G., Sotiropoulos, S. N., Jbabdi, S., Hernandez-Fernandez, M., Vallee, E., Vidaurre, D., Webster, M., McCarthy, P., Rorden, C., Daducci, A., Alexander, D. C., Zhang, H., Dragonu, I., Matthews, P. M., . . . Smith, S. M. (2018). Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage*, *166*, 400-424. https://doi.org/https://doi.org/10.1016/j.neuroimage.2017.10.034

Anderson, C. A., Pettersson, F. H., Clarke, G. M., Cardon, L. R., Morris, A. P., & Zondervan, K. T. (2010). Data quality control in genetic case-control association studies. *Nature Protocols*, *5*(9), 1564-1573. https://doi.org/10.1038/nprot.2010.116

Andlauer, T. F. M., & Nöthen, M. M. (2020). Polygenic scores for psychiatric disease: from research tool to clinical application. *Medizinische Genetik*, *32*(1), 39-45. https://doi.org/doi:10.1515/medgen-2020-2006

Ascher-Svanum, H., Zhu, B., Faries, D., Lacro, J. P., & Dolder, C. R. (2006). A prospective study of risk factors for nonadherence with antipsychotic medication in the treatment of schizophrenia. *The journal of clinical psychiatry*, *67*(7), 1114-1123. https://doi.org/10.4088/JCP.v67n0715

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

Berrandou, T.-E., Balding, D., & Speed, D. (2023). LDAK-GBAT: Fast and powerful gene-based association testing using summary statistics. *The American Journal of Human Genetics*, *110*(1), 23-29. https://doi.org/https://doi.org/10.1016/j.ajhg.2022.11.010

Buckley, P. F., Miller, B. J., Lehrer, D. S., & Castle, D. J. (2008). Psychiatric Comorbidities and Schizophrenia. *Schizophrenia Bulletin*, *35*(2), 383-402. https://doi.org/10.1093/schbul/sbn135

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., Duncan, L., Perry, J. R. B., Patterson, N., Robinson, E. B., Daly, M. J., Price, A. L., Neale, B. M., ReproGen, C., Psychiatric Genomics, C., & Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control, C. (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, *47*(11), 1236-1241. https://doi.org/10.1038/ng.3406

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., & O'Connell, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203-209.

Cai, J., Zhan, J., Arking, D. E., & Bader, J. S. (2023). Priors, population sizes, and power in genome-wide hypothesis tests. *BMC Bioinformatics*, *24*(1), 170. https://doi.org/10.1186/s12859-023-05261-9

Carbon, M., & Correll, C. U. (2014). Clinical predictors of therapeutic response to antipsychotics in schizophrenia. *Dialogues in Clinical Neuroscience*, *16*(4), 505-524. https://doi.org/10.31887/DCNS.2014.16.4/mcarbon

Choi, S. W., Mak, T. S.-H., & O'Reilly, P. F. (2020). Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols*, *15*(9), 2759-2772. https://doi.org/10.1038/s41596-020-0353-1

Cohrs, S. (2008). Sleep Disturbances in Patients with Schizophrenia. *CNS Drugs*, *22*(11), 939-962. https://doi.org/10.2165/00023210-200822110-00004

Correll, C. U., Brevig, T., & Brain, C. (2019). Exploration of Treatment-Resistant Schizophrenia Subtypes Based on a Survey of 204 US Psychiatrists. *Neuropsychiatric Disease and Treatment*, *15*, 3461-3473. https://doi.org/10.2147/NDT.S234813

Correll, C. U., & Schooler, N. R. (2020). Negative Symptoms in Schizophrenia: A Review and Clinical Guide for Recognition, Assessment, and Treatment. *Neuropsychiatr Dis Treat*, *16*, 519-534. https://doi.org/10.2147/ndt.S225643

