Aus der Klinik für Psychiatrie und Psychotherapie

der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Fine-tuning of QC and imputation pipeline *Ricopili* reveals a genetic overlap of panic disorder with neuroticism and depression.

Die Feinabstimmung der QC- und Imputations Pipeline *Ricopili* zeigt eine genetische Überlappung der Panikstörung mit Neurotizismus und Depression.

zur Erlangung des akademischen Grades

Doctor rerum medicinalium (Dr. rer. medic.)

vorgelegt der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

Von

Swapnil Awasthi

aus Lucknow (Indien)

Datum der Promotion: 30.06.2024

Content Page

List of abbreviations

Abbreviations	Definition
A	Adenine
AF	Allele frequency
ANOVA	Analysis of variance
BP	Bipolar disorder
BPD	Borderline personality disorder
C	Cytosine
CD-CV	Common disease - common variant hypothesis
DNA	Deoxyribonucleic acid
EA	East asian
EUR	European
FDR	False discovery rate
G	Guanine
GC	Genomic control
gIQ	General Intelligence quotient
GSA	Global screening array
GWAS	Genome-wide associations study
HRC	Haplotype reference consortium
IBD	Identity by descent
INFO score	Information score
LD	Linkage disequilibrium
LDSR	Linkage disequilibrium score regression
LOO-PRS	Leave-one-out Polygenic risk score
MAF	Minor allele frequency
MDD	Major depressive disorders
OR	Odds ratio
p	p value
pcorr	p value after correcting for multiple testing
PC	Principal component
PCA	Principal component analysis
PD	Panic disorder
PRS	Polygenic risk scores
QC	Quality control
Q–Q plot	Quantile quantile plot
rg	Genetic correlation
RNA	Ribonucleic acid
SCZ	Schizophrenia
SE	Standard error
SNP	Single nucleotide polymorphism
T	Thymine
HWE	Hardy-Weinberg Equilibrium
WISC-IV	Wechsler Intelligence Scale for Children - Fourth Edition

1) Abstract

A genome-wide association study (GWAS) is a standard study design for examining the association between genotype and disease status without knowledge of the underlying biological mechanisms. GWASs have led to the identification and classification of numerous variants associated with human traits. However, while this design has been widely used in contemporary genetic research, it is prone to technical biases and errors, which necessitate the development of a standardized workflow and analysis methodology.

To address this problem, we developed a simulation-based framework for calibrating GWAS pipelines. Using *Ricopili,* our comprehensive GWAS pipeline, we developed a standard common-variant analysis workflow and then demonstrated the pipeline's various functionalities and features. Furthermore, we demonstrated that this pipeline's current framework could be successfully used to perform full-scale analyses of genotype data, ranging from quality control to the downstream analysis of variants.

Furthermore, this thesis investigated the genetic architecture of panic disorder (PD) in six independent collections taken from four European countries. Given the comparably small total sample size of 2,147 cases and 7,760 controls, no genome-wide significant single nucleotide polymorphism (SNPs) were identified; however, we demonstrated a highly significant polygenic risk score (PRS) that explained up to 2.6% of the phenotypic variance. The SNP-based heritability for PD was estimated at 28.0–34.2%, and 135 out of the 255 most significant SNPs exhibited the same direction of effect in an independent replication sample ($p = 0.048$). In a combined meta-analysis, rs144783209 in the gene SMAD1 exhibited the strongest association ($Pcomb = 3.10 \times 10^{-7}$) with PD. A significant genomic correlation was detected with published GWAS results for major depressive disorder ($p = 0.025$), depressive symptoms ($p = 0.010$), and neuroticism ($p = 0.002$).

Moreover, in a distinct psychiatric phenotype, we found a highly significant genetic correlation (30–60%) between borderline personality disorder (998 cases and 1,545 controls) and three published adult psychiatric disorders, namely schizophrenia ($p = 4.37 \times 10^{-5}$), bipolar disorder (p

= 2.99 × 10^{-3}), and major depression (p = 1.04 × 10^{-3}). In a third analysis, we demonstrated that PRSs in the IMAGEN cohort (n = 1,475) derived from published GWASs of intelligence significantly explained 0.33–3.2% of the variance in general IQ.

In summary, our meta-analysis of PD represents a significant advancement in elucidating its genetic architecture, including the first SNP-based heritability estimate. We observed a notable genetic connection between PD and neuroticism. Additionally, the significant genetic correlation of borderline personality disorder (BPD) with other psychiatric disorders suggests that BPD shares underlying factors with these disorders, consistent with clinical observations. Finally, our research affirms the polygenic nature of general intelligence within the IMAGEN cohort.

2) Zusammenfassung

Die genomweite Assoziationsstudie (GWAS) ist eine Standardmethode, die Assoziationen von Allel- oder Genotyphäufigkeiten sogenannter SNPs (Single Nucleotide Polymorphisms) zu Fall/Kontroll-Verteilungen oder zu quantitativen Merkmalen untersucht. Während GWAS das Verständnis der genetischen Grundlage komplexer Merkmale stark vorantreibt, ist sie auch anfällig für technisch/statistische Verzerrungen und Fehler, die die Einhaltung eines hoch standardisierten Arbeitsablaufs und Analysemethodik erfordert.

Um diese zu erarbeiten, haben wir ein simulationsbasiertes Framework zur Kalibrierung von GWAS Pipelines entwickelt. Mit Hilfe unserer umfassenden GWAS-Pipeline *Ricopili* entwickelten wir einen standardisierte Analyseplan und demonstrieren die verschiedenen Funktionalitäten und Eigenschaften unserer Pipeline. Wir zeigen, wie mit dieser Pipeline erfolgreiche Analyse von Genotyp-Rohdaten durchgeführt werden können, angefangen von der Qualitätskontrolle bis hin zur endgültigen Assoziationsanalyse.

Wir untersuchen den Einfluss von häufigen genetischen Varianten auf die Panikstörung (PD) in einer Meta-Analyse von sechs unterschiedlichen Kohorten aus vier europäischen Ländern. Wie erwartet identifizierten wir bei einer vergleichsweise kleinen Gesamtstichprobengröße von 2,147 Patienten und 7,760 Kontrollen keine genomweit signifikanten Varianten, jedoch konnten wir einen signifikanten polygenen Risikoscore (PRS) nachweisen, der bis zu 2.6% der phänotypischen Varianz erklärt. Die SNP-basierte Heritabilität für PD schätzen wir auf 28.0-34.0 %. In einer unabhängigen Replikationskohorte zeigen 135 der 255 signifikantesten SNPs mit Schwellenwert die gleiche Effektrichtung wie in unserem Hauptdatensatz (signifikant mit p = 0.048). In der kombinierten Meta-Analyse zeigte sich rs144783209 im Gen SMAD1 als stärkste Gesamt - Assoziation ($p = 3.10 \times 10^{-7}$). Schliesslich konnten wir eine signifikante Korrelationen mit der Unipolaren Depression (Major Depressive Disorder - MDD) ($p = 1.04 \times 10^{-3}$), depressiven Symptomen ($p = 0.025$) und Neurotizismus ($p = 0.002$) finden.

Bei einer weiteren psychiatrischen Erkrankung fanden wir eine hochsignifikante genetische Korrelation (zwischen 30 und 60 %) zwischen der Borderline-Persönlichkeitsstörung (998 Fälle

und 1,545 Kontrollen) und den drei zentralen psychiatrischen Erkrankungen: Schizophrenie ($p = 4.37 \times 10^{-5}$), Bipolare Störung ($p = 2.99 \times 10^{-3}$) und Major Depression ($p = 1.04 \times 10^{-3}$). In einer dritten hier vorgestellten Analyse konnten wir polygene Riskoscores in der IMAGEN-Kohorte (n=1,475), aus GWASs der Intelligenz abgeleiten, die signifikant 0.33% bis. 3.2% der Varianz des allgemeinen IQ erklären.

Zusammenfassend stellt unsere Meta-Analyse von PD einen bedeutenden Fortschritt bei der Aufklärung seiner genetischen Architektur dar, einschließlich der ersten Schätzung der erblichen Veranlagung basierend auf SNPs. Wir haben eine bemerkenswerte genetische Verbindung zwischen PD und Neurotizismus beobachtet. Darüber hinaus legt die signifikante genetische Korrelation der Borderline-Persönlichkeitsstörung (BPS) mit anderen psychiatrischen Störungen nahe, dass BPS gemeinsame zugrunde liegende Faktoren mit diesen Störungen teilt, was mit klinischen Beobachtungen übereinstimmt. Schließlich bestätigt unsere Forschung die polygene Natur der allgemeinen Intelligenz innerhalb der IMAGEN-Kohorte.

3) Introduction

3.1 Genomics: Brief highlights

Human traits and disorders are influenced by both genetic and environmental factors. The complete genetic material of an organism is known as its **genome**. The genome is composed of **deoxyribonucleic acid (DNA),** a biomolecule that carries genetic information for trait functioning and development (Figure 3.1).

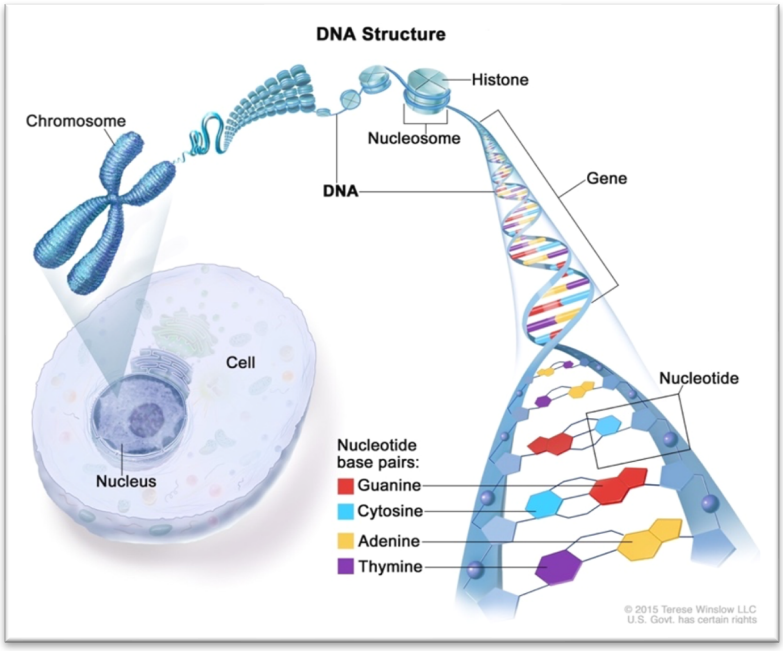

Figure 3.1a: DNA structure; The depiction illustrates the composition of DNA, featuring a chromosome, nucleosome, histone, gene, and nucleotide base pairs, including guanine, cytosine, adenine, and thymine. Additionally, it includes a cell with its nucleus [© *(copyright year as seen in the illustration)* Terese Winslow LLC, U.S. Govt. has certain rights: Permission granted].

DNA has a double-helix structure with **nucleotides** as its repeating units. Each nucleotide consists of a deoxyribose sugar, a phosphate group, and a nitrogenous base. Adenine (A) and guanine (G) are purines, while cytosine (C) and thymine (T) are pyrimidines (Figure 3.1). Adenine always pairs with thymine, and cytosine always pairs with guanine, forming **base pairs** and giving DNA its unique structure. The 146–147 base pair-long chain of DNA coils around a core of histone proteins to form **nucleosomes**, which are tightly packed into larger units called **chromosomes** (Figure 3.1). Human cells contain 23 pairs of chromosomes, with approximately 3.2 billion base pairs. Chromosomes carry **genes**, which are the fundamental units of heredity.

There are approximately 20,000 to 25,000 protein-coding genes in humans. This fraction of protein-encoding base pairs is only 1–2% of the total human genome. These genes contain instructions, which consist of triplets of nucleotides called **codons**. Each codon specifies a particular amino acid, the building blocks of proteins. Out of the 64 (4^3) possible codons, 61 code for amino acids, while the remaining three are stop codons.

Furthermore, the human genome is diploid, which means that it contains two sets of chromosomes inherited from each biological parent. As a result, different forms of the same gene, called **alleles**, can occur at a given genomic location or **locus**. These variations can be single-base or segment variations in the DNA sequence. The representation of these variations is referred to as the **genotype** of an individual at that locus. A **homozygous genotype** occurs when an individual inherits two identical alleles from their parents, while a **heterozygous genotype** occurs when an individual inherits two different alleles at a genomic locus. Crucially, approximately 99.9% of the genomes of all individuals are the same.

Variation within the genome can lead to differences in traits and disorders among individuals, and **heritability**, which refers to the proportion of the variation in a trait or disorder that can be attributed to genetic factors, plays a significant role in understanding the interplay between genetics and these traits and disorders.

The central dogma of molecular biology is a theory that describes the flow of genetic information from DNA to RNA (ribonucleic acid) to protein. It underlies the processes that govern the functioning and development of traits at a molecular level. Understanding genetics and genetic variations is crucial to comprehending human traits and disorders.

3.2 Monogenic disorders

Monogenic disorders are the result of a mutation in a single gene. Their inheritance pattern usually follows Mendel's laws (e.g., autosomal recessive), and thus, they are also termed Mendelian disorders. The effect size or penetrance of each variant is typically large; consequently, its frequency is usually driven low in a population by selective pressures. Huntington's disease is an example of a Mendelian disorder.

Linkage analysis is a study design used for identifying the causal gene associated with monogenic disorders or traits. In linkage analysis, the location of a disease-causing gene is identified by analyzing patterns of inheritance in related individuals. Usually, multiple families with a disorder (or trait of interest) are recruited, and their disease status and co-segregating genetic markers (e.g., short tandem repeat) are measured. Finally, the segregation patterns of these markers within the families are examined. A successful example of linkage analysis identified multiple CFTR gene mutations as the causal variants for cystic fibrosis (Kerem et al., 1989).

3.3 Complex traits

In contrast to Mendelian traits, complex traits are not explained by genetic variation in single genes. They have genetic and environmental components associated with them and are relatively more common than Mendelian disorders. Furthermore, the genetic component is usually spread across the genome – a phenomenon known as "polygenicity." Most psychiatric disorders are examples of common complex genetic disorders (Brainstorm Consortium., 2018), including diabetes mellitus (Xue et al., 2018) and Crohn's disease (Verstockt et al., 2018). Traits such as adult height (Yengo et al., 2018) and IQ (Savage et al., 2018) are examples of common/complex nonpsychiatric traits.

3.3.1 Common disease – common-variant hypothesis

The common disease – common-variant (CD-CV) hypothesis states that a common disorder is likely to be caused by genetic variation common in the population. The CD-CV hypothesis implies that a moderate-to-large number of variants across the genome contribute to disease risk and that each variant has relatively low penetrance and a small effect size (i.e., polygenicity). These common, polygenic traits are the antithesis of monogenic disorders, where variants usually have a low prevalence and a large effect.

3.3.2 Genome-wide association studies

Linkage analyses applied to common or complex disorders have typically been unsuccessful (Altmüller et al., 2001; Hirschhorn & Daly., 2015), which indicates that common disorders'

genetic architectures differ from those of Mendelian disorders. Genome-wide association studies (GWASs) have proven to be a successful tool for testing the CD-CV hypothesis. In principle, a GWAS compares the allele frequency (AF) of each single nucleotide polymorphism (SNP) across the human genome between cases and controls. In practice, this frequency–phenotype association is tested in a regression framework, which allows covariates like population stratification to be adjusted for. SNPs are the most prevalent form of genetic variation detected in the human genome, where approximately 88 million unique SNPs (1000 Genomes Project Consortium., 2015) have been identified in various worldwide populations. There are approximately 10 million SNPs in European ancestry with a minor AF (MAF) greater than 1% (1000 Genomes Project Consortium., 2015).

3.4 Technical and statistical aspects of GWASs

In the subsequent sections, a concise overview is provided of the fundamental procedures employed in genotype data analysis. These subsections delve into quality control measures, encompassing both technical and genomic aspects, with the aim of mitigating biases and minimizing the occurrence of false positive results. Furthermore, genotype imputation and association analysis are presented. Moreover, the significance of the Ricopili pipeline, which serves as an efficient tool for optimizing this entire workflow, is emphasized.

3.4.1 Technical quality control

While GWASs are a powerful tool for expanding our understanding of complex disorders, they are also prone to biases. If not considered, such biases can result in false-positive associations. While testing millions of markers across the genome, a minor bias can lead to a high type 1 error rate (i.e., rate of false positives). Proper quality control (QC) on individuals and genotypes (i.e., SNPs) can eliminate many of these biases. QC per individual includes checking consistencies between the sex predicted by the X chromosome and the ascertained sex, missing genotype rates per individual, and heterozygosity rates. The QC of SNPs consists of the missing rate per SNP, missing rate difference between cases and controls, deviation from Hardy–Weinberg equilibrium (HWE), and AF (variants with very low MAF, e.g., < 1%).

3.4.2 Genomic quality control

Genomic QC primarily corrects for a systemic bias attributed to AF differences due to different ancestries, which is known as **population stratification**. To identify such stratification, a multidimensional scaling method, such as principal component analysis (PCA), is used. One- and two-dimensional PCA plots are used to visualize the genetic heterogeneity in the study cohort (or across multiple cohorts), and subpopulations and outliers are separated or excluded from the analysis. The estimated PCs are also used to control for population stratification in downstream analysis.

Genomic QC is also used to address cryptic relatedness (the nonindependence of individuals) within a cohort. Identity by descent (IBD) is computed for each individual pair in a cohort, and first-, second-, and third-degree relatives are excluded from further analysis.

3.4.3 Tag SNPs and genotype imputation

Tag SNPs contains implicit information about linked contiguous stretches of DNA (also called **haplotypes**) shared between individuals, which are more substantial within single ancestries and inherited together. This nonrandom correlation or association between alleles located at distinct loci (genetic positions) on the same chromosome is known as **linkage disequilibrium (LD)**. LD is a natural correlation structure present between SNPs in physical proximity to each other on the human genome. In a GWAS, a moderate number of tag SNPs are genotyped on chip-based microarrays (e.g., Illumina and Affymetrix), thereby avoiding the extra cost of assaying all SNPs due to LD. Most modern genotyping chips contain in the order of 600,000 SNPs, tagging approximately 8 million SNPs in the European population through imputation, for example.

Imputation allows researchers to infer missing genotypes that are not directly genotyped. It exploits the LD information of tagged SNPs using large, publicly available reference panels, such as HapMap (International HapMap Consortium., 2003), 1000 Genomes (1000 Genomes Project Consortium, 2010), or the Haplotype Reference Consortium (HRC; McCarthy et al., 2016). Imputation allows meta-analyses to be performed between genotyped cohorts on various platforms, boosting a GWAS's power as well as enhancing fine mapping. This is achieved by increasing the number of SNPs up to 8 million from 600,000 directly genotyped on a modern

chip, such as the global screening array (GSA). There are many algorithms for genotype imputation, such as MaCH (Li., 2010), beagle (Browning et al., 2018), and impute2 (Howie et al., 2009), of which the output is a probabilistic estimate of genotype status.

In addition, several metrics exist for assessing the quality of genotype imputation. One such metric used for impression quality assessment is the information (INFO) score, which estimates how well the imputed genotypes correspond to the true genotypes on the reference genome. It compares the imputed allelic dosages, their variances, and the corresponding AFs with those in the reference data. The INFO score ranges from 0 to 1, where higher values indicate more reliable imputations and lower values indicate lower-quality genotype imputation.

3.4.4 Association analysis and meta-analysis

After comprehensive QC and imputation, the data are ready for association mapping. Traditionally, single SNP association analysis is performed, and the choice of statistical test depends on, among other variables, the type of phenotype. For quantitative phenotypes, a linear regression is typically performed, whereas for case-control phenotypes, logistic regression is mostly used. Regardless of phenotypes, how genotypes are encoded influences the power of the statistical analysis through altering the degrees of freedom in the test statistic. The genotypic-based test, where three genotype classes (two homozygous and one heterozygous) are combined and tested for association with the phenotype, has two degrees of freedom. By contrast, the additive, dominant, recessive or “heterozygote advantage” models use tests with one degree of freedom. The additive genetic model is frequently used in genetic research for testing associations and identifying various genetic risk factors, including dominant and recessive effects. However, it may not reveal the genetic association linked to heterozygote advantage models.

Principal components (PCs) generated from PCA analysis are used to control population stratification. If necessary, the regression is adjusted for clinical covariates (e.g., age, sex, clinical site, and batch effect). Many independent cohorts are usually meta-analyzed using an inverse variance weighted method, which implicitly gives more weight to larger studies with

higher imputation quality. A meta-analysis estimates the aggregated effect size, standard error, and associated *p* values of all the independent cohorts together.

3.4.5 Multiple testing and genome-wide significance

In general, a *p* value equal to or below 0.05 is considered significant (i.e., the alpha value) for a single statistical test. It implies that in a statistical test, there is a 5% probability of incorrectly rejecting the null hypothesis when it is actually true. If one conducts multiple hypothesis tests, the probability of incorrectly rejecting the null hypothesis naturally increases, and therefore, the burden of detecting a false positive is higher due to multiple testing. The various methods used to correct multiple testing problems in a GWAS are Bonferroni correction, the Benjamini–Hochberg false discovery rate (FDR), and permutation testing.

For GWASs in European and Asian populations, the most common alpha value is 5×10^{-08}. This is equivalent to a Bonferroni correction of $p = 0.05$ for 1 million independent common variants across the human genome. The number of independent tests in the African population is higher compared with that in the European population. Thus, a recommended alpha value is 1×10^{-08}. Pe'er et al. (2008) confirmed these thresholds for genome-wide significance with phenotype simulations in the International HapMap Consortium.

3.4.6 *Ricopili*: Our computational pipeline

A computational pipeline is a cohesive set of computational programs, tools, and software packages that are organized to perform specific tasks in a logical sequence. It operates by taking the output of one program and using it as input for the next program in a predefined order. The purpose of a pipeline is to make efficient use of computational resources and to streamline the execution of computational tasks. By using a pipeline, errors can be minimized, and a standardized approach can be established for processes that involve multiple programs (steps).

It is often helpful and common practice to assign a name (acronym) to a pipeline that performs specific tasks, as it assists its easy identification and use by others. In this thesis, we used the *Ricopili* pipeline, which integrates a set of tools and algorithms that perform technical QC, genomic QC, imputation, and statistical analysis for a GWAS (described Section 3.4).

3.5 Panic disorder

Panic disorder (PD) is a debilitating and severe form of anxiety disorder characterized by a sudden onset of fear and discomfort that lasts for several minutes or longer. PD has at least four anxiety-related symptoms, which might include "palpitations, diaphoresis, tremors, shortness of breath, suffocation, chest pain, chills, heat sensations, nausea, dizziness, depersonalization, paresthesia, derealization, fear of loss of control and fear of dying, numbness" (Virginia C., 2016, Pages 266-288 and Cackovic et al., 2020). The lifetime prevalence of PD is approximately 2–4%, while its lifetime morbid risk is 6% (Kessler et al., 2012). The age at which PD typically starts ranges from 26 to 35 years (Lijster et al., 2017). Constant panic attacks can further deteriorate the patient's health by hindering a healthy social life, leading to the development of pathological symptoms, such as long-term disabilities and agoraphobia (Hendriks et al., 2014).

3.5.1 Etiology of PD

The etiology of PD is only vaguely understood, with most research suggesting that the causes are a complex combination of genetic and environmental factors. The following subsections discuss PD's genetic risk factors, environmental risk factors, and neurobiology.

3.5.1.1 Genetic risk factors for PD

Twin and family studies of PD patients have reported heritability estimates of 40% (Hettema et al., 2001), implying that a significant genetic component contribute to PD liability (Schumacher & Deckert., 2010). Several linkage and association studies have been conducted to understand the mechanisms behind PD, but none have convincingly and robustly identified the underlying genetic factors. Furthermore, linkage analyses of PD have identified several chromosomal loci to the syndrome, including 1q, 2q, 4q31-q34, 7p, 9q, 12q, 13q, 14q, 15q, and 22q, but all of them have exhibited little consistency across studies (Na et al., 2011). This is in line with the general underperformance of linkage analyses for complex traits (see Section 3.2). Moreover, few PD GWASs have been published, and none have reported genome-wide significant SNPs. Erhardt et al. reported two SNPs, namely rs7309727 and rs11060369, in the TMEM132D gene on chromosome 12q24 associated with PD. Their study consisted of 909 cases and 915 controls of European ancestry (Erhardt et al., 2011), and they replicated these associations in five PD cohorts. The following year, Otowa et al. conducted a GWAS on 718 PD cases, and 1,717

controls and reported many putative associations (with variation contained in *BDKRB2* and *NPY5R*) in the Japanese population (Otowa et al., 2012). Still, none of these reported variants reached the genome-wide significance of 5×10^{-8}. A recent meta-analysis of combined anxiety and PD revealed the SNPs rs1709393 on an uncharacterized noncoding RNA locus on chromosomal 3q12.3 (with $p = 1.65 \times 10^{-8}$) and rs1067327 within the coding region of the gene CAMKMT on chromosome 2p21 (with $p = 2.86 \times 10^{-9}$; Otowa et al., 2016). However, due to the combined nature of the phenotype, these associations were not specific to PD.

3.5.1.2 Environmental risk factors for PD

A 40% heritability of PD indicates that approximately 60% of the risk is likely attributed to environmental factors. This understanding has been studied for many decades to gain insights into environmental risk factors for PD.

The disruption of early attachment, as supported by the Epidemiological Catchment Area Study (Tweed et al.,1989), plays a significant role in the development of PD. The study *(N* = 3,803) found that individuals whose mothers died before they turned 10 years of age were nearly seven times more likely to be diagnosed with agoraphobia accompanied by panic compared with those without a history of early maternal death. Similarly, individuals whose parents separated or divorced before the age of 10 years were also found to have an increased likelihood of being diagnosed with agoraphobia with panic. Extensive preclinical research has strongly indicated that early disruptions in the attachment between infants and their mothers result in enduring behavioral and biological changes, such as those observed in PD (Francis & Meaney et al., 1999 and Anisman et al., 1998).

Furthermore, bodily sensations are closely linked to the onset of PD (Roy et al., 2006), and anxiety sensitivity is one such factor that heightens these sensations (Reiss et al., 1980). The anxiety sensitivity index predicts the development of PD in several and diverse cohorts. This predictive power has been found to remain even after accounting for previous depression (Hayward et al., 2000, Maller & Reiss et al., 1992 and Ehlers et al., 1995). However, anxiety sensitivity is commonly associated with neuroticism and a propensity for experiencing negative emotions, as opposed to being a direct causal factor for PD (Roy et al., 2006).

3.5.1.3 Neurobiology of PD

Several compounds with distinct mechanisms of action have been identified to induce panic symptoms in patients but not in control subjects. Examples of such compounds include hyperosmolar sodium lactate, caffeine, isoproterenol, yohimbine, carbon dioxide, and cholecystokinin (Roy et al., 2006). Although relevant studies have not significantly contributed to the understanding of the biological mechanisms that underlie PD, they have elucidated the activation of a broader neural network associated with conditioned fear, which is currently proposed to be the neural network for PD (Gorman et al., 2000). The sensory input for the conditioned stimulus travels through the anterior thalamus to the lateral nucleus of the amygdala before being transmitted to its central nucleus. The amygdala's central nucleus serves as a central hub for information dissemination, coordinating autonomic and behavioral responses associated with PD.

3.5.2 Comorbidity of PD

Individuals with panic disorder often present with one or more concurrent medical illnesses. In a study conducted by Starcevic et al. in 1999, which included 88 individuals diagnosed with PD, it was found that 82% of patients were diagnosed with one or more Axis I disorders, and 49% of patients were diagnosed with one or more Axis II disorders.

The most prevalent Axis I disorder among these PD patients was anxiety disorder, specifically Generalized Anxiety Disorder (GAD), with prevalence rates ranging from 22% (as reported by Turki et al. in 2017) to 59% (as reported by Starcevic et al. in 1999). Major Depressive Disorders (MDD) also exhibited a high prevalence, with an approximate rate of around 50% in both the study by Starcevic et al. in 1999 and the research by Gorman et al. in 1996.

Regarding Axis II disorders, Cluster B (or emotional) personality disorders, including Antisocial Personality Disorder, Borderline Personality Disorder, and Histrionic Narcissistic Personality Disorder, showed prevalence rates ranging from 25% to 27%. On the other hand, Cluster C (anxious) personality disorders, comprising Avoidant Personality Disorder, Dependent Personality Disorder, and Obsessive-Compulsive Personality Disorder, exhibited prevalence

rates ranging from 15% to 40% (as reported by Ozkan & Altindag in 2005 and Starcevic et al. in 1999).

PD is also found to be associated with personality traits. In a study by Zugliani et al. (2017), a significant association was found between PD patients and higher neuroticism scores and lower extraversion scores compared to healthy controls.

PD also frequently co-occurs with other medical illnesses that share symptomatology, such as respiratory conditions (e.g., asthma and Chronic Obstructive Pulmonary Disease), cardiovascular conditions (e.g., hypertension and coronary heart disease), Irritable Bowel Syndrome, and Diabetes. This comorbidity is associated with a significant increase in PD patients, estimated to be at least 1.5 to twofold (Meuret et al. in 2020).

3.6 Borderline personality disorder

Patients with borderline personality disorder (BPD) experience unstable relationships, perception shifts, cognitive and self-image issues, and intense emotions as well as engage in impulsive and risky behaviors during emotional distress. BPD is a complex neuropsychiatric disorder with a lifetime prevalence of approximately 3% (Tomko et al., 2014). The typical age of onset of BPD (Biskin., 2015) is 18 years, but its symptoms (e.g., self-harm) can start as early as 12 years (Zanarini et al., 2008). BPD is more commonly diagnosed in female patients, who account for approximately 75% of cases (American Psychiatric Association., 2000). BPD is associated with high healthcare utilization and a chronic, severely debilitating clinical course (Bohus & Schmahl et al., 2007). Suicide rates range between 6% and 8%, and up to 90% of patients engage in nonsuicidal self-injurious behavior (Zanarini et al., 2018).

3.6.1 Etiology of BPD

The following subsections present BPD's genetic risk factors, environmental risk factors, and neurobiology.

3.6.1.1 Genetic risk factors for BPD

Distel et al. (2008) reported a heritability estimate of 42% in a twin study conducted across three different cohorts from The Netherlands, Belgium, and Australia, implying the existence of a genetic component in BPD. However, most available studies have focused on candidate genes and had relatively small sample sizes (Calati et al., 2013). Several of these studies have reported associations of genetic variants within genes such as tryptophan hydroxylase 1 and the serotonin 1B receptor, but these associations yielded insignificant results, as highlighted in a systematic review (meta-analysis) conducted by Amad et al. (2014).

Nevertheless, a GWAS by Lubke et al. (2014) identified a potential association between the SERINC5 gene, which plays a role in myelination, and affective instability, a key characteristic of manic phases of bipolar disorder (hereinafter "BP").

3.6.1.2 Environmental risk factors for BPD

Environmental factors explain approximately 58% of the variation in BPD, as reported by Distel et al. (2008). Adverse childhood experiences, including trauma and maltreatment, strongly contribute to the development of BPD (Zanarini et al., 1987 and Afifi et al., 2011). Separation from mothers at an early age (Crawford et al., 2009), abnormal attachment (Rogosch & Cicchetti., 2004), and delays in identity development during adolescence (Fonagy & Bateman., 2008) can also lead to personality pathology. Other childhood and adolescent disorders, such as depression, anxiety, and disruptive behavior disorders, increase people's predisposition to BPD (Helgeland et al., 2005).

3.6.1.3 Neurobiology of BPD

Altered functioning in specific brain regions, such as the medial prefrontal cortex, temporoparietal junction, posterior cingulate cortex, and precuneus, contributes to distorted self-perception and thoughts about others in individuals with BPD (Krause et al., 2014 and Beeney et al., 2016). Impulsivity in BPD involves changes in the reward and control circuits, mediated by the ventral striatum and prefrontal areas (Herbort et al., 2016 and Gunderson et al., 2018).

3.6.2 Comorbidity of BPD

BPD is a condition characterized by a high degree of comorbidity. In the Collaborative Longitudinal Personality Disorders Study conducted by McGlashan et al. in 2000, which included 175 BPD patients, it was found that individuals with BPD had an average of 4.1 lifetime Axis I comorbidities and 1.9 lifetime Axis II comorbidities (personality disorders).

The most common Axis I comorbidity in individuals with BPD is mood disorders, affecting approximately 95% of them (Shah & Zanarini, 2018). Within this category, MDD exhibits a prevalence ranging from 71% (Zanarini et al., 1998a) to 83% (McGlashan et al., 2000), while the comorbidity between BPD and BP is less prevalent and debated, as they have overlapping symptoms but distinct features specific to each. The comorbidity ranges from 9% (Zanarini et al., 2004) to 20% (Fornaro et al., 2014) between BPD and BP. In contrast, Tsanas et al. (2016) highlighted symptomatic differences between the two groups. Anxiety disorders, particularly Post-Traumatic Stress Disorder, are highly prevalent, co-occurring in approximately 51% of individuals with BPD (McGlashan et al., 2000 and Zanarini et al., 1998a), making them the second most common comorbidity. Psychotic disorders represent the least common Axis I comorbidity with BPD, with schizophrenia (SCZ) reported in 2% (Slotema et al., 2018) to 17% (Kingdon et al., 2010) of cases.

Among Axis II comorbidities, the most prevalent disorders in conjunction with BPD were from the anxious cluster (Cluster C personality disorders). Within this category, Avoidant Personality Disorder ranged from 43% to 47%, and Dependent Personality Disorder ranged from 16% to 51% (Zanarini et al., 1998b and McGlashan et al., 2000).

Moreover, a twin study (Distel et al., 2009) and a large population-based study (Gale et al., 2016) have suggested a genetic association between BPD's features and neuroticism – a recognized risk factor for BP and other psychiatric conditions.

3.7 Aims and hypotheses of the Dissertation

The sensitivity of GWASs to various biases and errors makes following a highly standardized workflow or analysis methodology critical. This dissertation incorporates four publications in the following two chapters (Chapter 4: Materials and methods and Chapter 5: Results), each of which has four subsections that correspond to each publication.

The first publication (**4.1 and 5.1**) presents *Ricopili*, a well-developed, highly standardized, and flexible GWAS analysis pipeline that we designed to analyze large genomic data sets. First, we demonstrate how our pipeline can detect and correct for biases and technical errors in simulated genotype data and develop a framework along the way. We implement this workflow and conduct a GWAS on 2,248 PD cases and 7,992 healthy controls to characterize the genetic variation associated with PD in **4.2 and 5.2**. This dissertation's aims are as follows:

1. To investigate and identify the genetic variation associated with PD;
2. To estimate the proportion of genetic variance within our cohort that can explain the development of PD;
3. To determine the extent to which causal genes overlap across different conditions and suggest potential common pathways or mechanisms underlying various other complex disorders;
4. To examine, through in silico functional analysis, the genes involved in regulating the causal biological pathways of PD, thereby providing insights into the molecular mechanisms that underlie PD development and potential therapeutic targets for intervention.

In addition, I present my contribution to deciphering the genetic architecture by estimating the extent of genetic overlap between BPD and schizophrenia (SCZ), bipolar disorder (BD), and Major Depressive Disorder (MDD) as detailed in 4.3 and 5.3. Furthermore, I examine the polygenic nature of general IQ, as outlined in 4.4 and 5.4.

4) Materials and methods

This chapter provides a comprehensive description of the materials used and the methods employed in conducting this research work. Section 4.1 begins by presenting a standardized GWAS methodology developed using the pipeline's features and modules, which are detailed in sub-sections 4.1.1 to 4.1.6. Additionally, sub-section 4.1.7 explains the incorporation of simulated data with artificial technical and genomic interventions that is used to test and fine-tune the methodology.

Section 4.2 provides an in-depth overview of the panic disorder (PD) cohorts, covering the methodology used for technical and genomic quality control, imputation, association analysis, leave-one-out polygenic risk score (LOO-PRS) analysis, replication, heritability estimation, genetic correlation and functional analysis.

In Section 4.3, the cohort of borderline personality disorder (BPD) is described, along with the materials and statistical methods used to estimate its genetic correlation with psychiatric disorders such as schizophrenia (SCZ), major depressive disorder (MDD), and bipolar disorder (BP).

Finally, section 4.4 introduces the IMAGEN cohorts, outlines how intelligence measures are inferred, and details the methodology used to generate polygenic scores.

4.1 Development of a standard GWAS analysis methodology (Lam and **Awasthi** et al., 2020)**:**

The simulated genotype cohorts along with the interventions listed in Table 4.1a were used to standardize and develop a GWAS workflow. The workflow, presented in Figure 4.1a, is representative of standard, common-variant GWASs and is applicable to a wide range of common disorders. Comprehensive QC (technical and genomic) can be performed by employing the pre-imputation and PCA modules of the *Ricopili* pipeline, in addition to some manual quality checks.

As depicted in the flowchart (Figure 4.1a), the initial QC module should be started only if the number of cases and controls both exceed 50. Excessive SNP loss (attributable to various influences, such as batch effect, batch sizes, and ancestral composition) should be addressed before proceeding further. In cases of high numbers of excluded individuals, their reported phenotypic information should be checked and corrected if possible (e.g., sex errors). Crucially, one should also check whether a general bias exists in the test statistics. This can be examined using lambda GC, which is estimated from the initial association analysis. **Lambda GC** is quantified as the ratio of the median of the empirically observed test statistics to the expected test statistic distribution. High lambda inflation (i.e., $\lambda_{GC} > 1.05$) should be controlled by applying a stricter filtering criterion or addressing sample heterogeneity and cryptic relatedness through the PCA module. In the initial association analysis, significant findings should be carefully examined and included only if they fail to meet more rigorous filtering criteria than the default (Subsection 4.1.1). For instance, SNPs with a missingness rate greater than 0.01 or a frequency lower than 5% could be excluded.

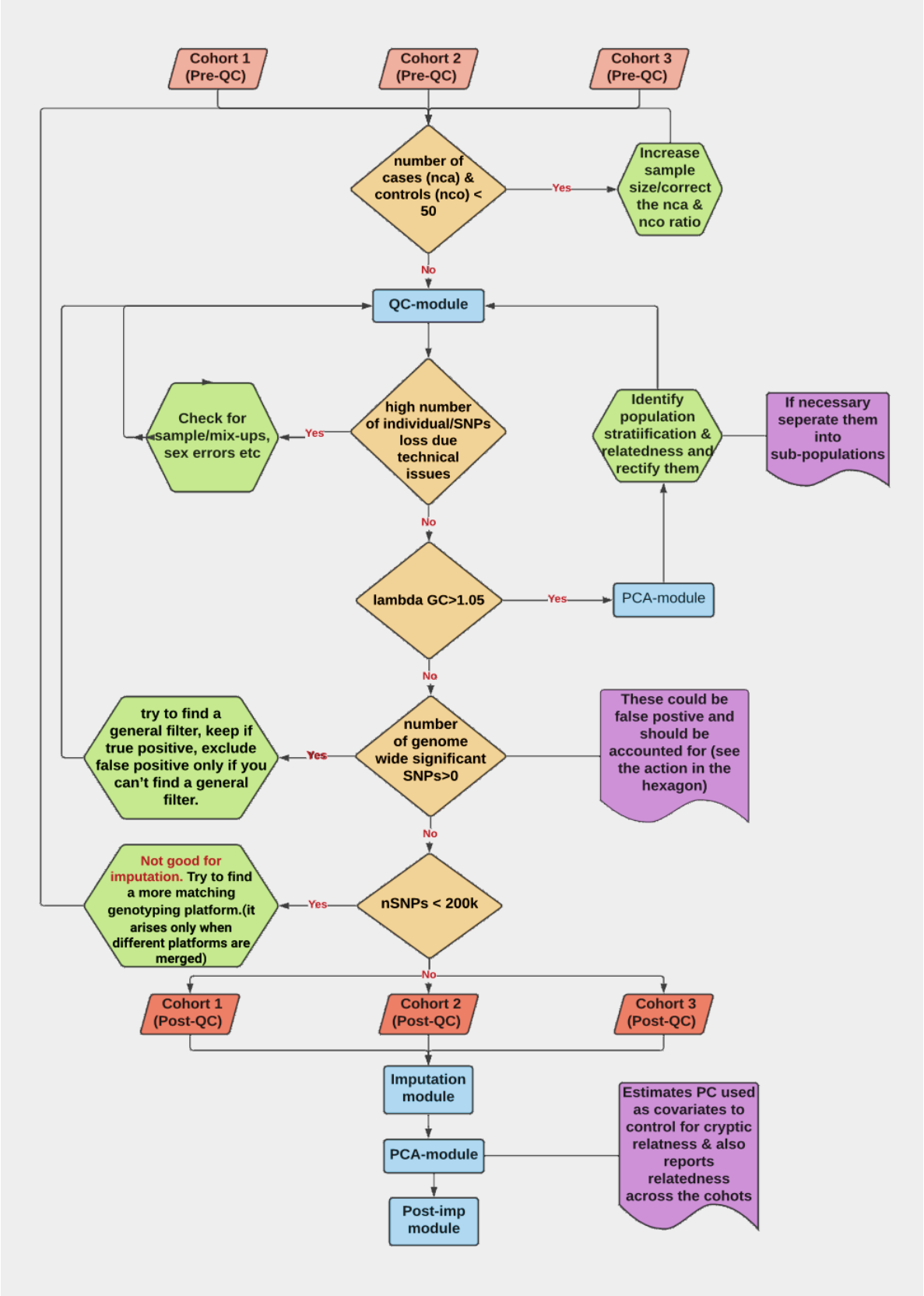

Figure 4.1a: Analysis flowchart developed and used throughout this project.

Finally, after automated and manual QC, the total number of SNPs in the QC must be at least 200,000 (preferably >400,000) to ensure robust imputation. Post-imputation PCA is performed on the best-guess genotype to compute fine-grained PCs, which are then used to control for population stratification in the final downstream analysis. When performing the PCA module on multiple cohorts, overlapping/related individuals across cohorts are identified; therefore, only nonrelated individuals are analyzed. Finally, the post-imputation module is used to perform association and meta-analyses.

4.1.1. Pre-imputation/QC module

By default, the QC module of the *Ricopili* pipeline performs QC for each SNP and individual through the following steps:

1. Excluding SNP missingness >0.05 (before the removal of individual subjects);
2. Excluding individuals with missing SNPs >0.02;
3. Excluding individuals with autosomal heterozygosity deviation (|Fhet| > 0.2);
4. Excluding individuals whose reported sex does not match the predicted sex by the chromosome X genotypes;
5. Excluding SNP with missingness >0.02;
6. Excluding SNPs with a difference in missingness >0.02 between patients and controls;
7. Excluding SNPs that deviate from HWE ($p < 10^{-10}$ in patients, $p < 10^{-6}$ in controls);
8. Excluding SNPs without valid association *p* values with the phenotypes (i.e., invariant SNPs).

This module also performs baseline association analysis without covariates to identify genome-wide inflation (λ_{GC}) due to technical artifacts.

4.1.2 PCA module

The PCA module addresses three purposes, namely the control of lambda inflation, calculation of ancestral components, and estimation of relatedness. These purposes are respectively outlined in Subsections 4.1.2.1, 4.1.2.2, and 4.1.2.3.

4.1.2.1 Control of lambda inflation

The PCA module generates two- and one-dimensional PC plots, which help to identify population outliers. These are useful quantities during the QC process as they allow lambda inflation due to population outliers to be adjusted for.

4.1.2.2 Calculation of ancestral components

The PCA module produces, by default, 20 PCs and identifies those that are significantly associated with the phenotype. These significant PCs are used to adjust for fine-grained population stratification; the first four PCs should be included regardless of their association with phenotypes.

4.1.2.3 Estimation of relatedness

IBD is evaluated between each pair of individuals within and across cohorts using a sufficiently large number of high-quality SNPs. This quantity is commonly known as **PiHat** and is defined as the sum of the probability of individuals sharing two homozygous alleles and half of the likelihood of them sharing heterozygous alleles. Identical twins or duplicated individuals will have a PiHat >0.9 (share 90+% of IBD); first-degree relatives will have a PiHat of approximately 0.5 (50% of IBD); and second- and third-degree relatives share approximately 25% and 12.5% of IBD, respectively. All pairs of individuals with PiHat >0.2 are identified. One individual is excluded among all pairs until there is no relatedness among any pair of individuals with PiHat >0.2. To maximize the power to detect associations, trios are preferentially retained, followed by cases, and finally individuals genotyped on a preferred platform specified by the user. This automated process can be modified manually.

4.1.3 Imputation module

This module implements the imputation of quality-controlled genotypes using a user-defined reference panel. The imputation module automates the following steps: First, it converts the cohort to hg19 (Human Build GRCh37) in the case of different genome builds (Hinrichs et al., 2006). Second, it aligns SNP names and alleles, removing SNPs whose alleles do not match the reference. Third, strand flips are adjusted for unambiguous (A/C and T/G) and ambiguous (A/T and G/C) SNPs. Fourth, from ambiguous SNPs, the module removes SNPs with frequencies

between 40% and 60%. SNPs with a frequency difference of more than 15% from the selected ancestral reference are also removed. For computational efficiency, the whole genome is split into genomic chunks of variable sizes with overlapping windows to keep the LD structure at the dividing lines. For example, in the imputation reference from the HRC, the data are split into 132 genomic chunks, but users are free to choose between a variety of chunk sizes. Within each genomic chunk, haplotype estimation (phasing) is performed with Eagle (Loh et al., 2016) and finally genotype imputation using Minimac3 (Das et al., 2016), although the user can also choose other imputation and phasing algorithms.