Darke, P., Cassidy, S., Catt, M., Taylor, R., Missier, P., & Bacardit, J. (2021). Curating a longitudinal research resource using linked primary care EHR data—a UK Biobank case study. *Journal of the American Medical Informatics Association*, *29*(3), 546-552. https://doi.org/10.1093/jamia/ocab260

Dragovic, M., & Hammond, G. (2005). Handedness in schizophrenia: a quantitative review of evidence. *Acta Psychiatrica Scandinavica*, *111*(6), 410-419. https://doi.org/https://doi.org/10.1111/j.1600-0447.2005.00519.x

Elliott, P., Peakman, T. C., & Biobank, o. b. o. U. (2008). The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *International Journal of Epidemiology*, *37*(2), 234-244. https://doi.org/10.1093/ije/dym276

Facal, F., & Costas, J. (2023). Polygenic risk scores for schizophrenia and treatment resistance: New data, systematic review and meta-analysis. *Schizophrenia Research*, *252*, 189-197. https://doi.org/https://doi.org/10.1016/j.schres.2023.01.012

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., Collins, R., & Allen, N. E. (2017). Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *American Journal of Epidemiology*, *186*(9), 1026-1034. https://doi.org/10.1093/aje/kwx246

Galderisi, S., Kaiser, S., Bitter, I., Nordentoft, M., Mucci, A., Sabé, M., Giordano, G. M., Nielsen, M. Ø., Glenthøj, L. B., Pezzella, P., Falkai, P., Dollfus, S., & Gaebel, W. (2021). EPA guidance on treatment of negative symptoms in schizophrenia. *European psychiatry*, *64*(1), e21-e21. https://doi.org/10.1192/j.eurpsy.2021.13

Haddad, P. M., Brain, C., & Scott, J. (2014). Nonadherence with antipsychotic medication in schizophrenia: challenges and management strategies. *Patient Related Outcome Measures*, *5*(null), 43-62. https://doi.org/10.2147/PROM.S42735

Haller, C. S., Padmanabhan, J. L., Lizano, P., Torous, J., & Keshavan, M. (2014). Recent advances in understanding schizophrenia. *F1000 prime reports*, *6*, 57-57. https://doi.org/10.12703/P6-57

Harrow, M., & Jobe, T. H. (2013). Does Long-Term Treatment of Schizophrenia With Antipsychotic Medications Facilitate Recovery? *Schizophrenia Bulletin*, *39*(5), 962-965. https://doi.org/10.1093/schbul/sbt034

Hartling, L., Abou-Setta, A. M., Dursun, S., Mousavi, S. S., Pasichnyk, D., & Newton, A. S. (2012). Antipsychotics in adults with schizophrenia: comparative effectiveness of first-generation versus second-generation medications: a systematic review and meta-analysis. *Ann Intern Med*, *157*(7), 498-511. https://doi.org/10.7326/0003-4819-157-7-201210020-00525

Hartz, S. M., Horton, A. C., Hancock, D. B., Baker, T. B., Caporaso, N. E., Chen, L.-S., Hokanson, J. E., Lutz, S. M., Marazita, M. L., McNeil, D. W., Pato, C. N., Pato, M. T., Johnson, E. O., & Bierut, L. J. (2018). Genetic correlation between smoking behaviors and schizophrenia. *Schizophrenia Research*, *194*, 86-90. https://doi.org/https://doi.org/10.1016/j.schres.2017.02.022

Henquet, C., Murray, R., Linszen, D., & van Os, J. (2005). The Environment and Schizophrenia: The Role of Cannabis Use. *Schizophrenia Bulletin*, *31*(3), 608-612. https://doi.org/10.1093/schbul/sbi027