Once finished, the module will have generated genotype probability files for markers with INFO > 0.1 and MAF > 0.005. Moreover, a ready-for-analysis "best-guess genotype" is also created through the module making a hard call to the genotype with a probability >0.8. The pipeline produces three variations of best-guess genotype files based on the following filtration criteria: (1) no additional filter; (2) missing rate <2%; (3) missing rate <1%; and MAF >5%.

For trio cohorts, *Ricopili* generates pseudo-controls from the nontransmitted alleles after phasing, using relatedness information from the families. For example, if the alleles of the case are A/A and those of the parents are A/G and A/G, then the nontransmitted alleles will be G/G for the pseudo-controls. The imputation of each affected offspring (probands) and the perfectly matched pseudo-control is then performed independently. For downstream analyses, controlling for population stratification is not necessary.

4.1.4 Post-imputation module

The post-imputation module performs a genome-wide association separately within the imputed dosage chunks (e.g., ~132 by default in HRC imputation) generated in the imputation module. This process is highly parallelized and computes association statistics for each genomic chunk for each study independently. Additional covariates can be adjusted for along with the PCs generated in the PCA modules. Naturally, this module also allows the testing of alternative phenotypes. The meta-analysis of multiple cohorts is conducted within each chunk and then combined genome-wide. The pipeline also incorporates genome-wide summary statistics from external sources. It uses PLINK (Purcell et al., 2007), METAL (Willer et al., 2010), and R

scripts (R Core Team., 2013) to achieve its tasks. The module produces publication-ready results, including whole genome summary statistics, Manhattan plots, Q–Q plots, forest- and regional-level plots, and Excel files.

4.1.5 Post-imputation module for generating polygenic risk scores

The post-imputation module is also used to estimate PRSs for each individual in a test cohort. It uses genetic loads from independent LD-clumped training data (see Subsection 4.1.6.2). By default, SNPs with a MAF <5% and an info score <0.9 are excluded. To create the PRS of each individual, the natural log of each variant's odds ratio (OR) is multiplied by the imputation/genotype probability of the risk allele. The resulting values are then summed over the whole genome to create a single PRS for each individual. This is computed separately over 10 *p*-value thresholds (5×10^{-8}, 1×10^{-6}, 1×10^{-4}, 0.001, 0.01, 0.05, 0.1, 0.2, 0.5, and 1.0) in the training data.

Next, the module tests the association of the risk scores with the phenotype and estimates the variance explained. The variance explained is calculated using Nagelkerke's R^2 by comparing scores generated from a full model (containing covariates and PRSs) and a reduced model (covariates only). The liability for the trait or disease within the population is represented as a continuum of PRSs. The individuals with higher risk scores are positioned at higher points on the liability scale, indicating an elevated predisposition to the disease and vice-versa. Diseases with higher prevalence encompass a broader range of PRSs above a certain threshold on the liability scale, while diseases with a lower prevalence may involve individuals with exceptionally high PRSs.

4.1.6 Additional modules of the *Ricopili* pipeline

The following section describes the array of additional analysis modules within the Ricopili pipeline, which greatly facilitate downstream GWAS analysis.

4.1.6.1 Reference builder

This module helps the first-time user to set up the publicly available genome reference (e.g., a 1000 Genomes reference panel) or those available by permission (e.g., HRC) in the *Ricopili*

format on their cluster. It also defines genomic chunks for roughly equal SNP numbers throughout the genome.

4.1.6.2 Clumper module

This module allows LD-clumping to be performed on summary statistics to produce independent regions. By default, the most significant SNP is retained by discarding all SNPs within 500 kb with a high LD of $r^2 \geq 0.1$. This is useful for assessing the number of independent SNPs associated with the phenotype. In the complicated and large major histocompatibility complex (MHC) region of chromosome 6, only the most significant SNP is retained.

4.1.6.3 Replication module

The *Ricopili* pipeline also provides a module for performing a replication analysis. The replication module allows users to calculate sign tests of the effect directions between discovery and replication data and the meta-analyses of top SNPs (e.g., $p < 1 \times 10^{-06}$).

4.1.6.4 Leave-one-out polygenic risk score

This module eases the leave-one-out (LOO) PRS analysis even for a larger number of cohorts. The pipeline automates the creation of training data by leaving out every single cohort one at a time. It then creates the PRS (using the method described in Section 4.1.5) for individuals in the excluded cohort (see Section 5.1.4).

4.1.7 Simulated data

We simulated genotype data using the freely available tool HAPGEN (Su., 2011), which produces genome-wide haplotypes based on the LD pattern of a reference panel of one's choice. For this project, we used the 1000 Genomes Phase 3 reference panel to simulate genotype data and generated one cohort for a European ancestry population (the EUR data set; 503 reference individuals) and another for an East Asian ancestry population (the EAS data set; 504 reference individuals). After simulations, both data sets consisted of 100,000 simulated individuals. We performed first-pass filtering by removing all variants with an MAF <0.005, duplicate variant identifiers, or genomic location matches. After QC, 11,015,883 variants remained for the EUR data set, and 9,952,334 variants remained for the EAS data set. To create data sets relevant for education and method development, we further modified the simulated data.

4.1.7.1 Description of data interference: Population stratification

From the HAPGEN simulated sets, we randomly chose 6,000 individuals of European ancestry (EUR) and 200 of East Asian ancestry (EAS). We extracted SNPs from the widely used Illumina OmniExpress platform. The EUR cohort was split into five subcohorts (N = 2,000, 1,000, 1,000, 1,000, and 1,000) and the EAS cohort was split into two subcohorts (N = 100). The case and control status of every individual were assigned randomly (50:50) in the EUR cohort and at ratios of 2:98 and 5:95 to the EAS. Finally, we merged the two EAS cohorts with two of the EUR cohorts (cohort 1: N = 2,100 and cohort 3: N = 1,100; Table 4.1a) to create artificially unbalanced population stratification.

4.1.7.2 Description of data interference: Technical errors

We selected cohorts 1 and 2 to contain technical errors with false-positive associations. To do so, we performed the following steps **separately for cases and controls** in both cohorts:

1. To generate **autosomal heterozygosity rate deviations** in individuals, heterozygous genotypes were substituted with homozygous genotypes for all SNPs in **10** selected probands.
2. To create **missingness per individual**, **100** probands were randomly selected, and an SNP-missing rate per individual of 0–10% was introduced (from a right-skewed distribution with a lower probability of higher missing rates).
3. To create **sex errors**, **10** (male/female) probands were chosen randomly, and their gender assignments were switched to the opposite sex.
4. To create **missingness per SNP**, **2% of all SNPs** were randomly selected to induce missing genotypes by choosing missingness rates between 0–10% (from a right-skewed distribution with a higher probability of low missing rates).
5. To create **Hardy–Weinberg disequilibrium per SNP**, **2% of all SNPs** were randomly selected to induce artificial excess homozygosity.
6. To create **false-positive associations**, **20** SNPs were randomly selected, and their alleles A and B were swapped **along with missingness being introduced** to provide a sign for a technical error.

Finally, cases and controls were merged back into a single cohort. Table 4.1 provides a comprehensive description of each dataset, along with a detailed account of the interference applied to them.

Table 4.1a: Description of the simulated data cohort.

	Data set	**Name**	***N***	**Cases**	**Controls**	**SNPs**	**Interference**
1	sim_sim1a_eur_sa_merge.miss	hap1a	2,100	985	1,115	547,764[1]	Population stratification (merging European with Asian cohort 1) and technical errors
2	sim_sim2a_eur_sa_merge.miss	hap2a	1,000	474	526	593,970	Only technical errors
3	hapgen_sample3a	hap3a	1,100	478	622	547,764[1]	Only population stratification (merging European with Asian cohort 1)
4	hapgen_sample4b	hap4a	1,000	483	517	593,970	Shares 10 overlapping individuals with cohort 5
5	hapgen_sample5a.ph	hap5a	1,000	516	484	593,970	–

[1] The number of SNPs are less than the other cohorts as these were formed by a merging Asian and European cohort, so we only took the overlapping SNPs between them.

4.2 Panic disorder cohort description (Forstner, **Awasthi**, and Wolf et al., 2019):

This PD GWAS consisted of 2,248 cases and 7,992 controls originating from four European countries, namely Germany, Denmark, Estonia, and Sweden. Moreover, the written and informed consent of all participants was received, as was the approval of local ethics committees. DSM-III-R, DSM-IV, or ICD-10 criteria were used to diagnose lifetime PD in all patients. Table 4.2a presents the sample sizes for each cohort, both for cases and controls.

Table 4.2a: Description of the panic disorder cohort.

PD Cases (*N* = sample size)	**Controls (*N* = sample size)**
Germany I (*N* = 492)	Heinz Nixdorf Recall study (*N* = 1,882; Schmermund et al., 2002)
Germany II (*N* = 251)	Munich-based community cohort (*N* = 538)
Germany III (*N* = 290)	Munich-based community cohort (*N* = 856)
Denmark (*N* = 254)	Danish (*N* = 1,034)
Estonia (*N* = 346)	Estonian (*N* = 1,065)
Sweden (*N* = 615)	Swedish (*N* = 2,617; Ripke et al., 2002)

4.2.1 SNP QC of six PD cohorts

QC was applied to each cohort independently using the parameters and methodology described in Section 4.1.1. To summarize, we used the pre-imputation module to exclude problematic SNPs and samples, and if the Q–Q plots indicated high lambda inflation, they were scrutinized for genetic biases (Section 4.1.2). Details regarding pre- and post-QC sample sizes, along with the number of exclusions due to technical issues per cohort, are provided in Table 4.2b. Manual adjustments were necessary for the Germany I, Germany III, and Sweden cohorts.

4.2.2 Genomic QC

Relatedness testing was conducted using the PCA module with a subset of 47,513 high-quality SNPs. This subset was obtained through LD pruning ($r^2 > 0.02$), a process that involves

removing SNPs in high linkage disequilibrium, reducing data redundancy. Additionally, SNP with low imputation quality (INFO score < 0.8), high missingness (SNP missingness > 0.01), and low minor allele frequency (MAF < 0.05) were excluded. The PCA module identified 8, 4, and 12 pairs of related or overlapping individuals within the Germany I, Germany II, and Sweden cohorts, respectively. Additionally, 12 pairs of individuals were found to be related across the German cohorts, with further details provided in Table 4.2b. Randomly, one member out of each pair (case–case/control–control) with π-hat > 0.2 was removed. If related individuals are cases and controls, we retain the cases and exclude the controls. The PCs were also estimated from the genotype data, and their phenotype association was tested using logistic regression. PCs 1–7, 11, 16, and 18 significantly influenced the genome-wide test statistics and were used in further analyses to adjust for population stratification.

Table 4.2b: Pre- and post-QC distribution of sample sizes and single nucleotide polymorphisms. This table provides a comprehensive overview of SNPs and individual exclusion details, including corresponding technical errors.

Cohort	Pre-QC	Post-QC	Exclusion Count and Reasons
Germany I	Nsnp = 219,166	Nsnp = 218,563	• 361 SNPs were excluded due to them being missing in more than 2% of individuals. • 86 SNPs were excluded due to missing differences > 0.020 between cases and controls. • Two SNPs were **manually excluded** due to significantly higher deviation of AFs compared with the reference genome.
	Ncas = 492, Ncon = 1,882	Ncas = 472, Ncon = 1,803	• Four cases and six controls were excluded due to missing SNPs > 2%. • Four controls were excluded due to the heterozygosity rate being outside +−0.20. • 12 cases and 61 controls were excluded due to sex violations. • Eight pairs of overlapping or related samples were found within this cohort, resulting in the exclusion of eight controls. • Four cases were excluded as they were overlapping or related to cases in Germany II (n = 2) and Germany III (n = 2).
Germany II	Nsnp = 296,835	Nsnp = 295,955	• 110 SNPs were excluded due to missingness > 0.2. • 770 SNPs were excluded due to missing differences > 0.020 between cases and controls.

	Ncas = 251, Ncon = 538	Ncas = 247, Ncon = 537	• Four cases were excluded as they were overlapping or related to cases in Germany I along with one control that was related to an individual in Germany III.
Germany III	Nsnp = 486,864	Nsnp = 485,410	• 621 SNPs were excluded due to missingness > 0.2. • 829 SNPs were excluded due to missing differences > 0.020 between cases and controls. • Two SNPs were excluded due to an HWE *p* value of less than 10^{-8} in cases. • Two SNPs were **manually excluded** for a significantly higher deviation of AF compared with the reference genome.
	Ncas = 290, Ncon = 856	Ncas = 280, Ncon = 855	• Four cases were excluded due to missing SNPs > 2%. • Four pairs of overlapping and related samples were found within this cohort, resulting in the exclusion of four cases. • Two cases were excluded as they were overlapping or related to cases in Germany I along with one control that was related to an individual in Germany II.
Denmark	Nsnp = 248,028	Nsnp = 232,069	• 5,955 SNPs were excluded due to missingness > 0.2. • 2,472 invariant SNPs were excluded. • 1,776 SNPs were excluded due to missing differences > 0.020 between cases and controls. • 5,734 SNPs were excluded due to an HWE *p* value of less than 10^{-6} in controls. • 22 SNPs were excluded due to an HWE *p* value of less than 10^{-8} in cases.
	Ncas = 254, Ncon = 1,034	Ncas = 248, Ncon = 970	• Two controls were excluded due to missing SNPs > 2%. • Two cases and 20 controls were excluded due to sex violations. • Two cases and 42 controls were identified as population outliers and were subsequently excluded.
Estonia	Nsnp = 247,451	Nsnp = 225,045	• 2,925 SNPs were excluded due to missingness > 0.2. • 11,553 SNPs were excluded due to missing differences > 0.020 between cases and controls. • 2,479 invariant SNPs were excluded. • 169 SNPs were excluded due to an HWE *p* value of less than 10^{-8} in cases. • 5,280 SNPs were excluded due to an HWE *p* value of less than 10^{-6} in controls.

	Ncas = 346, Ncon = 1,065	Ncas = 339, Ncon = 1,004	• 26 controls were excluded due to missing SNPs > 2%. • 20 controls were excluded due to sex violations. • Seven cases and 15 controls were identified as population outliers and subsequently excluded.
Sweden	Nsnp = 222,726	Nsnp = 213,578	• 341 SNPs were excluded due to missingness > 0.2. • 384 SNPs were excluded due to missing differences > 0.020 between cases and controls. • Seven invariant SNPs were excluded. • Three SNPs were excluded due to an HWE *p* value of less than 10^{-6} in controls. • 7,413 SNPs with a MAF of less than 1% were **manually excluded.**
	Ncas = 615, Ncon = 2,617	Ncas = 561, Ncon = 2,591	• 12 cases were excluded due to missing SNPs > 2%. • 21 pairs of overlapping and related samples were found within this cohort, resulting in the exclusion of five cases and 16 controls. • 37 cases and 10 controls were identified as population outliers and subsequently excluded.

4.2.3 Imputation

The imputation of quality-controlled data for all six cohorts was performed using IMPUTE2 (Howie et al., 2012) and SHAPEIT (Delaneau et al., 2011) as well as the pre-phasing/imputation stepwise approach. We used default parameters, a chunk size of 3 megabases (Mb), and the 1000 Genomes Project reference panel (release "v3.macGT1").

4.2.4 Association testing and meta-analysis.

Each marker was tested for associations with PD using an additive logistic regression model and by controlling for PCs 1–7, 11, 16, and 18 independently in six PD cohorts. The *p* value threshold of 5×10^{-8} was used to assign genome-wide significance. Furthermore, the six cohorts were meta-analyzed using METAL (Willer et al., 2010), averaging the genetic effects (ORs) weighted by inverse standard error (SE) of the effect.

4.2.5 Leave-one-out polygenic risk scoring

LOO polygenic risk scoring was performed on all six cohorts by excluding one of them and using the remaining meta-analytic summary statistics to compute the PRS (using the method described in Section 4.1.5) for the individuals of the left-out cohort. This was done using the LOO-PRS module of the *Ricopili* pipeline.

4.2.6 Replication

We used three independent European PD studies for the follow-up analysis. These included iPSYCH (Denmark, n = 905 cases, n = 3,620 controls); deCODE (Iceland, n = 547 cases, n = 220,285 controls); and NESDA/NTR (the Netherlands, n = 956 cases, n = 4,565 controls). The combined meta-analysis included 2,498 PD patients and 228,470 from the follow-up cohorts. The meta-analysis was performed using the inverse standard error-weighted OR combination. A binomial "sign test" was performed on the number of same-direction effects in the replication data to test for significant accumulation.

4.2.7 Heritability estimation and genetic correlation

Linkage disequilibrium score regression (Sullivan et al., 2015) was used to determine the SNP-based heritability of PD. To test PD's genetic overlap with a wide range of phenotypes, we took advantage of the online database and web interface LD HUB (Zheng et al., 2017), which is based on the LD score regression method. The linkage disequilibrium adjusted kinship (LDAK) model of the SumHer tool was used to confirm the results.

4.2.8 Functional analysis

To classify our risk variants associated with PD, we performed a gene-based test, gene-set enrichment, and tissue enrichment analyses using a gene analysis tool MAGMA (de Leeuw et al., 2015) implemented in online web interface FUMA (Watanabe et al., 2017).

4.3 Borderline personality disorder cohort (Witt et al., 2017):

The BPD cohort consisted of 1,075 cases and 1,675 controls. The cases and controls of European ancestry were obtained from multiple sites (Table 4.3a) and merged into a single cohort. Patients aged 16–65 years with a lifetime DSM-IV diagnosis of BPD were included.

Table 4.3a: Description of the borderline personality disorder cohort.

BPD Cases (*N* = sample size)	**Controls (*N* = sample size)**
1. Department of Psychiatry, Charité, Campus Benjamin Franklin, Berlin (*N* = 494)	1. Central Institute of Mental Health Mannheim (*N* = 1,583)
2. Department of Psychosomatic Medicine, Central Institute of Mental Health, Mannheim (*N* = 350)	2. University Medical Center Mainz (*N* = 92)
3. Department of Psychiatry and Psychotherapy, University Medical Center Mainz (*N* = 231)	

LD score regression (Sullivan et al., 2015) was used to estimate the genetic correlation of BPD with schizophrenia (SCZ), major depressive disorder (MDD), and bipolar disorder (BP). The GWAS summary statistics for these disorders were obtained from Psychiatric Genomics Consortium (PGC) publications:

1. SCZ: 33,640 cases and 43,456 controls (Ripke et al. 2014);
2. BP: 20,352 cases and 31,358 controls (Stahl et al. 2019);
3. MDD: 16,823 cases and 25,632 controls (Ripke et al 2013).

4.4 IMAGEN cohort (Kaminski et al., 2018):

IMAGEN is a large multicenter (longitudinal) neuroimaging and genetics study at eight locations in four different European countries, namely Germany, the United Kingdom, France, and Ireland (Schumann et al., 2010). Each European center enrolls at least 250 healthy adolescents aged 14 years, who are then tracked at the ages of 16, 19, and 22 years. The primary aim is to leverage

the substantial sample size ($N > 2{,}000$) to establish a robust statistical confirmation of whether the associations between brain structure, function, and genetic variance are direct (e.g., linked to substance abuse) or if they emerge as secondary consequences of the disorder (e.g., due to the effects of pharmacological drugs (Mascarell Maričić and Walter et al., 2020).

For this project, our cohort comprised 1,475 subjects (mean age = 14.43 years; SD = 0.45, including 765 female participants) selected from the IMAGEN study.

4.4.1 Intelligence measurement

We estimated a measure of general cognitive ability by performing PCA (Deary et al., 2010) over WISC-IV scores, consisting of matrix reasoning, block design, digit span backward and forward, similarities, and vocabulary (Wechsler., 2003). The first PC explained the largest proportion of variance and was used as an intelligence marker (general IQ; gIQ).

4.4.2 Polygenic risk scoring

GWAS results for childhood intelligence (Benyamin et al., 2014) on 7,989 individuals (1,380,159 SNPs) and human intelligence (Sniekers et al., 2017) on 78,308 adults (10,499,625 SNPs) were first LD-clumped using the method described in Subsection 4.1.6.2. The clumped GWAS results were further used to generate polygenic scores (using the method described in Section 4.1.5) for each individual in the IMAGEN cohort. To examine the association between polygenic scores and gIQ, we performed linear regression models using the polygenic scores as predictors.

5) Results

This chapter presents the results obtained from this dissertation work. Section 5 provides a summary of the standardized GWAS methodology and details the technical and genomic QC results that refined this methodology.

Section 5.2 presents the results of the QC, imputation, meta-analysis, replication, leave-one-out polygenic risk score (LOO-PRS) analysis, heritability estimation, genetic correlation, gene based, get set and tissue enrichment analysis.

In Section 5.3, the results of genetic correlation estimation between borderline personality disorder (BPD) and psychiatric disorders such as schizophrenia (SCZ), major depressive disorder (MDD), and bipolar disorder (BP) are discussed.

Finally, Section 5.4 presents the results of the association between intelligence measures and polygenic scores in the IMAGEN cohorts

5.1 Standard GWAS analysis methodology (Lam and **Awasthi** et al., 2020):

The flowchart in Figure 4.1a (Chapter 4 – Materials and methods) illustrates the standard GWAS methodology that was developed in this project and used throughout. The flowchart uses various modules of the *Ricopili* pipeline along with some manual inspections and corrections.

In summary, this process entailed conducting thorough QC (both technical and genomic) through using the pre-imputation and PCA modules of the *Ricopili* pipeline, supplemented with manual quality checks. These two modules were employed iteratively, in various sequences, and multiple times if necessary, until a reliable genotype data set was obtained. Additionally, subsequent steps such as imputation, association/meta-analysis, and downstream modules were executed to perform a comprehensive GWAS analysis using our pipeline.

5.1.1 Simulated cohorts

The methodology presented in Section 5.1 and depicted in **Figure 4.1a** was used to analyze the simulated cohorts 1–5. This section only includes the results for cohort 1, as the same process was extended to all other cohorts (2–5). Later sections showcase plots and figures obtained through using the modules described in Section 4.1.

5.1.2 QC on simulated cohort 1

This section demonstrates the use of *Ricopili's* QC module to identify and rectify the technical interventions and biases introduced in the simulated data. This module generates comprehensive QC reports that aid in high-quality data analysis and visualization. The illustrations from Figures 5.1 a–i were extracted from this QC report.

The flags (Figure 5.1a) obtained from the initial run of the QC module of *Ricopili* indicated that this cohort needed more work even after automatic QC. Green flags indicate that no double-check was necessary and the cohorts could thus proceed directly to imputation and meta-analysis. Cohorts with yellow flags needed to be carefully examined; if a sufficient explanation was provided, they could progress to the next steps. Lastly, red flags signal that additional filtering and repeated QC were necessary until all red flags were resolved.

Flags

Nr.	Flagname	value	yellow-th	red-th	flag	color
01)	nsnps-postqc	507493	250000	200000	0	green
02)	nsnps-postqc-per-platform(OMEX)	507493	(408380/789375)	(217882/979873)	0	green
03)	ncases-postqc	929	100	50	0	green
04)	ncontrols-postqc	1059	100	50	0	green
05)	case-control-ratio-postqc	0.8772	(0.25/4)	(0.167/6)	0	green
06)	nids-lost-ratio	0.05333	0.01	0.1	1	yellow
07)	n-nopt-postqc	0	0	10	0	green
08)	nids-sexcheck-ratio	0.0004762	0.005	0.025	0	green
09)	lambda-postqc	2.541	1.1	1.2	2	red
10)	nsnps-gws	29	0	1	2	red

Figure 5.1a: All flagged (red), warning (yellow), and nonflagged (green) technical issues with the genotypes. [Ref. Lam and Awasthi et al., 2020, Bioinformatics: Permission granted]

Figure 5.1b presents an exclusion table for individuals and low-quality SNPs. The table indicates that 112 individuals were excluded due to a low call rate, sex violation, and heterozygosity deviation. A total of 40,271 (7.4%) SNPs were excluded due to high HWE deviation and high missing rates. After QC, 507,493 SNPs were retained. However, Figure 5.1b presents 29 SNPs that were significantly associated with the phenotype even after GC correction (genomic control for lambda inflation). These false positives would carry through imputation, and thus, they were scrutinized in the following steps.

General Info

Size of Sample

Test	pre QC	post QC	exclusion-N
Cases,Controls,Missing	985,1115,0	929,1059,0	56,56,0
Males,Females,Unspec	1332,768,0	1262,726,0	70,42,0
SNPs	547764	507493	40271 (7.4%)

Exclusion overview

would have excluded 73 individuals without pre-filter (SNP-Missing 0.05)

Filter	N
SNPs: call rate < 0.950 (pre - filter)	107 (0.0%)
IDs: call rate (cases/controls) < 0.980	73 (36/37)
IDs: FHET outside +- 0.20 (cases/controls)	20 (10/10)
IDs: Sex violations -excluded- (N-tested)	26 (2100)
IDs: Sex warnings (undefined phenotype / ambiguous genotypes)	1 (0/1)
SNPs: call rate < 0.980	8051 (1.5%)
SNPs: missing difference > 0.020	5472 (1.0%)
SNPs: without valid association p-value (invariant)	0 (0.0%)
SNPs: HWE-controls < -6	16427 (3.0%)
SNPs: HWE-cases < -10	10818 (2.0%)
Warning: genomewide significant SNPs (autosomal/known)	29 (29/0)

Figure 5.1b: Size of sample (top panel): This provides a broad overview of the sample size and SNP distribution before and after QC, along with the number of exclusions. The sample is further described in detail, split into two different categories. First, it is divided by phenotypes (cases, controls, and missing data), and second, it is divided by gender (males, females, and unspecified). Exclusion overview (bottom panel): This provides a quantitative list of excluded SNPs and individuals (in the "N" column) along with the corresponding technical reasons for exclusion (in the "Filter" column). [Ref. Lam and Awasthi et al., 2020, Bioinformatics: Permission granted]

The Q–Q plot (Figure 5.1c) revealed inflated lambda values even after QC filters. These were at least partly driven by population stratification given the two genetically distinct populations (Figure 5.1d) in this cohort and/or sample overlap. Here, as we sought to restrict the analysis to

European ancestry, we chose a threshold that excluded the much smaller Asian subset (e.g., PCA1 = 0.01).

3.2 Manhattan-Plot - post-QC (QQplot with MAF 0.02)

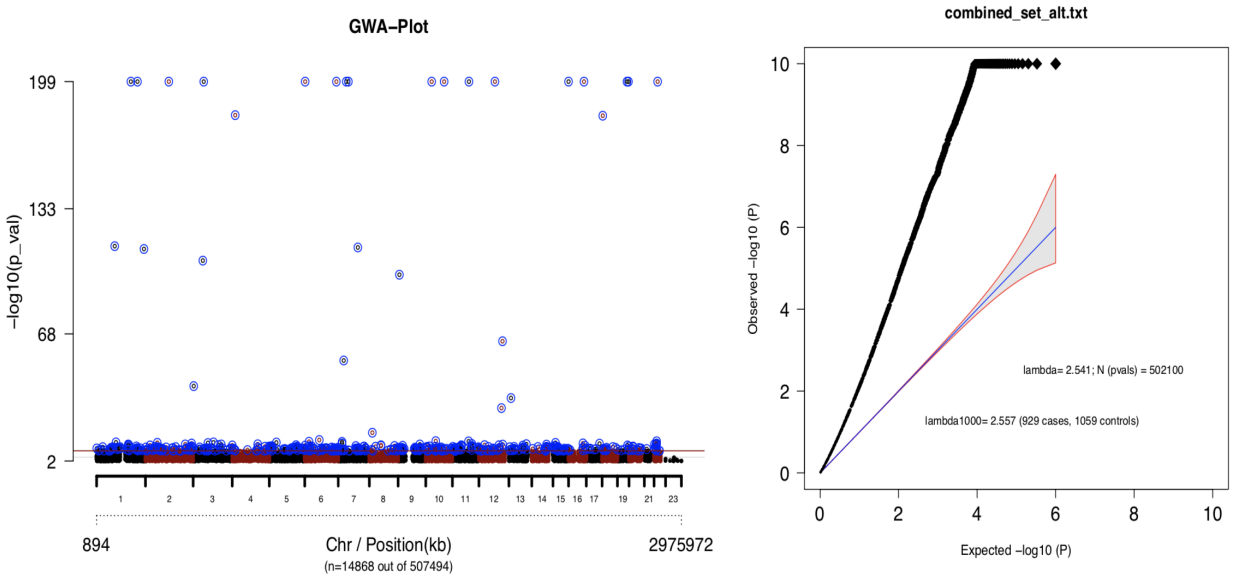

Figure 5.1c: Post-QC Manhattan plot (left); The y axis represents the −log of *p* values for variants with a *p* value < 0.001, presented along their genomic locations on the x axis (22 autosomes). Q–Q plot (right): The x axis represents expected −log10(*P*) values, while the y axis represents observed −log10(*P*) values. The gray shaded area surrounded by a red line indicates the 95% confidence interval under the null (no inflation). **Lambda** is the observed median χ^2 test statistic divided by the median expected χ^2 test statistic under the null ($p = 0.5$). Since lambda scales with sample sizes, it is informative to examine the rescaled lambda for 1,000 cases and 1,000 controls (i.e., lambda 1000; de Bakker., 2008). All SNPs with MAF > 0.02 were used to create the Q–Q plot [Ref. Lam and Awasthi et al., 2020, Bioinformatics: Permission granted].

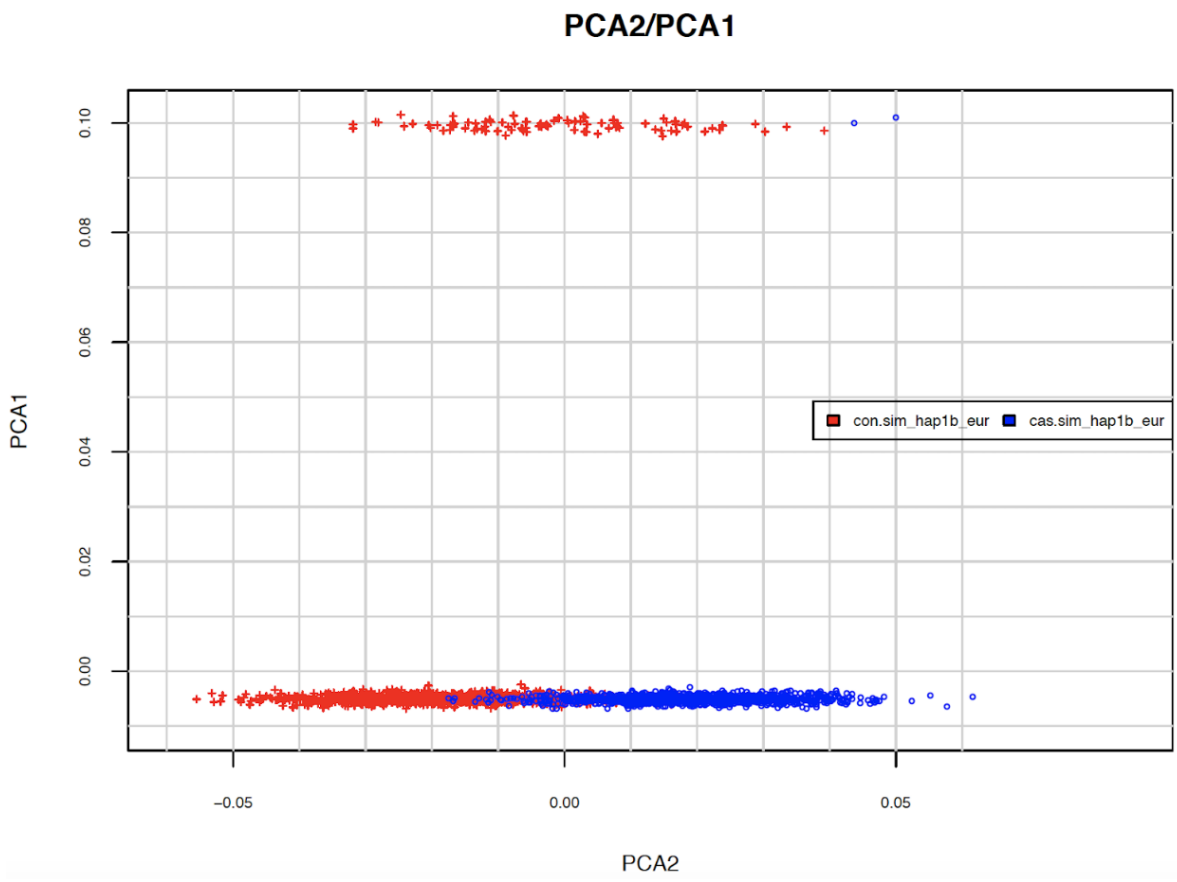

Figure 5.1d: PCA plot: x axis = principal component (PC) 1 and y axis = PC 2. Red dots represent cases and blue dots represent controls. Two distinct populations, European (PCA1 < 0.00), and East Asian (PCA1 > 0.08), are identified [Ref. Lam and Awasthi et al., 2020, Bioinformatics: Permission granted].

After we removed the Asian subset from the cohort, a significant drop in lambda inflation occurred (compare Figure 5.1 c and e). Still, there were 28 genome-wide significant SNPs, as listed in Figure 5.1f. In summary, some red QC flags from Figure 5.1a were resolved, but not all of them. In Figure 5.1f, we identified high missing rates indicated in columns **F_MISS** (overall proportion of missing genotype data), **F_MISS_A (**proportion of missing genotype data in cases), and **F_MISS_U (**proportion of missing genotype data in controls) for flagged SNPs and excluded them from further analysis.

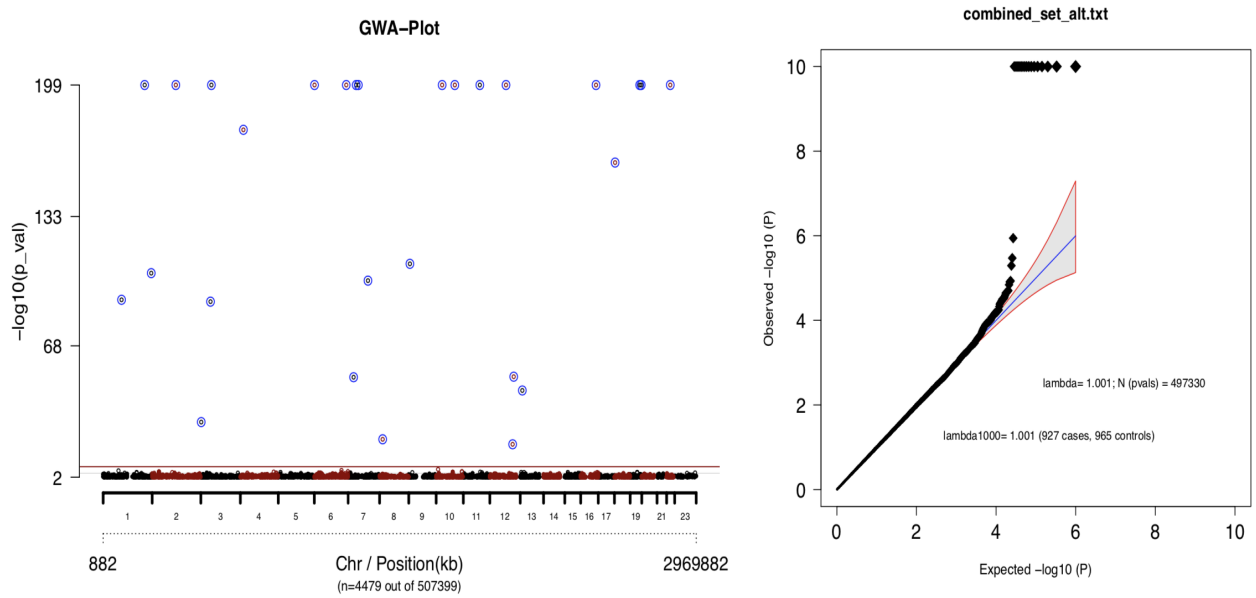

Figure 5.1e: Manhattan plot (left) and Q–Q plot (right) after removing population outliers. Compare it with Figure 5.2.c [Ref. Lam and Awasthi et al., 2020, Bioinformatics: Permission granted].

RSID	CHR	BP	A1	F_A	F_U	A2	P_ASSOC	OR	SE	MAF	F_MISS	F_MISS_A	F_MISS_U	F_MISS_DIFF	F_MISS_P	log(P)_HWE_cas	log(P)_HWE_con	log(P)_HWE_all
rs10862292	12	81711953	G	0.9277	0.06955	A	0	171.7	0.1277	0.4903	0.01586	0.0151	0.01658	-0.00148	0.8553	0	-0.0967	-246
rs7611432	3	53971360	G	0.9571	0.0385	T	0	557.7	0.1663	0.4884	0.01797	0.01834	0.01762	0.000719999999999998	1	-0.394	0	-200
rs4848495	2	118928678	C	0.9091	0.08281	A	0	110.8	0.1165	0.4882	0.01638	0.0151	0.01762	-0.00252	0.7196	-0.162	-0.875	-211
rs221795	7	100283261	C	0.6663	0.3158	T	1.797e-101	4.326	0.07003	0.4874	0.01638	0.01726	0.01554	0.00172	0.8569	-0.299	-0.301	-6.92
rs6940091	6	2437996	G	0.8615	0.1452	A	0	36.63	0.09413	0.4962	0.0185	0.01834	0.01865	-0.000310000000000001	1	-0.313	0	-115
rs37267	7	41337044	A	0.8571	0.1126	G	0	47.3	0.09888	0.4776	0.01903	0.01834	0.01969	-0.00135	0.8676	-0.0495	0	-133
rs7524120	1	92865160	G	0.6559	0.3224	A	7.507e-92	4.006	0.06965	0.486	0.0185	0.01726	0.01969	-0.00243	0.7353	-0.517	-0.0844	-7.73
rs11758794	6	161521680	G	0.9552	0.04198	A	0	487.1	0.1612	0.4925	0.0185	0.01187	0.02487	-0.013	0.04052	0	-0.395	-200
rs1417731	10	93635541	G	0.9071	0.08078	A	0	111.2	0.1168	0.4857	0.0185	0.01834	0.01865	-0.000310000000000001	1	0	-0.294	-203
rs7043262	9	6926738	G	0.676	0.3103	A	6.387e-110	4.638	0.07058	0.4889	0.01956	0.02265	0.01658	0.00607	0.4071	-0.0271	-0.705	-10
rs238112	18	3499767	A	0.7233	0.2802	G	8.939e-161	6.715	0.07318	0.4968	0.01691	0.01942	0.01451	0.00491	0.4771	-0.168	-0.0594	-16.9
rs1929254	10	30186233	T	0.7555	0.2466	C	3.847e-211	9.441	0.07626	0.4965	0.0185	0.01618	0.02073	-0.00455	0.4986	-0.184	0	-29.3
rs11234161	11	84469196	T	0.8812	0.1183	G	0	55.29	0.1016	0.4919	0.01903	0.01942	0.01865	0.00077	1	0	-0.0576	-144
rs12533618	7	52142647	A	0.9078	0.1111	G	0	78.8	0.1091	0.4992	0.01691	0.02265	0.0114	0.01125	0.07353	-0.161	-1.67	-184
rs696964	1	208598450	G	0.7563	0.2282	A	1.04e-227	10.5	0.07727	0.4863	0.01691	0.01942	0.01451	0.00491	0.4771	-0.492	-0.706	-26.6
rs16947913	16	78554218	C	0.8069	0.1837	A	0	18.57	0.08391	0.4884	0.01744	0.01942	0.01554	0.00388	0.5995	-0.47	-0.0391	-59.6
rs149150	3	3094216	G	0.5857	0.3976	A	1.954e-30	2.142	0.06685	0.4898	0.0185	0.01834	0.01865	-0.000310000000000001	1	-0.077	-1.04	-0.215
rs11715915	3	49455330	T	0.6625	0.3309	C	6.33e-91	3.969	0.06953	0.4933	0.01691	0.01726	0.01658	0.00068	1	-0.427	-0.146	-6.37
rs2157216	22	35344317	A	0.7968	0.1828	G	5.267e-308	17.54	0.08306	0.4834	0.01427	0.0151	0.01347	0.00163	0.8473	-0.0368	0	-60.6
rs1248060	12	114864252	C	0.5667	0.4195	T	2.856e-19	1.81	0.06634	0.4919	0.01744	0.01294	0.02176	-0.00882	0.1619	-0.697	-0.0487	-0.0333
rs4698491	4	16526736	A	0.7243	0.259	G	3.17e-177	7.516	0.07412	0.4876	0.01691	0.01402	0.01969	-0.00567	0.3761	-0.669	-0.303	-21.3
rs11772815	7	28391047	A	0.3713	0.6225	G	5.805e-53	0.3582	0.06777	0.4992	0.01691	0.0151	0.01865	-0.00355	0.5959	-0.447	-0.62	-0.676
rs168474	8	15599587	G	0.5728	0.4155	A	8.632e-22	1.886	0.0664	0.4927	0.01638	0.01402	0.01865	-0.00463	0.4723	-0.164	-0.457	-0.287
rs275581	19	48848274	C	0.2123	0.7712	A	4.63e-255	0.07997	0.0791	0.4987	0.0148	0.01942	0.01036	0.00906	0.1275	-0.312	-0.504	-35.9
rs558107	13	30172458	G	0.6175	0.3833	A	2.595e-46	2.597	0.06743	0.4984	0.01586	0.01294	0.01865	-0.00571	0.3607	-0.849	-0.167	-0.979
rs740032	12	120264341	G	0.6203	0.3684	A	3.096e-53	2.8	0.06779	0.4911	0.01691	0.02265	0.0114	0.01125	0.07353	-0.0517	-0.576	-1.11
rs10412597	19	56473189	G	0.2168	0.7592	T	5.888e-240	0.08779	0.0782	0.4935	0.01691	0.01726	0.01658	0.00068	1	-0.073	0	-37.3
rs12140273	1	241542880	C	0.6786	0.321	T	3.096e-105	4.465	0.0703	0.4962	0.0185	0.01834	0.01865	-0.000310000000000001	1	-1.02	-0.341	-8.7

Figure 5.1f: Summary statistics of genome-wide significant SNPs. **"RSID"**: SNP identifier; **"CHR"**: chromosome number; **"BP"**: base pair position of the SNP; **"A1"**: reference allele for the SNP and also the effect allele; **"F_A"**: allele frequency of the effect allele; **"F_U"**: allele frequency of the effect allele in controls; **"A2"**: alternative allele for the SNP; **"P_ASSOC"**: *p* value representing the statistical association of the SNP with the phenotype; **"OR"**: odds ratio (a measure of effect size); **"SE"**: standard error of the OR; **"MAF"**: minor allele frequency; **"F_MISS"**: overall proportion of missing genotype data; **"F_MISS_A"**: proportion of missing genotype data in cases; **"F_MISS_U"**: proportion of missing genotype data in controls; **"F_MISS_DIFF"**: difference in missing genotype data between cases and controls; **"F_MISS_P"**: *p* value representing the statistical significance of the difference in missing genotype data between cases and controls; **"log(P)_HWE_cas"**: log-transformed *p* value for the HWE test in cases; **"log(P)_HWE_con"**: log-transformed *p* value for the HWE test in controls; and **"log(P)_HWE_all"**: log-transformed *p* value for the HWE test in all samples (cases and controls combined) [Ref. Lam and Awasthi et al., 2020, Bioinformatics: Permission granted].

Flags

Nr.	Flagname	value	yellow-th	red-th	flag	color
01)	nsnps-postqc	503121	250000	200000	0	green
02)	nsnps-postqc-per-platform(OMEX)	503121	(408380/789375)	(217882/979873)	0	green
03)	ncases-postqc	927	100	50	0	green
04)	ncontrols-postqc	965	100	50	0	green
05)	case-control-ratio-postqc	0.9606	(0.25/4)	(0.167/6)	0	green
06)	nids-lost-ratio	0	0.01	0.1	0	green
07)	n-nopt-postqc	0	0	10	0	green
08)	nids-sexcheck-ratio	0.0005285	0.005	0.025	0	green
09)	lambda-postqc	1	1.1	1.2	0	green
10)	nsnps-gws	0	0	1	0	green

Figure 5.1g: Flagging after automatic and manual QC [Ref. Lam and Awasthi et al., 2020, Bioinformatics: Permission granted].

General Info

Size of Sample

Test	pre QC	post QC	exclusion-N
Cases,Controls,Missing	927,965,0	927,965,0	0,0,0
Males,Females,Unspec	1166,726,0	1166,726,0	0,0,0
SNPs	503121	503121	0 (0.0%)

Exclusion overview

would have excluded 0 individuals without pre-filter (SNP-Missing 0.05)

Filter	N
SNPs: call rate < 0.950 (pre - filter)	0 (0.0%)
IDs: call rate (cases/controls) < 0.980	0 (0/0)
IDs: FHET outside +- 0.20 (cases/controls)	0 (0/0)
IDs: Sex violations -excluded- (N-tested)	0 (1892)
IDs: Sex warnings (undefined phenotype / ambiguous genotypes)	1 (0/1)
SNPs: call rate < 0.980	0 (0.0%)
SNPs: missing difference > 0.020	0 (0.0%)
SNPs: without valid association p-value (invariant)	0 (0.0%)
SNPs: HWE-controls < -6	0 (0.0%)
SNPs: HWE-cases < -10	0 (0.0%)
Warning: genomewide significant SNPs (autosomal/known)	0 (0/0)

Figure 5.1h: SNP and individuals excluded due to various technical issues after automatic and manual QC [Ref. Lam and Awasthi et al., 2020, Bioinformatics: Permission granted].

Finally, Figures 5.1g and 5.1h indicate that with these various QC actions, we produced a data set with no noticeable biases. The 10 green QC flags (in Figures 5.1g) also support this. After QC, 507,493 SNPs remained, which was a sufficient number (i.e., greater than 200,000) for proper imputation. The previous sections have described the QC actions and results specific to cohort 1. The same process (also outlined in the flowchart in Figure 4.1a) was subsequently extended to all other cohorts (2–5).

5.1.3 Meta-analysis

The 7,149,025 SNPs imputed in the simulated cohorts were tested for association with the randomly generated null phenotype. As expected, no genome-wide significant marker was discovered (Figure 5.1i), and genome-wide inflation of p values was absent.

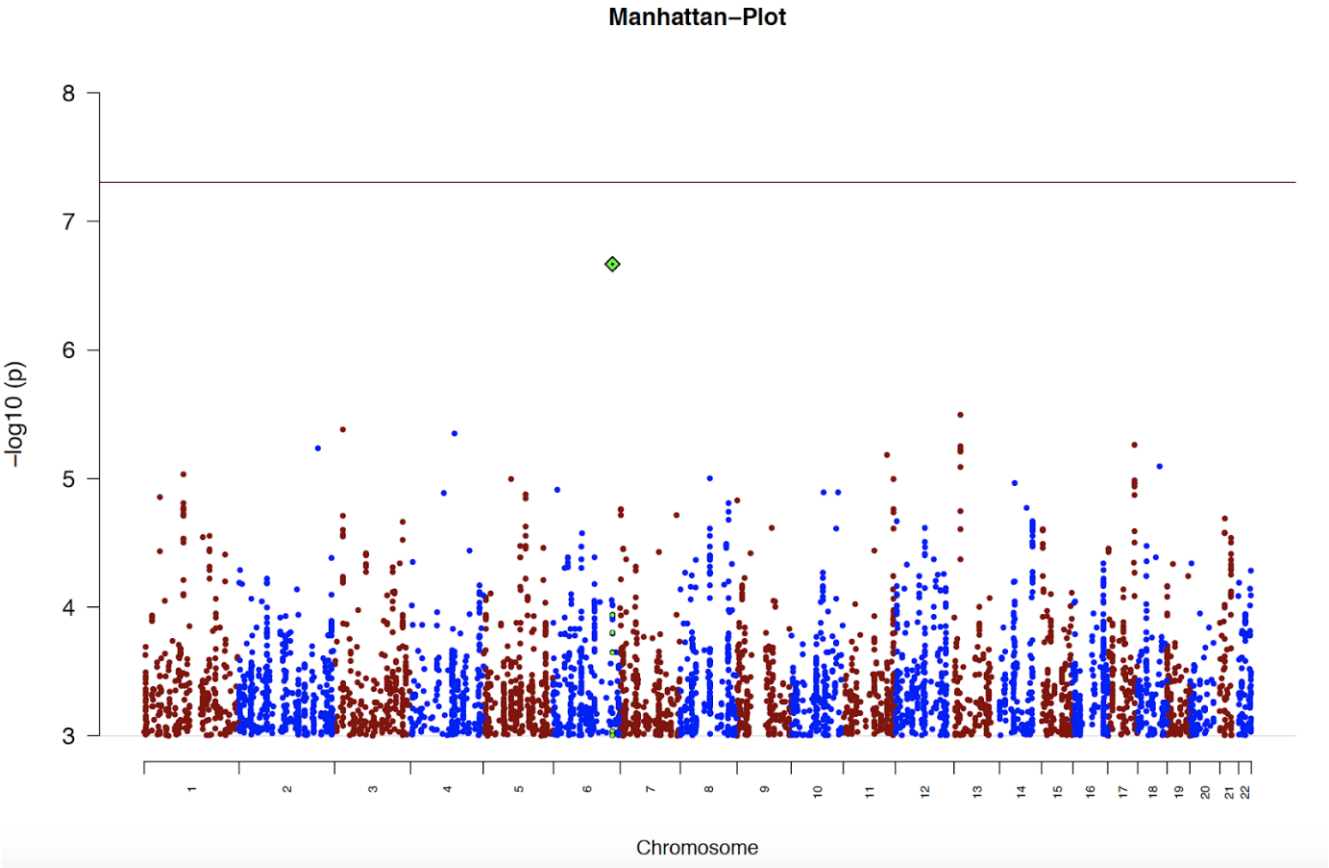

Figure 5.1i: Manhattan plot illustrating association results of 2,820 random cases and 2,962 random controls. The y axis represents the −log of *p* values for variants with $p < 0.001$, presented along their genomic locations on the x axis (22 autosomes). The green diamond is the lead variant in each locus, and the red line is genome-wide significance [Ref. Lam and Awasthi et al., 2020, Bioinformatics: Permission granted].

5.1.4 LOO-PRS analysis

PRSs for each simulated cohort were calculated through the GWAS meta-analysis of the remaining four cohorts. As expected for a simulated collection, the correlations (measured as Nagelkerke's r^2) between genetic risk scores and the random phenotypes displayed a seemingly random direction of effect with no statistical significance for any of the *p* value thresholds (Figure 5.1j).

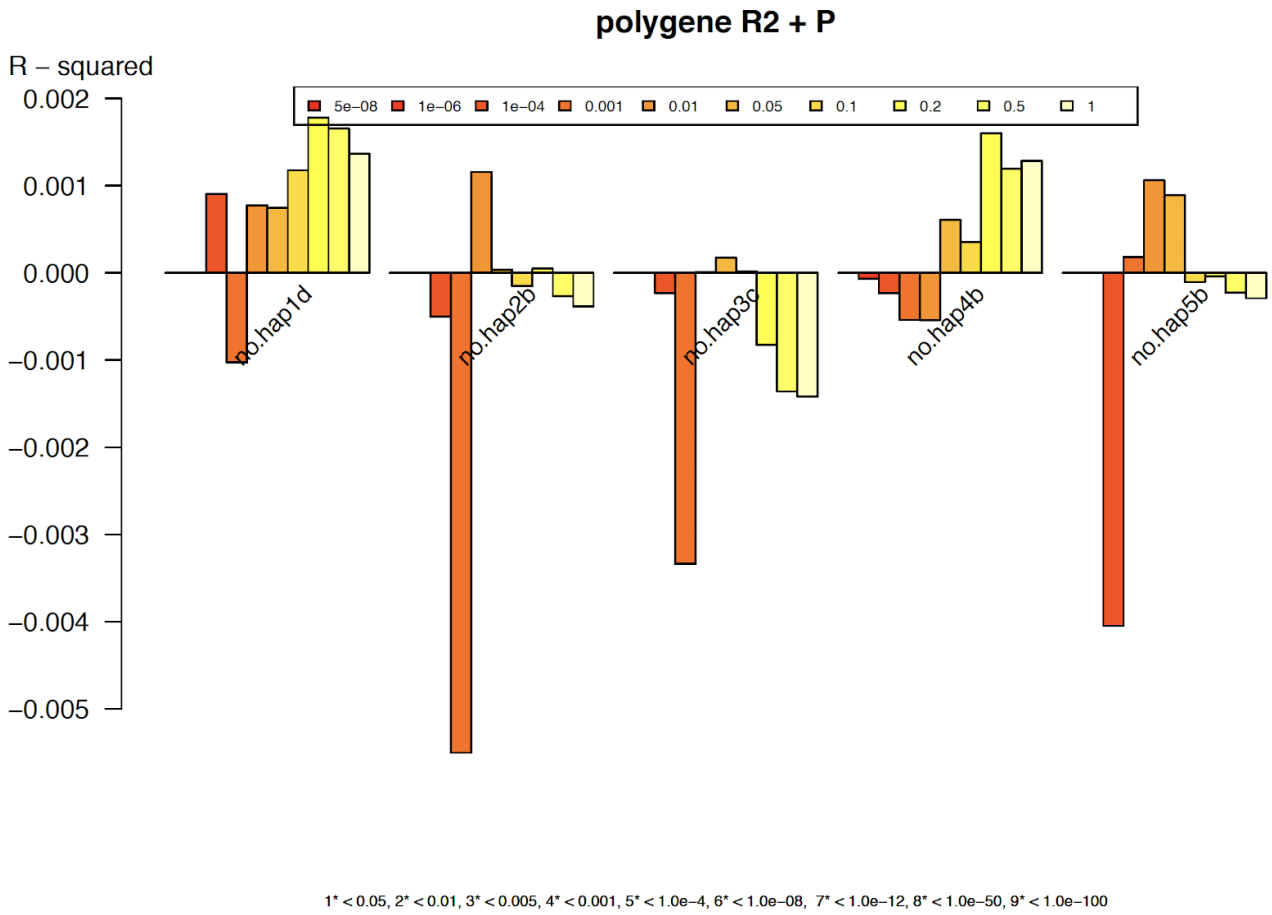

Figure 5.1j: Leave-one-out polygenic risk score: This figure illustrates the PRS results of each of the five simulated cohorts generated using GWAS meta-analysis effects of the remaining four cohorts as training data. PRSs were calculated at 10 different *p*-value thresholds indicated by respective color bars in the legend. The variance explained by PRSs is presented along the y axis. Not a single result in any category was statistically significant at $p < 0.05$ (and not a single absolute R^2 score reached more than 0.5% explained variance) [Ref. Lam and Awasthi et al., 2020, Bioinformatics: Permission granted].

As expected in simulated cohorts, the null phenotype yielded an SNP heritability that was indistinguishable from 0 (2.24% – SE of 7%).

5.1.6 LD score regression analysis: Genetic correlation (co-heritability)

No co-heritability existed between the association results with the null phenotype and the results from the SCZ and MDD GWAS.

5.1.7 Replication analysis

Sign analysis was performed between simulated cohort 4 (replication cohort) and discovery meta-analysis of the remaining four cohorts as the discovery set. In the binomial test, 60 out of 127 top associated SNPs (discovery p value $< 1 \times 10^{-4}$) exhibited the same direction effects, which were not distinguishable from randomness – as expected for a null phenotype.

5.2 Panic disorder GWAS (Forstner, **Awasthi** and Wolf et al., 2019):

5.2.1 QC and imputation

The thorough technical and genomic QC resulted in more than 200,000 SNPs in each cohort (Table 4.2b), which was a sufficient number for genotype imputation. Genotype imputation further increased the SNP count (by predictions based on the reference genome) to approximately 10 million in each cohort.

5.2.2 Meta-analysis results

None of the 8,757,275 high-quality SNPs revealed any significant association to PD in the meta-analysis (Figure 5.2a), and the genome-wide association signal exhibited no significant inflation (Q–Q plot in Figure 5.2b). A small intergenic deletion on chromosome 14 exhibited the highest significance ($p = 1.01 \times 10^{-7}$, OR = 1.64, MAF in controls = 0.05, MAF in cases = 0.07, imputation INFO score = 0.59).

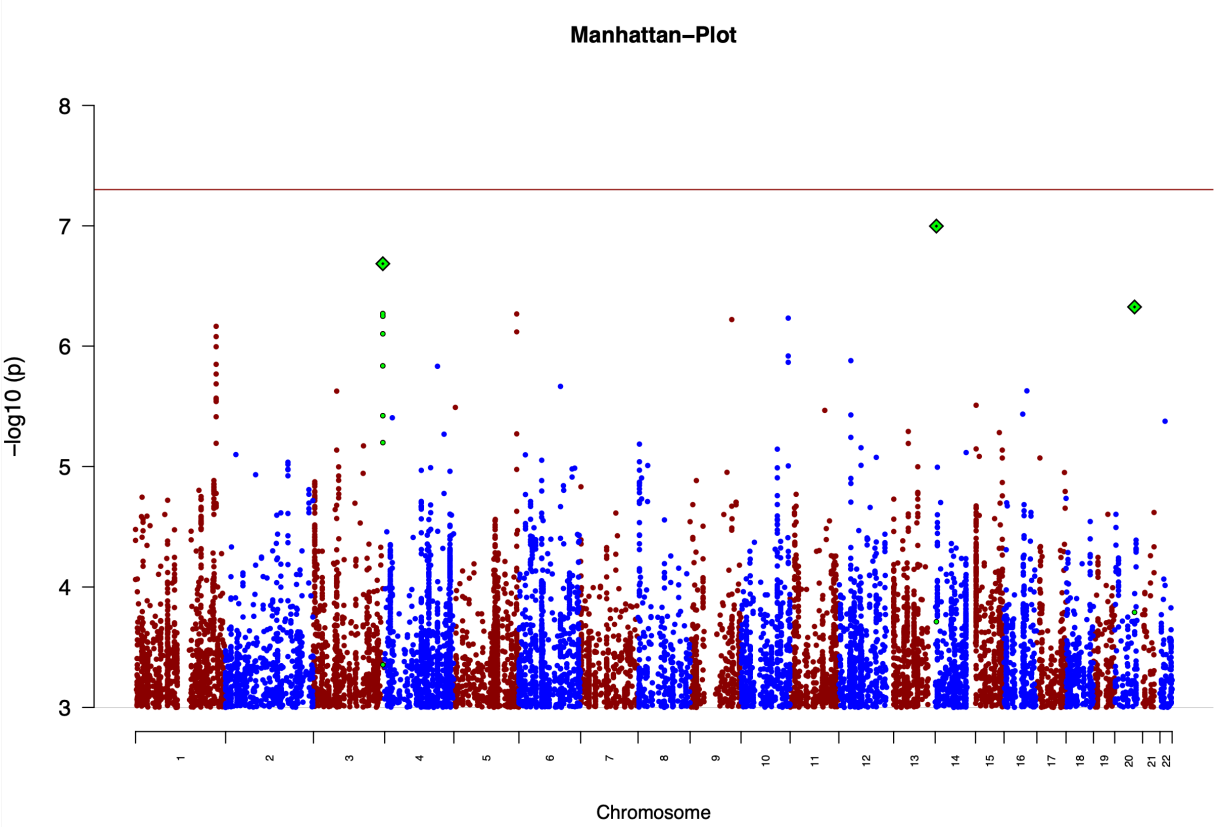

Figure 5.2a: Manhattan plot illustrating association results of 2,147 PD cases and 7,760 controls. The y axis represents −log of *p* values for all the tested variants, while their genomic location is along the x axis. The green diamond is the lead variant in each locus, and the red line is genome-wide [Ref. Andreas, Awasthi and Wolf et al., 2019, Molecular Psychiatry: Permission granted].

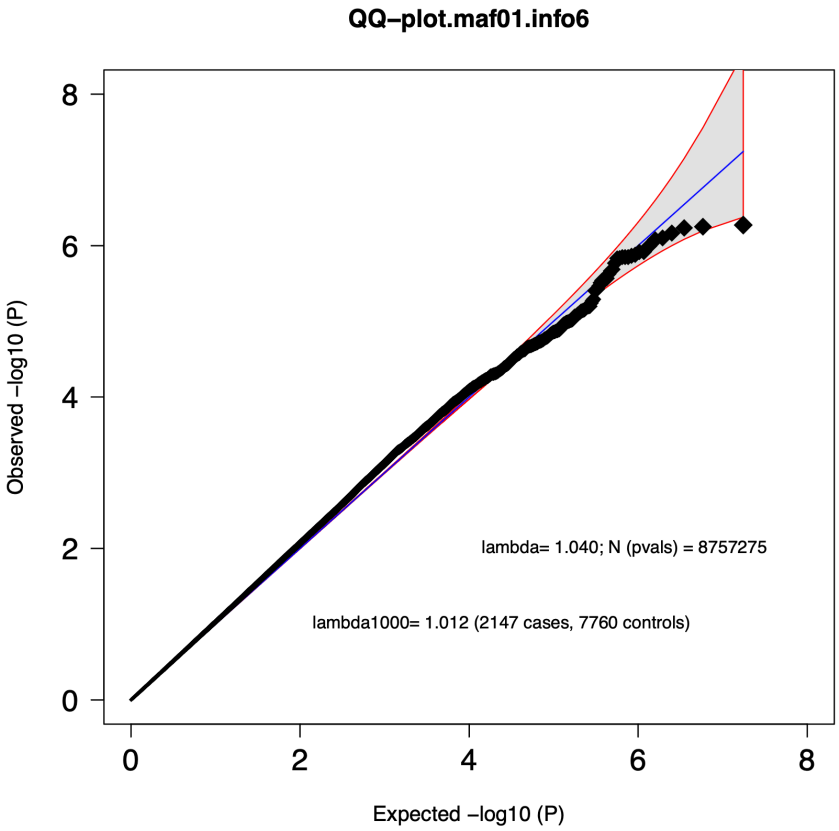

Figure 5.2b: Q–Q plot – the x axis represents expected −log10(*P*) values and the y axis represents observed −log10(*P*) values. The gray-shaded area surrounded by a red line indicates the 95% confidence interval under the null (no inflation). Lambda is the observed median χ^2 test statistic divided by the median expected χ^2 test statistic under the null ($p = 0.5$). Since lambda scales with sample sizes, it was informative to examine the rescaled lambda for 1,000 cases and 1,000 controls (i.e., lambda 1,000; de Bakker., 2008) [Ref. Andreas, Awasthi and Wolf et al., 2019, Molecular Psychiatry: Permission granted].

5.2.3 Replication analysis

By comparing the direction of effects between the replication and discovery results on the 243 SNPs with a discovery *p* value of $< 1 \times 10^{-4}$, we were able to demonstrate a nominally significant ($p = 0.048$) proportion of same-direction effects in 135 SNPs using the sign test.

The combined meta-analysis of the discovery PD and replication cohorts found no significant single SNP genome-wide association, with PD.SNP rs144783209 – located on chromosome 4/intron 1 of the SMAD1 gene – exhibiting the lowest *p* value ($Pcomb = 3.10 \times 10^{-7}$).

5.2.4 LOO-PRS analysis

In contrast to the null phenotype (Figure 5.1j and Section 5.1.4) the LOO-PRS analysis for PD reliably predicted disease status in all five cohorts (Figure 5.2c). For lower *p*-value thresholds, the analysis demonstrated the same direction of effect in all cohorts, with the maximum explained observed variance ranging from 0.8% (Swedish) to 2.6% (Germany II).

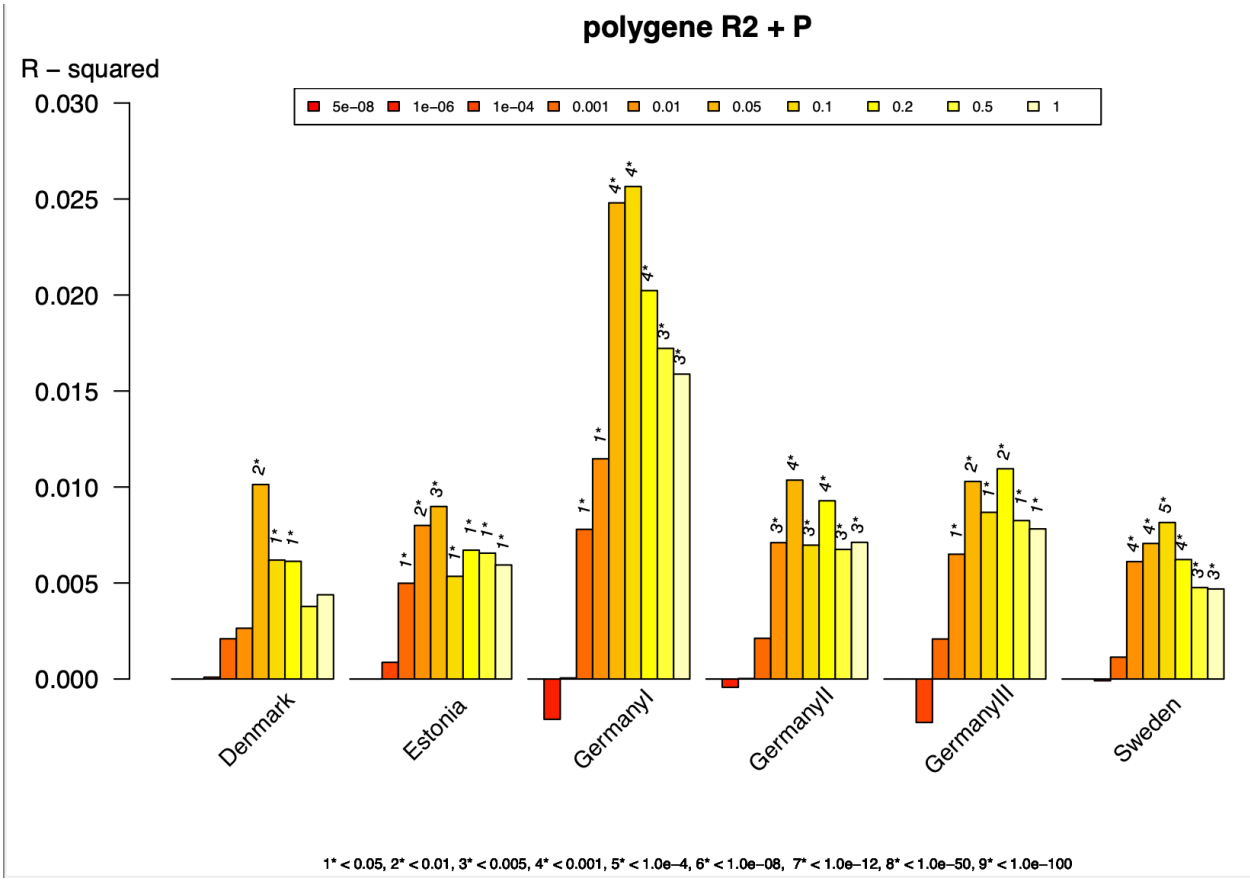

Figure 5.2c: Leave-one-out polygenic risk score (PRS) results of six GWAS cohorts using the GWAS results from the remaining five at 10 different *p*-value thresholds, as indicated by the respective color bars in the legend. The observed variance explained by PRS is along the x axis; * indicates the significance of variance over each bar [Ref. Andreas, Awasthi and Wolf et al., 2019, Molecular Psychiatry: Permission granted].

5.2.5 LD score regression analysis: SNP-based heritability

Using the LD score regression (SR) method, we observed that SNP-based heritability for PD significantly ranged from 28.0% (SD = 5.7%) for a lifetime prevalence of 2% to 34.2% (SD =

6.9%) for a lifetime prevalence of 4%. Using the SumHer/LDAK method, we confirmed these values with 36.3% (SD = 4.7%) and 44.0% (SD = 5.6%), respectively.

5.2.6 LD score regression analysis: Genetic correlation (co-heritability)

A significant genetic correlation of PD was observed with MDD (genetic correlation/*rg* = 0.431; SE = 0.134; *p* value corrected for multiple testing/*Pcorr* = 0.025); depressive symptoms (*rg* = 0.322; SE = 0.093; *pcorr* = 0.010); and neuroticism (*rg* = 0.316; SE = 0.082; *Pcorr* = 0.002). Moreover, a slightly significant positive genetic correlation was observed with anxiety disorder, posttraumatic stress disorder (PTSD), PGC cross-disorder analysis phenotype, and SCZ, whereas a significantly negative genetic correlation was observed with years of schooling as a surrogate marker for IQ (Figure 5.2d).

Furthermore, we estimated the co-heritability of PD with MDD in two settings – namely PD patients (1) with and (2) without comorbid MDD. Noteworthily, MDD exhibited a nominally significant genetic correlation with PD without MDD (*rg* = 0.415; SE = 0.209; *p* = 0.047), but it had no significant correlation with PD with MDD (*rg* = 0.662; SE = 0.422; *p* = 0.117). This supports the hypothesis of an observed co-heritability between PD and MDD that is largely driven by shared genetic variants and not by PD–MDD comorbidity. A disparity existed in sample sizes between these two settings, indicating the need for additional confirmation.

Moreover, using the SumHer/LDAK method, we confirmed the strong positive genetic correlation between PD and (i) MDD (*rg* = 0.208; SD = 0.065); (ii) depressive symptoms (*rg* = 0.275; SD = 0.092); and (iii) neuroticism (*rg* = 0.260; SD = 0.077) and other phenotypes.

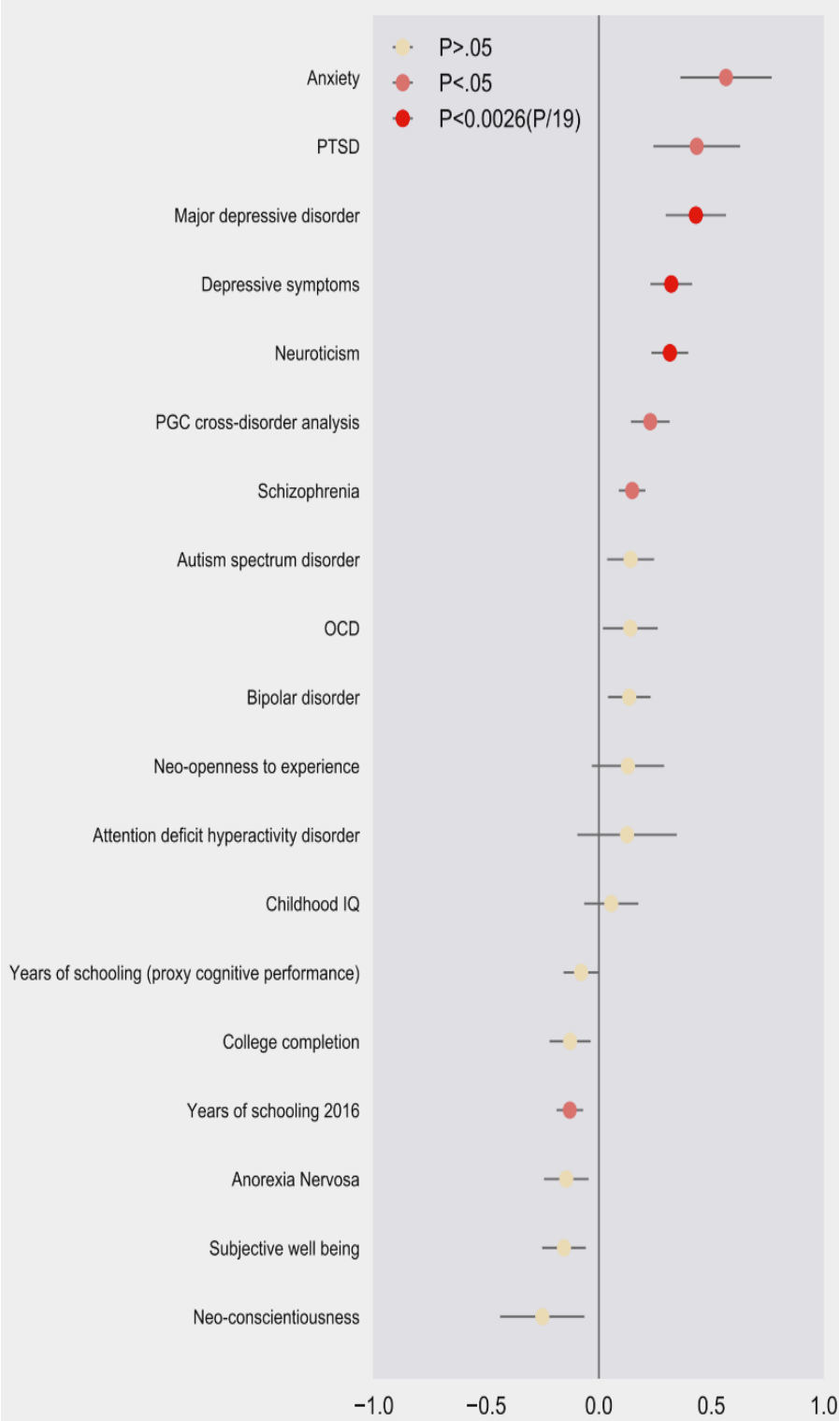

Figure 5.2d: Genetic correlations between PD and other phenotypes: For the 19 phenotypes on the right, the genetic correlation is denoted as a dot with a line as the standard error. The significance of each correlation is described in the upper left [Ref. Andreas, Awasthi and Wolf et al., 2019, Molecular Psychiatry: Permission granted].

5.2.7 Gene-based analysis

A gene-based analysis was performed using MAGMA on a total of 18,335 genes. In total, 42 genes were found to be significantly associated with PD at $p = 0.001$, but none of them reached significance after Bonferroni correction ($p < 0.05/18{,}335 = 2.73 \times 10^{-6}$).

5.2.8 Gene set analysis

A gene set analysis using gene-based results (Section 5.2.6) revealed 521 gene sets/pathways to be nominally significant. Again, however, none of the gene sets remained significant ($p < 0.05/10\ 891 = 4.59 \times 10^{-6}$) after Bonferroni correction.

5.2.9 Tissue enrichment analysis

Tissue enrichment analysis revealed an enrichment of associated genes expressed in brain tissues. Genes expressed in the cortex exhibited the most robust enrichment, followed by those expressed in the amygdala (Figure 5.2e). Here again, none of the 53 investigated tissues reached statistical significance after Bonferroni correction.

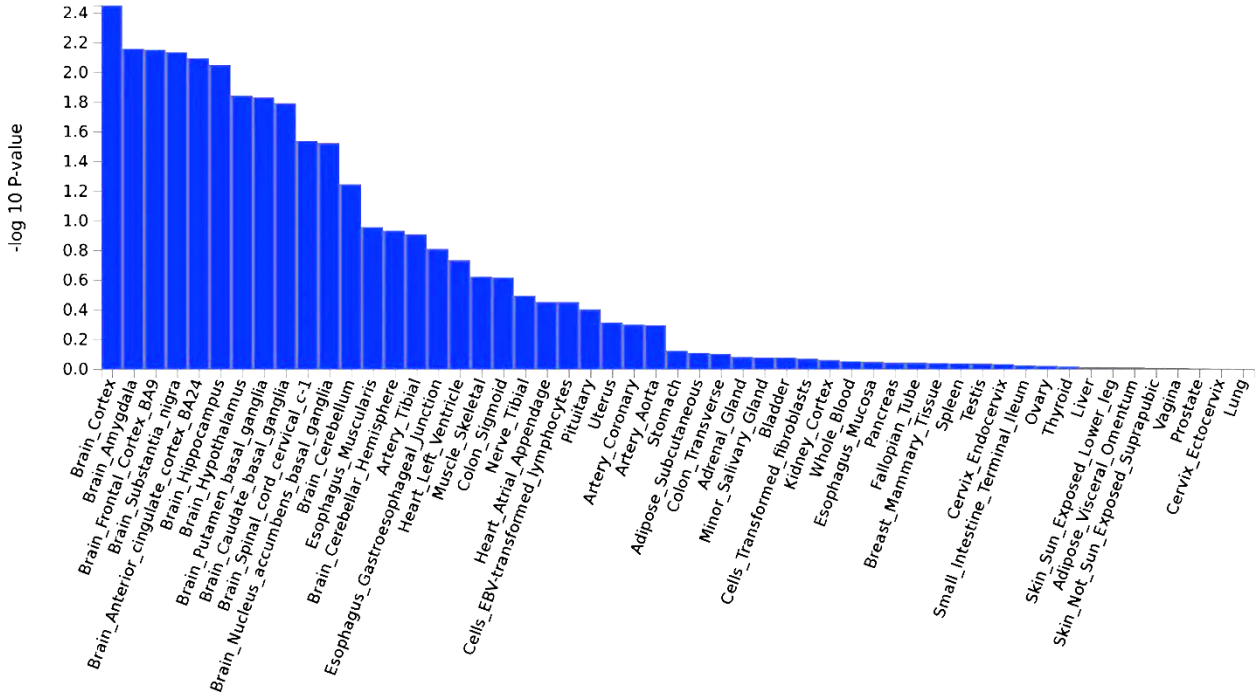

Figure 5.2e: Illustration of the results of tissue enrichment analysis implemented by MAGMA (embedded in FUMA), using GTEx data for 53 tissue types. Nominal −log10 *p* values are presented along the *y* axis. None of the investigated tissues exhibited a significant enrichment after correction for multiple testing [Ref. Andreas, Awasthi and Wolf et al., 2019, Molecular Psychiatry: Permission granted].

The following sections present findings from two additional research projects. Notably, I emphasize significant genetic insights within two complex traits – namely BPD (presented in Section 5.3) and general intelligence (presented in Section 5.4).

5.3 Borderline personality disorder (Witt et al., 2017): Genetic correlation with schizophrenia, major depressive disorder, and bipolar disorder

A total of 207 individuals were excluded due to technical and genomic QC, mostly due to genetic overlap and population outliers. The post-QC cohort consisted of 998 BPD cases (914 female/84 male) and 1,545 controls (868 female/677 male). These were finally imputed using the 1000 Genomes Project reference panel (Abecasis et al., 2010). As the most critical outcome, we were able to demonstrate significant genetic correlations of BPD with BP (rg = 0.28; SE = 0.094; $p = 2.99 \times 10^{-3}$), MDD (rg = 0.57; SE = 0.18; $p = 1.04 \times 10^{-3}$), and SCZ (rg = 0.34; SE = 0.082; $p = 4.37 \times 10^{-5}$).

5.4 General intelligence in the IMAGEN cohort (Kaminski et al., 2018): Association between gIQ and polygenic scores

PRSs were derived from a childhood intelligence GWAS (Benyamin et al., 2014) with 7,989 individuals as well as a human intelligence GWAS (Sniekers et al., 2017) with 78,308 participants. The former PRSs were significantly associated with gIQ at a *p*-value threshold of 0.1 (comprising 16,972 SNPs) and explained 0.33% of the phenotypic variance ($p = 1.7 \times 10^{-2}$). By contrast, the latter PRSs used a *p*-value threshold of 0.01 (leaving 5,636 SNPs) and increased the explained variance up to 3.2% ($p = 7.3 \times 10^{-8}$).

6) Discussion

GWASs have changed our understanding of a plethora of complex human traits in medical and nonmedical fields. Genome-wide summary statistics derived from GWASs are used for downstream analyses, which assist in understanding the biology of complex traits and the etiology of complex disorders. Thus, to support such progress, the National Institute of Health requests that such data be made publicly available (Paltoo et al., 2014). There are a growing number of resources in which one can find publicly available genome-wide data, such as the PGC (Sullivan et al., 2018), GWAS Atlas (Watanabe et al., 2019), UK BioBank (Sudlow et al., 2015), LocusZoom (Pruim et al., 2010), dbGAP (Mailman et al., 2007), and EGA (Lappalainen et al., 2015). However, conclusions drawn from these results will not be robust if the GWAS analysis was initially compromised. Even minor biases and errors can increase the rate of false-positive and -negative results (Finno et al., 2014). Therefore, it is essential to produce high-quality GWAS results to ensure a reliable clinical impact, such as through individual risk prediction or gene interaction pathways.

Our *Ricopili* pipeline and the best practice methodology are aimed at helping to produce robust GWAS results by addressing a multitude of real-world biases and errors throughout. The pipeline incorporates state-of-the-art tools and techniques used in GWASs to provide a coherent and streamlined workflow. It comprises four major modules—pre-imputation/QC, PCA, imputation, and post-imputation. Each module ensures the seamless progression of a GWAS, effectively identifying any biases that might compromise the integrity of the results and analysis. This efficient approach significantly reduces researchers' time investment by allowing them to focus on result interpretation and target discovery rather than laborious data management.

Furthermore, the *Ricopili* pipeline automates report generation and produces informative plots in each module, thus facilitating the early detection and correction of errors. For example, to ensure accurate representation of the population, it is crucial to perform LD pruning and remove long-range LD regions. Additionally, aligning genotypes to the reference build is essential, as failure to do so can lead to poorly imputed genotypes and inadequate signals. Analysts often overlook these steps, and an efficient pipeline like *Ricopili* helps to mitigate such issues.

This standardized procedure and reporting system also contribute to documentation, which is often a time-consuming task for analysts. Designed to be easily installed on high-performance clusters and adaptable to various job management systems (e.g., Slurm or qsub), the *Ricopili* pipeline enhances scalability and the efficient use of computational resources. These advantages bolster scientific rigor, promote reproducibility, and alleviate the computational burden on researchers, enabling them to devote more attention to the scientific aspects of their research.

However, the Ricopili pipeline also has some limitations. The pipeline is difficult to be installed on a normal computing machine. It usually deals with data and algorithms that are computationally expensive, requiring extensive computation resources, such as High-Performance Computing (HPC) Clusters. Initially, installing it on a new system and understanding all the algorithms, tools, and scripts could be daunting. However, comprehensive documentation of the pipeline are available online. Users should possess certain skills, such as working with a Unix operating system, to work smoothly with the pipeline.

Nonetheless, this work is a valuable learning resource and will hopefully encourage more researchers to perform genome-wide studies. The five HAPGEN-generated simulated data sets used to develop the methodology are publicly available for download, along with a detailed guiding tutorial. The simulated data will help to educate new users about *Ricopili* and GWASs as well as further motivate experienced users to develop new functionality using these data. This work has produced many protocols for performing and tuning various downstream analyses. Within this project, we tested this developed methodology on PD, BPD, and gIQ later in the project.

As expected, the GWAS of the to-date largest PD cohort did not reveal any genomic loci associated with PD due to its small sample size. However, it did reveal interesting genetic characteristics of the disorder. Furthermore, the LOO-PRS analysis significantly predicted cases and control in all subcohorts, ranging from 0.8% to 2.6% variance explained, which strongly supports the shared risk-variant consistency among these subcohorts. These results also confirm that uniform diagnostic criteria were applied to recruit the PD patients and that shared risk variants existed among the subcohorts. This phenotypic variance explained by common variants

is comparable to other complex psychiatric traits with similar sample sizes, such as SCZ (Ripke et al., 2013; Purcell et al., 2009). Additionally, we present the first estimate of SNP-based heritability (28–34%) from this cohort of 2,248 cases and 7,992 controls using the LDSR and SumHerr (LDAK) methods, which implied that common variants with small effect sizes influence a large proportion of PD susceptibility.

Furthermore, this study identified a strong positive genetic correlation of PD with MDD, depressive symptoms, and neuroticism, consistent with those frequently observed at clinics (Dold et al., 2017). Moreover, these results also supported the previously published finding of overlapping genetic risk factors between depression and anxiety (Demirkan et al., 2011). We performed an additional analysis of PD with/without MDD on 11,153 individuals (with the presence and absence of a lifetime history of MDD available) to assess the potential influence of MDD comorbidity on the reported genetic correlations. Noteworthily, this additional analysis revealed a significant positive correlation between MDD and PD without MDD. Quantitatively, the correlation between MDD and PD with MDD exhibited a stronger but statistically nonsignificant effect. These results suggested that MDD and PD's reported correlation is mostly independent of comorbidity, but they might have inflated the current estimate to a small extent.

Notably, the strongest correlation was found between PD and neuroticism, a trait that is highly correlated with many internalizing mental disorders. This is consistent with previous clinical findings of a possible relationship between PD and neuroticism (Võhma et al., 2010). This study also found nominally significant positive correlations of PD with anxiety disorder, PTSD, PGC cross-disorder, and SCZ as well as a negative correlation with years of schooling. These are expected to be replicated in larger cohorts in the future.

Moreover, the gene-based, gene set, and tissue expression analyses from MAGMA revealed no significant results after correcting multiple tests. However, the results implied that the genes tagged by variants in our present cohorts are enriched for expression in various brain-related tissues. Specifically, the strongest enrichment was observed for the brain cortex's genes, followed by the amygdala. These results support that PD's biological origin lies in the brain and

confirm the previously reported roles of the brain cortex and amygdala in the neural networks of anxiety and fear (Dresler et al., 2013; Pfleiderer et al., 2007; Kim et al., 2011).

While this study has provided insights into PD's etiology, it also had some limitations. First, the sample size could not point to specific variants significantly associated with PD. However, the positive sign test and polygenic results anticipated that PD's genetic association will become robust as the sample size grows in future studies. Second, this study cannot be generalized to a population other than that of European ancestry, but this limitation is addressed with various worldwide efforts in diverse GWAS ancestry studies.

In two related side projects, we were able to uncover valuable insights. Our BPD project revealed significant genetic overlaps with various psychiatric traits, such as BP ($rg = 0.28$) and SCZ ($rg = 0.34$) or MDD ($rg = 0.57$). BP's genetic correlation was the weakest, even though some diagnostic criteria for BPD overlap. The overlap between BP, SCZ, and MDD is consistent with previous genetic overlap observations of other psychiatric disorders (Lee et al., 2013).

In the last part of our project, we were able to replicate significant associations with gIQ with PRSs derived from childhood intelligence (Benyamin et al., 2014) and human intelligence (Sniekers et al., 2017) GWASs. These associations are consistent with studies that have reported a substantial heritable background of intellect in individuals along with environmental effects (Davies et al., 2016; Flynn 1987).

7) Conclusion

GWAS results could be misleading, even due to slight biases. In the future, only good-quality outcomes can ensure reliable clinical impacts of GWAS results. Our *Ricpoili* pipeline and best practice tutorial can address many of these biases as well as assist with the production of reliable results. This work will encourage and educate more researchers to perform clean and robust genome-wide studies. The developed methodology is expandable over a broad range of polygenic traits and will improve the overall quality of GWASs.

In this PhD project, the method was effectively employed to assess the genetic factors that influence PD, BPD, and gIQ. These three GWAS analyses lacked a well-powered sample size for single variant discoveries. However, we were first to report SNP-based heritability for PD and estimate a significant genetic correlation with depression and neuroticism. The BPD cohort analysis revealed a significant genetic overlap with BP, SCZ, and MDD. Furthermore, the findings from the PD and BPD analyses suggested that both traits are not discrete but rather have an etiological overlap with other personality and psychiatric disorders. Examinations of shared and nonshared clinical and genetic characteristics are critical for developing new and personalized treatments for PD, BPD, and other complex disorders. The final part of this thesis confirmed the polygenic characteristics of general intelligence in the IMAGEN cohort.

A deeper understanding of the intricacies of genomics and its role in complex traits and disorders unlocks new doors to understanding the human condition. Such studies not only enhance knowledge in this field but also transform the treatment of psychiatric disorders into a patient-specific, precision medicine approach. It is in these scientific findings that we find the power to improve lives and offer hope for a brighter future.

8) References

1000 Genomes Project Consortium, Abecasis, G. R., Altshuler, D., Auton, A., Brooks, L. D., Durbin, R. M., Gibbs, R. A., Hurles, M. E., & McVean, G. A. (2010). A map of human genome variation from population-scale sequencing. Nature, 467(7319), 1061–1073. https://doi.org/10.1038/nature09534

1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., & Abecasis, G. R. (2015). A global reference for human genetic variation. Nature, 526(7571), 68–74. https://doi.org/10.1038/nature15393

Afifi, T. O., Mather, A., Boman, J., Fleisher, W., Enns, M. W., Macmillan, H., & Sareen, J. (2011). Childhood adversity and personality disorders: results from a nationally representative population-based study. *Journal of psychiatric research*, *45*(6), 814–822. https://doi.org/10.1016/j.jpsychires.2010.11.008

Altmüller, J., Palmer, L. J., Fischer, G., Scherb, H., & Wjst, M. (2001). Genomewide scans of complex human diseases: true linkage is hard to find. American journal of human genetics, 69(5), 936–950. https://doi.org/10.1086/324069

Amad, A., Ramoz, N., Thomas, P., Jardri, R., & Gorwood, P. (2014). Genetics of borderline personality disorder: systematic review and proposal of an integrative model. *Neuroscience and biobehavioral reviews*, *40*, 6–19. https://doi.org/10.1016/j.neubiorev.2014.01.003

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.).