Howes, O. D., McCutcheon, R., Agid, O., de Bartolomeis, A., van Beveren, N. J. M., Birnbaum, M. L., Bloomfield, M. A. P., Bressan, R. A., Buchanan, R. W., Carpenter, W. T., Castle, D. J., Citrome, L., Daskalakis, Z. J., Davidson, M., Drake, R. J., Dursun, S., Ebdrup, B. H., Elkis, H., Falkai, P., . . . Correll, C. U. (2017). Treatment-Resistant Schizophrenia: Treatment Response and Resistance in Psychosis (TRRIP) Working Group Consensus Guidelines on Diagnosis and Terminology. *The American journal of psychiatry*, *174*(3), 216-229. https://doi.org/10.1176/appi.ajp.2016.16050503

Howes, O. D. D., & Murray, R. M. P. (2014). Schizophrenia: an integrated sociodevelopmental-cognitive model. *The Lancet (British edition)*, *383*(9929), 1677-1687. https://doi.org/10.1016/S0140-6736(13)62036-X

Hübel, C., Gaspar, H. A., Coleman, J. R. I., Hanscombe, K. B., Purves, K., Prokopenko, I., Graff, M., Ngwa, J. S., Workalemahu, T., O'Reilly, P. F., Bulik, C. M., Breen, G., Consortium, A. W. G. o. t. P. G., Meta-Analyses of, G., Insulin-related traits, c., Autism Working Group of the Psychiatric Genomics, C., Bipolar Disorder Working Group of the Psychiatric Genomics, C., Eating Disorders Working Group of the Psychiatric Genomics, C., Major Depressive Disorder Working Group of the Psychiatric Genomics, C., . . . International Headache Genetics, C. (2019). Genetic correlations of psychiatric traits with body composition and glycemic traits are sex- and age-dependent. *Nature Communications*, *10*(1), 5765. https://doi.org/10.1038/s41467-019-13544-0

Jian-Ping Zhang, M.D., Ph.D. ,, Delbert Robinson, M.D. ,, Jin Yu, M.S. ,, Juan Gallego, W. Wolfgang Fleischhacker, M.D. ,, Rene S. Kahn, M.D. ,, Benedicto Crespo-Facorro, M.D. ,, Javier Vazquez-Bourgon, M.D. ,, John M. Kane, M.D. ,, Anil K. Malhotra, M.D. ,, & Todd Lencz, Ph.D. (2019). Schizophrenia Polygenic Risk Score as a Predictor of Antipsychotic Efficacy in First-Episode Psychosis. *American Journal of Psychiatry*, *176*(1), 21-28. https://doi.org/10.1176/appi.ajp.2018.17121363

Johnson, E. C., Hatoum, A. S., Deak, J. D., Polimanti, R., Murray, R. M., Edenberg, H. J., Gelernter, J., Di Forti, M., & Agrawal, A. (2021). The relationship between cannabis and schizophrenia: a genetically informed perspective. *Addiction*, *116*(11), 3227-3234. https://doi.org/https://doi.org/10.1111/add.15534

Keshavan, M. S., Tandon, R., Boutros, N. N., & Nasrallah, H. A. (2008). Schizophrenia, "just the facts": What we know in 2008. Part 3: Neurobiology. *Schizophrenia Research*, *106*(2-3), 89-107. https://doi.org/10.1016/j.schres.2008.07.020

Lee, P. H., Baker, J. T., Holmes, A. J., Jahanshad, N., Ge, T., Jung, J. Y., Cruz, Y., Manoach, D. S., Hibar, D. P., Faskowitz, J., McMahon, K. L., de Zubicaray, G. I., Martin, N. H., Wright, M. J., Öngür, D., Buckner, R., Roffman, J., Thompson, P. M., & Smoller, J. W. (2016). Partitioning heritability analysis reveals a shared genetic basis of brain anatomy and schizophrenia. *Molecular Psychiatry*, *21*(12), 1680-1689. https://doi.org/10.1038/mp.2016.164

Lee, S. H., Wray, N. R., Goddard, M. E., & Visscher, P. M. (2011). Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet*, *88*(3), 294-305. https://doi.org/10.1016/j.ajhg.2011.02.002