Anisman, H., Zaharia, M. D., Meaney, M. J., & Merali, Z. (1998). Do early-life events permanently alter behavioral and hormonal responses to stressors? International journal of developmental neuroscience: the official journal of the International Society for Developmental Neuroscience, 16(3-4), 149–164. https://doi.org/10.1016/s0736-5748(98)00025-2

Beeney, J. E., Hallquist, M. N., Ellison, W. D., & Levy, K. N. (2016). Self-other disturbance in borderline personality disorder: Neural, self-report, and performance-based evidence. *Personality disorders*, *7*(1), 28–39. https://doi.org/10.1037/per0000127

Benyamin, B., Pourcain, B., Davis, O. S., Davies, G., Hansell, N. K., Brion, M. J., Kirkpatrick, R. M., Cents, R. A., Franić, S., Miller, M. B., Haworth, C. M., Meaburn, E., Price, T. S., Evans, D. M., Timpson, N., Kemp, J., Ring, S., McArdle, W., Medland, S. E., Yang, J., Harris, S.E., Liewald, D.C., Scheet, P., Xiao, X., Hudziak, J.J., de Geus, E.J., Wellcome Trust Case Control Consortium 2 (WTCCC2)., Jaddoe, V.W., Starr, J.M., Verhulst, F.C., Pennell, C., Tiemeier, H., Iacono, W.G., Palmer, L.J., Montgomery, G.W., Martin, N.G., Boomsma, D.I., Posthuma, D., McGue, M.,

Wright, M.J., Davey., Smith, G., Deary, I.J., Plomin, R., Visscher, P.M., & Visscher, P. M. (2014). Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. Molecular psychiatry, 19(2), 253–258. https://doi.org/10.1038/mp.2012.18

Biskin R. S. (2015). The Lifetime Course of Borderline Personality Disorder. *Canadian journal of psychiatry. Revue canadienne de psychiatrie*, *60*(7), 303–308. https://doi.org/10.1177/070674371506000702

Bohus, M., & Schmahl, C. (2007). Psychopathologie und Therapie der Borderline-Persönlichkeitsstörung [Psychopathology and treatment of borderline personality disorder]. *Der Nervenarzt*, *78*(9), 1069–1081. https://doi.org/10.1007/s00115-007-2341-x

Brainstorm Consortium, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, Duncan L, Escott-Price V, Falcone GJ, Gormley P, Malik R, Patsopoulos NA, Ripke S, Wei Z, Yu D, Lee PH, Turley P, Grenier-Boley B, Chouraki V, Kamatani Y, Berr C, Letenneur L, Hannequin D, Amouyel P, Boland A, Deleuze JF, Duron E, Vardarajan BN, Reitz C, Goate AM, Huentelman MJ, Kamboh MI, Larson EB, Rogaeva E, St George-Hyslop P, Hakonarson H, Kukull WA, Farrer LA, Barnes LL, Beach TG, Demirci FY, Head E, Hulette CM, Jicha GA, Kauwe JSK, Kaye JA, Leverenz JB, Levey AI, Lieberman AP, Pankratz VS, Poon WW, Quinn JF, Saykin AJ, Schneider LS, Smith AG, Sonnen JA, Stern RA, Van Deerlin VM, Van Eldik LJ, Harold D, Russo G, Rubinsztein DC, Bayer A, Tsolaki M, Proitsi P, Fox NC, Hampel H, Owen MJ, Mead S, Passmore P, Morgan K, Nöthen MM, Rossor M, Lupton MK, Hoffmann P, Kornhuber J, Lawlor B, McQuillin A, Al-Chalabi A, Bis JC, Ruiz A, Boada M, Seshadri S, Beiser A, Rice K, van der Lee SJ, De Jager PL, Geschwind DH, Riemenschneider M, Riedel-Heller S, Rotter JI, Ransmayr G, Hyman BT, Cruchaga C, Alegret M, Winsvold B, Palta P, Farh KH, Cuenca-Leon E, Furlotte N, Kurth T, Ligthart L, Terwindt GM, Freilinger T, Ran C, Gordon SD, Borck G, Adams HHH, Lehtimäki T, Wedenoja J, Buring JE, Schürks M, Hrafnsdottir M, Hottenga JJ, Penninx B, Artto V, Kaunisto M, Vepsäläinen S, Martin NG, Montgomery GW, Kurki MI, Hämäläinen E, Huang H, Huang J, Sandor C, Webber C, Muller-Myhsok B, Schreiber S, Salomaa V, Loehrer E, Göbel H, Macaya A, Pozo-Rosich P, Hansen T, Werge T, Kaprio J, Metspalu A, Kubisch C, Ferrari MD, Belin AC, van den Maagdenberg AMJM, Zwart JA, Boomsma D, Eriksson N, Olesen J, Chasman DI, Nyholt DR, Avbersek A, Baum L, Berkovic S, Bradfield J, Buono RJ, Catarino CB, Cossette P, De Jonghe P, Depondt C, Dlugos D, Ferraro TN, French J, Hjalgrim H, Jamnadas-Khoda J, Kälviäinen R, Kunz WS, Lerche H, Leu C, Lindhout D, Lo W, Lowenstein D, McCormack M, Møller RS, Molloy A, Ng PW, Oliver K, Privitera M, Radtke R, Ruppert AK, Sander T, Schachter S, Schankin C, Scheffer I, Schoch S, Sisodiya SM, Smith P, Sperling M, Striano P, Surges R, Thomas GN, Visscher F, Whelan CD, Zara F, Heinzen EL, Marson A, Becker F, Stroink H, Zimprich F, Gasser

T, Gibbs R, Heutink P, Martinez M, Morris HR, Sharma M, Ryten M, Mok KY, Pulit S, Bevan S, Holliday E, Attia J, Battey T, Boncoraglio G, Thijs V, Chen WM, Mitchell B, Rothwell P, Sharma P, Sudlow C, Vicente A, Markus H, Kourkoulis C, Pera J, Raffeld M, Silliman S, Boraska Perica V, Thornton LM, Huckins LM, William Rayner N, Lewis CM, Gratacos M, Rybakowski F, Keski-Rahkonen A, Raevuori A, Hudson JI, Reichborn-Kjennerud T, Monteleone P, Karwautz A, Mannik K, Baker JH, O'Toole JK, Trace SE, Davis OSP, Helder SG, Ehrlich S, Herpertz-Dahlmann B, Danner UN, van Elburg AA, Clementi M, Forzan M, Docampo E, Lissowska J, Hauser J, Tortorella A, Maj M, Gonidakis F, Tziouvas K, Papezova H, Yilmaz Z, Wagner G, Cohen-Woods S, Herms S, Julià A, Rabionet R, Dick DM, Ripatti S, Andreassen OA, Espeseth T, Lundervold AJ, Steen VM, Pinto D, Scherer SW, Aschauer H, Schosser A, Alfredsson L, Padyukov L, Halmi KA, Mitchell J, Strober M, Bergen AW, Kaye W, Szatkiewicz JP, Cormand B, Ramos-Quiroga JA, Sánchez-Mora C, Ribasés M, Casas M, Hervas A, Arranz MJ, Haavik J, Zayats T, Johansson S, Williams N, Dempfle A, Rothenberger A, Kuntsi J, Oades RD, Banaschewski T, Franke B, Buitelaar JK, Arias Vasquez A, Doyle AE, Reif A, Lesch KP, Freitag C, Rivero O, Palmason H, Romanos M, Langley K, Rietschel M, Witt SH, Dalsgaard S, Børglum AD, Waldman I, Wilmot B, Molly N, Bau CHD, Crosbie J, Schachar R, Loo SK, McGough JJ, Grevet EH, Medland SE, Robinson E, Weiss LA, Bacchelli E, Bailey A, Bal V, Battaglia A, Betancur C, Bolton P, Cantor R, Celestino-Soper P, Dawson G, De Rubeis S, Duque F, Green A, Klauck SM, Leboyer M, Levitt P, Maestrini E, Mane S, De-Luca DM, Parr J, Regan R, Reichenberg A, Sandin S, Vorstman J, Wassink T, Wijsman E, Cook E, Santangelo S, Delorme R, Rogé B, Magalhaes T, Arking D, Schulze TG, Thompson RC, Strohmaier J, Matthews K, Melle I, Morris D, Blackwood D, McIntosh A, Bergen SE, Schalling M, Jamain S, Maaser A, Fischer SB, Reinbold CS, Fullerton JM, Guzman-Parra J, Mayoral F, Schofield PR, Cichon S, Mühleisen TW, Degenhardt F, Schumacher J, Bauer M, Mitchell PB, Gershon ES, Rice J, Potash JB, Zandi PP, Craddock N, Ferrier IN, Alda M, Rouleau GA, Turecki G, Ophoff R, Pato C, Anjorin A, Stahl E, Leber M, Czerski PM, Cruceanu C, Jones IR, Posthuma D, Andlauer TFM, Forstner AJ, Streit F, Baune BT, Air T, Sinnamon G, Wray NR, MacIntyre DJ, Porteous D, Homuth G, Rivera M, Grove J, Middeldorp CM, Hickie I, Pergadia M, Mehta D, Smit JH, Jansen R, de Geus E, Dunn E, Li QS, Nauck M, Schoevers RA, Beekman AT, Knowles JA, Viktorin A, Arnold P, Barr CL, Bedoya-Berrio G, Bienvenu OJ, Brentani H, Burton C, Camarena B, Cappi C, Cath D, Cavallini M, Cusi D, Darrow S, Denys D, Derks EM, Dietrich A, Fernandez T, Figee M, Freimer N, Gerber G, Grados M, Greenberg E, Hanna GL, Hartmann A, Hirschtritt ME, Hoekstra PJ, Huang A, Huyser C, Illmann C, Jenike M, Kuperman S, Leventhal B, Lochner C, Lyon GJ, Macciardi F, Madruga-Garrido M, Malaty IA, Maras A, McGrath L, Miguel EC, Mir P, Nestadt G, Nicolini H, Okun MS,

Pakstis A, Paschou P, Piacentini J, Pittenger C, Plessen K, Ramensky V, Ramos EM, Reus V, Richter MA, Riddle MA, Robertson MM, Roessner V, Rosário M, Samuels JF, Sandor P, Stein DJ, Tsetsos F, Van Nieuwerburgh F, Weatherall S, Wendland JR, Wolanczyk T, Worbe Y, Zai G, Goes FS, McLaughlin N, Nestadt PS, Grabe HJ, Depienne C, Konkashbaev A, Lanzagorta N, Valencia-Duarte A, Bramon E, Buccola N, Cahn W, Cairns M, Chong SA, Cohen D, Crespo-Facorro B, Crowley J, Davidson M, DeLisi L, Dinan T, Donohoe G, Drapeau E, Duan J, Haan L, Hougaard D, Karachanak-Yankova S, Khrunin A, Klovins J, Kučinskas V, Lee Chee Keong J, Limborska S, Loughland C, Lönnqvist J, Maher B, Mattheisen M, McDonald C, Murphy KC, Nenadic I, van Os J, Pantelis C, Pato M, Petryshen T, Quested D, Roussos P, Sanders AR, Schall U, Schwab SG, Sim K, So HC, Stögmann E, Subramaniam M, Toncheva D, Waddington J, Walters J, Weiser M, Cheng W, Cloninger R, Curtis D, Gejman PV, Henskens F, Mattingsdal M, Oh SY, Scott R, Webb B, Breen G, Churchhouse C, Bulik CM, Daly M, Dichgans M, Faraone SV, Guerreiro R, Holmans P, Kendler KS, Koeleman B, Mathews CA, Price A, Scharf J, Sklar P, Williams J, Wood NW, Cotsapas C, Palotie A, Smoller JW, Sullivan P, Rosand J, Corvin A, Neale BM, Schott JM, Anney R, Elia J, Grigoroiu-Serbanescu M, Edenberg HJ, & Murray R. (2018). Analysis of shared heritability in common disorders of the brain. Science (New York, N.Y.), 360(6395), eaap8757. https://doi.org/10.1126/science.aap8757

Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. American journal of human genetics, 103(3), 338–348. https://doi.org/10.1016/j.ajhg.2018.07.015

Bulik-Sullivan, B. K., Loh, P. R., Finucane, H. K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson, N., Daly, M. J., Price, A. L., & Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature genetics, 47(3), 291–295. https://doi.org/10.1038/ng.3211

Cackovic C, Nazir S, Marwaha R. Panic Disorder (Attack) [Updated 2020 Jun 25]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2020 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK430973/.

Calati, R., Gressier, F., Balestri, M., & Serretti, A. (2013). Genetic modulation of borderline personality disorder: systematic review and meta-analysis. *Journal of psychiatric research*, *47*(10), 1275–1287. https://doi.org/10.1016/j.jpsychires.2013.06.002

Crawford, T. N., Cohen, P. R., Chen, H., Anglin, D. M., & Ehrensaft, M. (2009). Early maternal separation and the trajectory of borderline personality disorder symptoms *Development and psychopathology*, *21*(3), 1013–1030. https://doi.org/10.1017/S0954579409000546

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

Davies, G., Marioni, R. E., Liewald, D. C., Hill, W. D., Hagenaars, S. P., Harris, S. E., Ritchie, S. J., Luciano, M., Fawns-Ritchie, C., Lyall, D., Cullen, B., Cox, S. R., Hayward, C., Porteous, D. J., Evans, J., McIntosh, A. M., Gallacher, J., Craddock, N., Pell, J. P., Smith, D. J., Gale, C.R., & Deary, I. J. (2016). Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). Molecular psychiatry, 21(6), 758–767. https://doi.org/10.1038/mp.2016.45

Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences Nature reviews. Neuroscience, 11(3), 201–211. https://doi.org/10.1038/nrn2793

de Bakker, P. I., Ferreira, M. A., Jia, X., Neale, B. M., Raychaudhuri, S., & Voight, B. F. (2008). Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Human molecular genetics, 17(R2), R122–R128. https://doi.org/10.1093/hmg/ddn288

Delaneau, O., Marchini, J., & Zagury, J. F. (2011). A linear complexity phasing method for thousands of genomes. Nature Methods, 9(2), 179-81. https://doi.org/10.1038/nmeth.1785

de Leeuw, C. A., Mooij, J. M., Heskes, T., & Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS computational biology, 11(4), e1004219. https://doi.org/10.1371/journal.pcbi.1004219

Demirkan, A., Penninx, B. W., Hek, K., Wray, N. R., Amin, N., Aulchenko, Y. S., van Dyck, R., de Geus, E. J., Hofman, A., Uitterlinden, A. G., Hottenga, J. J., Nolen, W. A., Oostra, B. A., Sullivan, P. F., Willemsen, G., Zitman, F. G., Tiemeier, H., Janssens, A. C., Boomsma, D. I., van Duijn, C. M., & Middeldorp, C. M. (2011). Genetic risk profiles for depression and anxiety in adult and elderly cohorts. Molecular psychiatry, 16(7), 773–783. https://doi.org/10.1038/mp.2010.65

Distel, M. A., Trull, T. J., Derom, C. A., Thiery, E. W., Grimmer, M. A., Martin, N. G., Willemsen, G., & Boomsma, D. I. (2008). Heritability of borderline personality disorder features is similar across three countries. *Psychological medicine*, *38*(9), 1219–1229. https://doi.org/10.1017/S0033291707002024

Distel, M. A., Trull, T. J., Willemsen, G., Vink, J. M., Derom, C. A., Lynskey, M., Martin, N. G., & Boomsma, D. I. (2009). The five-factor model of personality and borderline personality disorder: a genetic analysis of comorbidity. *Biological psychiatry*, *66*(12), 1131–1138. https://doi.org/10.1016/j.biopsych.2009.07.017

Dold, M., Bartova, L., Souery, D., Mendlewicz, J., Serretti, A., Porcelli, S., Zohar, J., Montgomery, S., & Kasper, S. (2017). Clinical characteristics and treatment outcomes of patients with major depressive disorder and comorbid anxiety disorders - results from a European multicenter study. Journal of psychiatric research, 91, 1–13. https://doi.org/10.1016/j.jpsychires.2017.02.020

Dresler, T., Guhn, A., Tupak, S. V., Ehlis, A. C., Herrmann, M. J., Fallgatter, A. J., Deckert, J., & Domschke, K. (2013). Revise the revised? New dimensions of the neuroanatomical hypothesis of panic disorder Journal of neural transmission (Vienna, Austria : 1996), 120(1), 3–29. https://doi.org/10.1007/s00702-012-0811-1

Ehlers A. (1995). A 1-year prospective study of panic attacks: clinical course and factors associated with maintenance. *Journal of abnormal psychology*, *104*(1), 164–172. https://doi.org/10.1037//0021-843x.104.1.164

Erhardt, A., Akula, N., Schumacher, J., Czamara, D., Karbalai, N., Müller-Myhsok, B., Mors, O., Borglum, A., Kristensen, A. S., Woldbye, D. P., Koefoed, P., Eriksson, E., Maron, E., Metspalu, A., Nurnberger, J., Philibert, R. A., Kennedy, J., Domschke, K., Reif, A., Deckert, J., Otowa, T., Kawamura, Y., Kaiya, H., Okazaki, Y., Tanii, H., Tokunaga, K., Sasaki, T., Ioannidis, J.P., McMahon, F.J., & Binder, E. B. (2012). Replication and meta-analysis of TMEM132D gene variants in panic disorder. Translational psychiatry, 2(9), e156. https://doi.org/10.1038/tp.2012.85.

Erhardt, A., Czibere, L., Roeske, D., Lucae, S., Unschuld, P. G., Ripke, S., Specht, M., Kohli, M. A., Kloiber, S., Ising, M., Heck, A., Pfister, H., Zimmermann, P., Lieb, R., Pütz, B., Uhr, M., Weber, P., Deussing, J. M., Gonik, M., Bunck, M., Kebler, M.S., Frank, E., Hohoff, C., Domschke, K., Krakowitzky, P., Maier, W., Bandelow, B., Jacob, C., Deckert, J., Schreiber, S., Strohmaier, J., Nöthen, M., Cichon, S., Rietschel, M., Bettecken, T., Keck, M.E., Landgraf, R., Müller-Myhsok, B., Holsboer, F., & Binder, E.B. (2011). TMEM132D, a new candidate for anxiety phenotypes: evidence from human and mouse studies. Molecular psychiatry, 16(6), 647–663. https://doi.org/10.1038/mp.2010.41.

Finno, C. J., Aleman, M., Higgins, R. J., Madigan, J. E., & Bannasch, D. L. (2014). Risk of false positive genetic associations in complex traits with underlying population structure: a case study. Veterinary journal (London, England : 1997), 202(3), 543–549. https://doi.org/10.1016/j.tvjl.2014.09.013

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. Psychological Bulletin, 101(2), 171–191. https://doi.org/10.1037/0033-2909.101.2.171

Fonagy, P., & Bateman, A. (2008). The development of borderline personality disorder--a mentalizing model. *Journal of personality disorders*, *22*(1), 4–21. https://doi.org/10.1521/pedi.2008.22.1.4

Fornaro, M., Orsolini, L., Marini, S., De Berardis, D., Perna, G., Valchera, A., Ganança, L., Solmi, M., Veronese, N., & Stubbs, B. (2016). The prevalence and predictors of bipolar and borderline

personality disorders comorbidity: Systematic review and meta-analysis. *Journal of affective disorders*, *195*, 105–118. https://doi.org/10.1016/j.jad.2016.01.040

Forstner AJ, Awasthi S, Wolf C, Maron E, Erhardt A, Czamara D, Eriksson E, Lavebratt C, Allgulander C, Friedrich N, Becker J, Hecker J, Rambau S, Conrad R, Geiser F, McMahon FJ, Moebus S, Hess T, Buerfent BC, Hoffmann P, Herms S, Heilmann-Heimbach S, Kockum I, Olsson T, Alfredsson L, Weber H, Alpers GW, Arolt V, Fehm L, Fydrich T, Gerlach AL, Hamm A, Kircher T, Pané-Farré CA, Pauli P, Rief W, Ströhle A, Plag J, Lang T, Wittchen HU, Mattheisen M, Meier S, Metspalu A, Domschke K, Reif A, Hovatta I, Lindefors N, Andersson E, Schalling M, Mbarek H, Milaneschi Y, de Geus EJC, Boomsma DI, Penninx BWJH, Thorgeirsson TE, Steinberg S, Stefansson K, Stefansson H, Müller-Myhsok B, Hansen TF, Børglum AD, Werge T, Mortensen PB, Nordentoft M, Hougaard DM, Hultman CM, Sullivan PF, Nöthen MM, Woldbye DPD, Mors O, Binder EB, Rück C, Ripke S, Deckert J, Schumacher J. Genome-wide association study of panic disorder reveals genetic overlap with neuroticism and depression. Mol Psychiatry. 2019 Nov 11. doi: 10.1038/s41380-019-0590-2. Epub ahead of print. PMID: 31712720

Francis, D. D., & Meaney, M. J. (1999). Maternal care and the development of stress responses *Current opinion in neurobiology*, *9*(1), 128–134. https://doi.org/10.1016/s0959-4388(99)80016-6

Gale, C. R., Hagenaars, S. P., Davies, G., Hill, W. D., Liewald, D. C., Cullen, B., Penninx, B. W., International Consortium for Blood Pressure GWAS, CHARGE Consortium Aging and Longevity Group, Boomsma, D. I., Pell, J., McIntosh, A. M., Smith, D. J., Deary, I. J., & Harris, S. E. (2016). Pleiotropy between neuroticism and physical and mental health: findings from 108 038 men and women in UK Biobank. *Translational psychiatry*, *6*(4), e791. http://doi.org/10.1038/tp.2016.56

Gorman, J. M., & Coplan, J. D. (1996). Comorbidity of depression and panic disorder. *The Journal of clinical psychiatry*, *57 Suppl 10*, 34–43.

Gorman, J. M., Kent, J. M., Sullivan, G. M., & Coplan, J. D. (2000). Neuroanatomical hypothesis of panic disorder, Revised *The American Journal of Psychiatry*, *157*(4), 493–505. https://doi.org/10.1176/appi.ajp.157.4.493

Gunderson, J. G., Herpertz, S. C., Skodol, A. E., Torgersen, S., & Zanarini, M. C. (2018). Borderline personality disorder. *Nature reviews. Disease primers*, *4*, 18029. https://doi.org/10.1038/nrdp.2018.29

Hayward, C., Killen, J. D., Kraemer, H. C., & Taylor, C. B. (2000). Predictors of panic attacks in adolescents *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*(2), 207–214. https://doi.org/10.1097/00004583-200002000-00021

Helgeland, M. I., Kjelsberg, E., & Torgersen, S. (2005). Continuities between emotional and disruptive behavior disorders in adolescence and personality disorders in adulthood *The American Journal of Psychiatry*, *162*(10), 1941–1947. https://doi.org/10.1176/appi.ajp.162.10.1941

Hendriks, S. M., Spijker, J., Licht, C. M., Beekman, A. T., Hardeveld, F., de Graaf, R., Batelaan, N. M., & Penninx, B. W. (2014). Disability in anxiety disorders. Journal of affective disorders, 166, 227–233. https://doi.org/10.1016/j.jad.2014.05.006.

Hettema, J. M., Neale, M. C., & Kendler, K. S. (2001). A review and meta-analysis of the genetic epidemiology of anxiety disorders The American Journal of Psychiatry, 158(10), 1568–1578. https://doi.org/10.1176/appi.ajp.158.10.1568.

Hinrichs, A. S., Karolchik, D., Baertsch, R., Barber, G. P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T. S., Harte, R. A., Hsu, F., Hillman-Jackson, J., Kuhn, R. M., Pedersen, J. S., Pohl, A., Raney, B. J., Rosenbloom, K. R., Siepel, A., Smith, K. E., Sugnet, C. W., Sultan-Qurraie, A., Thomas, D.J., Trumbower, H., Weber, R.J., Weirauch, M., Zweig, A.S., Haussler, D., & Kent, W.J. (2006). The UCSC Genome Browser Database: update 2006. Nucleic acids research, 34(Database issue), D590–D598. https://doi.org/10.1093/nar/gkj144

Hirschhorn, J. N., & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. Nature reviews. Genetics, 6(2), 95–108. https://doi.org/10.1038/nrg1521

Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., & Abecasis, G. R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing Nature Genetics, 44(8), 955–959.

Howie, B. N., Donnelly, P., & Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies PLoS Genetics, 5(6), e1000529. https://doi.org/10.1371/journal.pgen.1000529

International HapMap Consortium (2003). The International HapMap Project Nature, 426(6968), 789–796. https://doi.org/10.1038/nature02168

International HapMap Consortium, Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., Pasternak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S,

Sallée C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, & Stewart, J. (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature, 449(7164), 851–861. https://doi.org/10.1038/nature06258

Kaminski JA, Schlagenhauf F, Rapp M, **Awasthi S,** Ruggeri B, Deserno L, Banaschewski T, Bokde ALW, Bromberg U, Büchel C, Quinlan EB, Desrivières S, Flor H, Frouin V, Garavan H, Gowland P, Ittermann B, Martinot JL, Martinot MP, Nees F, Orfanos DP, Paus T, Poustka L, Smolka MN, Fröhner JH, Walter H, Whelan R, Ripke S, Schumann G, Heinz A; IMAGEN consortium. Epigenetic variance in dopamine D2 receptor: a marker of IQ malleability? Transl Psychiatry. 2018 Aug 30;8(1):169. doi: 10.1038/s41398-018-0222-7. PMID: 30166545; PMCID: PMC6117339

Kerem, B., Rommens, J. M., Buchanan, J. A., Markiewicz, D., Cox, T. K., Chakravarti, A., Buchwald, M., & Tsui, L. C. (1989). Identification of the cystic fibrosis gene: genetic analysis. Science (New York, N.Y.), 245(4922), 1073–1080. https://doi.org/10.1126/science.2570460

Kessler, R. C., Petukhova, M., Sampson, N. A., Zaslavsky, A. M., & Wittchen, H. (2012). Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States

International Journal of Methods in Psychiatric Research, 21(3), 169–184. https://doi.org/10.1002/mpr.1359.

Kim, M. J., Loucks, R. A., Palmer, A. L., Brown, A. C., Solomon, K. M., Marchante, A. N., & Whalen, P. J. (2011). The structural and functional connectivity of the amygdala: from normal emotion to pathological anxiety Behavioural brain research, 223(2), 403–410. https://doi.org/10.1016/j.bbr.2011.04.025

Kingdon, D. G., Ashcroft, K., Bhandari, B., Gleeson, S., Warikoo, N., Symons, M., Taylor, L., Lucas, E., Mahendra, R., Ghosh, S., Mason, A., Badrakalimuthu, R., Hepworth, C., Read, J., & Mehta, R. (2010). Schizophrenia and borderline personality disorder: similarities and differences in the experience of auditory hallucinations, paranoia, and childhood trauma. *The Journal of nervous and mental disease*, *198*(6), 399–403. https://doi.org/10.1097/NMD.0b013e3181e08c27

Krause-Utz, A., Elzinga, B. M., Oei, N. Y., Paret, C., Niedtfeld, I., Spinhoven, P., Bohus, M., & Schmahl, C. (2014). Amygdala and Dorsal Anterior Cingulate Connectivity during an Emotional Working Memory Task in Borderline Personality Disorder Patients with Interpersonal Trauma History. *Frontiers in human neuroscience*, *8*, 848. https://doi.org/10.3389/fnhum.2014.00848

Lam M, Awasthi S, Watson HJ, Goldstein J, Panagiotaropoulou G, Trubetskoy V, Karlsson R, Frei O, Fan CC, De Witte W, Mota NR, Mullins N, Brügger K, Lee SH, Wray NR, Skarabis N, Huang H, Neale B, Daly MJ, Mattheisen M, Walters R, Ripke S. RICOPILI: Rapid Imputation for COnsortias PIpeLIne. Bioinformatics. 2020 Feb 1;36(3):930-933. doi: 10.1093/bioinformatics/btz633. PMID: 31393554; PMCID: PMC7868045

Lappalainen, I., Almeida-King, J., Kumanduri, V., Senf, A., Spalding, J. D., Ur-Rehman, S., Saunders, G., Kandasamy, J., Caccamo, M., Leinonen, R., Vaughan, B., Laurent, T., Rowland, F., Marin-Garcia, P., Barker, J., Jokinen, P., Torres, A. C., de Argila, J. R., Llobet, O. M., Medina, I., Puy, M.S., Alberich, M., de la Torre, S., Navarro, A., Paschall, J., & Flicek, P. (2015). The European Genome-phenome Archive of human data consented for biomedical research. *Nature genetics*, *47*(7), 692–695. https://doi.org/10.1038/ng.3312

Lee, S. H., Ripke, S., Neale, B. M., Faraone, S. V., Purcell, S. M., Perlis, R. H., Mowry, B. J., Thapar, A., Goddard, M. E., Witte, J. S., Absher, D., Agartz, I., Akil, H., Amin, F., Andreassen, O. A., Anjorin, A., Anney, R., Anttila, V., Arking, D. E., Asherson P, Azevedo MH, Backlund L, Badner JA, Bailey AJ, Banaschewski T, Barchas JD, Barnes MR, Barrett TB, Bass N, Battaglia A, Bauer M, Bayés M, Bellivier F, Bergen SE, Berrettini W, Betancur C, Bettecken T, Biederman J, Binder EB, Black DW, Blackwood DH, Bloss CS, Boehnke M, Boomsma DI, Breen G, Breuer R, Bruggeman R, Cormican P, Buccola NG, Buitelaar JK, Bunney WE, Buxbaum JD, Byerley WF, Byrne EM, Caesar S, Cahn W, Cantor RM, Casas M, Chakravarti A, Chambert K, Choudhury K, Cichon S,

Cloninger CR, Collier DA, Cook EH, Coon H, Cormand B, Corvin A, Coryell WH, Craig DW, Craig IW, Crosbie J, Cuccaro ML, Curtis D, Czamara D, Datta S, Dawson G, Day R, De Geus EJ, Degenhardt F, Djurovic S, Donohoe GJ, Doyle AE, Duan J, Dudbridge F, Duketis E, Ebstein RP, Edenberg HJ, Elia J, Ennis S, Etain B, Fanous A, Farmer AE, Ferrier IN, Flickinger M, Fombonne E, Foroud T, Frank J, Franke B, Fraser C, Freedman R, Freimer NB, Freitag CM, Friedl M, Frisén L, Gallagher L, Gejman PV, Georgieva L, Gershon ES, Geschwind DH, Giegling I, Gill M, Gordon SD, Gordon-Smith K, Green EK, Greenwood TA, Grice DE, Gross M, Grozeva D, Guan W, Gurling H, De Haan L, Haines JL, Hakonarson H, Hallmayer J, Hamilton SP, Hamshere ML, Hansen TF, Hartmann AM, Hautzinger M, Heath AC, Henders AK, Herms S, Hickie IB, Hipolito M, Hoefels S, Holmans PA, Holsboer F, Hoogendijk WJ, Hottenga JJ, Hultman CM, Hus V, Ingason A, Ising M, Jamain S, Jones EG, Jones I, Jones L, Tzeng JY, Kähler AK, Kahn RS, Kandaswamy R, Keller MC, Kennedy JL, Kenny E, Kent L, Kim Y, Kirov GK, Klauck SM, Klei L, Knowles JA, Kohli MA, Koller DL, Konte B, Korszun A, Krabbendam L, Krasucki R, Kuntsi J, Kwan P, Landén M, Långström N, Lathrop M, Lawrence J, Lawson WB, Leboyer M, Ledbetter DH, Lee PH, Lencz T, Lesch KP, Levinson DF, Lewis CM, Li J, Lichtenstein P, Lieberman JA, Lin DY, Linszen DH, Liu C, Lohoff FW, Loo SK, Lord C, Lowe JK, Lucae S, MacIntyre DJ, Madden PA, Maestrini E, Magnusson PK, Mahon PB, Maier W, Malhotra AK, Mane SM, Martin CL, Martin NG, Mattheisen M, Matthews K, Mattingsdal M, McCarroll SA, McGhee KA, McGough JJ, McGrath PJ, McGuffin P, McInnis MG, McIntosh A, McKinney R, McLean AW, McMahon FJ, McMahon WM, McQuillin A, Medeiros H, Medland SE, Meier S, Melle I, Meng F, Meyer J, Middeldorp CM, Middleton L, Milanova V, Miranda A, Monaco AP, Montgomery GW, Moran JL, Moreno-De-Luca D, Morken G, Morris DW, Morrow EM, Moskvina V, Muglia P, Mühleisen TW, Muir WJ, Müller-Myhsok B, Murtha M, Myers RM, Myin-Germeys I, Neale MC, Nelson SF, Nievergelt CM, Nikolov I, Nimgaonkar V, Nolen WA, Nöthen MM, Nurnberger JI, Nwulia EA, Nyholt DR, O'Dushlaine C, Oades RD, Olincy A, Oliveira G, Olsen L, Ophoff RA, Osby U, Owen MJ, Palotie A, Parr JR, Paterson AD, Pato CN, Pato MT, Penninx BW, Pergadia ML, Pericak-Vance MA, Pickard BS, Pimm J, Piven J, Posthuma D, Potash JB, Poustka F, Propping P, Puri V, Quested DJ, Quinn EM, Ramos-Quiroga JA, Rasmussen HB, Raychaudhuri S, Rehnström K, Reif A, Ribasés M, Rice JP, Rietschel M, Roeder K, Roeyers H, Rossin L, Rothenberger A, Rouleau G, Ruderfer D, Rujescu D, Sanders AR, Sanders SJ, Santangelo SL, Sergeant JA, Schachar R, Schalling M, Schatzberg AF, Scheftner WA, Schellenberg GD, Scherer SW, Schork NJ, Schulze TG, Schumacher J, Schwarz M, Scolnick E, Scott LJ, Shi J, Shilling PD, Shyn SI, Silverman JM, Slager SL, Smalley SL, Smit JH, Smith EN, Sonuga-Barke EJ, St Clair D, State M, Steffens M, Steinhausen HC, Strauss JS, Strohmaier J, Stroup TS, Sutcliffe JS, Szatmari

P, Szelinger S, Thirumalai S, Thompson RC, Todorov AA, Tozzi F, Treutlein J, Uhr M, van den Oord EJ, Van Grootheest G, Van Os J, Vicente AM, Vieland VJ, Vincent JB, Visscher PM, Walsh CA, Wassink TH, Watson SJ, Weissman MM, Werge T, Wienker TF, Wijsman EM, Willemsen G, Williams N, Willsey AJ, Witt SH, Xu W, Young AH, Yu TW, Zammit S, Zandi PP, Zhang P, Zitman FG, Zöllner S, Devlin B, Kelsoe JR, Sklar P, Daly MJ, O'Donovan MC, Craddock N, Sullivan PF, Smoller JW, Kendler KS, & Wray NR International Inflammatory Bowel Disease Genetics Consortium (IIBDGC) (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nature genetics, 45(9), 984–994. https://doi.org/10.1038/ng.2711

Lijster, J. M., Dierckx, B., Utens, E. M., Verhulst, F. C., Zieldorff, C., Dieleman, G. C., & Legerstee, J. S. (2017). The Age of Onset of Anxiety Disorders. *Canadian journal of psychiatry. Revue canadienne de psychiatrie*, *62*(4), 237–246. https://doi.org/10.1177/0706743716640757

Li, Y., Willer, C. J., Ding, J., Scheet, P., & Abecasis, G. R. (2010). MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genetic epidemiology, 34(8), 816–834. https://doi.org/10.1002/gepi.20533

Loh, P.-R., Palamara, P. F., & Price, A. L. (2016). Fast and accurate long-range phasing in a UK Biobank cohort. Nature Genetics, 48(7), 811–816. https://doi.org/10.1038/ng.3571

Lubke, G. H., Laurin, C., Amin, N., Hottenga, J. J., Willemsen, G., van Grootheest, G., Abdellaoui, A., Karssen, L. C., Oostra, B. A., van Duijn, C. M., Penninx, B. W., & Boomsma, D. I. (2014). Genome-wide analyses of borderline personality features. *Molecular psychiatry*, *19*(8), 923–929. https://doi.org/10.1038/mp.2013.109

Mailman, M. D., Feolo, M., Jin, Y., Kimura, M., Tryka, K., Bagoutdinov, R., Hao, L., Kiang, A., Paschall, J., Phan, L., Popova, N., Pretel, S., Ziyabari, L., Lee, M., Shao, Y., Wang, Z. Y., Sirotkin, K., Ward, M., Kholodov, M., Zbicz, K., Beck, J., Kimelman, M., Shevelev, S., Preuss, D., Yaschenko, E., Graeff, A., Ostell, J., & Sherry, S. T. (2007). The NCBI dbGaP database of genotypes and phenotypes. *Nature genetics*, *39*(10), 1181–1186. https://doi.org/10.1038/ng1007-1181

Maller, R. G., & Reiss, S. (1992). Anxiety sensitivity in 1984 and panic attacks in 1987. *Journal of Anxiety Disorders, 6*(3), 241–247. https://doi.org/10.1016/0887-6185(92)90036-7

Mascarell Maričić, L., Walter, H., Rosenthal, A., Ripke, S., Quinlan, E. B., Banaschewski, T., Barker, G. J., Bokde, A. L. W., Bromberg, U., Büchel, C., Desrivières, S., Flor, H., Frouin, V., Garavan, H., Itterman, B., Martinot, J. L., Martinot, M. P., Nees, F., Orfanos, D. P., Paus, T., … IMAGEN consortium (2020). The IMAGEN study: a decade of imaging genetics in adolescents. *Molecular psychiatry*, *25*(11), 2648–2671. https://doi.org/10.1038/s41380-020-0822-5

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., Kang, H. M., Fuchsberger, C., Danecek, P., Sharp, K., Luo, Y., Sidore, C., Kwong, A., Timpson, N., Koskinen, S., Vrieze, S., Scott, L. J., Zhang, H., Mahajan, A., Veldink, J., Peters U, Pato C, van Duijn CM, Gillies CE, Gandin I, Mezzavilla M, Gilly A, Cocca M, Traglia M, Angius A, Barrett JC, Boomsma D, Branham K, Breen G, Brummett CM, Busonero F, Campbell H, Chan A, Chen S, Chew E, Collins FS, Corbin LJ, Smith GD, Dedoussis G, Dorr M, Farmaki AE, Ferrucci L, Forer L, Fraser RM, Gabriel S, Levy S, Groop L, Harrison T, Hattersley A, Holmen OL, Hveem K, Kretzler M, Lee JC, McGue M, Meitinger T, Melzer D, Min JL, Mohlke KL, Vincent JB, Nauck M, Nickerson D, Palotie A, Pato M, Pirastu N, McInnis M, Richards JB, Sala C, Salomaa V, Schlessinger D, Schoenherr S, Slagboom PE, Small K, Spector T, Stambolian D, Tuke M, Tuomilehto J, Van den Berg LH, Van Rheenen W, Volker U, Wijmenga C, Toniolo D, Zeggini E, Gasparini P, Sampson MG, Wilson JF, Frayling T, de Bakker PI, Swertz MA, McCarroll S, Kooperberg C, Dekker A, Altshuler D, Willer C, Iacono W, Ripatti S, Soranzo N, Walter K, Swaroop A, Cucca F, Anderson CA, Myers RM, Boehnke M, McCarthy MI & Durbin R. Haplotype Reference Consortium (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nature genetics, 48(10), 1279–1283. https://doi.org/10.1038/ng.3643

McGlashan, T. H., Grilo, C. M., Skodol, A. E., Gunderson, J. G., Shea, M. T., Morey, L. C., Zanarini, M. C., & Stout, R. L. (2000). The Collaborative Longitudinal Personality Disorders Study: baseline Axis I/II and II/II diagnostic co-occurrence. *Acta psychiatrica Scandinavica*, *102*(4), 256–264. https://doi.org/10.1034/j.1600-0447.2000.102004256.x

Na, H. R., Kang, E. H., Lee, J. H., & Yu, B. H. (2011). The genetic basis of panic disorder Journal of Korean Medical Science, 26(6), 701–710. https://doi.org/10.3346/jkms.2011.26.6.701.

Otowa, T., Hek, K., Lee, M., Byrne, E. M., Mirza, S. S., Nivard, M. G., Bigdeli, T., Aggen, S. H., Adkins, D., Wolen, A., Fanous, A., Keller, M. C., Castelao, E., Kutalik, Z., Van der Auwera, S., Homuth, G., Nauck, M., Teumer, A., Milaneschi, Y., Hottenga, J. J., Direk, N., Hofman, A., Uitterlinden, A., Mulder, C.L., Henders, A.K., Medland, S.E., Gordon, S., Heath, A.C., Madden, P.A., Pergadia, M.L., van der Most, P.J., Nolte, I.M., van Oort, F.V., Hartman, C.A., Oldehinkel, A.J., Preisig, M., Grabe, H.J., Middeldorp, C.M., Penninx, B.W., Boomsma, D., Martin, N.G., Montgomery, G., Maher, B.S., van den Oord, E.J., Wray, N.R., Tiemeier, H., & Hettema, J. M. (2016). Meta-analysis of genome-wide association studies of anxiety disorders. Molecular psychiatry, 21(10), 1391–1399. https://doi.org/10.1038/mp.2015.197.

Otowa, T., Kawamura, Y., Nishida, N., Sugaya, N., Koike, A., Yoshida, E., Inoue, K., Yasuda, S., Nishimura, Y., Liu, X., Konishi, Y., Nishimura, F., Shimada, T., Kuwabara, H., Tochigi, M., Kakiuchi, C., Umekage, T., Miyagawa, T., Miyashita, A., Shimizu, E., Akiyoshi, J., Someya, T.,

Kato, T., Yoshikawa, T., Kuwano, R., Kasai, K., Kato, N., Kaiya, H., Tokunaga, K., Okazaki, Y., Tanii, H., & Sasaki, T. (2012). Meta-analysis of genome-wide association studies for panic disorder in the Japanese population. Translational psychiatry, 2(11), e186. https://doi.org/10.1038/tp.2012.89

Ozkan, M., & Altindag, A. (2005). Comorbid personality disorders in subjects with panic disorder: do personality disorders increase clinical severity?. *Comprehensive psychiatry*, *46*(1), 20–26. https://doi.org/10.1016/j.comppsych.2004.07.015

Paltoo, D. N., Rodriguez, L. L., Feolo, M., Gillanders, E., Ramos, E. M., Rutter, J. L., Sherry, S., Wang, V. O., Bailey, A., Baker, R., Caulder, M., Harris, E. L., Langlais, K., Leeds, H., Luetkemeier, E., Paine, T., Roomian, T., Tryka, K., Patterson, A., Green, E. D., & National Institutes of Health Genomic Data Sharing Governance Committees (2014). Data use under the NIH GWAS data sharing policy and future directions. *Nature genetics*, *46*(9), 934–938. https://doi.org/10.1038/ng.3062

Pe'er, I., Yelensky, R., Altshuler, D., & Daly, M. J. (2008). Estimation of the multiple testing burden for genomewide association studies of nearly all common variants Genetic epidemiology, 32(4), 381–385. https://doi.org/10.1002/gepi.20303

Pfleiderer, B., Zinkirciran, S., Arolt, V., Heindel, W., Deckert, J., & Domschke, K. (2007). fMRI amygdala activation during a spontaneous panic attack in a patient with panic disorder The World Journal of Biological Psychiatry: The Official Journal of the World Federation of Societies of Biological Psychiatry, 8(4), 269–272. https://doi.org/10.1080/15622970701216673

Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., Boehnke, M., Abecasis, G. R., & Willer, C. J. (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics (Oxford, England)*, *26*(18), 2336–2337. https://doi.org/10.1093/bioinformatics/btq419

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., Bakker, P.D., Daly, M., & Sham, P. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics, 81 3, 559-75. DOI: https://doi.org/10.1086/519795

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., & Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature, 460(7256), 748–752. https://doi.org/10.1038/nature08185

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

Reiss, S. (1980). Pavlovian conditioning and human fear: An expectancy model *Behavior Therapy, 11*(3), 380–396. https://doi.org/10.1016/S0005-7894(80)80054-2

Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J. L., Kähler, A. K., Akterin, S., Bergen SE, Collins AL, Crowley JJ, Fromer M, Kim Y, Lee SH, Magnusson PK, Sanchez N, Stahl EA, Williams S, Wray NR, Xia K, Bettella F, Borglum AD, Bulik-Sullivan BK, Cormican P, Craddock N, de Leeuw C, Durmishi N, Gill M, Golimbet V, Hamshere ML, Holmans P, Hougaard DM, Kendler KS, Lin K, Morris DW, Mors O, Mortensen PB, Neale BM, O'Neill FA, Owen MJ, Milovancevic MP, Posthuma D, Powell J, Richards AL, Riley BP, Ruderfer D, Rujescu D, Sigurdsson E, Silagadze T, Smit AB, Stefansson H, Steinberg S, Suvisaari J, Tosato S, Verhage M, Walters JT; Multicenter Genetic Studies of Schizophrenia Consortium, Levinson DF, Gejman PV, Kendler KS, Laurent C, Mowry BJ, O'Donovan MC, Owen MJ, Pulver AE, Riley BP, Schwab SG, Wildenauer DB, Dudbridge F, Holmans P, Shi J, Albus M, Alexander M, Campion D, Cohen D, Dikeos D, Duan J, Eichhammer P, Godard S, Hansen M, Lerer FB, Liang KY, Maier W, Mallet J, Nertney DA, Nestadt G, Norton N, O'Neill FA, Papadimitriou GN, Ribble R, Sanders AR, Silverman JM, Walsh D, Williams NM, Wormley B; Psychosis Endophenotypes International Consortium, Arranz MJ, Bakker S, Bender S, Bramon E, Collier D, Crespo-Facorro B, Hall J, Iyegbe C, Jablensky A, Kahn RS, Kalaydjieva L, Lawrie S, Lewis CM, Lin K, Linszen DH, Mata I, McIntosh A, Murray RM, Ophoff RA, Powell J, Rujescu D, Van Os J, Walshe M, Weisbrod M, Wiersma D; Wellcome Trust Case Control Consortium 2, Donnelly P, Barroso I, Blackwell JM, Bramon E, Brown MA, Casas JP, Corvin AP, Deloukas P, Duncanson A, Jankowski J, Markus HS, Mathew CG, Palmer CN, Plomin R, Rautanen A, Sawcer SJ, Trembath RC, Viswanathan AC, Wood NW, Spencer CC, Band G, Bellenguez C, Freeman C, Hellenthal G, Giannoulatou E, Pirinen M, Pearson RD, Strange A, Su Z, Vukcevic D, Donnelly P, Langford C, Hunt SE, Edkins S, Gwilliam R, Blackburn H, Bumpstead SJ, Dronov S, Gillman M, Gray E, Hammond N, Jayakumar A, McCann OT, Liddle J, Potter SC, Ravindrarajah R, Ricketts M, Tashakkori-Ghanbaria A, Waller MJ, Weston P, Widaa S, Whittaker P, Barroso I, Deloukas P, Mathew CG, Blackwell JM, Brown MA, Corvin AP, McCarthy MI, Spencer CC, Bramon E, Corvin AP, O'Donovan MC, Stefansson K, Scolnick E, Purcell S, McCarroll SA, Sklar P, Hultman CM, & Sullivan PF.. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nature Genetics, 45(10), 1150-1159. https://doi.org/10.1038/ng.2742

Ripke, S., Wray, N. R., Lewis, C. M., Hamilton, S. P., Weissman, M. M., Breen, G., Byrne, E. M., Blackwood, D. H., Boomsma, D. I., Cichon, S., Heath, A. C., Holsboer, F., Lucae, S., Madden, P. A., Martin, N. G., McGuffin, P., Muglia, P., Noethen, M. M., Penninx, B. P., Pergadia ML, Potash JB, Rietschel M, Lin D, Müller-Myhsok B, Shi J, Steinberg S, Grabe HJ, Lichtenstein P,

Magnusson P, Perlis RH, Preisig M, Smoller JW, Stefansson K, Uher R, Kutalik Z, Tansey KE, Teumer A, Viktorin A, Barnes MR, Bettecken T, Binder EB, Breuer R, Castro VM, Churchill SE, Coryell WH, Craddock N, Craig IW, Czamara D, De Geus EJ, Degenhardt F, Farmer AE, Fava M, Frank J, Gainer VS, Gallagher PJ, Gordon SD, Goryachev S, Gross M, Guipponi M, Henders AK, Herms S, Hickie IB, Hoefels S, Hoogendijk W, Hottenga JJ, Iosifescu DV, Ising M, Jones I, Jones L, Jung-Ying T, Knowles JA, Kohane IS, Kohli MA, Korszun A, Landen M, Lawson WB, Lewis G, Macintyre D, Maier W, Mattheisen M, McGrath PJ, McIntosh A, McLean A, Middeldorp CM, Middleton L, Montgomery GM, Murphy SN, Nauck M, Nolen WA, Nyholt DR, O'Donovan M, Oskarsson H, Pedersen N, Scheftner WA, Schulz A, Schulze TG, Shyn SI, Sigurdsson E, Slager SL, Smit JH, Stefansson H, Steffens M, Thorgeirsson T, Tozzi F, Treutlein J, Uhr M, van den Oord EJ, Van Grootheest G, Völzke H, Weilburg JB, Willemsen G, Zitman FG, Neale B, Daly M, Levinson DF, & Sullivan, P. F. (2013). A mega-analysis of genome-wide association studies for major depressive disorder. Molecular psychiatry, 18(4), 497–511. https://doi.org/10.1038/mp.2012.21

Rogosch, F. A., & Cicchetti, D. (2004). Child maltreatment and emergent personality organization: perspectives from the five-factor model *Journal of abnormal child psychology*, *32*(2), 123–145. https://doi.org/10.1023/b:jacp.0000019766.47625.40

Roy-Byrne, P. P., Craske, M. G., & Stein, M. B. (2006). Panic disorder. *Lancet (London, England)*, *368*(9540), 1023–1032. https://doi.org/10.1016/S0140-6736(06)69418-X

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C. A., Nagel, M., Awasthi, S., Barr, P. B., Coleman, J., Grasby, K. L., Hammerschlag, A. R., Kaminski, J. A., Karlsson, R., Krapohl, E., Lam, M., Nygaard, M., Reynolds, C. A., Trampush, J. W., Young, H., Zabaneh D, Hägg S, Hansell NK, Karlsson IK, Linnarsson S, Montgomery GW, Muñoz-Manchado AB, Quinlan EB, Schumann G, Skene NG, Webb BT, White T, Arking DE, Avramopoulos D, Bilder RM, Bitsios P, Burdick KE, Cannon TD, Chiba-Falek O, Christoforou A, Cirulli ET, Congdon E, Corvin A, Davies G, Deary IJ, DeRosse P, Dickinson D, Djurovic S, Donohoe G, Conley ED, Eriksson JG, Espeseth T, Freimer NA, Giakoumaki S, Giegling I, Gill M, Glahn DC, Hariri AR, Hatzimanolis A, Keller MC, Knowles E, Koltai D, Konte B, Lahti J, Le Hellard S, Lencz T, Liewald DC, London E, Lundervold AJ, Malhotra AK, Melle I, Morris D, Need AC, Ollier W, Palotie A, Payton A, Pendleton N, Poldrack RA, Räikkönen K, Reinvang I, Roussos P, Rujescu D, Sabb FW, Scult MA, Smeland OB, Smyrnis N, Starr JM, Steen VM, Stefanis NC, Straub RE, Sundet K, Tiemeier H, Voineskos AN, Weinberger DR, Widen E, Yu J, Abecasis G, Andreassen OA, Breen G, Christiansen L, Debrabant B, Dick DM, Heinz A, Hjerling-Leffler J, Ikram MA, Kendler KS, Martin NG, Medland SE, Pedersen NL, Plomin R, Polderman TJC, Ripke S, van der

Sluis S, Sullivan PF, Vrieze SI, Wright MJ, & Posthuma, D. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nature genetics, 50(7), 912–919. https://doi.org/10.1038/s41588-018-0152-6

Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014 Jul 24;511(7510):421-7. doi: 10.1038/nature13595. Epub 2014 Jul 22. PMID: 25056061; PMCID: PMC4112379

Schmermund, A., Moehlenkamp, S., Stang, A., Grönemeyer, D., Seibel, R., Hirche, H., Mann, K., Siffert, W., Lauterbach, K., Siegrist, J., Jöckel, K., & Erbel, R. (2002). Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL Study Risk Factors, Evaluation of Coronary calcium, and lifestyle American Heart Journal, 144 2, 212-8. https://doi.org/10.1067/mhj.2002.123579

Schumann, G., Loth, E., Banaschewski, T., Barbot, A., Barker, G., Büchel, C., Conrod, P. J., Dalley, J. W., Flor, H., Gallinat, J., Garavan, H., Heinz, A., Itterman, B., Lathrop, M., Mallik, C., Mann, K., Martinot, J. L., Paus, T., Poline, J. B., Robbins, T. W., Rietschel, M., Reed, L., Smolka, M., Spanagel, R., Speiser, C., Stephens, D.N., Ströhle, A., Struve, M., & IMAGEN consortium (2010). The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. Molecular psychiatry, 15(12), 1128–1139. https://doi.org/10.1038/mp.2010.4

Schumacher, J., & Deckert, J. (2010). Serotonin transporter polymorphisms and panic disorder. Genome medicine, 2(6), 40. https://doi.org/10.1186/gm161.

Shah, R., & Zanarini, M. C. (2018). Comorbidity of Borderline Personality Disorder: Current Status and Future Directions. *The Psychiatric clinics of North America*, *41*(4), 583–593. https://doi.org/10.1016/j.psc.2018.07.009

Slotema, C. W., Blom, J. D., Niemantsverdriet, M. B. A., Deen, M., & Sommer, I. E. C. (2018). Comorbid Diagnosis of Psychotic Disorders in Borderline Personality Disorder: Prevalence and Influence on Outcome. *Frontiers in psychiatry*, *9*, 84. https://doi.org/10.3389/fpsyt.2018.00084

Sniekers, S., Stringer, S., Watanabe, K., Jansen, P. R., Coleman, J., Krapohl, E., Taskesen, E., Hammerschlag, A. R., Okbay, A., Zabaneh, D., Amin, N., Breen, G., Cesarini, D., Chabris, C. F., Iacono, W. G., Ikram, M. A., Johannesson, M., Koellinger, P., Lee, J. J., Magnusson, P., McGue, M., Miller, M.B., Ollier, W.E.R., Payton. A., Pendleton, N., Plomin, R., Rietveld, C.A., Tiemeier, H., van Duijn, C. M., Posthuma, D., & Posthuma, D. (2017). Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. Nature genetics, 49(7), 1107–1112. https://doi.org/10.1038/ng.3869

Stahl, E. A., Breen, G., Forstner, A. J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J., Gaspar, H. A., de Leeuw, C. A., Steinberg, S., Pavlides, J., Trzaskowski, M., Byrne, E. M., Pers, T. H., Holmans, P. A., Richards, A. L., Abbott, L., Agerbo, E., Akil H, Albani D, Alliey-Rodriguez N, Als TD, Anjorin A, Antilla V, Awasthi S, Badner JA, Bækvad-Hansen M, Barchas JD, Bass N, Bauer M, Belliveau R, Bergen SE, Pedersen CB, Bøen E, Boks MP, Boocock J, Budde M, Bunney W, Burmeister M, Bybjerg-Grauholm J, Byerley W, Casas M, Cerrato F, Cervantes P, Chambert K, Charney AW, Chen D, Churchhouse C, Clarke TK, Coryell W, Craig DW, Cruceanu C, Curtis D, Czerski PM, Dale AM, de Jong S, Degenhardt F, Del-Favero J, DePaulo JR, Djurovic S, Dobbyn AL, Dumont A, Elvsåshagen T, Escott-Price V, Fan CC, Fischer SB, Flickinger M, Foroud TM, Forty L, Frank J, Fraser C, Freimer NB, Frisén L, Gade K, Gage D, Garnham J, Giambartolomei C, Pedersen MG, Goldstein J, Gordon SD, Gordon-Smith K, Green EK, Green MJ, Greenwood TA, Grove J, Guan W, Guzman-Parra J, Hamshere ML, Hautzinger M, Heilbronner U, Herms S, Hipolito M, Hoffmann P, Holland D, Huckins L, Jamain S, Johnson JS, Juréus A, Kandaswamy R, Karlsson R, Kennedy JL, Kittel-Schneider S, Knowles JA, Kogevinas M, Koller AC, Kupka R, Lavebratt C, Lawrence J, Lawson WB, Leber M, Lee PH, Levy SE, Li JZ, Liu C, Lucae S, Maaser A, MacIntyre DJ, Mahon PB, Maier W, Martinsson L, McCarroll S, McGuffin P, McInnis MG, McKay JD, Medeiros H, Medland SE, Meng F, Milani L, Montgomery GW, Morris DW, Mühleisen TW, Mullins N, Nguyen H, Nievergelt CM, Adolfsson AN, Nwulia EA, O'Donovan C, Loohuis LMO, Ori APS, Oruc L, Ösby U, Perlis RH, Perry A, Pfennig A, Potash JB, Purcell SM, Regeer EJ, Reif A, Reinbold CS, Rice JP, Rivas F, Rivera M, Roussos P, Ruderfer DM, Ryu E, Sánchez-Mora C, Schatzberg AF, Scheftner WA, Schork NJ, Shannon Weickert C, Shehktman T, Shilling PD, Sigurdsson E, Slaney C, Smeland OB, Sobell JL, Søholm Hansen C, Spijker AT, St Clair D, Steffens M, Strauss JS, Streit F, Strohmaier J, Szelinger S, Thompson RC, Thorgeirsson TE, Treutlein J, Vedder H, Wang W, Watson SJ, Weickert TW, Witt SH, Xi S, Xu W, Young AH, Zandi P, Zhang P, Zöllner S; eQTLGen Consortium; BIOS Consortium, Adolfsson R, Agartz I, Alda M, Backlund L, Baune BT, Bellivier F, Berrettini WH, Biernacka JM, Blackwood DHR, Boehnke M, Børglum AD, Corvin A, Craddock N, Daly MJ, Dannlowski U, Esko T, Etain B, Frye M, Fullerton JM, Gershon ES, Gill M, Goes F, Grigoroiu-Serbanescu M, Hauser J, Hougaard DM, Hultman CM, Jones I, Jones LA, Kahn RS, Kirov G, Landén M, Leboyer M, Lewis CM, Li QS, Lissowska J, Martin NG, Mayoral F, McElroy SL, McIntosh AM, McMahon FJ, Melle I, Metspalu A, Mitchell PB, Morken G, Mors O, Mortensen PB, Müller-Myhsok B, Myers RM, Neale BM, Nimgaonkar V, Nordentoft M, Nöthen MM, O'Donovan MC, Oedegaard KJ, Owen MJ, Paciga SA, Pato C, Pato MT, Posthuma D, Ramos-Quiroga JA, Ribasés M, Rietschel M, Rouleau GA, Schalling M, Schofield PR, Schulze TG,

Serretti A, Smoller JW, Stefansson H, Stefansson K, Stordal E, Sullivan PF, Turecki G, Vaaler AE, Vieta E, Vincent JB, Werge T, Nurnberger JI, Wray NR, Di Florio A, Edenberg HJ, Cichon S, Ophoff RA, Scott LJ, Andreassen OA, Kelsoe J, Sklar P & Bipolar Disorder Working Group of the Psychiatric Genomics Consortium (2019). Genome-wide association study identifies 30 loci associated with bipolar disorder. Nature genetics, 51(5), 793–803. https://doi.org/10.1038/s41588-019-0397-8

Starcevic, V., Bogojevic, G., Marinkovic, J., & Kelin, K. (1999). Axis I and axis II comorbidity in panic/agoraphobic patients with and without suicidal ideation. *Psychiatry research*, *88*(2), 153–161. https://doi.org/10.1016/s0165-1781(99)00078-5

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., & Psychiatric Genomics Consortium (2018). Psychiatric Genomics: An Update and an Agenda. *The American journal of psychiatry*, *175*(1), 15–27. https://doi.org/10.1176/appi.ajp.2017.17030283

Su, Z., Marchini, J., & Donnelly, P. (2011). HAPGEN2: simulation of multiple disease SNPs. Bioinformatics (Oxford, England), 27(16), 2304–2305. https://doi.org/10.1093/bioinformatics/btr341

Tomko, R. L., Trull, T. J., Wood, P. K., & Sher, K. J. (2014). Characteristics of borderline personality disorder in a community sample: comorbidity, treatment utilization, and general functioning. *Journal of personality disorders*, *28*(5), 734–750. https://doi.org/10.1521/pedi_2012_26_093

Tsanas, A., Saunders, K. E., Bilderbeck, A. C., Palmius, N., Osipov, M., Clifford, G. D., Goodwin, G. M., & De Vos, M. (2016). Daily longitudinal self-monitoring of mood variability in bipolar disorder and borderline personality disorder. *Journal of affective disorders*, *205*, 225–233. https://doi.org/10.1016/j.jad.2016.06.065

Tweed, J. L., Schoenbach, V. J., George, L. K., & Blazer, D. G. (1989). The effects of childhood parental death and divorce on six-month history of anxiety disorders *The British Journal of Psychiatry: The Journal of Mental Science*, *154*, 823–828. https://doi.org/10.1192/bjp.154.6.823

Turki, M., Thabet, J., Charfi, N., Mezghanni, M., Bouali, M., Zouari, N., Zouari, L., Maalej, M. (2017). Generalized anxiety disorder comorbidities: Panic and depressive disorder. *European Psychiatry, 41*(S1), S486-S486. doi:10.1016/j.eurpsy.2017.01.580

Verstockt, B., Smith, K. G., & Lee, J. C. (2018). Genome-wide association studies in Crohn's disease: Past, present, and future. Clinical & translational immunology, 7(1), e1001. https://doi.org/10.1002/cti2.1001

Virginia C. Stoffel, Soo Lee, Jaclyn K. Schwartz, George T. Grossberg, CHAPTER 16 - Mental health and common psychiatric disorders associated with aging, Editor(s): Karen Frank Barney, Chair Emerita, Margaret A. Perkinson, Occupational Therapy with Aging Adults, Mosby, 2016, Pages 266-288, ISBN 9780323067768, https://doi.org/10.1016/B978-0-323-06776-8.00025-6.

Võhma, U., Aluoja, A., Vasar, V., Shlik, J., & Maron, E. (2010). Evaluation of personality traits in panic disorder using Swedish universities Scales of Personality. Journal of anxiety disorders, 24(1), 141–146. https://doi.org/10.1016/j.janxdis.2009.10.001

Watanabe, K., Stringer, S., Frei, O., Umićević Mirkov, M., de Leeuw, C., Polderman, T., van der Sluis, S., Andreassen, O. A., Neale, B. M., & Posthuma, D. (2019). A global overview of pleiotropy and genetic architecture in complex traits. *Nature genetics*, *51*(9), 1339–1348. https://doi.org/10.1038/s41588-019-0481-0

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

Wechsler D. Wechsler Intelligence Scale for Children: Fourth Edition (WISC-IV) San Antonio: Pearson; 2003

Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics (Oxford, England). 2010 Sep;26(17):2190-2191. https://doi.org/10.1093/bioinformatics/btq340

Witt SH, Streit F, Jungkunz M, Frank J, **Awasthi S**, Reinbold CS, Treutlein J, Degenhardt F, Forstner AJ, Heilmann-Heimbach S, Dietl L, Schwarze CE, Schendel D, Strohmaier J, Abdellaoui A, Adolfsson R, Air TM, Akil H, Alda M, Alliey-Rodriguez N, Andreassen OA, Babadjanova G, Bass NJ, Bauer M, Baune BT, Bellivier F, Bergen S, Bethell A, Biernacka JM, Blackwood DHR, Boks MP, Boomsma DI, Børglum AD, Borrmann-Hassenbach M, Brennan P, Budde M, Buttenschøn HN, Byrne EM, Cervantes P, Clarke TK, Craddock N, Cruceanu C, Curtis D, Czerski PM, Dannlowski U, Davis T, de Geus EJC, Di Florio A, Djurovic S, Domenici E, Edenberg HJ, Etain B, Fischer SB, Forty L, Fraser C, Frye MA, Fullerton JM, Gade K, Gershon ES, Giegling I, Gordon SD, Gordon-Smith K, Grabe HJ, Green EK, Greenwood TA, Grigoroiu-Serbanescu M, Guzman-Parra J, Hall

LS, Hamshere M, Hauser J, Hautzinger M, Heilbronner U, Herms S, Hitturlingappa S, Hoffmann P, Holmans P, Hottenga JJ, Jamain S, Jones I, Jones LA, Juréus A, Kahn RS, Kammerer-Ciernioch J, Kirov G, Kittel-Schneider S, Kloiber S, Knott SV, Kogevinas M, Landén M, Leber M, Leboyer M, Li QS, Lissowska J, Lucae S, Martin NG, Mayoral-Cleries F, McElroy SL, McIntosh AM, McKay JD, McQuillin A, Medland SE, Middeldorp CM, Milaneschi Y, Mitchell PB, Montgomery GW, Morken G, Mors O, Mühleisen TW, Müller-Myhsok B, Myers RM, Nievergelt CM, Nurnberger JI, O'Donovan MC, Loohuis LMO, Ophoff R, Oruc L, Owen MJ, Paciga SA, Penninx BWJH, Perry A, Pfennig A, Potash JB, Preisig M, Reif A, Rivas F, Rouleau GA, Schofield PR, Schulze TG, Schwarz M, Scott L, Sinnamon GCB, Stahl EA, Strauss J, Turecki G, Van der Auwera S, Vedder H, Vincent JB, Willemsen G, Witt CC, Wray NR, Xi HS; Bipolar Disorders Working Group of the Psychiatric Genomics Consortium; Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium; Schizophrenia Working Group of the Psychiatric Genomics Consortium, Tadic A, Dahmen N, Schott BH, Cichon S, Nöthen MM, Ripke S, Mobascher A, Rujescu D, Lieb K, Roepke S, Schmahl C, Bohus M, Rietschel M. Genome-wide association study of borderline personality disorder reveals genetic overlap with bipolar disorder, major depression and schizophrenia. Transl Psychiatry. 2017 Jun 20;7(6): e1155. doi: 10.1038/tp.2017.115. PMID: 28632202

Xue, A., Wu, Y., Zhu, Z., Zhang, F., Kemper, K. E., Zheng, Z., Yengo, L., Lloyd-Jones, L. R., Sidorenko, J., Wu, Y., eQTLGen Consortium, McRae, A. F., Visscher, P. M., Zeng, J., & Yang, J. (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

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., & GIANT Consortium (2018). Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. Human molecular genetics, 27(20), 3641–3649. https://doi.org/10.1093/hmg/ddy271

Zanarini, M. C., Frankenburg, F. R., Dubo, E. D., Sickel, A. E., Trikha, A., Levin, A., & Reynolds, V. (1998a). Axis I comorbidity of borderline personality disorder. *The American journal of psychiatry*, *155*(12), 1733–1739. https://doi.org/10.1176/ajp.155.12.1733

Zanarini, M. C., Frankenburg, F. R., Dubo, E. D., Sickel, A. E., Trikha, A., Levin, A., & Reynolds, V. (1998b). Axis II comorbidity of borderline personality disorder. *Comprehensive psychiatry*, *39*(5), 296–302. https://doi.org/10.1016/s0010-440x(98)90038-4

Zanarini, M. C., Frankenburg, F. R., Hennen, J., Reich, D. B., & Silk, K. R. (2004). Axis I comorbidity in patients with borderline personality disorder: 6-year follow-up and prediction of time to

remission. *The American journal of psychiatry*, *161*(11), 2108–2114. https://doi.org/10.1176/appi.ajp.161.11.2108

Zanarini, M. C., Frankenburg, F. R., Reich, D. B., Fitzmaurice, G., Weinberg, I., & Gunderson, J. G. (2008). The 10-year course of physically self-destructive acts reported by borderline patients and axis II comparison subjects. Acta psychiatrica Scandinavica, 117(3), 177–184. https://doi.org/10.1111/j.1600-0447.2008.01155.x

Zanarini, M. C., Williams, A. A., Lewis, R. E., Reich, R. B., Vera, S. C., Marino, M. F., Levin, A., Yong, L., & Frankenburg, F. R. (1997). Reported pathological childhood experiences associated with the development of borderline personality disorder. *The American journal of psychiatry*, *154*(8), 1101–1106. Zheng, J., Erzurumluoglu, A. M., Elsworth, B. L., Kemp, J. P., Howe, L., Haycock, P. C., Hemani, G., Tansey, K., Laurin, C., Early Genetics and Lifecourse Epidemiology (EAGLE) Eczema Consortium, Pourcain, B. S., Warrington, N. M., Finucane, H. K., Price, A. L., Bulik-Sullivan, B. K., Anttila, V., Paternoster, L., Gaunt, T. R., Evans, D. M., & Neale, B. M. (2017). LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics (Oxford, England)*, *33*(2), 272–279. https://doi.org/10.1093/bioinformatics/btw613

Zugliani, M. M., Martin-Santos, R., Nardi, A. E., & Freire, R. C. (2017). Personality Traits in Panic Disorder Patients With and Without Comorbidities. *The Journal of nervous and mental disease*, *205*(11), 855–858. https://doi.org/10.1097/NMD.0000000000000745

9) Statutory Declaration (Eidesstattliche Versicherung)

"I, **Swapnil Awasthi**, by personally signing this document in lieu of an oath, hereby affirm that I prepared the submitted dissertation on the topic **Fine-tuning of QC and imputation pipeline *Ricopili* reveals a genetic overlap of panic disorder with neuroticism and depression (Die Feinabstimmung der QC- und Imputations Pipeline *Ricopili* zeigt eine genetische Überlappung der Panikstörung mit Neurotizismus und Depression)** and, independently and without the support of third parties, and that I used no other sources and aids than those stated.

All parts that are based on the publications or presentations of other authors, either in letter or in spirit, are specified as such in accordance with the citing guidelines. The sections on methodology (in particular regarding practical work, laboratory regulations, statistical processing) and results (in particular regarding figures, charts and tables) are exclusively my responsibility.

Furthermore, I declare that I have correctly marked all of the data, the analyses, and the conclusions generated from data obtained in collaboration with other persons, and that I have correctly marked my own contribution and the contributions of other persons (cf. declaration of contribution). I have correctly marked all texts or parts of texts that were generated in collaboration with other people.
My contributions to any publications related to this dissertation correspond to those stated in the below joint declaration made together with the supervisor. All publications created within the scope of the dissertation comply with the guidelines of the ICMJE (International Committee of Medical Journal Editors; http://www.icmje.org/ on authorship. In addition, I declare that I shall comply with the regulations of Charité – Universitätsmedizin Berlin on ensuring good scientific practice.
I declare that I have not yet submitted this dissertation in identical or similar form to another Faculty.
The significance of this statutory declaration and the consequences of a false statutory declaration under criminal law (Sections 156, 161 of the German Criminal Code) are known to me."

Date Signature

10) Declaration of your own contribution to the publications

Swapnil Awasthi contributed the following to the below-listed publications:

Publication 1: Lam, M*., **Awasthi, S***., Watson, H. J., Goldstein, J., Panagiotaropoulou, G., Trubetskoy, V., Karlsson, R., Frei, O., Fan, C. C., De Witte, W., Mota, N. R., Mullins, N., Brügger, K., Lee H., Wray, N., Skarabis, N., Huang, H., Neale, B., Daly, M., Mattheissen, M., Walters, R., Ripke, S., RICOPILI: Rapid Imputation for COnsortias PIpeLIne, Bioinformatics, 2019.

Contribution (please set out in detail):
I tested the various modules of the *Ricopili* pipeline described under sections *"Pre-imputation/QC" (2.2), "Principal components analysis" (2.3), "Imputation" (2.4), "Post-imputation" (2.5), and "Additional utility modules" (2.6)*, especially for calculating polygenic risk scoring and replication analysis. I also introduced the interventions to the simulated genotype to mimic real-world genotype data discussed under section *"Availability of simulated GWAS data" (2.7)* of the manuscript. I independently prepared the framework for common variant analysis using the modules from 2.2 to 2.6 and the simulated data (2.7). I developed a tutorial based on a common variant framework and many other tutorials (relevant to this project) that constitute a large part of the supplementary document. Moreover, I was also involved with testing the pipeline's portability on various clusters *(section 2.8; "Cluster portability")* with other authors and independently developing the web app *(section 2.9; "RICOPILI web app")* for the users who have less experience with Unix operating systems. Finally, I was involved in the writing of the manuscript along with the first author.

Data underlying all ***tables (1 to 11)*** and all ***figures (1 to 46)*** from the supplement[2] were produced with my analyses, furthermore I created those tables and plots myself with the exception of figure 33.

Publication 2: Forstner AJ*, **Awasthi S***, Wolf C*, Maron E, Erhardt A, Czamara D, Eriksson E, Lavebratt C, Allgulander C, Friedrich N, Becker J, Hecker J, Rambau S, Conrad R, Geiser F, McMahon FJ, Moebus S, Hess T, Buerfent BC, Hoffmann P, Herms S, Heilmann-Heimbach S, Kockum I, Olsson T, Alfredsson L, Weber H, Alpers GW, Arolt V, Fehm L, Fydrich T, Gerlach AL, Hamm A, Kircher T, Pané-Farré CA, Pauli P, Rief W, Ströhle A, Plag J, Lang T, Wittchen HU, Mattheisen M, Meier S, Metspalu A, Domschke K, Reif A, Hovatta I, Lindefors N, Andersson E, Schalling M, Mbarek H,

[2] The journal did not allow for figures and tables in the main text.

Milaneschi Y, de Geus EJC, Boomsma DI, Penninx BWJH, Thorgeirsson TE, Steinberg S, Stefansson K, Stefansson H, Müller-Myhsok B, Hansen TF, Børglum AD, Werge T, Mortensen PB, Nordentoft M, Hougaard DM, Hultman CM, Sullivan PF, Nöthen MM, Woldbye DPD, Mors O, Binder EB, Rück C, Ripke S, Deckert J, Schumacher J.. Genome-wide association study of panic disorder reveals genetic overlap with neuroticism and depression, Molecular Psychiatry, 2019.

Contribution (please set out in detail):

I was the core analyst for this project. My contribution includes preparing an analysis plan, quality control, imputation *(Results; "Quality control and imputation"),* and genotype-phenotype association in six case-control panic disorder (PD) cohorts *(Results; "Single-marker association analysis").* I conducted the meta-analysis *(Figure 1: Manhattan plot and Result's "Single-marker association analysis")* and downstream analyses for all main outcome parameters, such as estimating SNP-based heritability using various methods *(Results; "LD Score regression analyses" and "SumHer/LDAK analysis")*, testing PD's co-heritability with different psychiatric and non-psychiatric traits *(Results; "LD Score regression analyses" and Figure 3)* gene-based and gene set analysis *(Results; "MAGMA: gene-based analysis," "MAGMA: gene-set and tissue expression enrichment analyses" and Figure 4)*. I investigated the replication of the top findings in several independent studies mentioned in the results section *"Follow-up analysis" and Table 1*. I performed a leave-one-out polygenic risk analysis *(Results; "Polygenic risk score analysis" and Figure. 2)* and was involved in writing the manuscript along with the other first author.

I created all ***figures (1 to 4)*** of this manuscript, which also resulted from my own analysis. ***Table 1*** in the manuscript is also a direct outcome of my analysis. Moreover, I also created the data and plot in ***supplementary figure 1*** and the data generated from my analysis are presented in ***supplementary tables 2,3 and 4.***

Publication 3: Witt SH, Streit F, Jungkunz M, Frank J, **Awasthi S**, Reinbold CS, Treutlein J, Degenhardt F, Forstner AJ, Heilmann-Heimbach S, Dietl L, Schwarze CE, Schendel D, Strohmaier J, Abdellaoui A, Adolfsson R, Air TM, Akil H, Alda M, Alliey-Rodriguez N, Andreassen OA, Babadjanova G, Bass NJ, Bauer M, Baune BT, Bellivier F, Bergen S, Bethell A, Biernacka JM, Blackwood DHR, Boks MP, Boomsma DI, Børglum AD, Borrmann-Hassenbach M, Brennan P, Budde M, Buttenschøn HN, Byrne EM, Cervantes P, Clarke TK, Craddock N, Cruceanu C, Curtis D, Czerski PM, Dannlowski U, Davis T, de Geus EJC, Di Florio A, Djurovic S, Domenici E, Edenberg HJ, Etain B, Fischer SB, Forty L, Fraser C, Frye MA, Fullerton JM, Gade K, Gershon ES, Giegling I, Gordon SD, Gordon-Smith K, Grabe HJ, Green

EK, Greenwood TA, Grigoroiu-Serbanescu M, Guzman-Parra J, Hall LS, Hamshere M, Hauser J, Hautzinger M, Heilbronner U, Herms S, Hitturlingappa S, Hoffmann P, Holmans P, Hottenga JJ, Jamain S, Jones I, Jones LA, Juréus A, Kahn RS, Kammerer-Ciernioch J, Kirov G, Kittel-Schneider S, Kloiber S, Knott SV, Kogevinas M, Landén M, Leber M, Leboyer M, Li QS, Lissowska J, Lucae S, Martin NG, Mayoral-Cleries F, McElroy SL, McIntosh AM, McKay JD, McQuillin A, Medland SE, Middeldorp CM, Milaneschi Y, Mitchell PB, Montgomery GW, Morken G, Mors O, Mühleisen TW, Müller-Myhsok B, Myers RM, Nievergelt CM, Nurnberger JI, O'Donovan MC, Loohuis LMO, Ophoff R, Oruc L, Owen MJ, Paciga SA, Penninx BWJH, Perry A, Pfennig A, Potash JB, Preisig M, Reif A, Rivas F, Rouleau GA, Schofield PR, Schulze TG, Schwarz M, Scott L, Sinnamon GCB, Stahl EA, Strauss J, Turecki G, Van der Auwera S, Vedder H, Vincent JB, Willemsen G, Witt CC, Wray NR, Xi HS; Bipolar Disorders Working Group of the Psychiatric Genomics Consortium; Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium; Schizophrenia Working Group of the Psychiatric Genomics Consortium, Tadic A, Dahmen N, Schott BH, Cichon S, Nöthen MM, Ripke S, Mobascher A, Rujescu D, Lieb K, Roepke S, Schmahl C, Bohus M, Rietschel M. Genome-wide association study of borderline personality disorder reveals genetic overlap with bipolar disorder, major depression and schizophrenia, Translational Psychiatry, 2017.

Contribution (please set out in detail):

I conducted quality, genomic control, imputation, and genome-wide association analysis of borderline personality disorder (BPD) cohorts, which consisted of 998 BPD cases and 1545 controls *(Results section "Single marker analysis," figure 1 and figure 2)*. I also computed genetic overlap with MDD, SCZ, and BP, the primary outcome of this study *(mentioned under Results section "LD-score regression")*. Further, I calculated the PRS for MDD, SCZ, and bipolar for each individual in the borderline cohorts and tested PRS's discriminative ability to predict BPD cases and controls *(under results section "Polygenic risk score" and figure 3)*. Moreover, along with providing the plots from my part of the investigation, I critically reviewed the manuscript.

I created ***figures 1 and 2*** based on my independent analysis. ***Figure 3*** was generated by the data produced by my analysis. All the ***supplementary figures (1 to 8)*** are created by me and depict the data produced from my analysis. The ***supplementary tables 2 and 4*** present the data I produced through my independent analysis.

Publication 4: Kaminski, J. A., Schlagenhauf, F., Rapp, M., **Awasthi, S.**, Ruggeri, B., Deserno, L., Banaschewski, T., Bokde, A. L. W., Bromberg, U., Büchel, C., Quinlan, E. B., Desrivières, S., Flor, H., Frouin, V., Garavan, H., Gowland, P., Ittermann, B., Martinot, J. L., Martinot, M. P., Nees, F., Orfanos, D. P., Paus, T., Poustka, L., Smolka, M. N., Fröhner, J. H., Walter H., Whelan, R., Ripke, S., Schumann, G., Heinz, A.; IMAGEN consortium. Epigenetic variance in dopamine D2 receptor: a marker of IQ malleability? Translational Psychiatry. 2018.

Contribution (please set out in detail):

I conducted quality control and imputation of the genotypic data of the 1475 European subjects from the IMAGEN project *(described section "Genetics").* I calculated polygenic scores with two formerly published studies, childhood intelligence GWAS (Benyamin et al., 2014) and human intelligence GWAS (Sniekers et al., 2017), for each individual in the cohort *(presented in figure. 2)***.** I also performed the whole genome epigenetic (methylation sites) association of the general IQ, which helped prioritize the methylation sites to test their association with general IQ *(mentioned in section "Epigenetics" and Table 3 of the supplementary text)***.** I critically reviewed the manuscript and provided plots based on his analysis.

Figures 1 and 2 of the manuscript and supplementary ***figures 3, 5*** and ***table 2,3*** of the supplementary were created by the data generated by my own independent analysis.

_____

Signature, date, and stamp of first supervising university professor / lecturer

_____

Signature of doctoral candidate

11) Publication documents (Publikationschrift)

11.1 Publication 1: (Lam and **Awasthi** et al., 2020)

Publication	RICOPILI: Rapid Imputation for Consortias Pipeline
Jornal	BIOINFORMATICS
Date of Submission	18/04/2019
Rank/ Total	6/78
Total Cites	95,300
Impact factor	5.481
Eigenfactor Score	0.201110

Journal Data Filtered By: **Selected JCR Year: 2017** Selected Editions: SCIE,SSCI
Selected Categories: **"BIOCHEMICAL RESEARCH METHODS"** Selected
Category Scheme: WoS

Gesamtanzahl: 78 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	NATURE METHODS	54,686	26.919	0.243170
2	Nature Protocols	36,821	12.423	0.086550
3	CURRENT OPINION IN BIOTECHNOLOGY	14,009	8.380	0.024860
4	BRIEFINGS IN BIOINFORMATICS	4,731	6.302	0.015010
5	LAB ON A CHIP	29,513	5.995	0.051340
6	BIOINFORMATICS	95,300	5.481	0.201110
7	ACS Synthetic Biology	3,112	5.316	0.014110
8	Journal of Biological Engineering	756	5.256	0.001600
9	MOLECULAR & CELLULAR PROTEOMICS	17,761	5.232	0.046930
10	BIOCONJUGATE CHEMISTRY	15,194	4.485	0.021530
11	Plant Methods	2,026	4.269	0.004470
12	METHODS	19,646	3.998	0.024790
13	PLoS Computational Biology	23,758	3.955	0.082790
14	JOURNAL OF PROTEOME RESEARCH	21,459	3.950	0.041220
15	Journal of Biophotonics	2,723	3.768	0.006190
16	New Biotechnology	2,343	3.733	0.004740
17	Journal of Proteomics	9,432	3.722	0.025360
18	JOURNAL OF CHROMATOGRAPHY A	61,361	3.716	0.050920
19	BIOLOGICAL PROCEDURES ONLINE	679	3.581	0.000770
20	Proteomics Clinical Applications	1,972	3.567	0.004640
21	PROTEOMICS	14,902	3.532	0.023220
22	Clinical Proteomics	661	3.516	0.002120
23	Biotechnology Journal	4,515	3.507	0.008950
24	Expert Review of Proteomics	1,702	3.489	0.003440
25	Journal of Breath Research	1,606	3.489	0.003060
26	Biomedical Optics Express	8,120	3.482	0.022750
27	ANALYTICAL AND BIOANALYTICAL CHEMISTRY	28,970	3.307	0.042400
28	CYTOMETRY PART A	4,152	3.260	0.007940
29	Acta Crystallographica Section D-Structural Biology	20,390	3.099	0.028000
30	Drug Testing and Analysis	2,606	2.993	0.005100
31	JOURNAL OF NEUROSCIENCE METHODS	15,861	2.668	0.017520
32	JALA	989	2.632	0.002390

Bioinformatics, 36(3), 2020, 930–933
doi: 10.1093/bioinformatics/btz633
Advance Access Publication Date: 8 August 2019
Applications Note

Genome analysis

RICOPILI: Rapid Imputation for COnsortias PIpeLIne

Max Lam [1,2,3,4,5,†], Swapnil Awasthi,[1,6,†], Hunna J. Watson[7,8,9], Jackie Goldstein[1,2], Georgia Panagiotaropoulou[1,6], Vassily Trubetskoy[1,6], Robert Karlsson [10], Oleksander Frei[11], Chun-Chieh Fan[11], Ward De Witte[12], Nina R. Mota [12,13], Niamh Mullins[14], Kim Brügger[15], S. Hong Lee[16], Naomi R. Wray[17,18], Nora Skarabis[6], Hailiang Huang[1,2], Benjamin Neale[1,2], Mark J. Daly[1,2], Manuel Mattheisen[19,20,21], Raymond Walters[1,2] and Stephan Ripke[1,2,6,*]

[1]Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA, [2]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Cambridge, MA 02114, USA, [3]Division of Psychiatry Research, The Zucker Hillside Hospital, Northwell Health, Glen Oaks, NY 11004, USA, [4]Research Division, Institute of Mental Health Singapore, Singapore 539747, Singapore, [5]Human Genetics, Genome Institute of Singapore, Singapore 138672, Singapore, [6]Department of Psychiatry and Psychotherapy, Charité - Universitätsmedizin, 10117 Berlin, Germany, [7]Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [8]Division of Paediatrics, School of Medicine, The University of Western Australia, Perth, WA 6009, Australia, [9]School of Psychology, Curtin University, Perth, WA 6102, Australia, [10]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, 171 77 Stockholm, Sweden, [11]NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, Ullevål Hospital, N-0424 Oslo, Norway, [12]Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, [13]Department of Psychiatry, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, 6525 GA Nijmegen, The Netherlands, [14]Pamela Sklar Division of Psychiatric Genomics, Department of Genetics and Genomic Science, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA, [15]Department of Informatics, University of Bergen, 5020 Bergen, Norway, [16]Australian Centre for Precision Health, University of South Australia Cancer Research Institute, University of South Australia, Adelaide, South Australia 5000, Australia, [17]Institute for Molecular Bioscience, The University of Queensland, [18]Queensland Brain Institute, The University of Queensland, Brisbane, QLD 4072, Australia, [19]Department of Psychiatry, Psychosomatics and Psychotherapy, Center of Mental Health, University Hospital Würzburg, Würzburg, Germany, [20]Department of Clinical Neuroscience, Centre for Psychiatric Research, Karolinska Institutet, Stockholm, Sweden and [21]Department of Biomedicine, Aarhus University, Aarhus, Denmark

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the first two authors contributed equally.
Associate Editor: Russell Schwartz

Received on April 18, 2019; revised on July 31, 2019; editorial decision on August 2, 2019; accepted on August 7, 2019

Abstract

Summary: Genome-wide association study (GWAS) analyses, at sufficient sample sizes and power, have successfully revealed biological insights for several complex traits. RICOPILI, an open-sourced Perl-based pipeline was developed to address the challenges of rapidly processing large-scale multi-cohort GWAS studies including quality control (QC), imputation and downstream analyses. The pipeline is computationally efficient with portability to a wide range of high-performance computing environments. RICOPILI was created as the Psychiatric Genomics Consortium pipeline for GWAS and adopted by other users. The pipeline features (i) technical and genomic QC in case-control and trio cohorts, (ii) genome-wide phasing and imputation, (iv) association analysis, (v) meta-analysis, (vi) polygenic risk scoring and (vii) replication analysis. Notably, a major differentiator from other GWAS pipelines, RICOPILI leverages on automated parallelization and cluster job management approaches for rapid production of imputed genome-wide data. A comprehensive meta-analysis of simulated GWAS data has been incorporated demonstrating each step of the pipeline. This includes all the associated visualization plots, to allow ease of data interpretation and manuscript preparation. Simulated GWAS datasets are also packaged with the pipeline for user training tutorials and developer work.

Availability and implementation: RICOPILI has a flexible architecture to allow for ongoing development and incorporation of newer available algorithms and is adaptable to various HPC environments (QSUB, BSUB, SLURM and others). Specific links for genomic resources are either directly provided in this paper or via tutorials and external links. The central location hosting scripts and tutorials is found at this URL: https://sites.google.com/a/broadinstitute.org/RICOPILI/home
Contact: sripke@broadinstitute.org
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Genome-wide association studies (GWASs) have enabled the discovery of genetic variants underlying a plethora of complex traits (https://www.ebi.ac.uk/gwas/diagram). GWASs have highlighted previously unknown biological mechanisms associated with complex diseases and traits (Breen *et al.*, 2016). The Psychiatric Genomics Consortium (PGC) (http://www.med.unc.edu/pgc) the largest umbrella organization for psychiatric genetics (Sullivan *et al.*, 2018)—have made possible to advance the objectives of (i) revealing biological insights of psychiatric illness, (ii) informing clinical practice and (iii) presenting new therapeutic targets through sheer number of cohorts for GWASs across various psychiatric traits (Breen *et al.*, 2016; Sullivan *et al.*, 2012). The exponential availability of cohorts requires efficient, consistent and standardized approaches for various aspects of GWAS data management and analysis. Here, we introduce RICOPILI, the pipeline that automates rapid GWAS analysis workflow across various PGC workgroups. The pipeline is state-of-art, constantly incorporating latest available GWAS computational techniques and methods. With open-sourced simulated GWAS datasets and training tutorials packaged with the pipeline, RICOPILI is ideal for those contributing to large-scale genetic studies.

1.1 Comparison with other GWAS quality control and imputation pipelines

To our understanding, RICOPILI is the only open-sourced GWAS pipeline allowing secure data management, efficient data processing and downstream analysis scalable on both desktop and cluster environment. First, RICOPILI features an integrated quality control (QC), imputation and association analysis within its framework. Second, RICOPILI allows more than one imputation approach and reference panel to be utilized within its framework. Furthermore, computer intensive imputation can be processed locally within a closed cluster system. Third, the RICOPILI framework allows scalable processing of GWAS data, from a single CPU, to a cluster set up, or even within the cloud-based systems.

We compare RIOCPILI to existing available GWAS processing pipelines in Supplementary Table S1. All other tools focus on specific stages of GWAS analysis and do not provide the comprehensive features of RICOPILI. In the ensuing sections we will further highlight and discuss the features and functions of RICOPILI.

2 Design and implementation

2.1 Pipeline description

RICOPILI automates and integrates standard GWAS analysis methods, allowing for automated cluster submission and parallelization. The pipeline unifies standard software for its functions and implements best data analysis practices, provides sensible default settings while permitting the user to flexibly customize filters, thresholds and job resources as required. The optimization of cluster resources allows computations and visualizations to be completed quickly without significant user intervention. Written predominantly in Perl and R, the pipeline is organized according to analysis modules. Each module runs in its entirety via a single command line. The main module functions include:

- Pre-imputation technical QC;
- Principal components analysis (PCA) and relatedness estimation;
- Genome-wide imputation of genotype probabilities and generation of best guess genotypes in PLINK format (Purcell *et al.*, 2007);
- Downstream analyses, including GWAS, meta-analysis and polygenic risk scoring;
- Harmonizing large imputation reference panels (such as 1000 Genomes and the Haplotype Reference Consortium) to fit the architecture of RICOPILI.

RICOPILI takes dataset with unfiltered genotype calls, through trait association analysis, multi-cohort meta-analysis, linkage disequilibrium (LD) score regression (Bulik-Sullivan *et al.*, 2015), conditional analysis, replication analysis and polygenic risk scoring (Supplementary Fig. S1). Little intermediate interaction is required, allowing for efficient standardized analysis of genome-wide data and results. Standardized file naming conventions are designed to optimize overview and analysis record tracking within large-scale genetic projects. Publication-ready data visualizations and reports (in PDF and Excel format) permits easy evaluation of the results. Simulated datasets are also available with the pipeline for training and development purposes. In the ensuing sections, we describe the main components of the pipeline.

2.2 Pre-imputation/QC

The pre-imputation/QC module (Supplementary Section S1) consists of the following general steps (Supplementary Fig. S2):

- Inferring the genotyping chip;
- Standardizing file names and sample identifiers, incorporating chip information and ensuring that sample IDs across distinct cohorts are unique while keeping original sample IDs intact;
- Carrying out technical sample and variant QC procedures: RICOPILI will assign red, yellow and green flags to various QC parameters to help with the decision if a cohort needs further work before going into the following modules (Supplementary Fig. S1.1).

Detailed sample and variant filtering reports provide diagnostics to identify possible QC issues and solutions. Quality controlled datasets are saved separately for downstream analysis.

2.3 Principal components analysis

The PCA module (Supplementary Section S2) fulfils two objectives (Supplementary Fig. S3):

- Identify and remove duplicated or related samples for case-control and trio cohorts;
- Assess ancestral outliers and population stratification with EIGENSTRAT (Price *et al.*, 2006);
- Principal component scores are computed and could be utilized for visualization or as covariates to adjust for population structure in downstream post-imputation GWAS.

2.4 Imputation

RICOPILI automates computationally costly genotype imputation with an optimized routine for high-performance computing (HPC) environments (Supplementary Section S3 and Fig. S4). This module aligns genotype data to the imputation reference, pre-phases haplotypes and executes imputation. Users have the option to:

- Impute genotypes to the 1000 Genomes (1000 Genomes Project Consortium et al., 2015) or Haplotype Reference Consortium panel (McCarthy *et al.*, 2016);
- Perform pre-phasing with Eagle (Loh *et al.*, 2016) or SHAPEIT (Delaneau et al., 2011);
- Perform imputation with IMPUTE (Bycroft *et al.*, 2018; Howie *et al.*, 2009) or Minimac (Das *et al.*, 2016; Howie *et al.*, 2012).

RICOPILI allows for automated data preparation, alignment and sharing with public imputation servers (https://docs.google.com/document/d/18dupvU4kw11slREc1TUfwQwhO_eI0n_MeKVpwi4HLNA/) [e.g. Michigan (https://imputationserver.sph.umich.edu/index.html#!pages/home), Sanger (https://imputation.sanger.ac.uk/)], and reintegration of the results back into the RICIOPILI data structure. This is especially beneficial if an HPC environment is not accessible, and imputation by third party services has been approved by the user's local Institutional Review Board (IRB). More importantly with larger reference panels, such as the HRC and TopMed imputation panels becoming available but not directly accessible, RICOPILI allows such resources to be utilized.