Lewis, C. M., & Vassos, E. (2020). Polygenic risk scores: from research tools to clinical instruments. *Genome Med*, *12*(1), 44. https://doi.org/10.1186/s13073-020-00742-5

Lichtenstein, P., Yip, B. H., Björk, C., Pawitan, Y., Cannon, T. D., Sullivan, P. F., & Hultman, C. M. (2009). Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *The Lancet*, *373*(9659), 234-239. https://doi.org/10.1016/S0140-6736(09)60072-6

Littlejohns, T. J., Holliday, J., Gibson, L. M., Garratt, S., Oesingmann, N., Alfaro-Almagro, F., Bell, J. D., Boultwood, C., Collins, R., Conroy, M. C., Crabtree, N., Doherty, N., Frangi, A. F., Harvey, N. C., Leeson, P., Miller, K. L., Neubauer, S., Petersen, S. E., Sellors, J., . . . Allen, N. E. (2020). The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nature Communications*, *11*(1), 2624. https://doi.org/10.1038/s41467-020-15948-9

Lu, Y., Pouget, J. G., Andreassen, O. A., Djurovic, S., Esko, T., Hultman, C. M., Metspalu, A., Milani, L., Werge, T., & Sullivan, P. F. (2018). Genetic risk scores and family history as predictors of schizophrenia in Nordic registers. *Psychological Medicine*, *48*(7), 1201-1208. https://doi.org/10.1017/S0033291717002665

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., Pendlington, Z. M., Welter, D., Burdett, T., Hindorff, L., Flicek, P., Cunningham, F., & Parkinson, H. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res*, *45*(D1), D896-d901. https://doi.org/10.1093/nar/gkw1133

Mäkinen, J., Miettunen, J., Isohanni, M., & Koponen, H. (2008). Negative symptoms in schizophrenia-A review. *Nordic journal of psychiatry*, *62*(5), 334-341. https://doi.org/10.1080/08039480801959307

Marees, A. T., de Kluiver, H., Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., & Derks, E. M. (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int J Methods Psychiatr Res*, *27*(2), e1608. https://doi.org/10.1002/mpr.1608

Martínez-Ortega, J. M., Carretero, M. D., Gutiérrez-Rojas, L., Díaz-Atienza, F., Jurado, D., & Gurpegui, M. (2011). Winter birth excess in schizophrenia and in non-schizophrenic psychosis: Sex and birth-cohort differences. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *35*(7), 1780-1784. https://doi.org/https://doi.org/10.1016/j.pnpbp.2011.07.002

McGrath, J. J., Burne, T. H., Féron, F., Mackay-Sim, A., & Eyles, D. W. (2010). Developmental Vitamin D Deficiency and Risk of Schizophrenia: A 10-Year Update. *Schizophrenia Bulletin*, *36*(6), 1073-1078. https://doi.org/10.1093/schbul/sbq101

Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Xu, J., Bartsch, A. J., Jbabdi, S., Sotiropoulos, S. N., Andersson, J. L. R., Griffanti, L., Douaud, G., Okell, T. W., Weale, P., Dragonu, I., Garratt, S., Hudson, S., Collins, R., Jenkinson, M., . . . Smith, S. M. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature Neuroscience*, *19*(11), 1523-1536. https://doi.org/10.1038/nn.4393

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

Ollier, W., Sprosen, T., & Peakman, T. (2005). UK Biobank: from concept to reality. *Pharmacogenomics*, *6*(6), 639-646. https://doi.org/10.2217/14622416.6.6.639

Pardiñas, A. F., Smart, S. E., Willcocks, I. R., Holmans, P. A., Dennison, C. A., Lynham, A. J., Legge, S. E., Baune, B. T., Bigdeli, T. B., Cairns, M. J., Corvin, A., Fanous, A. H., Frank, J., Kelly, B., McQuillin, A., Melle, I., Mortensen, P. B., Mowry, B. J., Pato, C. N., . . . Consortium, t. S. W. G. o. t. P. G. (2022). Interaction Testing and Polygenic Risk Scoring to Estimate the Association of Common Genetic Variants With Treatment Resistance in Schizophrenia. *JAMA Psychiatry*, *79*(3), 260-269. https://doi.org/10.1001/jamapsychiatry.2021.3799

Patel, K. R., Cherian, J., Gohil, K., & Atkinson, D. (2014). Schizophrenia: overview and treatment options. *P t*, *39*(9), 638-645.