The imputation output files are a set of genotype probabilities for all markers and 'best-guess' genotype hardcall files filtered on imputation quality and minor allele frequency. Hard call genotypes are available in three levels (hardcall with genotype probability >0.8, otherwise missing): (i) no further filter, (ii) lightly filtered (missingness <0.02) and (iii) filtered with strict criteria (missingness <0.01; MAF $>5\%$).

RICOPILI allows the creation of case-pseudo-controls to handle imputation and association procedures for trios.

2.5 Post-imputation

The post-imputation module (Supplementary Section S4 and Fig. S5) performs association analysis using imputed dosage files, meta-analysis via METAL (Willer *et al.*, 2010), conditional analysis, polygenic risk scoring, LD score regression (Bulik-Sullivan *et al.*, 2015) and replication analysis. Covariates (e.g. age, sex, principal components from PCA) and alternative phenotypes, including quantitative traits may be incorporated within the post-imputation module. Automated 'clumping' of genome-wide significant single nucleotide polymorphisms to facilitate identification of independently associated genetic loci. Publication-ready reports and visualizations such as Manhattan plots, QQ-plots, forest plots, annotated region plots and polygenic risk distributions are generated by the module as well. It is notable that genome-wide summary statistics as well as input statistics for various Manhattan and QQ-plots, as well as clumped summary statistics are automatically made available in the distribution/folder as part of the pipeline. These could then be utilized for downstream and follow-on analysis (https://docs.google.com/document/d/1jiD25BYjPAO-TLRAPkYSspiovn8wiQ29ZmZv9Pe2I2U/) (e.g. GCTA; Yang *et al.*, 2011, Spredixcan; Barbeira *et al.*, 2018 and FUSION; Gusev *et al.*, 2016) for the GWAS results.

2.6 Additional utility modules

RICOPILI allows for additional features and modules (see Supplementary Information). Including, (i) reference builder: builds reference data for genotype imputation from publicly accessible reference panels (Supplementary Fig. S6), (ii) replication of GWAS: using external summary data or those generated by RICOPILI and (iii) polygenic leave-one-out analysis: where each input dataset is used as a hold out and polygenic risk prediction is done iteratively across hold out data. All helper scripts and modules are saved in a *centralized location* specified by the user within a folder called rp_bin/and logging files with *_info suffix are also available.

2.7 Availability of simulated GWAS data

To allow new users to familiarize themselves with RICOPILI and experienced users to develop new functionality for the pipeline, we simulated freely available GWAS data using HAPGEN (Su *et al.*, 2011) (Supplementary Section S6). The dataset comprises 6200 'individuals' across ~600 000 markers based on the Illumina OmniExpress, a widely used genotyping platform. For training and development purposes, population stratification, cross-sample relatedness and technical errors were introduced to the simulated data. The sample is separated into five datasets 'HapGen5' packaged with RICOPILI (https://docs.google.com/document/d/1ux_FbwnvSzaiBVEwgS7eWJoYlnc_o0YHFb07SPQsYjI/). Data description and results are described in further detail in Extended Data Analysis and User Guide.

2.8 Cluster portability

RICOPILI is portable (https://docs.google.com/document/d/14aaoeT5hF541I8hHsDAL_42oyvlHRC5FWR7gir4xco/) to various LINUX-based HPC environments {e.g. BSUB (https://docs.google.com/document/d/1fNFnC3-rBZkmtH47Je_yUfGatB9qhDGi9HtMSA3_MPw/), QSUB (https://docs.google.com/document/d/1oY5IA4a6yG_pmbvWJC8A6MTzjYoGzVlqQ_aXUwWCl8I/), SLURM, GCP [Google Cloud Platform (https://docs.google.com/document/d/115NAaH6c8_C6Gn7D5JTldfW0CMGwUOMhxqd15thku-E/)]} (Supplementary Section S7). Support for Docker (https://hub.docker.com/r/bruggerk/ricopili; https://github.com/vtrubets/ricopili_docker) implementation of RICOPILI is also underway. In the absence of an HPC environment, RICOPILI can use the full potential of multi-core machines with parallel optimization. Regular updates and maintenance of the pipeline are carried out to incorporate the latest advances in genetic association methods. Ongoing support includes an active user forum (https://groups.google.com/forum/#!forum/ricopili-user-group), support website (https://sites.google.com/a/broadinstitute.org/ricopili/home) and detailed tutorials written by current RICOPILI analysts (consult footnotes).

2.9 RICOPILI web app

RICOPILI is now usable via browser on a cluster backed by Google Cloud: http://34.74.48.153. Here the user does not need any UNIX knowledge. Naturally the user needs to make sure that IRB allows for uploading genotype data to third party computer environments.

3 Discussion

RICOPILI has supported the analytical capability of the PGC, encompassing over 800 investigators internationally. The consortium is a testament to collaborative science that has unified much of the field and collated data collections, and enabled rapid progress in uncovering the genetic and biological basis of psychiatric disorders. RICOPILI addresses the need for a rapid computational pipeline for GWAS that integrates leading bioinformatics resources and produces publication-ready outputs. The PGC has reported GWAS studies in high-impact publications, most of which featured RICOPILI as the main analysis pipeline—including the seminal report identifying 108 GWAS loci for schizophrenia (Ripke et al., 2014). The pipeline has been adapted across various consortia, with 112 analysts performing rapid computation for GWAS to date. For this reason, we introduce RICOPILI to an audience of principal investigators, academics, analysts and all personnel tasked with determining the common variation underlying complex, heritable diseases and traits.

Acknowledgements

Computing and network infrastructure was provided by various sources, including SURFsara (Genetic Cluster Computer; LISA cluster) and the Stanley Center for Psychiatric Research at the Broad Institute of MIT and Harvard.

Funding

The PGC has received major funding from the US National Institutes of Health (NIH) [U01 MH109528 and U01 MH1095320]; M.L. received funding support from National Medical Research Council, Singapore [NMRC/TCR/003; MH095: 003/008-1014; NMRC/CG/004/2013; NMRC/SEEDFD/019/2017]; S.R. received funding support from NIH/NIMH [U01MH109528

01 (Ripke, PI) 2016-2021 'Psychiatric Genomics Consortium: Finding Actionable Variation'], Broad Institute of MIT and Harvard [691099; 'Psychiatric Genomics Consortium'] and the German Research Foundation [DFG, RI2846/1-1].

Conflict of Interest: none declared.

References

Barbeira,A.N. *et al.* (2018) Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.*, **9**, 1825.

Breen,G. *et al.* (2016) Translating genome-wide association findings into new therapeutics for psychiatry. *Nat. Neurosci.*, **19**, 1392–1396.

Bulik-Sullivan,B.K. *et al.* (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.*, **47**, 291–295.

Bycroft,C. *et al.* (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature*, **562**, 203–209.

Das,S. *et al.* (2016) Next-generation genotype imputation service and methods. *Nat. Genet.*, **48**, 1284–1287.

Delaneau,O. *et al.* (2011) A linear complexity phasing method for thousands of genomes. *Nat. Methods*, **9**, 179–181.

Gusev,A. *et al.* (2016) Integrative approaches for large-scale transcriptome--wide association studies. *Nat. Genet.*, **48**, 245–252.

Howie,B. *et al.* (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.*, **44**, 955–959.

Howie,B.N. *et al.* (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.*, **5**, e1000529.

Loh,P.R. *et al.* (2016) Fast and accurate long-range phasing in a UK Biobank cohort. *Nat. Genet.*, **48**, 811–816.

McCarthy,S. *et al.* (2016) A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.*, **48**, 1279–1283.

Price,A.L. *et al.* (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909.

Purcell,S. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.

Ripke,S. *et al.* (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, **511**, 421–427.

Su,Z. *et al.* (2011) HAPGEN2: simulation of multiple disease SNPs. *Bioinformatics*, **27**, 2304–2305.

Sullivan,P.F. *et al.* (2012) Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat. Rev. Genet.*, **13**, 537–551.

Sullivan,P.F. *et al.* (2018) Psychiatric genomics: an update and an agenda. *Am. J. Psychiatry*, **175**, 15–27.

Willer,C.J. *et al.* (2010) METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics*, **26**, 2190–2191.

Yang,J. *et al.* (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76–82.

1000 Genomes Project Consortium *et al.* (2015) A global reference for human genetic variation. *Nature*, **536**, 68–74.

11.2 Publication 2: (Forstner, **Awasthi** and Wolf et al., 2019)

Publication	Genome-wide association study of panic disorder reveals genetic overlap with neuroticism and depression
Jornal	MOLECULAR PSYCHIATRY
Date	29/05/2019
Rank/ Total	7/142
Total Cites	18,460
Impact factor	11.640
Eigenfactor Score	0.047200

Journal Data Filtered By: **Selected JCR Year: 2017** Selected Editions: SCIE,
Selected Categories: **"PSYCHIATRY"** Selected Category
Scheme: WoS

Gesamtanzahl: 142 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	World Psychiatry	4,055	30.000	0.010540
2	JAMA Psychiatry	8,414	16.642	0.044550
3	Lancet Psychiatry	3,223	15.233	0.015210
4	AMERICAN JOURNAL OF PSYCHIATRY	42,369	13.391	0.037870
5	PSYCHOTHERAPY AND PSYCHOSOMATICS	3,597	13.122	0.005520
6	BIOLOGICAL PSYCHIATRY	42,494	11.982	0.056910
7	MOLECULAR PSYCHIATRY	18,460	11.640	0.047200
8	JOURNAL OF NEUROLOGY NEUROSURGERY AND PSYCHIATRY	29,695	7.144	0.032980
9	SCHIZOPHRENIA BULLETIN	15,697	6.944	0.027700
10	NEUROPSYCHOPHARMACOLOGY	24,537	6.544	0.042870
11	JOURNAL OF CHILD PSYCHOLOGY AND PSYCHIATRY	18,604	6.486	0.023410
12	JOURNAL OF THE AMERICAN ACADEMY OF CHILD AND ADOLESCENT PSYCHIATRY	19,482	6.250	0.019260
13	ADDICTION	18,607	5.953	0.028990
14	BRITISH JOURNAL OF PSYCHIATRY	24,481	5.867	0.022960
15	Epidemiology and Psychiatric Sciences	950	5.684	0.003550
16	PSYCHOLOGICAL MEDICINE	23,080	5.475	0.039400
17	JOURNAL OF PSYCHIATRY & NEUROSCIENCE	2,989	5.182	0.004700
18	AUSTRALIAN AND NEW ZEALAND JOURNAL OF PSYCHIATRY	6,624	5.084	0.008440
19	DEPRESSION AND ANXIETY	7,923	5.043	0.015870
20	ACTA PSYCHIATRICA SCANDINAVICA	12,498	4.984	0.010890
21	JOURNAL OF PSYCHOPHARMACOLOGY	5,808	4.738	0.010900
22	PSYCHONEUROENDOCRINOLOGY	16,507	4.731	0.030420
23	Translational Psychiatry	5,384	4.691	0.021220
24	BIPOLAR DISORDERS	5,070	4.490	0.007870
25	CURRENT OPINION IN PSYCHIATRY	3,675	4.266	0.006830
26	JOURNAL OF CLINICAL PSYCHIATRY	18,677	4.247	0.020820
27	CNS DRUGS	4,364	4.206	0.007540
28	PROGRESS IN NEURO-PSYCHOPHARMACOLOGY & BIOLOGICAL PSYCHIATRY	9,823	4.185	0.013170
29	EUROPEAN NEUROPSYCHOPHARMACOLOGY	6,920	4.129	0.015110
29	EUROPEAN PSYCHIATRY	4,876	4.129	0.007890
31	JOURNAL OF PSYCHIATRIC RESEARCH	14,397	4.000	0.022480

Forstner AJ, Awasthi S, Wolf C, et al. Genome-wide association study of panic disorder reveals genetic overlap with neuroticism and depression. *Mol Psychiatry*. 2021;26(8):4179-4190. doi: 10.1038/s41380-019-0590-2

Forstner AJ, Awasthi S, Wolf C, et al. Genome-wide association study of panic disorder reveals genetic overlap with neuroticism and depression. *Mol Psychiatry*. 2021;26(8):4179-4190.
doi: 10.1038/s41380-019-0590-2

Forstner AJ, Awasthi S, Wolf C, et al. Genome-wide association study of panic disorder reveals genetic overlap with neuroticism and depression. *Mol Psychiatry*. 2021;26(8):4179-4190.
doi: 10.1038/s41380-019-0590-2

Forstner AJ, Awasthi S, Wolf C, et al. Genome-wide association study of panic disorder reveals genetic overlap with neuroticism and depression. *Mol Psychiatry*. 2021;26(8):4179-4190.
doi: 10.1038/s41380-019-0590-2

Forstner AJ, Awasthi S, Wolf C, et al. Genome-wide association study of panic disorder reveals genetic overlap with neuroticism and depression. *Mol Psychiatry*. 2021;26(8):4179-4190.
doi: 10.1038/s41380-019-0590-2

Forstner AJ, Awasthi S, Wolf C, et al. Genome-wide association study of panic disorder reveals genetic overlap with neuroticism and depression. *Mol Psychiatry*. 2021;26(8):4179-4190.
doi: 10.1038/s41380-019-0590-2

Forstner AJ, Awasthi S, Wolf C, et al. Genome-wide association study of panic disorder reveals genetic overlap with neuroticism and depression. *Mol Psychiatry*. 2021;26(8):4179-4190.
doi: 10.1038/s41380-019-0590-2

Forstner AJ, Awasthi S, Wolf C, et al. Genome-wide association study of panic disorder reveals genetic overlap with neuroticism and depression. *Mol Psychiatry*. 2021;26(8):4179-4190.
doi: 10.1038/s41380-019-0590-2

Forstner AJ, Awasthi S, Wolf C, et al. Genome-wide association study of panic disorder reveals genetic overlap with neuroticism and depression. *Mol Psychiatry*. 2021;26(8):4179-4190.
doi: 10.1038/s41380-019-0590-2

Forstner AJ, Awasthi S, Wolf C, et al. Genome-wide association study of panic disorder reveals genetic overlap with neuroticism and depression. *Mol Psychiatry*. 2021;26(8):4179-4190.
doi: 10.1038/s41380-019-0590-2

Forstner AJ, Awasthi S, Wolf C, et al. Genome-wide association study of panic disorder reveals genetic overlap with neuroticism and depression. *Mol Psychiatry*. 2021;26(8):4179-4190. doi: 10.1038/s41380-019-0590-2

Forstner AJ, Awasthi S, Wolf C, et al. Genome-wide association study of panic disorder reveals genetic overlap with neuroticism and depression. *Mol Psychiatry*. 2021;26(8):4179-4190.
doi: 10.1038/s41380-019-0590-2

11.3 Publication 3: (Witt et al., 2017)

Publication	Genome-wide association study of borderline personality disorder reveals genetic overlap with bipolar disorder, major depression and schizophrenia
Jornal	Translational Psychiatry
Date	16/02/2017
Rank/ Total	16/142
Total Cites	2,862
Impact factor	5.538
Eigenfactor Score	0.01375

Journal Data Filtered By: **Selected JCR Year: 2015** Selected Editions: SCIE Selected Categories: "**PSYCHIATRY**" Selected Category Schema: WoS Selected Open Access

Gesamtzahl:142

Rank	Journal name	Total Citations	Journal Impact factor	Eigenfactor
1	World Psychiatry	2,410	20.205	0.00686
2	JAMA Psychiatry	4,034	14.417	0.02158
3	AMERICAN JOURNAL OF PSYCHIATRY	41,752	13.505	0.04654
4	MOLECULAR PSYCHIATRY	15,797	13.314	0.0444
5	BIOLOGICAL PSYCHIATRY	42,289	11.212	0.07406
6	SCHIZOPHRENIA BULLETIN	14,839	7.757	0.02869
7	PSYCHOTHERAPY AND PSYCHOSOMATICS	2,808	7.632	0.00463
8	JOURNAL OF THE AMERICAN ACADEMY OF CHILD AND ADOLESCENT PSYCHIATRY	18,227	7.182	0.02318
9	BRITISH JOURNAL OF PSYCHIATRY	22,682	7.06	0.02504
10	JOURNAL OF CHILD PSYCHOLOGY AND PSYCHIATRY	16,111	6.615	0.02469
11	JOURNAL OF NEUROLOGY NEUROSURGERY AND PSYCHIATRY	26,259	6.431	0.03284
12	NEUROPSYCHOPHARMACOLOGY	22,869	6.399	0.04998
13	ACTA PSYCHIATRICA SCANDINAVICA	12,243	6.128	0.01226
14	Lancet Psychiatry	459	5.756	0.00151
15	JOURNAL OF PSYCHIATRY & NEUROSCIENCE	2,690	5.57	0.00551
16	Translational Psychiatry	2,862	5.538	0.01375
17	PSYCHOLOGICAL MEDICINE	20,304	5.491	0.03655
18	JOURNAL OF CLINICAL PSYCHIATRY	18,062	5.408	0.02312
19	DEPRESSION AND ANXIETY	6,362	5.004	0.01753
20	ADDICTION	16,558	4.972	0.03134
21	CNS DRUGS	3,755	4.91	0.00814
22	BIPOLAR DISORDERS	5,191	4.882	0.00957
23	PSYCHONEUROENDOCRINOLOGY	13,117	4.704	0.0276
24	JOURNAL OF PSYCHIATRIC RESEARCH	13,108	4.465	0.02474
25	SCHIZOPHRENIA RESEARCH	19,988	4.453	0.03502

OPEN

Citation: Transl Psychiatry (2017) 7, e1155; doi:10.1038/tp.2017.115

www.nature.com/tp

ORIGINAL ARTICLE

Genome-wide association study of borderline personality disorder reveals genetic overlap with bipolar disorder, major depression and schizophrenia

SH Witt[1,123], F Streit[1,123], M Jungkunz[2,3], J Frank[1], S Awasthi[4], CS Reinbold[5], J Treutlein[1], F Degenhardt[6,7], AJ Forstner[5,6,7,8], S Heilmann-Heimbach[6], L Dietl[9], CE Schwarze[10], D Schendel[1], J Strohmaier[1], A Abdellaoui[11], R Adolfsson[12], TM Air[13], H Akil[14], M Alda[15], N Alliey-Rodriguez[16], OA Andreassen[17,18], G Babadjanova[19], NJ Bass[20], M Bauer[21], BT Baune[13], F Bellivier[22], S Bergen[23], A Bethell[24], JM Biernacka[25], DHR Blackwood[26], MP Boks[27], DI Boomsma[11], AD Børglum[28,29,30], M Borrmann-Hassenbach[31], P Brennan[32], M Budde[33,34], HN Buttenschøn[35], EM Byrne[36], P Cervantes[37], T-K Clarke[26], N Craddock[38], C Cruceanu[39], D Curtis[40,41], PM Czerski[42], U Dannlowski[43,44], T Davis[13], EJC de Geus[11], A Di Florio[45], S Djurovic[46,47], E Domenici[48], HJ Edenberg[49,50], B Etain[51], SB Fischer[5], L Forty[45], C Fraser[45], MA Frye[52], JM Fullerton[53,54], K Gade[33,34], ES Gershon[16], I Giegling[55], SD Gordon[56], K Gordon-Smith[57], HJ Grabe[58], EK Green[59], TA Greenwood[60], M Grigoroiu-Serbanescu[61], J Guzman-Parra[62], LS Hall[26,63], M Hamshere[38], J Hauser[42], M Hautzinger[64], U Heilbronner[34], S Herms[5,6,7], S Hitturlingappa[13], P Hoffmann[5,6,7], P Holmans[38], J-J Hottenga[11], S Jamain[51,65], I Jones[24], LA Jones[57], A Juréus[23], RS Kahn[66], J Kammerer-Ciernioch[67], G Kirov[38], S Kittel-Schneider[68], S Kloiber[69,70,71], SV Knott[57], M Kogevinas[72], M Landén[23,73], M Leber[74], M Leboyer[75], QS Li[76], J Lissowska[77], S Lucae[71], NG Martin[56,78], F Mayoral-Cleries[62], SL McElroy[79], AM McIntosh[26,80], JD McKay[81], A McQuillin[82], SE Medland[56], CM Middeldorp[11], Y Milaneschi[83], PB Mitchell[84,85], GW Montgomery[86], G Morken[87,88], O Mors[89,90], TW Mühleisen[91,92], B Müller-Myhsok[39,93,94], RM Myers[95], CM Nievergelt[60], JI Nurnberger[96], MC O'Donovan[97], LMO Loohuis[98], R Ophoff[99], L Oruc[100], MJ Owen[97], SA Paciga[101], BWJH Penninx[83], A Perry[57], A Pfennig[21], JB Potash[102], M Preisig[103], A Reif[68], F Rivas[62], GA Rouleau[104,105], PR Schofield[53,54], TG Schulze[1,33,34,106,107], M Schwarz[108], L Scott[109], GCB Sinnamon[110], EA Stahl[111,112], J Strauss[69], G Turecki[37], S Van der Auwera[58], H Vedder[108], JB Vincent[113], G Willemsen[11], CC Witt[114], NR Wray[36,115], HS Xi[116], Bipolar Disorders Working Group of the Psychiatric Genomics Consortium, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium, A Tadic[117,118], N Dahmen[118], BH Schott[4,119], S Cichon[6,91,92,120], MM Nöthen[6,7], S Ripke[4,121,122], A Mobascher[118], D Rujescu[55], K Lieb[118], S Roepke[9], C Schmahl[2], M Bohus[3] and M Rietschel[1]

Borderline personality disorder (BOR) is determined by environmental and genetic factors, and characterized by affective instability and impulsivity, diagnostic symptoms also observed in manic phases of bipolar disorder (BIP). Up to 20% of BIP patients show comorbidity with BOR. This report describes the first case–control genome-wide association study (GWAS) of BOR, performed in one of the largest BOR patient samples worldwide. The focus of our analysis was (i) to detect genes and gene sets involved in BOR and (ii) to investigate the genetic overlap with BIP. As there is considerable genetic overlap between BIP, major depression (MDD) and schizophrenia (SCZ) and a high comorbidity of BOR and MDD, we also analyzed the genetic overlap of BOR with SCZ and MDD. GWAS, gene-based tests and gene-set analyses were performed in 998 BOR patients and 1545 controls. Linkage disequilibrium score regression was used to detect the genetic overlap between BOR and these disorders. Single marker analysis revealed no significant association after correction for multiple testing. Gene-based analysis yielded two significant genes: *DPYD* ($P = 4.42 \times 10^{-7}$) and *PKP4* ($P = 8.67 \times 10^{-7}$); and gene-set analysis yielded a significant finding for exocytosis (GO:0006887, $P_{FDR} = 0.019$; FDR, false discovery rate). Prior studies have implicated *DPYD*, *PKP4* and exocytosis in BIP and SCZ. The most notable finding of the present study was the genetic overlap of BOR with BIP ($r_g = 0.28$ [$P = 2.99 \times 10^{-3}$]), SCZ ($r_g = 0.34$ [$P = 4.37 \times 10^{-5}$]) and MDD ($r_g = 0.57$ [$P = 1.04 \times 10^{-3}$]). We believe our study is the first to demonstrate that BOR overlaps with BIP, MDD and SCZ on the genetic level. Whether this is confined to transdiagnostic clinical symptoms should be examined in future studies.

Translational Psychiatry (2017) **7,** e1155; doi:10.1038/tp.2017.115; published online 20 June 2017

INTRODUCTION

Borderline personality disorder (BOR; for the sake of readability, we have decided to use the rather unconventional abbreviation 'BOR' for Borderline Personality Disorder and the abbreviation 'BIP' for Bipolar Disorder) is a complex neuropsychiatric disorder with a lifetime prevalence of around 3%.[1] Untreated cases often have a chronic and severely debilitating clinical course.[1] BOR affects up to 20% of all psychiatric inpatients, and is associated with high health-care utilization. BOR therefore represents a substantial socio-economic burden.[2,3]

Correspondence: Dr SH Witt, Central Institute of Mental Health, Department of Genetic Epidemiology in Psychiatry, University of Heidelberg, 68159 Mannheim, Germany.
E-mail: Stephanie.Witt@zi-mannheim.de
[123]These authors contributed equally to this work.
A full list of affiliations appears at the end of the paper.
Received 16 February 2017; accepted 10 April 2017

BOR is characterized by affective instability, emotional dysregulation and poor interpersonal functioning.[3] Suicide rates in BOR range between 6 and 8%, and up to 90% of patients engage in non-suicidal self-injurious behavior.[4] Other prototypical features include high-risk behaviors and impulsive aggression. Current theories view dysfunctions in emotion processing, social interaction and impulsivity as core psychological mechanisms of BOR.[5]

To date, genetic research into BOR has been limited. Available genetic studies have involved small samples and focused on candidate genes, while no genome-wide association study (GWAS) of BOR patients has yet been performed.[6] However, Lubke *et al*.[7] conducted a GWAS of borderline personality features using data from three cohorts comprising $n = 5802$, $n = 1332$ and $n = 1301$ participants, respectively. Using the borderline subscale of the Personality Assessment Inventory (PAI-BOR), four borderline personality features (affect instability, identity problems, negative relations and self-harm) were assessed. The most promising signal in the combined analysis of two samples was for seven SNPs in the gene *SERINC5*, which encodes a protein involved in myelination. Two of the SNPs could be replicated in the third sample. Interestingly, here, the effect was highest for the affect instability items, that is, features that are key characteristics of manic phases of bipolar disorder (BIP).

Understanding of the pathogenesis of BOR remains limited. Both environmental and genetic factors are known to have a role in BOR etiology. Familial aggregation has been demonstrated,[8,9] and heritability estimates from twin studies range from 35 to 65%, with higher heritability estimates being obtained with self-ratings.[10–12]

The potential comorbidity between BOR and BIP is part of an ongoing debate. For example, Fornaro *et al*.[13] report substantial comorbidity of ~20% with BIP, whereas Tsanas *et al*.[14] find clear symptomatic differences between these two diagnostic groups. BOR displays an overlap of some symptoms with BIP, such as affective instability. In contrast, features such as dissociative symptoms, a feeling of chronic emptiness and identity disturbances are specific to BOR.[15] To date, no twin or family study has generated conclusive results concerning a genetic overlap between the two disorders.[16,17] However, a twin study[18] and a large-population-based study using polygenic risk score analyses[19] indicate a genetic overlap between borderline personality features and neuroticism, an established risk factor for BIP and other psychiatric disorders.[20]

To the best of our knowledge, the present study represents the first case–control GWAS in BOR, and was performed in one of the largest BOR patient samples worldwide. Given the limited heritability and the expected complex genetic architecture of BOR, the sample is too small to generate significant results for single markers. Instead, the main aim of the investigation was to detect (i) genes and gene sets with a potential involvement in BOR; and (ii) potential genetic overlap with BIP. As a substantial overlap of common risk variants exists between BIP and schizophrenia (SCZ), and to a lesser extent between BIP and major depressive disorder (MDD), and as there is also a high comorbidity of BOR and MDD, a further aim of the study was to determine whether any observed genetic overlap between BOR and BIP, MDD and SCZ was driven by disorder-specific genetic factors using linkage disequilibrium (LD)-score regression and polygenic risk scores (PRS).

MATERIALS AND METHODS

Participants

The present sample comprised 1075 BOR patients and 1675 controls.[21] All the participants provided written informed consent before inclusion. The study was approved by the respective local ethics committees.

The patients were recruited at the following German academic institutions: Department of Psychosomatic Medicine, Central Institute of Mental Health, Mannheim ($n = 350$); Department of Psychiatry and Psychotherapy, University Medical Center Mainz ($n = 231$); and the Department of Psychiatry, Charité, Campus Benjamin Franklin, Berlin ($n = 494$). Inclusion criteria for patients were: age 16 to 65 years; Central European ancestry; and a lifetime DSM-IV diagnosis of BOR. The control sample comprised 1583 unscreened blood donors from Mannheim, and 92 subjects recruited by the University Medical Center Mainz.

Clinical assessment

The diagnoses of BOR were assigned according to DSM-IV criteria and on the basis of structured clinical interviews. The diagnostic criteria for BOR were assessed using the German version of the IPDE[22] or the SKID-II.[23] All the diagnostic interviews were conducted by trained and experienced raters. BOR patients with a comorbid diagnosis of BIP or SCZ assessed with SKID-I[23] were excluded.

Genotyping

Automated genomic DNA extraction was performed using the chemagic Magnetic Separation Module I (Chemagen Biopolymer-Technologie, Baesweiler, Germany). Genotyping was performed using the Infinium PsychArray-24 Bead Chip (Illumina, San Diego, CA, USA).

Quality control and imputation

A detailed description of the quality control and imputation procedures is provided elsewhere.[24]

Briefly, quality control parameters for the exclusion of subjects and single-nucleotide polymorphisms (SNPs) were: subject missingness > 0.02; autosomal heterozygosity deviation ($|Fhet| > 0.2$); SNP missingness > 0.02; difference in SNP missingness between cases and controls > 0.02; and SNP Hardy–Weinberg equilibrium ($P < 10^{-6}$ in controls; $P < 10^{-10}$ in cases).

Genotype imputation was performed using the pre-phasing/imputation stepwise approach in IMPUTE2/SHAPEIT (default parameters and a chunk size of 3 Mb),[25,26] using the 1000 Genomes Project reference panel (release 'v3.macGT1').[27]

Relatedness testing and population structure analysis were performed using a SNP subset that fulfilled strict quality criteria (INFO > 0.8, missingness $< 1\%$, minor allele frequency > 0.05), and which had been subjected to LD pruning ($r^2 > 0.02$). This subset comprised 63 854 SNPs. In cryptically related subjects, one member of each pair ($\hat{\pi} > 0.2$) was removed at random following the preferential retention of cases over controls. Principal components (PCs) were estimated from genotype data (see Supplementary Figures 1–6), and phenotype association was tested using logistic regression. The impact of the PCs on genome-wide test statistics was assessed using λ.

Association analysis

Including the first four PCs as covariates, an additive logistic regression model was used to test single marker associations, as implemented in PLINK.[28] The *P*-value threshold for genome-wide significance was set at 5×10^{-8}.

Gene-based analysis

To determine whether genes harbored an excess of variants with small *P*-values, a gene-based test was performed with MAGMA Version 1.04 (http://ctg.cncr.nl/software/magma)[29] using genotyped markers only, filtered with a minor allele frequency $> 1\%$ ($n = 284\,220$). This test uses summary data and takes LD between variants into account. SNPs within ± 10 kb of the gene boundary were assigned to each gene. Obtained *P*-values were Bonferroni-corrected for the number of tested genes ($n = 17\,755$, $P = 2.8 \times 10^{-6}$).

Gene-set analysis

Gene-set-based analysis was implemented using genotyped markers only, filtered as above. As in the gene-based analysis, SNPs within ± 10 kb of the gene boundary were assigned to each gene. Gene-set analyses were carried out using Gene Ontology (GO, http://software.broadinstitute.org/gsea/msigdb/) terms.

The discovery gene-set-based analysis was carried out using i-GSEA4GWASv2 (http://gsea4gwas-v2.psych.ac.cn/).[30] The size of the gene sets was restricted to 20–200 genes, and the major histocompatibility complex region was excluded. In total, 674 gene sets were tested. The results were

adjusted for multiple testing using false discovery rate (FDR). To validate the significant finding, the respective gene set was investigated with (i) GSA-SNP, using the *P*-value of the second-best SNP in each gene (https://gsa.muldas.org)[31] and (ii) MAGMA using summary data and a nominal *P*-value threshold of $P < 0.05$.

LD-score regression

To investigate a possible genetic overlap between BOR and SCZ, BIP and MDD, LD-score regression was performed.[32] Genetic correlations between BOR and (i) BIP, (ii) SCZ and (iii) MDD were calculated[33] using the result files of the Psychiatric Genomics Consortium (PGC) meta-analyses for SCZ (33 640 cases and 43 456 controls),[34] BIP (20 352 cases and 31 358 controls)[35] and MDD (16 823 cases and 25 632 controls).[35] There was no overlap in cases or controls of the present BOR GWAS sample with the PGC samples.

Polygenic risk score

To determine the impact of polygenic risk on BOR and subgroups (that is, BOR with and without MDD), PRS were calculated for each subject based on the above-mentioned PGC data sets.

To obtain a highly informative SNP set with minimal statistical noise, the following were excluded: low frequency SNPs (minor allele frequency < 0.1); low-quality variants (imputation INFO < 0.9) and indels. Subsequently, these SNPs were clumped discarding markers within 500 kb of, and in high LD ($r^2 \geqslant 0.1$) with, another more significant marker. From the major histocompatibility complex region, only one variant with the strongest significance was retained. PRS were calculated as described elsewhere.[36] This involved *P*-value thresholds 5×10^{-8}, 1×10^{-6}, 1×10^{-4}, 0.001, 0.01, 0.05, 0.1, 0.2, 0.5 and 1.0, and multiplication of the natural logarithm of the odds ratio of each variant by the imputation probability for the risk allele. The resulting values were then totaled. For each subject, this resulted in one PRS for SCZ, MDD and BIP for each *P*-value threshold.

In a first step, the association of the PRS for BIP, SCZ and MDD with BOR case–control status was analyzed using standard logistic regression and by including the four PCs as covariates. For each *P*-value threshold, the proportion of variance explained (Nagelkerke's R^2) in BOR case–control status was computed by comparison of a full model (covariates+PRS) score to a reduced model (covariates only).

For further exploratory analysis, the $P < 0.05$ PRS for each disorder was selected (that is, including all markers that reached nominal significance in the training samples). To determine whether the different scores contribute independently to the case–control status, a regression including the PRS for MDD, SCZ and BIP and the four PCs was computed. In a secondary analysis, two further models were computed. These included the PRS for BIP and the PRS of either MDD or SCZ, while controlling for the four PCs.

Furthermore, PRS were analyzed by differentiating between controls, and patients with or without comorbid MDD. For each PRS, a linear model was computed using the PRS as a dependent variable, disease state as an independent variable and the four PCs as covariates. Differences between groups were assessed using *post hoc* tests (Bonferroni-corrected).

RESULTS

Sample characteristics

Genetic quality control led to the exclusion of 207 subjects. Reasons for exclusion were: (i) insufficient data quality (low call rate), $n = 6$; (ii) relatedness, $n = 63$; and (iii) population outlier status, $n = 138$. After quality control, the sample comprised 998 BOR cases (914 female/84 male) and 1545 controls (868 female/677 male). Mean age for cases was 29.58 years (range: 18–65 years, standard deviation (s.d. = 8.64)). Mean age for controls was 44.19 years (range: 18–72 years, s.d. = 13.24; details see Supplementary Table 1). Of the 998 cases, 666 had comorbid lifetime MDD, and 262 did not (data missing for 40 cases).

Single marker analysis

A total of 10 736 316 single markers were included in the analysis. As expected for GWAS on a complex psychiatric disorder with the current sample size, the single marker analysis revealed no significant hit after correction for multiple testing (see Figures 1 and 2). The most significant marker was rs113507694 in *DPPA3* on chromosome 12 ($P = 2.01 \times 10^{-07}$; odds ratio = 0.35, minor allele frequency = 0.03, INFO = 0.59). Single markers with $P < 1 \times 10^{-5}$ are listed in Supplementary Table 2.

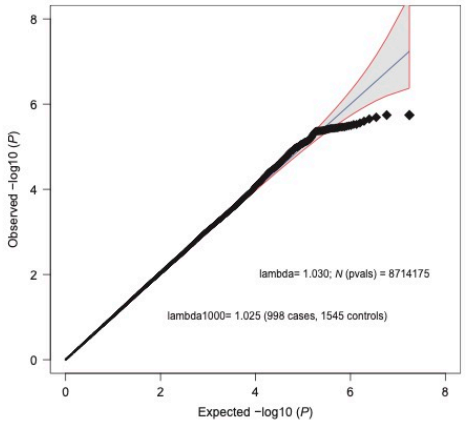

Figure 1. Quantile–Quantile plot. Quantile–Quantile plot of the case–control analysis (998 cases; 1545 controls) showing expected and observed −log10 *P*-values. The shaded region indicates the 95% confidence interval of expected *P*-values under the null hypothesis.

Gene-based analysis

In the gene-based analysis, a total of 17 755 genes were tested. Two genes showed significant association with BOR after correction for multiple testing: the gene coding for Plakophilin-4 on chromosome 2 (*PKP4*; $P = 8.24 \times 10^{-7}$); and the gene coding for dihydropyrimidine dehydrogenase on chromosome 1 (*DPYD*, $P = 1.20 \times 10^{-6}$). The most significant genes ($P < 5 \times 10^{-4}$) are listed in Table 1. The top hit of the previous GWAS of borderline personality features, *SERINC5,* achieved nominal significance in the present study ($P_{uncorrected} = 0.016$).

Gene-set analysis

Gene-set analysis with i-GSEA4GWASv2 revealed one significant gene set: exocytosis (GO: 0006887; $P_{FDR} = 0.019$). Of 25 genes in this gene set, 22 were mapped with variants and 15 showed nominally significant associations. Details on significant and nonsignificant genes in this gene set are provided in Supplementary Table 3. All gene sets with $P_{uncorrected} < 0.01$ are shown in Table 2. A technical replication analysis with GSA-SNP and MAGMA confirmed the gene-set exocytosis (GSA-SNP: $P_{uncorrected} = 2.32 \times 10^{-4}$; MAGMA: $P_{uncorrected} = 0.056$).

LD-score regression

Significant genetic correlations with BOR were found for BIP ($r_g = 0.28$; s.e. = 0.094; $P = 2.99 \times 10^{-3}$), MDD ($r_g = 0.57$; s.e. = 0.18; $P = 1.04 \times 10^{-3}$) and SCZ ($r_g = 0.34$; s.e. = 0.082; $P = 4.37 \times 10^{-5}$). A meta-analytic comparison revealed no significant differences between the correlations (all $P > 0.13$).

Polygenic risk score

PRS analysis revealed significant associations with BOR for the PRS of BIP, MDD and SCZ. SCZ PRS were significant for all investigated thresholds. BIP and MDD scores were significant for all PRS that

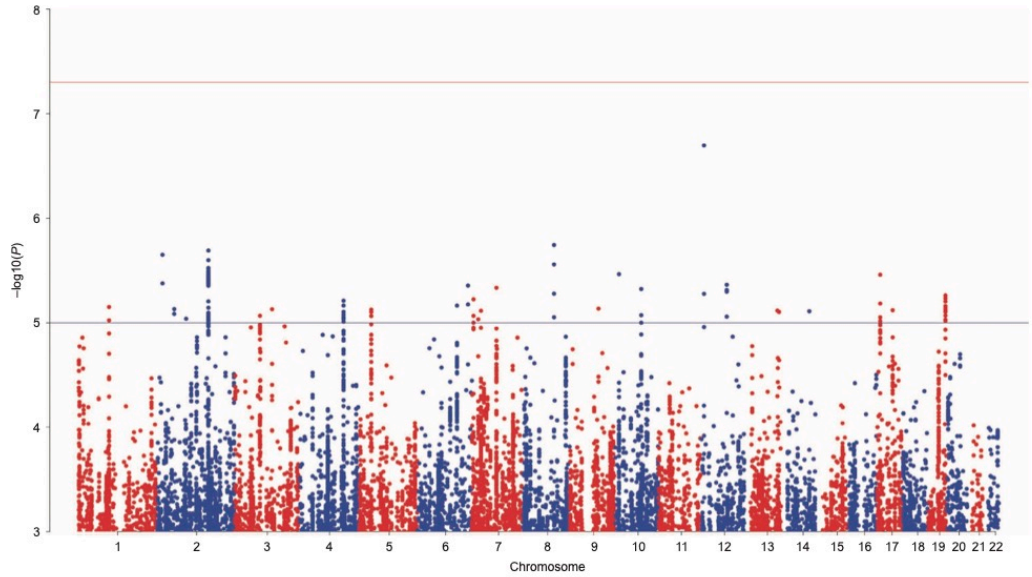

Figure 2. Manhattan plot showing association results. Manhattan plot of the case–control analysis (998 cases; 1545 controls). For each single-nucleotide polymorphism (SNP), the chromosomal position is shown on the *x* axis, and the –log10 *P*-value on the *y* axis. The red line indicates genome-wide significance ($P < 5 \times 10^{-8}$) and the blue line indicates suggestive evidence for association ($P < 1 \times 10^{-5}$).

Table 1. Results of the gene-based analysis using MAGMA

GENE	*CHR*	*START*	*STOP*	N_{SNPS}	N_{PARAM}	Z_{STAT}	P
PKP4	2	159303476	159547941	21	13	4.7924	8.24×10^{-7}
DPYD	1	97533299	98396615	105	68	4.7162	1.20×10^{-6}
GRAMD1B	11	123315191	123508478	34	28	3.8856	5.10×10^{-5}
STX8	17	9143788	9489275	38	33	3.7984	7.28×10^{-5}
BMP2	20	6738745	6770910	7	6	3.588	1.67×10^{-4}
TRAF3IP1	2	239219185	239319541	11	8	3.5389	2.01×10^{-4}
ZP3	7	76016841	76081388	9	7	3.5037	2.29×10^{-4}
PINX1	8	10612473	10707394	19	11	3.5034	2.30×10^{-4}
GTF3C4	9	135535728	135575471	4	4	3.4851	2.46×10^{-4}
DNAH1	3	52340335	52444513	11	8	3.4543	2.76×10^{-4}
YKT6	7	44230577	44263893	6	3	3.3841	3.57×10^{-4}
CCSER1	4	91038684	92533370	111	78	3.3804	3.62×10^{-4}
LRRC59	17	48448594	48484914	8	6	3.3716	3.74×10^{-4}
TMEM71	8	133712191	133782914	9	8	3.3668	3.80×10^{-4}
BAP1	3	52425020	52454121	3	3	3.345	4.11×10^{-4}
AQR	15	35138552	35271995	8	6	3.3299	4.34×10^{-4}
FGFR1	8	38258656	38336352	12	10	3.3162	4.56×10^{-4}

Abbreviations: CHR, chromosome; N_{PARAM}, number of parameters used in the model; N_{SNPS}, number of single-nucleotide polymorphisms; *P*, *P*-value of gene; Z_{STAT}, z-value of the gene. Most significant genes ($P < 5 \times 10^{-4}$) in the gene-based analysis and their chromosomal position. Genes in bold font were significant after correction for multiple testing.

included SNPs with *P*-values higher than 0.0001 and 0.001, respectively (see Supplementary Table 4). The share of variance explained in BOR case–control status (Nagelkerke's R^2) by the respective PRS was up to 0.86% for BIP; up to 3.1% for SCZ; and up to 2.1% for MDD (see Figure 3 and Supplementary Table 4).

Simultaneous addition of the PRS for SCZ, BIP and MDD (threshold $P < 0.05$) to the regression model explained 4.4% of the variance (Nagelkerke's R^2) in BOR case–control status. The PRS for SCZ and the PRS for MDD were significant predictors ($P = 9.78 \times 10^{-9}$ and $P = 1.9 \times 10^{-7}$, respectively). The PRS for BIP was not a significant predictor in this model ($P = 0.28$).

A secondary analysis was then performed including (i) BIP PRS with MDD PRS and (ii) BIP PRS with SCZ PRS. Here, BIP PRS explained variance independently of MDD PRS ($P = 0.0067$), but not of SCZ PRS ($P = 0.11$).

Differentiation between cases with and without comorbid MDD and controls revealed significant effects of BOR diagnosis on PRS for BIP, SCZ and MDD (all $P < 0.001$, see Figure 4). *Post hoc* analyses revealed no differences in PRS for the BIP, SCZ or MDD PRS of the BOR subgroup with comorbid MDD compared with the BOR subgroup without MDD (all $P > 0.5$).

Compared with controls, PRS for SCZ and MDD were significantly increased in the BOR subgroups with and without comorbid MDD (all $P < 0.001$). The PRS for BIP only showed a significant difference to controls in the BOR subgroup with comorbid MDD ($P < 0.001$, see Figure 4).

DISCUSSION

The present study is the first case–control GWAS of BOR. As expected, no genome-wide significant association was found for any single marker. In the gene-based test, however, two genes achieved genome-wide significance: dihydropyrimidine dehydrogenase (*DPYD*) and Plakophilin-4 (*PKP4*). *DPYD* encodes a pyrimidine catabolic enzyme, which is the initial and rate-limiting factor in the pathway of uracil and thymidine catabolism. Genetic deficiency of this enzyme results in an error in pyrimidine metabolism.[37] This is associated with thymine–uraciluria and an increased risk of toxicity in cancer patients receiving 5-fluorouracil chemotherapy (http://www.ncbi.nlm.nih.gov/gene/1806). Recent PGC meta-analyses revealed an association between *DPYD* and SCZ and BIP.[34,38,39] *DPYD* contains a binding site for the micro-RNA miR-137, which has previously been associated with schizophrenia,[40] and a previous exome-sequencing study reported two putative functional *de novo* variants in *DPYD* in cases with SCZ.[41] *PKP4* is involved in the regulation of cell adhesion and cytoskeletal organization.[42] In pathway analyses of PGC GWAS data, cell adhesion was associated with BIP,[43] and SCZ,[44] whereas cell junction was implicated in MDD, as well as in an integrative pathway analysis of all three disorders.[45]

SERINC5, which was the top hit of the previous GWAS of Borderline personality features,[7] achieved nominal significance in the present study. The protein SERINC5 incorporates serine into newly forming membrane lipids, and is enriched in myelin in the brain.[46] Previous research suggests that decreased myelination is associated with a reduced capacity for social interaction.[7,47]

The gene-set analyses yielded significant results for exocytosis. In neuronal synapses, exocytosis is triggered by an influx of calcium and critically underlies synaptic signaling. Dysregulated neuronal signaling and exocytosis are core features of neurodevelopmental psychiatric disorders such as the autism spectrum disorders and intellectual disability.[48,49] Moreover, recent findings from large meta-analyses have implicated dysregulated neuronal signaling and exocytosis in the molecular mechanisms of BIP, SCZ and MDD.[48,50,51] These processes may now represent promising starting points for further research into BOR.

The most interesting finding of this study is that BOR showed a genetic overlap with BIP, SCZ and MDD. Notably, BIP did not show a higher correlation with BOR ($r_g = 0.28$) than SCZ ($r_g = 0.34$) or MDD ($r_g = 0.57$). In view of the present sample size, these values must be viewed with caution. A more accurate estimation of these correlations will require calculations in larger cohorts.

Although comorbid BIP was excluded in the present BOR patients, the possibility that the observed genetic overlap between BOR and BIP was at least partly attributable to misdiagnosis cannot be excluded. However, an alternative explanation appears more likely, that is, that disorders currently categorized as BOR and BIP share a common genetic background,

Table 2. Results of the gene-set analysis

Gene-set name	*Number of genes*	*P-value*	*FDR P-value*
GO: EXOCYTOSIS	**25**	**0.001**	**0.019**
GO: RESPONSE TO ORGANIC SUBSTANCE	30	0.002	0.173
GO: BRAIN DEVELOPMENT	51	0.003	0.888
GO: HORMONE METABOLIC PROCESS	30	0.003	0.511
GO: PROTEIN C TERMINUS BINDING	73	0.003	0.536
GO: LYSOSOME	53	0.007	0.785
GO: LYTIC VACUOLE	53	0.007	0.785
GO: MULTI-ORGANISM PROCESS	143	0.007	0.920

Abbreviations: FDR, false discovery rate; GO, Gene Ontology; *P*-value, gene-set *P*-value. Most significant gene sets (uncorrected $P < 0.01$) in the gene-set analysis with i-GSEA4GWASv2 are listed. Gene sets in bold font were significant after correction for multiple testing.

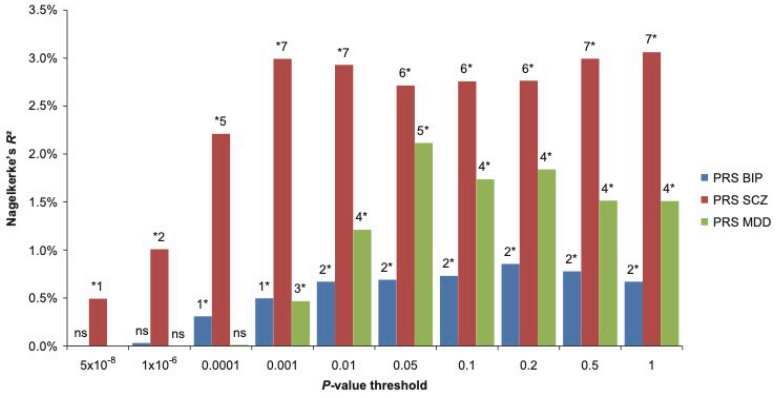

Figure 3. Polygenic risk score analysis. The proportion of variance explained in case–control status (*y* axis; Nagelkerke's R^2) by the PRS for BIP, SCZ and MDD is depicted for the different *P*-value cutoffs used in the calculation of the PRS. Principal components were included in the models to control for population stratification. 1*, $P < 0.05$; 2*, $P < 0.001$; 3*, $P < 1 \times 10^{-4}$; 4*, $P < 1 \times 10^{-6}$; 5*, $P < 1 \times 10^{-8}$; 6*, $P < 1 \times 10^{-10}$; 7*, $P < 1 \times 10^{-12}$. BIP, bipolar disorder; MDD, major depressive disorder; NS, nonsignificant; PRS, polygenic risk score; SCZ, schizophrenia.

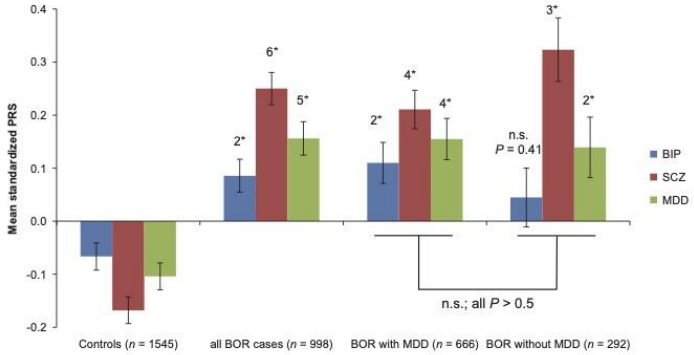

Figure 4. Polygenic risk score analysis in subgroups. Mean z-standardized PRS and standard error (s.e.) for BIP, SCZ and MDD are shown in the control group, all cases, and in cases with and without comorbid MDD. PRS with a *P*-value threshold of $P = 0.05$ were selected for this comparison and principal components were included in the models to control for population stratification. The numbers at the top of each bar indicate the significance of the difference in the respective PRS in comparison with the control group. 1*, $P < 0.05$; 2*, $P < 0.001$; 3*, $P < 1 \times 10^{-4}$; 4*, $P < 1 \times 10^{-6}$; 5*, $P < 1 \times 10^{-8}$; 6*, $P < 1 \times 10^{-10}$; 7*, $P < 1 \times 10^{-12}$. BIP, bipolar disorder; BOR, borderline personality disorder; MDD, major depressive disorder; NS, nonsignificant; PRS, polygenic risk score; SCZ, schizophrenia.

and they also do so with SCZ and MDD. This hypothesis is supported by the present observation of a genetic overlap between BOR and SCZ, two disorders that are rarely misdiagnosed by psychiatrists, despite the presence of common psychotic symptoms.

An explanation could also be that the genetic commonality between BOR and BIP, SCZ, and MDD might be due to a common effect of MDD. Prior to the introduction of DSM-IV, a history of MDD was required for a diagnosis of BIP, and MDD has a high prevalence in patients with SCZ (25-85%).[52,53] Therefore, the MDD genetic risk variants that are common to BOR, BIP, and SCZ may be responsible for the observed overlap. For this reason, we conducted two further analyses. First, we compared PRS of BIP, SCZ and MDD in subsamples of BOR patients with (~60%) and without comorbid MDD. Here, no differences in any of the PRS were found. Second, we performed a joint analysis of PRS of BIP, SCZ and MDD in a logistic regression analysis in BOR patients vs controls. Here, no differences were found in any of the PRS. Second, we performed a joint analysis of the PRS of BIP, SCZ and MDD in a logistic regression analysis in BOR patients vs controls. Here, both the SCZ and the MDD risk score explained variance in BOR case–control status independently. Secondary analysis revealed that the BIP risk score explained variance independently of the MDD risk score but not of the SCZ risk score. These results indicate that comorbidity with MDD does not explain the genetic overlap between BOR and BIP, SCZ and MDD. However, the training sets differ in terms of their power to detect underlying risk variants, and therefore the derived PRS differ in terms of the variance they can explain.

It must be noted, that in the PGC-BIP, -SCZ and -MDD samples, controls are partly overlapping. However, it is unlikely that this drives the genetic correlation of BOR with those disorders as the overlap of controls in these samples is rather small (under 10%).[54] Also, the joint logistic regression analysis demonstrated that polygenic risk for SCZ and MDD contributed independently to the BOR risk (see above).

The present study had several limitations. First, despite being one of the largest BOR samples available worldwide, the sample size was small in terms of the estimation of heritability. Replication of the present results is warranted in larger, independent cohorts. This should include the investigation of non-European samples. Second, no information was available on the presence of common clinical features such as psychotic symptoms and affect instability. This precluded detailed analysis of the identified genetic overlap. Future studies in larger cohorts should also investigate more detailed phenotypes, including comorbid axis I and axis II disorders, such as addiction and personality disorders, respectively. Third, the observation that psychiatric patients often establish non-random relationships with persons affected by the same or another psychiatric disorder,[55] and therefore have offspring with a higher genetic risk for psychiatric disorders, might contribute to the observed genetic correlation of BOR with BIP, SCZ and MDD. However, the LD-score method does not investigate the impact of assortative mating.[32] Therefore, assessment of the degree to which this phenomenon may have influenced the genetic correlation estimates was beyond the scope of the present study.

Despite these limitations, the results indicate that neither comorbidity with MDD nor risk variants that are exclusive to MDD explain the genetic overlap between BOR and BIP, SCZ and MDD. Future investigations of larger data sets for BOR and other psychiatric disorders are warranted to refine the analysis of shared and specific genetic risk.

Future studies are warranted to delineate the communalities and specificities of the respective disorders.

CONCLUSION

In summary, the present study is the first GWAS of patients diagnosed with BOR. The results suggest promising novel genes and a novel pathway for BOR, and demonstrate that, rather than being a discrete entity, BOR has an etiological overlap with the major psychoses. The genetic overlap with BIP is consistent with the observation that some diagnostic criteria for BOR overlap with those for BIP. The overlap between BOR and SCZ and MDD is consistent with previous observations of genetic overlap of other psychiatric disorders.[56] Given that BOR patients display specific clinical symptoms not observed in patients with other psychiatric disorders, knowledge of shared and non-shared genetic and clinical features will be important for the development of personalized treatment approaches.

CONFLICT OF INTEREST

JIN Jr is an investigator for Assurex and a consultant for Janssen. AT has received consultancy fees from Janssen and Novartis. The remaining authors declare no conflict of interest.

ACKNOWLEDGMENTS

We thank all patients and control subjects for their participation. We thank the KFO 256 workgroup of the CIMH, and Bipolar Disorder Working Group-, the Major Depression Working Group- and the Schizophrenia Working Group of the Psychiatric Genomics Consortium (PGC-BIP, PGC-MDD, PGC-SCZ) for providing access to the relevant data. The Romanian sample was funded by UEFISCDI, Romania, grant no. 89/2012 to MGS. The CoLaus|PsyCoLaus study was and is supported by research grants from GlaxoSmithKline, the Faculty of Biology and Medicine of Lausanne and the Swiss National Science Foundation (grants 3200B0–105993, 3200B0-118308, 33CSCO-122661, 33CS30-139468 and 33CS30-148401). The study was supported by the German Federal Ministry of Education and Research (BMBF) through the Integrated Network IntegraMent (Integrated Understanding of Causes and Mechanisms in Mental Disorders), under the auspices of the e:Med Programme (grant 01ZX1314A to MMN and SC; grant 01ZX1314G to MR). The study was supported by the German Research Foundation (DFG; grant FOR2107; RI908/11-1 to MR; WI3429/3-1 to SHW; NO246/10-1 to MMN; DA1151/5-1 to UD; KFO 256 BO 1487/12-1 to MB;.SFB 779 TP A08 to BHS).

REFERENCES

1 Tomko RL, Trull TJ, Wood PK, Sher KJ. Characteristics of borderline personality disorder in a community sample: comorbidity, treatment utilization, and general functioning. *J Pers Disord* 2014; **28**: 734–750.
2 Bohus M, Schmahl C. [Psychopathology and treatment of borderline personality disorder]. *Nervenarzt* 2007; **78**: 1069–1080.
3 Lieb K, Zanarini MC, Schmahl C, Linehan MM, Bohus M. Borderline personality disorder. *Lancet* 2004; **364**: 453–461.
4 Zanarini MC, Frankenburg FR, Reich DB, Fitzmaurice G, Weinberg I, Gunderson JG. The 10-year course of physically self-destructive acts reported by borderline patients and axis II comparison subjects. *Acta Psychiatr Scand* 2008; **117**: 177–184.
5 Tsanas A, Saunders KE, Bilderbeck AC, Palmius N, Osipov M, Clifford GD *et al.* Daily longitudinal self-monitoring of mood variability in bipolar disorder and borderline personality disorder. *J Affect Disord* 2016; **205**: 225–233.
6 Calati R, Gressier F, Balestri M, Serretti A. Genetic modulation of borderline personality disorder: systematic review and meta-analysis. *J Psychiatr Res* 2013; **47**: 1275–1287.
7 Lubke GH, Laurin C, Amin N, Hottenga JJ, Willemsen G, van Grootheest G *et al.* Genome-wide analyses of borderline personality features. *Mol Psychiatry* 2014; **19**: 923–929.
8 Gunderson JG, Zanarini MC, Choi-Kain LW, Mitchell KS, Jang KL, Hudson JI. Family study of borderline personality disorder and its sectors of psychopathology. *Arch Gen Psychiatry* 2011; **68**: 753–762.
9 Torgersen S, Lygren S, Oien PA, Skre I, Onstad S, Edvardsen J *et al.* A twin study of personality disorders. *Compr Psychiatry* 2000; **41**: 416–425.
10 Distel MA, Willemsen G, Ligthart L, Derom CA, Martin NG, Neale MC *et al.* Genetic covariance structure of the four main features of borderline personality disorder. *J Pers Disord* 2010; **24**: 427–444.
11 Kendler KS, Myers J, Reichborn-Kjennerud T. Borderline personality disorder traits and their relationship with dimensions of normative personality: a web-based cohort and twin study. *Acta Psychiatr Scand* 2011; **123**: 349–359.
12 Reichborn-Kjennerud T, Ystrom E, Neale MC, Aggen SH, Mazzeo SE, Knudsen GP *et al.* Structure of genetic and environmental risk factors for symptoms of DSM-IV borderline personality disorder. *JAMA Psychiatry* 2013; **70**: 1206–1214.
13 Fornaro M, Orsolini L, Marini S, De Berardis D, Perna G, Valchera A *et al.* The prevalence and predictors of bipolar and borderline personality disorders comorbidity: systematic review and meta-analysis. *J Affect Disord* 2016; **195**: 105–118.
14 Ghaemi SN, Dalley S, Catania C, Barroilhet S. Bipolar or borderline: a clinical overview. *Acta Psychiatr Scand* 2014; **130**: 99–108.
15 Ghaemi SN, Barroilhet S. Confusing borderline personality with severe bipolar illness. *Acta Psychiatr Scand* 2015; **132**: 281–282.
16 Loranger AW, Oldham JM, Tulis EH. Familial transmission of DSM-III borderline personality disorder. *Arch Gen Psychiatry* 1982; **39**: 795–799.
17 Pope HG Jr., Jonas JM, Hudson JI, Cohen BM, Gunderson JG. The validity of DSM-III borderline personality disorder. A phenomenologic, family history, treatment response, and long-term follow-up study. *Arch Gen Psychiatry* 1983; **40**: 23–30.
18 Distel MA, Trull TJ, Willemsen G, Vink JM, Derom CA, Lynskey M *et al.* The five-factor model of personality and borderline personality disorder: a genetic analysis of comorbidity. *Biol Psychiatry* 2009; **66**: 1131–1138.
19 Gale CR, Hagenaars SP, Davies G, Hill WD, Liewald DC, Cullen B *et al.* Pleiotropy between neuroticism and physical and mental health: findings from 108 038 men and women in UK Biobank. *Transl Psychiatry* 2016; **6**: e791.
20 Malouff JM, Thorsteinsson EB, Rooke SE, Schutte NS. Alcohol involvement and the Five-Factor model of personality: a meta-analysis. *J Drug Educ* 2007; **37**: 277–294.
21 Witt S, Dukal H, Hohmeyer C, Radosavljevic-Bjelic S, Schendel D, Frank J *et al.* Biobank of Psychiatric Diseases Mannheim - BioPsy. *Open J Bioresour* 2016; **3**: e2.
22 Loranger A, Sartorius N, Andreoli A, Berger P, Buchheim P, Chanabasavanna S. *German Version of the International Personality Disorder Examination: IPDE*. WHO: Geneva, Switzerland, 1998.
23 First MB, Spitzer RL, Robert L, Gibbon M, Williams JB. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version*, Patient Edition. Biometrics Research, New York State Psychiatric Institute: New York, NY, USA, 2002.
24 Consortium SWGotPG. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**: 421–427.
25 Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3 Genes Genomes Genetics* 2011; **1**: 457–470.
26 Delaneau O, Marchini J, Zagury J-F. A linear complexity phasing method for thousands of genomes. *Nat Methods* 2012; **9**: 179–181.
27 Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA *et al.* A map of human genome variation from population-scale sequencing. *Nature* 2010; **467**: 1061–1073.
28 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
29 de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 2015; **11**: e1004219.
30 Zhang K, Cui S, Chang S, Zhang L, Wang J. i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res* 2010; **38**(Web Server issue): W90–W95.
31 Nam D, Kim J, Kim SY, Kim S. GSA-SNP: a general approach for gene set analysis of polymorphisms. *Nucleic Acids Res* 2010; **38**(Web Server issue): W749–W754.
32 Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C*et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015; **47**: 291–295.
33 Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015; **47**: 1236–1241.
34 Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**: 421–427.
35 Abstracts of the XXIIIrd World Congress of Psychiatric Genetics (WCPG): final symposia and penary abstracts. *Eur Neuropsychopharmacol* 2015; 13 & 23.
36 International Schizophrenia C, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**: 748–752.
37 Edwards L, Gupta R, Filipp FV. Hypermutation of DPYD deregulates pyrimidine metabolism and promotes malignant progression. *Mol Cancer Res* 2016; **14**: 196–206.
38 Duan J, Shi J, Fiorentino A, Leites C, Chen X, Moy W *et al.* A rare functional noncoding variant at the GWAS-implicated MIR137/MIR2682 locus might confer risk to schizophrenia and bipolar disorder. *Am J Hum Genet* 2014; **95**: 744–753.
39 Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kahler AK, Akterin S *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet* 2013; **45**: 1150–1159.
40 Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* 2011; **43**: 969–976.
41 Xu B, Ionita-Laza I, Roos JL, Boone B, Woodrick S, Sun Y *et al. De novo* gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat Genet* 2012; **44**: 1365–1369.
42 Keil R, Schulz J, Hatzfeld M. p0071/PKP4, a multifunctional protein coordinating cell adhesion with cytoskeletal organization. *Biol Chem* 2013; **394**: 1005–1017.
43 Kao CF, Chen HW, Chen HC, Yang JH, Huang MC, Chiu YH *et al.* Identification of susceptible loci and enriched pathways for bipolar II disorder using genome-wide association studies. *Int J Neuropsychopharmacol* 2016; **19**: pyw064.
44 Zhang Z, Yu H, Jiang S, Liao J, Lu T, Wang L *et al.* Evidence for association of cell adhesion molecules pathway and NLGN1 polymorphisms with schizophrenia in Chinese Han population. *PLoS ONE* 2015; **10**: e0144719.
45 Network, Pathway Analysis Subgroup of Psychiatric Genomics C. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat Neurosci* 2015; **18**: 199–209.
46 Krueger WH, Gonye GE, Madison DL, Murray KE, Kumar M, Spoerel N *et al.* TPO1, a member of a novel protein family, is developmentally regulated in cultured oligodendrocytes. *J Neurochem* 1997; **69**: 1343–1355.
47 Liu J, Dietz K, DeLoyht JM, Pedre X, Kelkar D, Kaur J *et al.* Impaired adult myelination in the prefrontal cortex of socially isolated mice. *Nat Neurosci* 2012; **15**: 1621–1623.

48 Cupertino RB, Kappel DB, Bandeira CE, Schuch JB, da Silva BS, Muller D *et al.* SNARE complex in developmental psychiatry: neurotransmitter exocytosis and beyond. *J Neural Transm (Vienna)* 2016; **123**: 867–883.
49 Pescosolido MF, Gamsiz ED, Nagpal S, Morrow EM. Distribution of disease-associated copy number variants across distinct disorders of cognitive development. *J Am Acad Child Adolesc Psychiatry* 2013; **52**: 414–430.e14.
50 Zhao Z, Xu J, Chen J, Kim S, Reimers M, Bacanu SA *et al.* Transcriptome sequencing and genome-wide association analyses reveal lysosomal function and actin cytoskeleton remodeling in schizophrenia and bipolar disorder. *Mol Psychiatry* 2015; **20**: 563–572.
51 Sullivan PF, de Geus EJ, Willemsen G, James MR, Smit JH, Zandbelt T *et al.* Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Mol Psychiatry* 2009; **14**: 359–375.
52 Bosanac P, Castle DJ. Schizophrenia and depression. *Med J Australia* 2013; **199**(6 Suppl): S36–S39.
53 Buckley PF, Miller BJ, Lehrer DS, Castle DJ. Psychiatric comorbidities and schizophrenia. *Schizophr Bull* 2009; **35**: 383–402.
54 Anttila V, Bulik-Sullivan B, Finucane HK, Bras J, Duncan L, Escott-Price V *et al.* Analysis of shared heritability in common disorders of the brain. *bioRciv* 2016.
55 Nordsletten AE, Larsson H, Crowley JJ, Almqvist C, Lichtenstein P, Mataix-Cols D. Patterns of nonrandom mating within and across 11 major psychiatric disorders. *JAMA Psychiatry* 2016; **73**: 354–361.
56 Cross-Disorder Group of the Psychiatric Genomics C, Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 2013; **45**: 984–994.

Supplementary Information accompanies the paper on the Translational Psychiatry website (http://www.nature.com/tp)

[1]Central Institute of Mental Health, Department of Genetic Epidemiology in Psychiatry, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany; [2]Central Institute of Mental Health, Clinic of Psychosomatic and Psychotherapeutic Medicine, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany; [3]Central Institute of Mental Health, Institute for Psychiatric and Psychosomatic Psychotherapy (IPPP)/Psychosomatic Medicine and Psychotherapy, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany; [4]Department of Psychiatry and Psychotherapy, Charité Universitätsmedizin Berlin, Campus Mitte, Berlin, Germany; [5]Human Genomics Research Group, Department of Biomedicine, University of Basel, Basel, Switzerland; [6]Institute of Human Genetics, University of Bonn, Bonn, Germany; [7]Life and Brain Center, Department of Genomics, University of Bonn, Bonn, Germany; [8]Department of Psychiatry (UPK), University of Basel, Basel, Switzerland; [9]Department of Psychiatry, Charité-Universitätsmedizin Berlin, Campus Benjamin Franklin, Berlin, Germany; [10]Department of Clinical Psychology and Psychotherapy, University of Heidelberg, Heidelberg, Germany; [11]Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; [12]Department of Clinical Sciences, Psychiatry, Umeå University, Umeå, Sweden; [13]Discipline of Psychiatry, University of Adelaide, Adelaide, SA, Australia; [14]Molecular and Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI, USA; [15]Department of Psychiatry, Dalhousie University, Halifax, NS, Canada; [16]Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL, USA; [17]Division Mental Health and Addiction, Oslo University Hospital, Oslo, Norway; [18]NORMENT, University of Oslo, Oslo, Norway; [19]Institute of Pulmonology, Russian State Medical University, Moscow, Russian Federation; [20]Division of Psychiatry, University College London, London, UK; [21]Department of Psychiatry and Psychotherapy, University Hospital Carl Gustav Carus, Dresden, Germany; [22]Inserm, U1144, AP-HP, GH Saint-Louis, Département de Psychiatrie et de Médecine Addictologique, Paris, France; [23]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; [24]National Center for Mental Health, Cardiff University, Cardiff, UK; [25]Health Sciences Research, Mayo Clinic, Rochester, MN, USA; [26]Division of Psychiatry, University of Edinburgh, Edinburgh, UK; [27]Urain Center Rudolf Magnus, Department of Psychiatry, University Medical Center Utrecht, Utrecht, The Netherlands; [28]Department of Biomedicine, Aarhus University, Aarhus, Denmark; [29]iSEQ, Centre for Integrative Sequencing, Aarhus University, Aarhus, Denmark; [30]iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Aarhus, Denmark; [31]Medical and Quality Assurance, Clinics of Upper Bavaria, Munich, Germany; [32]Genetic Epidemiology Group, International Agency for Research on Cancer, Lyon, France; [33]Department of Psychiatry and Psychotherapy, University Medical Center Göttingen, Goettingen, Germany; [34]Medical Center of the University of Munich, Campus Innenstadt, Institute of Psychiatric Phenomics and Genomics (IPPG), Munich, Germany; [35]Translational Neuropsychiatry Unit, Department of Clinical Medicine, Aarhus University, Aarhus, Denmark; [36]Queensland Brain Institute, The University of Queensland, Brisbane, QLD, Australia; [37]Department of Psychiatry, McGill University, Montreal, QC, Canada; [38]Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, UK; [39]Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany; [40]Centre for Psychiatry, Queen Mary University of London, London, UK; [41]UCL Genetics Institute, University College London, London, UK; [42]Laboratory of Psychiatric Genetics, Department of Psychiatry, Poznan University of Medical Sciences, Poznan, Poland; [43]Department of Psychiatry, University of Marburg, Marburg, Germany; [44]Department of Psychiatry, University of Münste, Münster, Germany; [45]Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, UK; [46]Department of Medical Genetics, Oslo University Hospital Ullevål, Oslo, Norway; [47]NORMENT, KG Jebsen Centre for Psychosis Research, Department of Clinical Science, University of Bergen, Bergen, Norway; [48]Centre for Integrative Biology, Università degli Studi di Trento, Trento, Italy; [49]Indiana University School of Medicine, Department of Biochemistry and Molecular Biology, Indianapolis, IN, USA; [50]Indiana University School of Medicine, Department of Medical and Molecular Genetics, Indianapolis, IN, USA; [51]Faculté de Médecine, Université Paris Est, Créteil, France; [52]Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN, USA; [53]School of Medical Sciences, University of New South Wales, Sydney, NSW, Australia; [54]Neuroscience Research Australia, Sydney, NSW, Australia; [55]Department of Psychiatry, University of Halle, Halle, Germany; [56]Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia; [57]Department of Psychological Medicine, University of Worcester, Worcester, UK; [58]Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany; [59]School of Biomedical and Healthcare Sciences, Plymouth University Peninsula Schools of Medicine and Dentistry, Plymouth, UK; [60]Department of Psychiatry, University of California San Diego, La Jolla, CA, USA;

[61]Biometric Psychiatric Genetics Research Unit, Alexandru Obregia Clinical Psychiatric Hospital, Bucharest, Romania; [62]Mental Health Department, Biomedicine Institute, University Regional Hospital, Málaga, Spain; [63]Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK; [64]Department of Psychology, Eberhard Karls Universität Tübingen, Tubingen, Germany; [65]Inserm U955, Psychiatrie Translationnelle, Créteil, France; [66]University Medical Center Utrecht, Division of Neuroscience, Department of Psychiatry, Utrecht, The Netherlands; [67]Center of Psychiatry Weinsberg, Weinsberg, Germany; [68]Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt am Main, Frankfurt am Main, Germany; [69]Centre for Addiction and Mental Health, Toronto, ON, Canada; [70]Department of Psychiatry, University of Toronto, Toronto, ON, Canada; [71]Max Planck Institute of Psychiatry, Munich, Germany; [72]Center for Research in Environmental Epidemiology (CREAL), Barcelona, Spain; [73]Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, Sweden; [74]Clinic for Psychiatry and Psychotherapy, University Hospital Cologne, Cologne, Germany; [75]Inserm U955, Translational Psychiatry Laboratory, AP-HP, DHU PePSY, Department of Psychiatry, Université Paris Est, Créteil, France; [76]Janssen Research and Development, LLC, Neuroscience Therapeutic Area, Titusville, NJ, USA; [77]M. Sklodowska-Curie Cancer Center and Institute of Oncology, Cancer Epidemiology and Prevention, Warsaw, Poland; [78]School of Psychology, The University of Queensland, Brisbane, QLD, Australia; [79]Lindner Center of HOPE, Research Institute, Mason, OH, USA; [80]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK; [81]Genetic Cancer Susceptibility Group, International Agency for Research on Cancer, Lyon, France; [82]Division of Psychiatry, University College London, London, UK; [83]VU University Medical Center and GGZ inGeest, Department of Psychiatry, Amsterdam, The Netherlands; [84]School of Psychiatry, University of New South Wales, Sydney, NSW, Australia; [85]Black Dog Institute, Sydney, NSW, Australia; [86]Institute for Molecular Biology, University of Queensland, Brisbane, QLD, Australia; [87]Department of Neuroscience, Norwegian University of Science and Technology, Trondheim, Norway; [88]Department of Psychiatry, St Olavs University Hospital, Trondheim, Norway; [89]Risskov, Psychosis Research Unit, Aarhus University Hospital, Aarhus, Denmark; [90]iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Department of Clinical Medicine, Aarhus University, Aarhus, Denmark; [91]Research Center Juelich, Institute of Neuroscience and Medicine (INM-1), Juelich, Germany; [92]Division of Medical Genetics, University of Basel, Basel, Switzerland; [93]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany; [94]University of Liverpool, Liverpool, UK; [95]HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA; [96]Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN, USA; [97]MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK; [98]Center for Neurobehavioral Genetics, University of California Los Angeles, Los Angeles, CA, USA; [99]University Medical Center Utrecht, Division of Brain Research, Utrecht, The Netherlands; [100]Psychiatry Clinic, Clinical Center University of Sarajevo, Sarajevo, Bosnia-Herzegovina; [101]Pfizer Global Research and Development, Human Genetics and Computational Biomedicine, Groton, CT, USA; [102]Department of Psychiatry, University of Iowa, Iowa City, IA, USA; [103]Department of Psychiatry, Psychiatric University Hospital of Lausanne, Lausanne, Switzerland; [104]Department of Neurology and Neurosurgery, Faculty of Medicine, McGill University, Montreal, QC, Canada; [105]Montreal Neurological Institute and Hospital, Montreal, QC, Canada; [106]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, Baltimore, MD, USA; [107]NIMH Division of Intramural Research Programs, Human Genetics Branch, Bethesda, MD, USA; [108]Psychiatric Center Nordbaden, Wiesloch, Germany; [109]Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA; [110]School of Medicine and Dentistry, James Cook University, Townsville, QLD, Australia; [111]Broad Institute of MIT and Harvard, Medical and Population Genetics, Cambridge, MA, USA; [112]Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA; [113]Centre for Addiction and Mental Health, Molecular Neuropsychiatry and Development Laboratory, Toronto, ON, Canada; [114]Department of Anaesthesiology and Operative Intensive Care, University Hospital Mannheim, Medical Faculty Mannheim/Heidelberg University, Mannheim, Germany; [115]Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia; [116]Pfizer Global Research and Development, Computational Sciences Center of Emphasis, Cambridge, MA, USA; [117]AGAPLESION Elisabethenstift gGmbh, Department of Psychiatry, Psychosomatics and Psychotherapy, Darmstadt, Germany; [118]University Medical Center, Department of Psychiatry and Psychotherapy, Mainz, Germany; [119]Leibniz Institute for Neurobiology, Magdeburg, Germany; [120]Department of Biomedicine, University of Basel, Basel, Switzerland; [121]Stanley Center for Psychiatric Research and Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA and [122]Analytic and Translational Genetics Unit, Department of Medicine, Harvard Medical School, Massachusetts General Hospital, Boston, MA, USA

11.4 Publication 4: (Kaminski et al., 2018)

Publication	Epigenetic variance in dopamine D2 receptor: a marker of IQ malleability?
Jornal	Translational Psychiatry
Date	19/09/2017
Rank/ Total	37/281
Total Cites	3,895
Impact factor	4.730
Eigenfactor Score	0.018500

Journal Data Filtered By: **Selected JCR Year: 2016** Selected Editions: SCIE,SSCI
Selected Categories: **"PSYCHIATRY"** Selected Category
Scheme: WoS
Gesamtanzahl: 281 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	World Psychiatry	3,153	26.561	0.009520
1	World Psychiatry	3,153	26.561	0.009520
3	JAMA Psychiatry	6,112	15.307	0.035710
3	JAMA Psychiatry	6,112	15.307	0.035710
5	AMERICAN JOURNAL OF PSYCHIATRY	41,446	14.176	0.042640
5	AMERICAN JOURNAL OF PSYCHIATRY	41,446	14.176	0.042640
7	MOLECULAR PSYCHIATRY	17,452	13.204	0.049670
8	Lancet Psychiatry	1,636	11.588	0.007800
8	Lancet Psychiatry	1,636	11.588	0.007800
10	BIOLOGICAL PSYCHIATRY	41,859	11.412	0.067400
11	PSYCHOTHERAPY AND PSYCHOSOMATICS	3,245	8.964	0.005080
11	PSYCHOTHERAPY AND PSYCHOSOMATICS	3,245	8.964	0.005080
13	SCHIZOPHRENIA BULLETIN	15,093	7.575	0.030320
13	SCHIZOPHRENIA BULLETIN	15,093	7.575	0.030320
15	JOURNAL OF NEUROLOGY NEUROSURGERY AND PSYCHIATRY	27,955	7.349	0.034720
16	ACTA PSYCHIATRICA SCANDINAVICA	12,783	6.790	0.012800
16	ACTA PSYCHIATRICA SCANDINAVICA	12,783	6.790	0.012800
18	JOURNAL OF THE AMERICAN ACADEMY OF CHILD AND ADOLESCENT PSYCHIATRY	18,262	6.442	0.019860
18	JOURNAL OF THE AMERICAN ACADEMY OF CHILD AND ADOLESCENT PSYCHIATRY	18,262	6.442	0.019860
20	NEUROPSYCHOPHARMACOLOGY	23,920	6.403	0.046670
21	BRITISH JOURNAL OF PSYCHIATRY	23,445	6.347	0.024990
21	BRITISH JOURNAL OF PSYCHIATRY	23,445	6.347	0.024990
23	JOURNAL OF CHILD PSYCHOLOGY AND PSYCHIATRY	16,890	6.226	0.023820
23	JOURNAL OF CHILD PSYCHOLOGY AND PSYCHIATRY	16,890	6.226	0.023820
25	ADDICTION	17,772	5.789	0.030290
25	ADDICTION	17,772	5.789	0.030290
27	JOURNAL OF CLINICAL PSYCHIATRY	19,061	5.291	0.023050
27	JOURNAL OF CLINICAL PSYCHIATRY	19,061	5.291	0.023050
29	PSYCHOLOGICAL MEDICINE	21,253	5.230	0.038380
29	PSYCHOLOGICAL MEDICINE	21,253	5.230	0.038380
31	JOURNAL OF PSYCHIATRY & NEUROSCIENCE	2,759	5.165	0.004970

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
31	JOURNAL OF PSYCHIATRY & NEUROSCIENCE	2,759	5.165	0.004970
33	DEPRESSION AND ANXIETY	7,146	4.971	0.016970
33	DEPRESSION AND ANXIETY	7,146	4.971	0.016970
35	PSYCHONEUROENDOCRINOLOGY	14,409	4.788	0.028830
36	Neuropsychiatry	149	4.778	0.000740
37	Translational Psychiatry	3,895	4.730	0.018500
38	INTERNATIONAL JOURNAL OF NEUROPSYCHOPHARMACOLOGY	6,082	4.712	0.015310
39	BIPOLAR DISORDERS	5,323	4.531	0.009660
40	CNS DRUGS	4,211	4.394	0.008200
41	Epidemiology and Psychiatric Sciences	653	4.246	0.002840
41	Epidemiology and Psychiatric Sciences	653	4.246	0.002840
43	EUROPEAN NEUROPSYCHOPHARMACOLOGY	6,575	4.239	0.015920
44	PROGRESS IN NEURO-PSYCHOPHARMACOLOGY & BIOLOGICAL PSYCHIATRY	9,740	4.187	0.016310
45	JOURNAL OF PSYCHIATRIC RESEARCH	13,460	4.183	0.024570
45	JOURNAL OF PSYCHIATRIC RESEARCH	13,460	4.183	0.024570
47	JOURNAL OF PSYCHOPHARMACOLOGY	5,518	4.179	0.012020
48	Journal of Behavioral Addictions	459	4.134	0.001520
48	Journal of Behavioral Addictions	459	4.134	0.001520
50	AUSTRALIAN AND NEW ZEALAND JOURNAL OF PSYCHIATRY	5,976	4.036	0.008490
50	AUSTRALIAN AND NEW ZEALAND JOURNAL OF PSYCHIATRY	5,976	4.036	0.008490
52	CURRENT OPINION IN PSYCHIATRY	3,287	4.020	0.007070
52	CURRENT OPINION IN PSYCHIATRY	3,287	4.020	0.007070
54	SCHIZOPHRENIA RESEARCH	19,578	3.986	0.034860
54	SCHIZOPHRENIA RESEARCH	19,578	3.986	0.034860
56	PSYCHOSOMATIC MEDICINE	11,959	3.863	0.011140
56	PSYCHOSOMATIC MEDICINE	11,959	3.863	0.011140
58	Current Psychiatry Reports	3,005	3.746	0.008460
58	Current Psychiatry Reports	3,005	3.746	0.008460
60	WORLD JOURNAL OF BIOLOGICAL PSYCHIATRY	2,057	3.658	0.005250
61	CNS SPECTRUMS	2,114	3.589	0.002760
62	EUROPEAN ARCHIVES OF PSYCHIATRY AND CLINICAL NEUROSCIENCE	3,647	3.569	0.005490
63	INTERNATIONAL JOURNAL OF EATING DISORDERS	7,865	3.567	0.009590
63	INTERNATIONAL JOURNAL OF EATING DISORDERS	7,865	3.567	0.009590
65	Frontiers in Psychiatry	2,670	3.532	0.012130
65	Frontiers in Psychiatry	2,670	3.532	0.012130

Selected JCR Year: 2016; Selected Categories: "PSYCHIATRY"

Kaminski et al. *Translational Psychiatry* (2018)8:169
DOI 10.1038/s41398-018-0222-7

Translational Psychiatry

ARTICLE Open Access

Epigenetic variance in dopamine D2 receptor: a marker of IQ malleability?

Jakob A. Kaminski[1,2], Florian Schlagenhauf[1,3], Michael Rapp[1,4], Swapnil Awasthi[1,5], Barbara Ruggeri[6], Lorenz Deserno[1,3,7], Tobias Banaschewski[8], Arun L. W. Bokde[9], Uli Bromberg[10], Christian Büchel[10], Erin Burke Quinlan[11], Sylvane Desrivières[12], Herta Flor[12,13], Vincent Frouin[14], Hugh Garavan[15], Penny Gowland[16], Bernd Ittermann[17], Jean-Luc Martinot[18], Marie-Laure Paillère Martinot[19], Frauke Nees[8,12], Dimitri Papadopoulos Orfanos[14], Tomáš Paus[20], Luise Poustka[8,21], Michael N. Smolka[22], Juliane H. Fröhner[22], Henrik Walter[1], Robert Whelan[23], Stephan Ripke[1,5], Gunter Schumann[11] and Andreas Heinz[1], the IMAGEN consortium

Abstract

Genetic and environmental factors both contribute to cognitive test performance. A substantial increase in average intelligence test results in the second half of the previous century within one generation is unlikely to be explained by genetic changes. One possible explanation for the strong malleability of cognitive performance measure is that environmental factors modify gene expression via epigenetic mechanisms. Epigenetic factors may help to understand the recent observations of an association between dopamine-dependent encoding of reward prediction errors and cognitive capacity, which was modulated by adverse life events. The possible manifestation of malleable biomarkers contributing to variance in cognitive test performance, and thus possibly contributing to the "missing heritability" between estimates from twin studies and variance explained by genetic markers, is still unclear. Here we show in 1475 healthy adolescents from the IMaging and GENetics (IMAGEN) sample that general IQ (gIQ) is associated with (1) polygenic scores for intelligence, (2) epigenetic modification of *DRD2* gene, (3) gray matter density in striatum, and (4) functional striatal activation elicited by temporarily surprising reward-predicting cues. Comparing the relative importance for the prediction of gIQ in an overlapping subsample, our results demonstrate neurobiological correlates of the malleability of gIQ and point to equal importance of genetic variance, epigenetic modification of DRD2 receptor gene, as well as functional striatal activation, known to influence dopamine neurotransmission. Peripheral epigenetic markers are in need of confirmation in the central nervous system and should be tested in longitudinal settings specifically assessing individual and environmental factors that modify epigenetic structure.

Introduction

Genetic variance is known to explain a substantial part of variability in cognitive capacity[1–5]. The largest available study describes that polygenic scores (i.e., those common genetic variants that are most strongly associated with test performance in previous studies) explain up to 4.8%[4] of the variance of general intelligence quotient IQ (gIQ). A more recent larger but not yet peer reviewed study, shows up to 5.4% of variance explained[5]. On the other hand, environmental factors have a significant impact on general cognitive capacity, as indicated by the strong rise in average IQ performance following the decades after World War II[6,7]. According to Flynn et al.[7] the change ranged from 5 to 25 IQ points (eg. 0.3 to 1.7 standard deviation (SD)) within one generation. This change appears to be too strong to be explained by genetic changes. While various environmental factors

Correspondence: Jakob A. Kaminski (jakob.kaminski@charite.de)
[1]Department of Psychiatry and Psychotherapy, Campus Charité Mitte, Charité – Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany
[2]Berlin Institute of Health (BIH), Kapelle Ufer 2, 10117 Berlin, Germany
Full list of author information is available at the end of the article.

(e.g. changes in the educational system, overall stress experience, nutrition, etc.) might contribute to this so-called Flynn effect, those factors should act via changes in neurobiological systems relevant for cognition. Possible neurobiological factors that mediate this effect and link genotype with complex traits like cognition are (1) epigenetic markers including methylation count, (2) cortical architecture of the brain evaluated using magnetic resonance imaging (MRI), and (3) the functioning of the brain explored in vivo with functional MRI (fMRI). Those malleable markers might as well contribute to the "missing heritability" that is present between variance explained by accumulating single-nucleotide polymorphisms (SNPs; 4.8% based on polygenic scoring[5]), estimates of genomic similarities between individuals (~20% SNP heritability[5]), and based on heritability estimates from twin studies (50–70%[2,3]). Here we aim to explore individual variance in gIQ that can be accounted for by neurobiological markers of cognitive performance and describe the interplay of mechanisms, including epigenetic variance that may contribute to individual malleability in cognitive capacity.

Several lines of evidence suggest that gIQ is associated with the architecture of the brain measured as cortical volume and thickness[8] explaining up to 16% of the variance in right insula. Beyond cortical findings, the architecture[9] and volume of subcortical structures have been associated with cognitive capacity explaining between 2.4% in striatum[10] and up to 4.2% in caudate volume[11]. The importance of subcortical structures is further underpinned by the finding that training in reasoning alters resting state connectivity between subcortical and cortical brain areas, including striatum, parietal, and prefrontal areas[12]. This is highly plausible given the relevance of cortico-striatal networks implicated in executive function and goal-directed behavior[13–15]. In line with this, dopamine synthesis capacity in the ventral striatum has been associated with frontal cortical and striatal functional activation during goal-directed vs. habitual decision-making as well as IQ[16,17], in accordance with the well-known role of dopamine in cognition and decision-making[18–20]. A readily available proxy for dopaminergic neurotransmission is the well-known reward anticipation signal that can be measured with fMRI[21,22]. Dopaminergic neurotransmission is partly heritable, but also substantially modulated by environmental factors[23–27]. An emerging field that could potentially link the abovementioned environmental factors and dopaminergic neurotransmission is epigenetic modulation, which can help to explain individual malleability. Finding possible links between epigenetic changes, reward signaling, and cognitive capacity in adolescents might contribute further evidence for long-lasting neurobiological correlates of environmental effects, including stress exposure, as already observed in rodents (for a review see Meaney et al.[28]).

The aforementioned candidate markers for neurobiological underpinnings of cognitive capacity have been assessed before[29], however, their relative importance has not been tested in a cumulative fashion. Moreover, the interplay between genetic variance and possible neurobiological underpinnings of individual difference in cognitive capacity, including epigenetic markers is not known in detail. Data from the IMaging and GENetics (IMAGEN) consortium, offer a well-characterized sample to study these topics. With experts from a variety of fields, we aimed at contributing a broader insight into that research question.

Therefore, we measured associations between cognitive capacity (gIQ) and polygenic scores, epigenetic markers of the dopaminergic system, gray matter density in striatum, and striatal activation during reward processing, in a large sample of healthy adolescents, and we quantified their relative contribution to interindividual differences in IQ. We addressed the following research questions:

1. Do two different polygenic scores, which have previously been associated with cognitive capacity[4,30], replicate in our sample?
2. Are there epigenetic markers (i.e., methylation count) of the dopaminergic system that show associations with gIQ?
3. Can we replicate previous findings[9–11] of a correlation between gray matter density in bilateral striatum and gIQ?
4. Can the previously observed association between functional activation of the ventral striatum (BOLD-signal) and IQ[17,31] be replicated in a large sample of adolescents?

In a subset of individuals for whom we have complete data, we evaluated the relative contribution of each of the aforementioned predictors for gIQ, assessed possible interactions of genetic variance with our other predictors, and performed model comparison for combinations of predictors.

Materials and methods

Participants

We used a sample of 1475 adolescents (mean age = 14.43 years; SD = 0.45, 765 female participants) from the large multicenter imaging and genetics study (IMAGEN[32]) with available data from neuropsychological assessment, functional imaging, and genetic data. The study is intended to investigate the genetic and neurobiological basis of individual variability in psychological traits, and their relation to the development of frequent neuropsychiatric disorders. Recruiting took place at eight different sites (Germany, United Kingdom, France, and Ireland). We therefore included site as a covariate in all

analysis in order to account for variance introduced by center-specific variations. We excluded subjects with contraindication for MRI scans as well as serious medical conditions. Each local ethics committee approved the study. Subjects and their parents provided informed assent and consent, respectively.

Intelligence measure

In previous work we started out with a focus on the fluid and crystallized IQ and stress exposure[17,31]. An abundant body of work on cognitive capacity and neurobiological correlates is based on a general factor derived from principal component analysis (PCA)[29]. On the other hand there is considerable criticism of constructing general factors[33] regarding cognitive test performance and the authors have voiced similar concerns elsewhere[34]. PCA gains general information at the expense of specific information associated with the Wechsler's intelligence scale (WISC) IV subscales. Calculating a general factor based on PCA (for a review see Deary et al.[29]) does not necessarily invalidate the nature of the original scales; instead, dimensionality reduction allows for capturing variance that is common to a variety of subscales. In this study, we therefore performed PCA in order to derive a measure of general cognitive ability from WISC IV subtests, comprising matrix reasoning, block design, digitspan backward and forward, similarities, and vocabulary[35]. The first principal component (gIQ) explained a large proportion of the variance (variance explained = 0.49) and was used for further analyses as a marker for gIQ (see Table 1 in the supplement). For a more fine-grained view we explored WISC IV subscales associations with biological markers calculating a correlation matrix in an overlapping subsample (see supplementary Table 7).