Robertson, I., Cheung, A., & Fan, X. (2019). Insomnia in patients with schizophrenia: current understanding and treatment options. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *92*, 235-242. https://doi.org/https://doi.org/10.1016/j.pnpbp.2019.01.016

Rössler, W., Joachim Salize, H., van Os, J., & Riecher-Rössler, A. (2005). Size of burden of schizophrenia and psychotic disorders. *European Neuropsychopharmacology*, *15*(4), 399-409. https://doi.org/https://doi.org/10.1016/j.euroneuro.2005.04.009

Ruderfer, D. M., Charney, A. W., Readhead, B., Kidd, B. A., Kähler, A. K., Kenny, P. J., Keiser, M. J., Moran, J. L., Hultman, C. M., Scott, S. A., Sullivan, P. F., Purcell, S. M., Dudley, J. T., & Sklar, P. (2016). Polygenic overlap between schizophrenia risk and antipsychotic response: a genomic medicine approach. *Lancet Psychiatry*, *3*(4), 350-357. https://doi.org/10.1016/s2215-0366(15)00553-2

Rybakowski, J. K. (2023). Application of Antipsychotic Drugs in Mood Disorders. *Brain Sciences*, *13*(3), 414. https://www.mdpi.com/2076-3425/13/3/414

Saha, S., Chant, D., Welham, J., & McGrath, J. (2005). A Systematic Review of the Prevalence of Schizophrenia. *PLOS Medicine*, *2*(5), e141. https://doi.org/10.1371/journal.pmed.0020141

Schoeler, T., Speed, D., Porcu, E., Pirastu, N., Pingault, J.-B., & Kutalik, Z. (2023). Participation bias in the UK Biobank distorts genetic associations and downstream analyses. *Nature Human Behaviour*, *7*(7), 1216-1227. https://doi.org/10.1038/s41562-023-01579-9

Smith, S. M., Douaud, G., Chen, W., Hanayik, T., Alfaro-Almagro, F., Sharp, K., & Elliott, L. T. (2021). An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature Neuroscience*, *24*(5), 737-745. https://doi.org/10.1038/s41593-021-00826-4

Speed, D., Cai, N., Johnson, M. R., Nejentsev, S., Balding, D. J., & the, U. C. (2017). Reevaluation of SNP heritability in complex human traits. *Nature Genetics*, *49*(7), 986-992. https://doi.org/10.1038/ng.3865

Speed, D., Holmes, J., & Balding, D. J. (2020). Evaluating and improving heritability models using summary statistics. *Nature Genetics*, *52*(4), 458-462. https://doi.org/10.1038/s41588-020-0600-y

Squassina, A., Meloni, A., Chillotti, C., & Pisanu, C. (2019). Zinc finger proteins in psychiatric disorders and response to psychotropic medications. *Psychiatric Genetics*, *29*(5). https://journals.lww.com/psychgenetics/fulltext/2019/10000/zinc_finger_proteins_in_psychiatric_disorders_and.2.aspx

Stahl, E. A., Breen, G., Forstner, A. J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J. R. I., Gaspar, H. A., de Leeuw, C. A., Steinberg, S., Pavlides, J. M. W., Trzaskowski, M., Byrne, E. M., Pers, T. H., Holmans, P. A., Richards, A. L., Abbott, L., . . . the Bipolar Disorder Working Group of the Psychiatric Genomics, C. (2019).

Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature Genetics*, *51*(5), 793-803. https://doi.org/10.1038/s41588-019-0397-8

Stroganov, O., Fedarovich, A., Wong, E., Skovpen, Y., Pakhomova, E., Grishagin, I., Fedarovich, D., Khasanova, T., Merberg, D., Szalma, S., & Bryant, J. (2022). Mapping of UK Biobank clinical codes: Challenges and possible solutions. *PLOS ONE*, *17*(12), e0275816. https://doi.org/10.1371/journal.pone.0275816

Sud, A., Kinnersley, B., & Houlston, R. S. (2017). Genome-wide association studies of cancer: current insights and future perspectives. *Nature Reviews Cancer*, *17*(11), 692-704. https://doi.org/10.1038/nrc.2017.82

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

Sullivan, P. F., Agrawal, A., Bulik, C. M., Andreassen, O. A., Børglum, A. D., Breen, G., Cichon, S., Edenberg, H. J., Faraone, S. V., Gelernter, J., Mathews, C. A., Nievergelt, C. M., Smoller, J. W., & O'Donovan, M. C. (2018). Psychiatric Genomics: An Update and an Agenda. *Am J Psychiatry*, *175*(1), 15-27. https://doi.org/10.1176/appi.ajp.2017.17030283

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), 1187-1192. https://doi.org/10.1001/archpsyc.60.12.1187

Sun, Y., Hu, D., Liang, J., Bao, Y.-P., Meng, S.-Q., Lu, L., & Shi, J. (2015). Association between variants of zinc finger genes and psychiatric disorders: Systematic review and meta-analysis. *Schizophrenia Research*, *162*(1), 124-137. https://doi.org/https://doi.org/10.1016/j.schres.2015.01.036

Tandon, R., Keshavan, M. S., & Nasrallah, H. A. (2008). Schizophrenia, "Just the Facts" What we know in 2008. 2. Epidemiology and etiology. *Schizophrenia Research*, *102*(1), 1-18. https://doi.org/10.1016/j.schres.2008.04.011

Tandon, R., Nasrallah, H. A., & Keshavan, M. S. (2009). Schizophrenia, "just the facts" 4. Clinical features and conceptualization. *Schizophrenia Research*, *110*(1), 1-23. https://doi.org/https://doi.org/10.1016/j.schres.2009.03.005

Tandon, R., Nasrallah, H. A., & Keshavan, M. S. (2010). Schizophrenia, "Just the Facts" 5. Treatment and prevention Past, present, and future. *Schizophrenia Research*, *122*(1), 1-23. https://doi.org/https://doi.org/10.1016/j.schres.2010.05.025

Trastulla, L., Moser, S., Jiménez-Barrón, L. T., Andlauer, T. F. M., von Scheidt, M., Budde, M., Heilbronner, U., Papiol, S., Teumer, A., Homuth, G., Falkai, P., Völzke, H., Dörr, M., Schulze, T. G., Gagneur, J., Iorio, F., Müller-Myhsok, B., Schunkert, H., & Ziller, M. J. (2023). Distinct genetic liability profiles define clinically relevant patient strata across common diseases. *medRxiv*. https://doi.org/10.1101/2023.05.10.23289788

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

Turgut, A. O., & Koca, D. (2024). The effects of case/control ratio and sample size on genome-wide association studies: A simulation study. *Veterinary Medicine and Science*, *10*(3), e1444. https://doi.org/https://doi.org/10.1002/vms3.1444

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(1), 59. https://doi.org/10.1038/s43586-021-00056-9

van Tol, M.-J., van der Meer, L., Bruggeman, R., Modinos, G., Knegtering, H., & Aleman, A. (2014). Voxel-based gray and white matter morphometry correlates of hallucinations in schizophrenia: The superior temporal gyrus does not stand alone. *NeuroImage: Clinical*, *4*, 249-257. https://doi.org/https://doi.org/10.1016/j.nicl.2013.12.008

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

Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*, *8*(1), 1826. https://doi.org/10.1038/s41467-017-01261-5

Wierenga, L. M., Doucet, G. E., Dima, D., Agartz, I., Aghajani, M., Akudjedu, T. N., Albajes-Eizagirre, A., Alnæs, D., Alpert, K. I., Andreassen, O. A., Anticevic, A., Asherson, P., Banaschewski, T., Bargallo, N., Baumeister, S., Baur-Streubel, R., Bertolino, A., Bonvino, A., Boomsma, D. I., . . . Tamnes, C. K. (2022). Greater male than female variability in regional brain structure across the lifespan. *Human Brain Mapping*, *43*(1), 470-499. https://doi.org/https://doi.org/10.1002/hbm.25204

Wimberley, T., Gasse, C., Meier, S. M., Agerbo, E., MacCabe, J. H., & Horsdal, H. T. (2017). Polygenic Risk Score for Schizophrenia and Treatment-Resistant Schizophrenia. *Schizophrenia Bulletin*, *43*(5), 1064-1069. https://doi.org/10.1093/schbul/sbx007

Wray, N. R., Lin, T., Austin, J., McGrath, J. J., Hickie, I. B., Murray, G. K., & Visscher, P. M. (2021). From Basic Science to Clinical Application of Polygenic Risk Scores: A Primer. *JAMA Psychiatry*, *78*(1), 101-109. https://doi.org/10.1001/jamapsychiatry.2020.3049

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

Wray, N. R., Yang, J., Goddard, M. E., & Visscher, P. M. (2010). The Genetic Interpretation of Area under the ROC Curve in Genomic Profiling. *PLOS Genetics*, *6*(2), e1000864. https://doi.org/10.1371/journal.pgen.1000864

Wu, Y., Byrne, E. M., Zheng, Z., Kemper, K. E., Yengo, L., Mallett, A. J., Yang, J., Visscher, P. M., & Wray, N. R. (2019). Genome-wide association study of medication-use and associated disease in the UK Biobank. *Nat Commun*, *10*(1), 1891. https://doi.org/10.1038/s41467-019-09572-5

Xue, A., Wu, Y., Zhu, Z., Zhang, F., Kemper, K. E., Zheng, Z., Yengo, L., Lloyd-Jones, L. R., Sidorenko, J., Wu, Y., Agbessi, M., Ahsan, H., Alves, I., Andiappan, A., Awadalla, P., Battle, A., Beutner, F., Bonder, Marc J., Boomsma, D., . . . e, Q. C. (2018). Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nature Communications*, *9*(1), 2941. https://doi.org/10.1038/s41467-018-04951-w

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E., & Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, *42*(7), 565-569. https://doi.org/10.1038/ng.608

Yang, J., Wray, N. R., & Visscher, P. M. (2010). Comparing apples and oranges: equating the power of case-control and quantitative trait association studies. *Genetic Epidemiology*, *34*(3), 254-257. https://doi.org/https://doi.org/10.1002/gepi.20456

Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., Frayling, T. M., Hirschhorn, J., Yang, J., & Visscher, P. M. (2018). Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum Mol Genet*, *27*(20), 3641-3649. https://doi.org/10.1093/hmg/ddy271

Zamani, F., & Doustkhah, E. (2022). Chapter 1 - An introduction to benzodiazepines and benzothiazepines. In F. Zamani & E. Doustkhah (Eds.), *Benzodiazepine-Based Drug Discovery* (pp. 1-8). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-824516-3.00007-0

Zhan, N., Sham, P. C., So, H. C., & Lui, S. S. Y. (2023). The genetic basis of onset age in schizophrenia: evidence and models. *Front Genet*, *14*, 1163361. https://doi.org/10.3389/fgene.2023.1163361

Zhang, Q., Privé, F., Vilhjálmsson, B., & Speed, D. (2021). Improved genetic prediction of complex traits from individual-level data or summary statistics. *Nature Communications*, *12*(1), 4192. https://doi.org/10.1038/s41467-021-24485-y