Genetics

For building a polygenic score, we obtained summary statistics from two large genome-wide association studies: Benyamin et al.[30] report associations with childhood intelligence[30] on 17 989 individuals and 1 380 159 SNPs; Sniekers et al.[4] provide a meta-analysis and report associations of common variants with intelligence in a maximum of 78 308 adults and children and included 10 499 625 SNPs.

We performed linkage disequilibrium (LD) pruning and "clumped" the summary statistics, discarding variants within 500 kb of, and in $r^2 \geq 0.1$ with, another (more significant) marker. After pruning we had 70 568 LD-independent SNPs for the score by Benyamin et al.[30] and 86 330 for the score according to Sniekers et al.[4]. For both scores we performed risk profile scores (RPS) of our sample described for a range of *p*-value thresholds (5×10^{-8}, 1×10^{-6}, 1×10^{-4}, 0.001, 0.01, 0.05, 0.1, 0.2, 0.5, and 1.0), multiplying the logistic regression (i.e., the natural log of the odds ratio) of each variant by the imputation probability for the "risk" allele in each individual. The resulting values were summed over each individual, so that each individual had a whole-genome RPS for further analysis. We aimed at replicating the association of the polygenic score with gIQ in a sample of 1388 subjects with sufficient data quality.

Epigenetics

Global blood DNA methylation levels were assessed by hybridizing DNA samples to the Infinium Human Methylation 450 Bead Chip (Illumina: http://www.illumina.com/products/methylation_450_beadchip_kits.html), following the manufacturer's protocol. Unlike polygenic scores based on multiple SNPs, no epigenetic score exists for intelligence. With respect to epigenome-wide data, our sample size is far too small to find effects on an epigenome-wide association study (EWAS) level. Therefore, we focused on epigenetic markers potentially affecting dopamine-dependent neural encoding of reward anticipation in the striatum and planned Bonferroni correction for multiple testing.

The FDb.InfiniumMethylation.hg19 package for R (https://bioconductor.org/packages/release/data/annotation/manuals/FDb.InfiniumMethylation.hg19/man/FDb.InfiniumMethylation.hg19.pdf) was used to tag candidate gene to its nearest CG site. We extracted the start coordinate for our candidate probes from this package. This start coordinate was then used to go up and down 50 kb to create a region file. We assessed methylation count in CG site from the following genes involved in dopamine metabolism and neurotransmission: tyrosine hydroxylase (*TH*); DOPA decarboxylase (*DDC*); catechol-*O*-methyl transferase (*COMT*); dopamine transporter 1 (*SLC6A3*); dopamine receptor D1 (*DRD1*); and dopamine receptor D2 (DRD2) resulting in 24 CG sites. We focused on D1 and D2 receptors because they are the most abundant dopamine receptors in the brain with expression in regions relevant for motor, limbic, and neuroendocrine functioning[36]. D3, D4, and D5 mRNAs are one to two orders of magnitude lower than that of the D1 or D2[37]. We think that in addition to D1 and D2, D3, 4, or 5 only provide limited further insight for possible markers of gIQ. Nonetheless, for a more comprehensive view, we include an exploratory search for D3, D4, and D5 receptor gene and tested for association with gIQ in the Supplement. Epigenetic data with sufficient quality and corresponding data on gIQ was available for 817 subjects.

Magnetic resonance imaging

Structural MRI

Subjects were scanned in 3T-MRI-Scanners from different manufacturers (Bruker, General Electric, Philips and Siemens[32]). We controlled for variance accounted for

by scanning site using dummy coded variables[38]. We used high resolution T1-weighted three-dimensional magnetization prepared rapid gradient echo sequence based on the ADNI protocol (http://www.loni.ucla.edu/ADNI/Cores/index.shtml). Gray matter density was estimated, including age, gender, and total intracranial volume as covariates of no interest. Mean striatal gray matter density was extracted from anatomical masks using the WFU-Pick atlas[39] comprising bilateral striatum (http://fmri.wfubmc.edu/software/pickatlas) as an anatomic voxel mask with the individual Brain Atlases tool in SPM (IBASPM 71). Structural imaging data with sufficient quality and corresponding data on gIQ were available for 1401 subjects.

Task details

During fMRI subjects performed a modified version of the well-known monetary incentive delay (MID) task[22,40]. The MID assesses how quickly the subject can react to a reward-indicating vs. neutral cues and pull a trigger to hit a target (with the left or right index finger). The cue was followed by a variable anticipation interval. Then the subjects were asked to push a button with their left or right index finger in order to hit an appearing target. If the subject is able to hit the target, following a reward-indicating (but not a neutral) cue, he or she scores points (Fig. 1 in SI).

Functional MRI

Due to our a priori hypotheses of an association between intelligence measures with activation of the ventral striatum during reward anticipation in the MID task[22], we tested our research question in a region of interest (ROI). For this a literature-based mask[41] of the bilateral ventral striatum was used in order to test for effects of individual signal change. We extracted the mean beta-values from the main effect in the abovementioned volume of interest of the ventral striatum from contrast images estimating the BOLD-signal change during anticipation of big and small vs. no reward (Table 5 in SI). This signal is considered as an estimate of temporal difference errors elicited by temporarily surprising reward-predicting cues, which are related to phasic dopaminergic neurotransmission (Fig. 1 in SI)[22,42]. For further details concerning scanning parameters, preprocessing, and single subject statistics please refer to the Supplement.

Statistical analysis

We assumed that for a sufficient power of $1-\beta = 80\%$ ($\alpha = 5\%$) and a small effect sizes ranging from 2.4%[10] in previous structural imaging studies and 4.8%[4] for previous polygenic scores, according to Hulley et al.[43] we would need a total sample size of 161–324 subjects. To estimate the variance explained by two different polygenic scores in our sample, we were able to calculate linear regression models with gIQ as dependent variable and polygenic scores as predictors in $n = 1388$ subjects. Additional covariates included age, gender, and principal components from our prior PCA, which account for population stratification and tested two different polygenic scores, therefore we chose a significance level of $p = 0.025$.

For epigenetic markers, linear regression models were fitted for each marker in a combined sample of 817 subjects. Age, gender, and site as well as first two principle components of estimated differential cell counts and wave information were included into linear regression models as variables of no interest. We plotted a correlation matrix for all candidate markers in order to explore associations between methylation count in each CG site. Correlations between candidate CG sites revealed that most regions were independent markers (Fig. 1b). As candidate markers appeared to be rather independent, we decided to apply Bonferroni correction to rigorously correct our results for multiple comparisons resulting in a significance level of $p = 2 \times 10^{-3}$.

We applied linear regression to estimate the correlation between gIQ and bilateral gray matter density in striatum of 1401 subjects with sufficient imaging quality. We accounted for variance from the following variables of no interest: age, gender, site, and total brain volume.

To statistically evaluate associations between bilateral ventral striatal reward anticipation signal (BOLD-signal) and gIQ, we used multiple linear regression controlling for age, gender, and site in a sample of 1475 subjects. For imaging parameters, we used split-half cross-validation on two subsets. A significance level of $p = 0.05$ was chosen.

For explorative analysis of gIQ and whole-brain associations with BOLD-signal during reward anticipation, we computed linear regression models at each voxel, using ordinary least squares. Due to spatial auto-correlation we used whole-brain family-wise error correction ($p = 0.05$) applying random field theory as implemented in statistical parametric mapping software (SPM 8) in $n = 1475$ subjects.

For our best predictors, we estimated variance explained and obtained 95% bootstrapped confidence intervals from 1000 randomly drawn samples to evaluate reliability of our results.

For further analysis, we choose to partial out variance from variables of no interest by calculating separate regression models of our nuisance variables on our predictors. For the polygenic scores we regressed out variance accounted for by age, gender, and principal component from our prior PCA. For epigenetic markers, we accounted for age, gender, site, first two principle components of estimated cell count, and wave information. For structural MRI, we regressed out variance from age, gender, site, and total brain volume. For fMRI, we

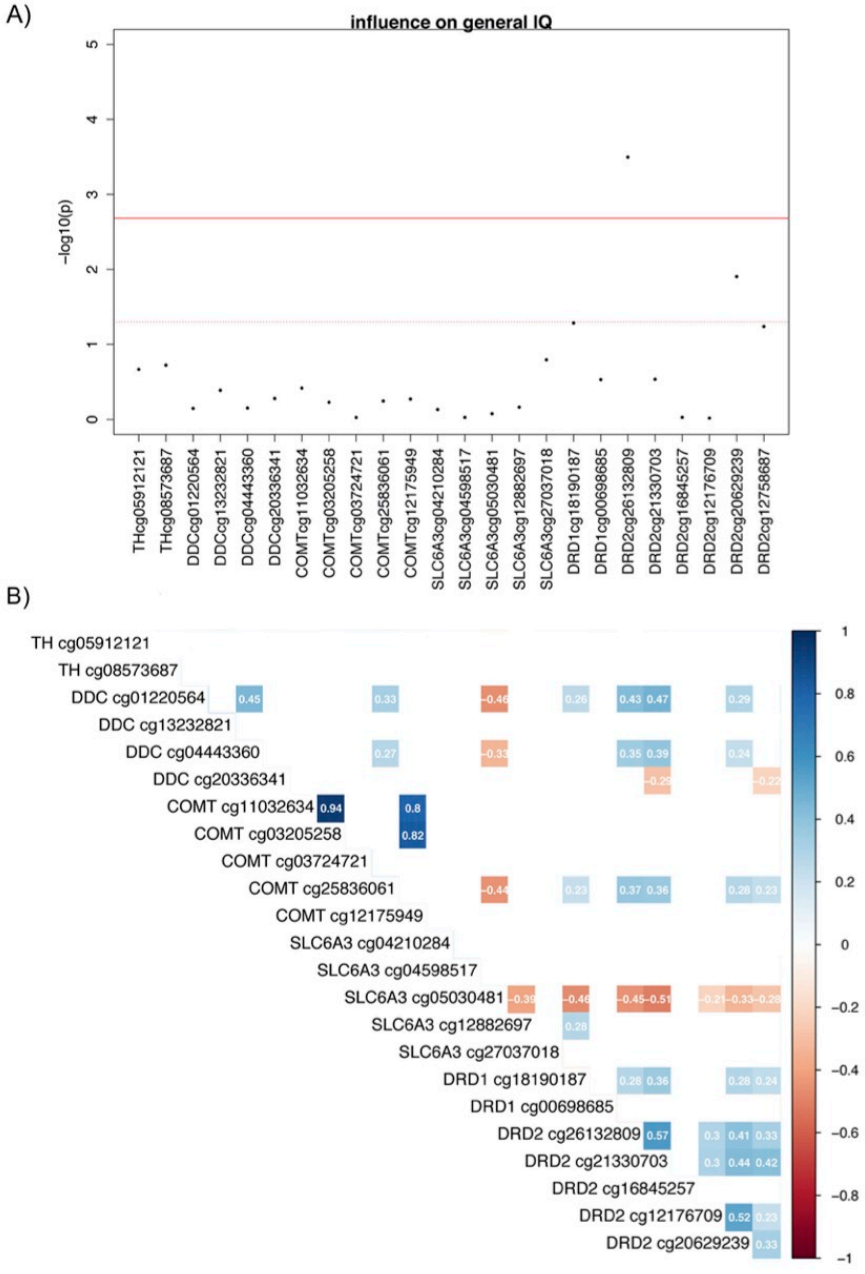

Fig. 1 (See legend on next page.)

Fig. 1 Association between methylation count in dopaminergic candidate markers and general IQ in $n = 817$ subjects. a Plot of negative decadic logarithm of *p*-values for association of methylation count in CG site 50 kb pairs up- and downstream from dopaminergic candidate markers. Candidate markers were tyrosine hydroxylase (*TH*), DOPA decarboxylase (*DDC*), catechol-*O*-methyl transferase (*COMT*), dopamine transporter 1 (*SLC6A3*), dopamine receptor 1 (*DRD1*), and dopamine receptor 2 (*DRD2*). Among 24 identified CG sites we found significant associations of epigenetic candidate markers for dopamine D2 receptor (cg26132809) involved in dopamine neurotransmission with general IQ correcting for age, gender, study site, wave information, and variability in cell type. The red line marks *p*-value threshold for multiple comparison correction for each CG site ($p < 2 \times 10^{-3}$) and the dashed line for $p < 0.05$. **b** Correlation matrix of epigenetic candidate markers involved in dopaminergic neurotransmission. Only correlation indices are displayed at a significance level of $p < 0.01$ (i.e., $R > 0.2$). Correlation coefficients are color-coded (positive correlation blue, negative correlation red)

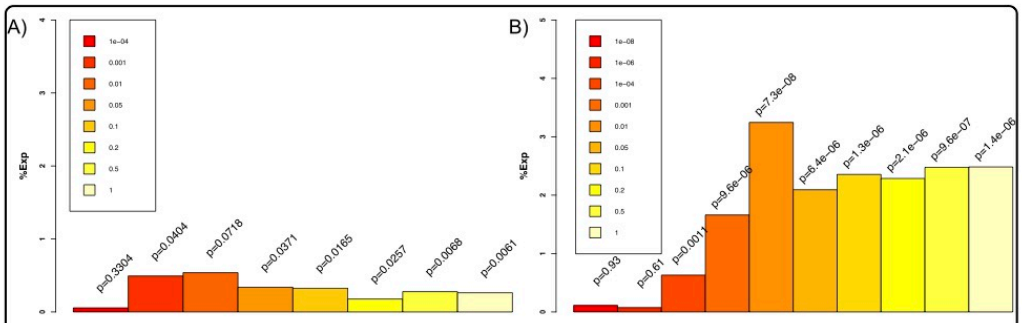

Fig. 2 Variance explained (%Exp) of two different polygenic scores predicting general IQ. Each bar represents variance explained for a given set of multi-SNP predictors at a given *p*-value threshold that is color-coded. The color code is described in the legend within the plot and represents *p*-value thresholds for inclusion of SNPs. On top of the bars *p*-values for association with gIQ for each polygenic score are reported. **a** The left panel shows the polygenic score derived from Benyamin et al.[30]. **b** The right panel shows the polygenic score derived from Sniekers et al.[4]

accounted for age, gender, and site. To explore possible interrelatedness between residuals of our variables, we calculated a correlation matrix.

We calculated one multiple linear regression model on residuals of our predictors in an overlapping subsample of 755 subjects with gIQ as independent variable and BOLD-signal, gray matter density, polygenic score, and epigenetic candidate marker as predictors. In order to estimate the effect size, we calculated standardized parameter estimates (beta) in a multiple regression model, which assumes standardized predictors and dependent variables (variance equals one). The standardized parameter estimate (beta) indicates how many SDs gIQ will change, per SD change in the predictor variable. We used the lavaan package[44] in combination with the SemPlot package[45] in R 3.2.4 for illustration purposes.

Although we did not primarily hypothesize interaction effects, we calculated interaction terms in order to explore the interplay between our variables. We focused on possible interactions of genetic effects on epigenetic, structural MRI and BOLD-signal resulting in three interaction terms (gene × epigenetic marker, gene × structural MRI, and gene × BOLD-signal). Correcting for multiple comparisons, we considered a significance level of $p = 0.017$ (i.e., $p = 0.05$ divided by the number of interaction terms).

Finally, we formally described and compared different combinations of our predictors using model comparison. With an exhaustive search for all combinations between our predictors we wanted to find the best model explaining gIQ. We choose to compare a set of all combinations, resulting in 15 models. We performed model comparison based on difference in Bayesian information criterion (BIC), which is known to penalize for models with larger numbers of parameters more strongly, resulting in a parsimonious model.

All probability values for the abovementioned tests are reported non-directional (two-tailed).

If not stated differently statistical tests were performed using R version 3.2.4

Results

Association between IQ and polygenic scores

With respect to our first research question regarding the influence of genetics on gIQ, we observed that the polygenic score by Benyamin et al.[30] at a *p*-threshold of 0.1 comprising 16 972 SNPs was significantly associated with gIQ (0.33% variance explained, degrees of freedom (df) = 1376, $p = 1.7 \times 10^{-2}$; Fig. 2a, and Table 2 in SI).

With respect to the score provided by Sniekers et al., we found the maximal proportion of variance explained with

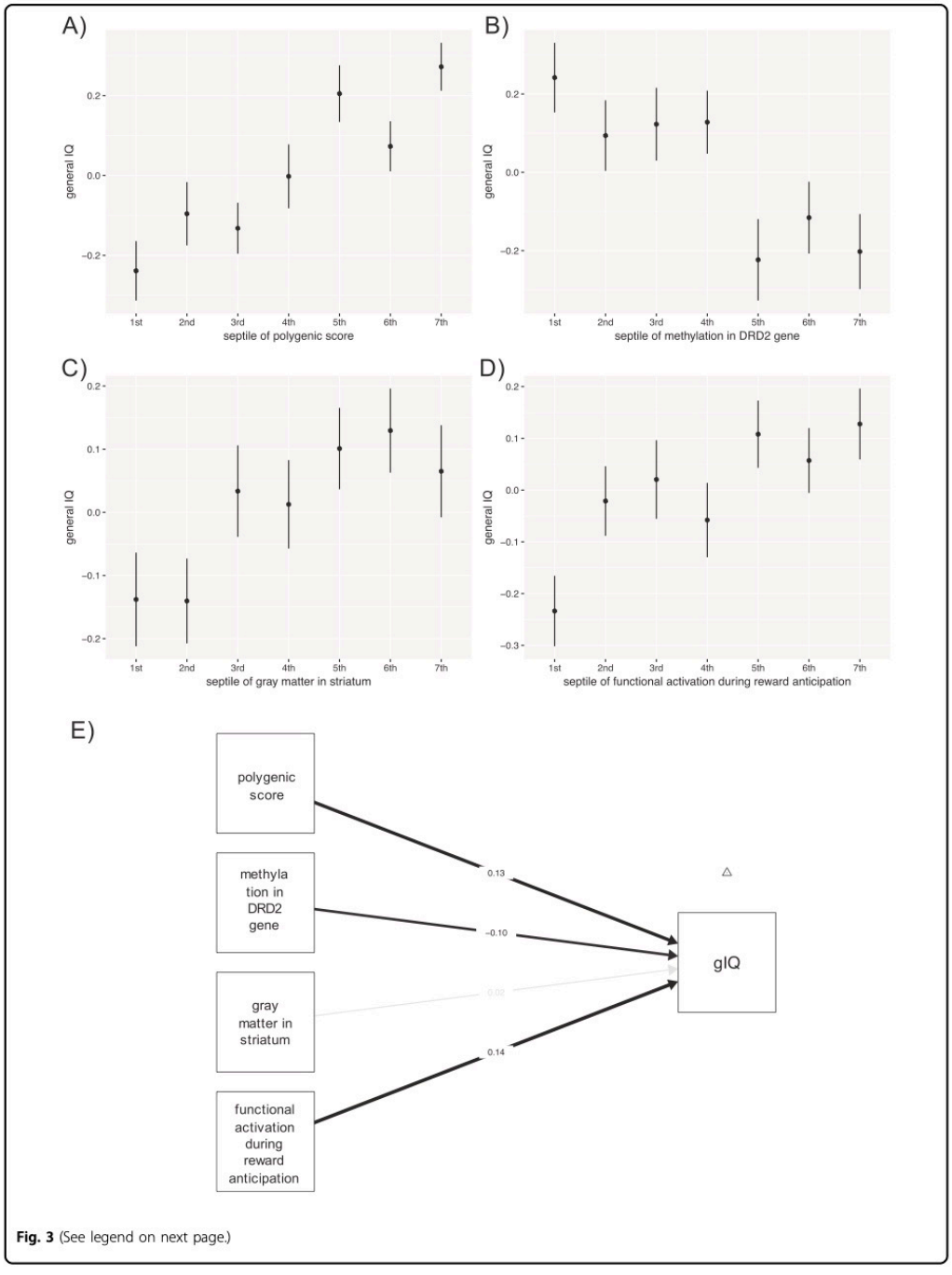

Fig. 3 (See legend on next page.)

Fig. 3 Candidate markers predicting general IQ. For display purpose, we grouped individuals into septiles of the candidate markers and plotted the mean phenotypic value (here general IQ) for each quantile on the *y*-axis[52]. Error bars indicate standard error of the mean. **a** General IQ can be predicted using polygenic score from Sniekers et al.[4] at a *p*-threshold of 0.01 comprising 5636 SNPs explaining 3.2% of variance (df = 1376; $p = 7.3 \times 10^{-8}$; correcting for age, gender, study site, principal components from imputation, and genetic strata). **b** Here we display association with the marker with the lowest *p*-value (methylation count in dopamine D2-receptor gene, *DRD2* cg26132809) among our candidate markers. We grouped individuals into septiles of their methylation level (higher septile rank indicating higher probability of methylation) and plotted those septiles against mean general IQ score on the *y*-axis. General IQ is negatively correlated with candidate marker for dopamine neurotransmission in our regression model (2.7% of variance explained, df = 803, $p = 3.18 \times 10^{-4}$ correcting for age, gender, study site, wave information, and variability in cell type) indicating that higher methylation count, which is considered as downregulation of transcription of DRD2 receptor, is related to lower IQ scores. **c** Gray matter density in bilateral striatum was used to group individuals into septiles. We plotted gray matter density against general IQ and found 0.71% variance explained (df = 1399, $p = 1.7 \times 10^{-3}$), correcting for age, gender, site, and total brain volume. **d** Here we plot general IQ by reward anticipation signal (BOLD-signal) in region of interest (ROI). We grouped individuals into septiles of beta parameter estimates (BOLD-signal) and plotted mean general IQ for each quantile on the *y*-axis for display purposes. General IQ is positively correlated with functional activation of the ventral striatum (1.4% of variance explained, df = 1463, $p = 4.11 \times 10^{-6}$; correcting for gender, age, and study site). **e** Regression model illustrating neurobiological correlates of general IQ in an overlapping sample of $n = 755$. A multiple linear regression model with general IQ (gIQ) as outcome variable was estimated with the residuals of the following predictors: polygenic score (from Sniekers et al.), methylation in *DRD2* gene, gray matter in striatum, and functional activation during reward anticipation. The whole model was significant with an adjusted $R^2 = 0.04$ (df = 750, $p = 3.3 \times 10^{-7}$). On the edges, we display the standardized parameter estimates for each predictor (beta) describing how many standard deviations the dependent variable (gIQ) will change, per standard deviation increase in the predictor variable. With respect to the different predictors, we could replicate previous findings that the established polygenic score (including 5636 SNPs significant at a *p*-threshold of 0.01) shows an association with general IQ (beta = 0.13, $p = 2.8 \times 10^{-4}$). We find variance in methylation count in our candidate CG site (*DRD2* cg26132809) that is negatively associated with general IQ (beta = −0.10, $p = 6.2 \times 10^{-3}$), indicating that higher methylation (lower gene activity) being associated with lower gIQ. In this subsample gray matter density in striatum was not associated with gIQ (beta = 0.02, $p = 0.5$). BOLD-signal change during reward anticipation significantly predicts cognitive capacity (beta = 0.14, $p = 9.4 \times 10^{-05}$)

3.2% ($p = 7.3 \times 10^{-8}$, df = 1376, bootstrapped confidence interval (CI): 1.64–5.43%) using a score comprising 5636 SNPs significant at a *p*-threshold of 0.01 (Figs. 2b and 3a, and Table 2 in SI).

Association between IQ and epigenetic markers of dopaminergic neurotransmission

Regarding our second research question, we found significant effects of methylation on gIQ in a CG site in the *DRD2* gene (cg26132809), which survived Bonferroni correction for multiple testing (2.7% variance explained, df = 803, $p = 3.18 \times 10^{-4}$, bootstrapped CI: 0.01–2.94%; Figs. 1a and 3b, and Table 3 in SI).

Association between IQ and gray matter density in striatum (MRI)

We found a positive correlation between gray matter density in bilateral striatum with gIQ (0.71% variance explained, df = 1399, $p = 1.7 \times 10^{-3}$, bootstrapped CI: 0.09–1.87%; Fig. 3c). Using split-half cross-validation, we could confirm this finding (d_0: 0.81%, df = 698, $p = 2.2 \times 10^{-2}$; d_1: 0.64%, df = 699, $p = 3.2 \times 10^{-2}$).

Association between IQ and ventral striatal BOLD-signal (fMRI)

In accordance with our third research question, we found that in our ROI, the ventral striatum, beta parameter estimates of reward anticipation (BOLD-signal) showed a significant positive association with gIQ (1.4% variance explained, df = 1463, $p = 4.1 \times 10^{-6}$, bootstrapped CI: 0.49–3.24%; Fig. 3d). This finding was confirmed using split-half cross-validation (d_0: 1.4%, df = 726, $p = 1.1 \times 10^{-3}$; d_1: 1.7%, df = 725, $p = 4.4 \times 10^{-4}$).

In an exploratory analysis we also observed that gIQ was positively correlated with functional activation during reward anticipation (BOLD-signal) in a large network outside of the ventral striatum as well in frontal and temporal regions (see Fig. 2 and Table 6 both in SI).

Influence of genetics, epigenetics, gray matter, and striatal activation on IQ

Calculating one regression model with residuals of our candidate markers, including the polygenic score (by Sniekers et al.), epigenetic finding (*DRD2* cg26132809), gray matter in striatum, and striatal activation, we observed that an increase in the polygenic score of one SD leads to a beta = 0.13 change in gIQ ($p = 2.8 \times 10^{-4}$; Fig. 3b). Increase in methylation count in candidate CG site (*DRD2* cg26132809) was associated with a decrease in gIQ (beta = −0.1, $p = 6.2 \times 10^{-3}$). There is a positive effect of BOLD-signal change during reward anticipation predicting gIQ (beta = 0.14, $p = 9.4 \times 10^{-5}$). In this additional analysis gray matter density in striatum showed no significant association with gIQ (beta = 0.02, $p = 0.5$; Fig. 3b). Calculating a correlation matrix, we found no significant association between the neurobiological predictors (Table 1). Exploring possible non-additive effects, we found no significant interaction between the polygenic scores (Snieker et al. and Benyamin et al.) and our candidate markers (epigenetics, gray matter, and striatal activation). We conducted an exhaustive search

Table 1 Correlation matrix for predictors of overlapping sample of $n = 755$

	gIQ	BOLD	Epigenetic	Polygenic score
BOLD	0.14****			
Epigenetic	−0.10**	0		
Polygenic score	0.13****	−0.03	−0.03	
Gray matter	0.03	0	−0.03	0.01

The correlation coefficients are based on linear regressions on residuals partialling out variance from variables of no interest
gIQ general IQ
BOLD, functional activation during reward anticipation; epigenetic, methylation in CG site *DRD2* cg26132809; polygenic score, polygenic score including 5636 SNPs significant at a *p*-threshold of 0.01 from Sniekers et al.; gray matter, gray matter density in striatum
Significant levels (two-tailed) ****$p < 0.0001$; ***$p < 0.001$; **$p < 0.01$, *$p < 0.05$

Table 2 Top six models of model comparison, among all possible combinations of 15 models

Model	ΔBIC	df
gIQ ~ polygenic score + epigenetic + BOLD	0	751
gIQ ~ polygenic score + BOLD	1.02	752
gIQ ~ polygenic score + epigenetic + BOLD + gray matter	6.16	750
gIQ ~ epigenetic + BOLD	6.73	752
gIQ ~ polygenic score + BOLD + gray matter	7.08	751
gIQ ~ BOLD	8.24	753

From top to bottom we display the models starting with the lowest Bayesian information criterion (BIC). We used the overlapping sample of $n = 755$ and residuals of our predictors (partialling out variance from variables of no interest)
df degrees of freedom, *gIQ* general IQ
ΔBIC, difference in Bayesian Information Criterion compared to the best model: ΔBIC = 0; BOLD, functional activation during reward anticipation in striatum; epigenetic, methylation in CG site of DRD2 gene cg26132809; polygenic score, polygenic score including 5636 SNPs significant at a *p*-threshold of 0.01 from Sniekers et al.; gray matter, gray matter density in striatum

for possible combinations of predictors and found the lowest BIC for a model comprising polygenic score by Sniekers et al., methylation count in the *DRD2* gene (cg26132809) and BOLD-signal during reward anticipation in the ventral striatum (Table 2). This model explained 4.81% of variance ($p = 1.05 \times 10^{-7}$, df = 751, bootstrapped CI: 2.22–9.04%). We calculated an additional correlation matrix in order to explore our candidate markers association with WISC IV subscales (see supplementary Table 7).

Discussion

Individual differences in intelligence have a substantial heritable background, while strong increases in test performance across the world in the last decades also point to strong environmental effects[7,46]. Using a polygenic score previously associated with cognitive capacity in children[30] and a novel score[4] tested in children and adults, we were able to replicate significant associations with gIQ, with the score based on 78 308 adults[4] performing better than the one based on a sample of 17 989 children[30]. The striking difference between heritability estimates derived from twin and adoption studies (around 50–70%[2,3]) and variance explained by common genetic polymorphisms (around 5%[5]) for many traits has been labeled "the case of the missing heritability". For example, regarding the partially heritable and polygenic human trait "body height", polygenic scores account for 10% of the variance[47], much below the high heritability estimates derived from monozygotic twin studies (heritability estimates around 85%[3]). So the discrepancy between variance explained by polygenic risk scores and twin studies may simply be due to the fact that polygenic risk scores only include common genetic polymorphisms and do not assess effects of rare gene variants. On the other hand, if epigenetic variation is transmitted to the offspring, as has been shown for some stress-related epigenetic effects[28], a more or less substantial part of the presumably genetic background regarding IQ test results may indeed be due to epigenetic factors (and hence environmental effects including social exclusion or discrimination stress).

Searching for neurobiological markers associated with dopaminergic neurotransmission in light of studies linking this system to cognitive capacity[11,17,18] we found significant associations between methylation of *DRD2* gene (cg26132809) and gIQ. Epigenetic control of gene expression is modulated by environmental factors such as stress exposure to the individual or in some cases parental generation[48]. Stress exposure and further environmental factors also strongly modulate dopaminergic neurotransmission, with relations to epigenetic modification unexplored. In line with previous findings[11] we found gIQ to be related to gray matter density in striatum. These observations suggest a striatal contribution to the malleability of cognitive capacity[49].

The association between ventral striatal activation and gIQ was found to be robust using split-half cross-validation as well as estimation of bootstrapped confidence intervals. In the MID task, temporarily unpredicted presentation of reward-associated stimuli elicit functional activation of the ventral striatum (BOLD-signal), which was previously associated with dopamine release measured indirectly by displacement of radio ligands of dopamine D2 receptors in this brain area[21]. Unlike in studies directly quantifying the size of the reward prediction error using computational modeling[31,50], in the MID task, the size of the temporal error in the prediction of reward-anticipatory cues cannot be individually computed. Although not limited to the ventral striatum, finding the strongest effect in this region suggests that dopamine-dependent encoding of reward-anticipatory

cues and prediction errors contribute to cognitive flexibility and rapid decision-making.

Calculating a regression model in a subsample of 755 subjects with fully available data for all predictors we found polygenic score of Sniekers et al.[4], epigenetic markers of the *DRD2* gene and the ventral striatal BOLD-signal were significantly associated with gIQ with a similar effect size. In combined assessment in this subsample, gray matter density in striatum did not show a significant effect. Exploring the interrelatedness of our candidate markers, we found no significant association, thus pointing to rather independent sources of variance for gIQ. The abovementioned polygenic scores (Sniekers et al. and Benyamin et al.) did not show a significant interaction effect with our candidate markers (epigenetics, gray matter and striatal activation) in our sample. In order to formally describe and compare possible models that predict gIQ, comparison of all possible combinations of our predictors resulted in a model comprising BOLD-signal during reward anticipation in the ventral striatum, methylation count in the *DRD2* gene and the polygenic score. Altogether, the winning model points to a rather independent contribution of variance of dopaminergic neurotransmission to variance in gIQ on the one hand and genetic differences on the other hand.

Limitations of our study include the sample size, which is rather large for neuroimaging studies but exceedingly small for explorative genetic and epigenetic studies. This is reflected in rather large CIs when applying bootstrap procedures in epigenetic markers. Furthermore, our sample size is too small for the detection of epigenome-wide markers. Therefore, the DNA methylation score was limited to CG sites in selected dopaminergic genes, with only a single CG site emerging as significant. It is highly likely that a more comprehensive DNA methylation analysis would have identified more epigenetic loci, which are associated with IQ score. Despite the relatively small sample size, we were able to replicate effects of polygenic score on gIQ derived from large samples[4,30]. Further limitations include that our epigenetic markers are assessed in peripheral blood. They may not reflect variance in brain tissue and have to be validated in studies with methods directly accessing tissue in the central nervous system[51]. The cross-sectional design of the study does not allow any statement concerning causality. Further studies with a longitudinal design in possible quasi experimental settings are warranted.

Taken together, our findings suggest that both functional activation of the reward system, epigenetic control of dopaminergic neurotransmission, and genetic markers contribute to gIQ. Of note, the effect sizes studied are small but in the same range as previous studies (2.4%[10] in previous structural imaging studies and up to 4.8%[4] for previous polygenic scores). Eventually, it is fundamental for the understanding of cognitive capacity that we find variable neurobiological correlates of gIQ. Variance of methylation count in our epigenetic candidate marker and individual differences in ventral striatal activation during reward anticipation seem to be independent predictors and do not show a relation with genetic correlates. Observing an association between epigenetic markers and neural signatures of gIQ should encourage further studies exploring mechanisms that mediate genetic and environmental effects on the neurobiological correlates of cognitive functions.

Acknowledgements
The IMAGEN study receives research funding from the European Community's Sixth Framework Programme (LSHM-CT-2007-037286).

Author details
[1]Department of Psychiatry and Psychotherapy, Campus Charité Mitte, Charité – Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany. [2]Berlin Institute of Health (BIH), Kapelle Ufer 2, 10117 Berlin, Germany. [3]Max-Planck-Institute for Human Cognitive and Brainsciences, Stephanstraße 1a, 04103 Leipzig, Germany. [4]Social and Preventive Medicine, University of Potsdam, Am Neuen Palais 10, 14469 Potsdam, Germany. [5]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [6]Centre for Neuroimaging Sciences, Institute of Psychiatry, Psychology & Neuroscience, King's Colleges, London, UK. [7]Department of Child and Adolescent Psychiatry, Psychotherapy and Psychosomatics, University of Leipzig, Liebigstrße 20a, 04103 Leipzig, Germany. [8]Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Square J5, 68159 Mannheim, Germany. [9]Discipline of Psychiatry, School of Medicine and Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland. [10]University Medical Centre Hamburg-Eppendorf, House W34, 3.OG, Martinistr. 52, 20246 Hamburg, Germany. [11]Medical Research Council—Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College, London, UK. [12]Department of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Square J5, Mannheim, Germany. [13]Department of Psychology, School of Social Sciences, University of Mannheim, 68131 Mannheim, Germany. [14]Neurospin, Commissariat à l'Energie Atomique, CEA-Saclay Center, Paris, France. [15]Departments of Psychiatry and Psychology, University of Vermont, Burlington, VT 05405, USA. [16]Sir Peter Mansfield Imaging Centre School of Physics and Astronomy, University of Nottingham, University Park, Nottingham, UK. [17]Physikalisch-Technische Bundesanstalt (PTB), Abbestr. 2-12, Berlin, Germany. [18]Institut National de la Santé et de la Recherche Médicale, INSERM Unit 1000 "Neuroimaging & Psychiatry", University Paris Sud, University Paris Descartes—Sorbonne Paris Cité, Maison de Solenn, Paris, France. [19]Institut National de la Santé et de la Recherche Médicale, INSERM Unit 1000 "Neuroimaging & Psychiatry", University Paris Sud, University Paris Descartes—Sorbonne Paris Cité; and AP-HP, Department of Adolescent Psychopathology and Medicine, Maison de Solenn, Cochin Hospital, Paris, France. [20]Rotman Research Institute, Baycrest and Departments of Psychology and Psychiatry, University of Toronto, Toronto, ON M6A 2E1, Canada. [21]Department of Child and Adolescent Psychiatry and Psychotherapy, Medical University of Vienna, Vienna, Austria. [22]Department of Psychiatry and Neuroimaging Center, Technische Universität Dresden, Dresden, Germany. [23]Department of Psychology, University College Dublin, Dublin, Ireland

Conflict of interest
The authors declare that they have no conflict of interest.

12) Curriculum vitae (Lebenslauf)

My curriculum vitae does not appear in the electronic version of my paper for reasons of data protection.

13) Publication list (Publikationsliste)

13.1 First or co-first author

1. Lam M, **Awasthi S**, Watson HJ, Goldstein J, Panagiotaropoulou G, Trubetskoy V, Karlsson R, Frei O, Fan CC, De Witte W, Mota NR, Mullins N, Brügger K, Lee SH, Wray NR, Skarabis N, Huang H, Neale B, Daly MJ, Mattheisen M, Walters R, Ripke S. RICOPILI: Rapid Imputation for COnsortias PIpeLIne. Bioinformatics. 2020 Feb 1;36(3):930-933. doi: 10.1093/bioinformatics/btz633. PMID: 31393554; PMCID: PMC7868045. [IF: 5.6]
2. Forstner AJ, **Awasthi S,** Wolf C, Maron E, Erhardt A, Czamara D, Eriksson E, Lavebratt C, Allgulander C, Friedrich N, Becker J, Hecker J, Rambau S, Conrad R, Geiser F, McMahon FJ, Moebus S, Hess T, Buerfent BC, Hoffmann P, Herms S, Heilmann-Heimbach S, Kockum I, Olsson T, Alfredsson L, Weber H, Alpers GW, Arolt V, Fehm L, Fydrich T, Gerlach AL, Hamm A, Kircher T, Pané-Farré CA, Pauli P, Rief W, Ströhle A, Plag J, Lang T, Wittchen HU, Mattheisen M, Meier S, Metspalu A, Domschke K, Reif A, Hovatta I, Lindefors N, Andersson E, Schalling M, Mbarek H, Milaneschi Y, de Geus EJC, Boomsma DI, Penninx BWJH, Thorgeirsson TE, Steinberg S, Stefansson K, Stefansson H, Müller-Myhsok B, Hansen TF, Børglum AD, Werge T, Mortensen PB, Nordentoft M, Hougaard DM, Hultman CM, Sullivan PF, Nöthen MM, Woldbye DPD, Mors O, Binder EB, Rück C, Ripke S, Deckert J, Schumacher J. Genome-wide association study of panic disorder reveals genetic overlap with neuroticism and depression. Mol Psychiatry. 2019 Nov 11. doi: 10.1038/s41380-019-0590-2. Epub ahead of print. PMID: 31712720. [IF: 12.38]

13.2 Co-authors

1. Nenadić I, Meller T, Schmitt S, Stein F, Brosch K, Mosebach J, Ettinger U, Grant P, Meinert S, Opel N, Lemke H, Fingas S, Förster K, Hahn T, Jansen A, Andlauer TFM, Forstner AJ, Heilmann-Heimbach S, Hall ASM, **Awasthi S**, Ripke S, Witt SH, Rietschel M, Müller-Myhsok B, Nöthen MM, Dannlowski U, Krug A, Streit F, Kircher T. Polygenic risk for schizophrenia and schizotypal traits in non-clinical subjects. Psychol Med. 2020 Aug 6:1-11. doi: 10.1017/S0033291720002822. Epub ahead of print. PMID: 32758327. [IF:5.64]
2. Lett TA, Vogel BO, Ripke S, Wackerhagen C, Erk S, Awasthi S, Trubetskoy V, Brandl EJ, Mohnke S, Veer IM, Nöthen MM, Rietschel M, Degenhardt F, Romanczuk-Seiferth N, Witt SH, Banaschewski T, Bokde ALW, Büchel C, Quinlan EB, Desrivières S, Flor H, Frouin V, Garavan H, Gowland P, Ittermann B, Martinot JL, Martinot MP, Nees F, Papadopoulos-Orfanos D, Paus T, Poustka L, Fröhner JH, Smolka MN, Whelan R, Schumann G, Tost H, Meyer-Lindenberg A,

Heinz A, Walter H; IMAGEN consortium. Cortical Surfaces Mediate the Relationship Between Polygenic Scores for Intelligence and General Intelligence. Cereb Cortex. 2020 Apr 14;30(4):2707-2718. doi: 10.1093/cercor/bhz270. PMID: 31828294; PMCID: PMC7175009. [IF:5.04]

3. Kogelman LJA, Esserlind AL, Francke Christensen A, Awasthi S, Ripke S, Ingason A, Davidsson OB, Erikstrup C, Hjalgrim H, Ullum H, Olesen J, Folkmann Hansen T; DBDS Genomic Consortium, The International Headache Genetics Consortium. Migraine polygenic risk score associates with efficacy of migraine-specific drugs. Neurol Genet. 2019 Oct 24;5(6):e364. doi: 10.1212/NXG.0000000000000364. PMID: 31872049; PMCID: PMC6878840. [IF: 3.5]
4. Spalthoff R, Degenhardt F, **Awasthi S**, Heilmann-Heimbach S, Besteher B, Gaser C, Ripke S, Nöthen MM, Nenadić I. Effects of a neurodevelopmental genes based polygenic risk score for schizophrenia and single gene variants on brain structure in non-clinical subjects: A preliminary report. Schizophr Res. 2019 Oct;212:225-228. doi: 10.1016/j.schres.2019.07.061. Epub 2019 Aug 6. PMID: 31395486.[IF: 4.56]
5. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, Mattheisen M, Wang Y, Coleman JRI, Gaspar HA, de Leeuw CA, Steinberg S, Pavlides JMW, Trzaskowski M, Byrne EM, Pers TH, Holmans PA, Richards AL, Abbott L, Agerbo E, Akil H, Albani D, Alliey-Rodriguez N, Als TD, Anjorin A, Antilla V, **Awasthi S,** Badner JA, Bækvad-Hansen M, Barchas JD, Bass N, Bauer M, Belliveau R, Bergen SE, Pedersen CB, Bøen E, Boks MP, Boocock J, Budde M, Bunney W, Burmeister M, Bybjerg-Grauholm J, Byerley W, Casas M, Cerrato F, Cervantes P, Chambert K, Charney AW, Chen D, Churchhouse C, Clarke TK, Coryell W, Craig DW, Cruceanu C, Curtis D, Czerski PM, Dale AM, de Jong S, Degenhardt F, Del-Favero J, DePaulo JR, Djurovic S, Dobbyn AL, Dumont A, Elvsåshagen T, Escott-Price V, Fan CC, Fischer SB, Flickinger M, Foroud TM, Forty L, Frank J, Fraser C, Freimer NB, Frisén L, Gade K, Gage D, Garnham J, Giambartolomei C, Pedersen MG, Goldstein J, Gordon SD, Gordon-Smith K, Green EK, Green MJ, Greenwood TA, Grove J, Guan W, Guzman-Parra J, Hamshere ML, Hautzinger M, Heilbronner U, Herms S, Hipolito M, Hoffmann P, Holland D, Huckins L, Jamain S, Johnson JS, Juréus A, Kandaswamy R, Karlsson R, Kennedy JL, Kittel-Schneider S, Knowles JA, Kogevinas M, Koller AC, Kupka R, Lavebratt C, Lawrence J, Lawson WB, Leber M, Lee PH, Levy SE, Li JZ, Liu C, Lucae S, Maaser A, MacIntyre DJ, Mahon PB, Maier W, Martinsson L, McCarroll S, McGuffin P, McInnis MG, McKay JD, Medeiros H, Medland SE, Meng F, Milani L, Montgomery GW, Morris DW, Mühleisen TW, Mullins N, Nguyen H, Nievergelt CM, Adolfsson AN, Nwulia EA, O'Donovan C, Loohuis LMO, Ori APS, Oruc L, Ösby U, Perlis RH, Perry A, Pfennig A, Potash JB, Purcell SM, Regeer EJ, Reif A, Reinbold CS, Rice JP, Rivas F,

Rivera M, Roussos P, Ruderfer DM, Ryu E, Sánchez-Mora C, Schatzberg AF, Scheftner WA, Schork NJ, Shannon Weickert C, Shehktman T, Shilling PD, Sigurdsson E, Slaney C, Smeland OB, Sobell JL, Søholm Hansen C, Spijker AT, St Clair D, Steffens M, Strauss JS, Streit F, Strohmaier J, Szelinger S, Thompson RC, Thorgeirsson TE, Treutlein J, Vedder H, Wang W, Watson SJ, Weickert TW, Witt SH, Xi S, Xu W, Young AH, Zandi P, Zhang P, Zöllner S; eQTLGen Consortium; BIOS Consortium, Adolfsson R, Agartz I, Alda M, Backlund L, Baune BT, Bellivier F, Berrettini WH, Biernacka JM, Blackwood DHR, Boehnke M, Børglum AD, Corvin A, Craddock N, Daly MJ, Dannlowski U, Esko T, Etain B, Frye M, Fullerton JM, Gershon ES, Gill M, Goes F, Grigoroiu-Serbanescu M, Hauser J, Hougaard DM, Hultman CM, Jones I, Jones LA, Kahn RS, Kirov G, Landén M, Leboyer M, Lewis CM, Li QS, Lissowska J, Martin NG, Mayoral F, McElroy SL, McIntosh AM, McMahon FJ, Melle I, Metspalu A, Mitchell PB, Morken G, Mors O, Mortensen PB, Müller-Myhsok B, Myers RM, Neale BM, Nimgaonkar V, Nordentoft M, Nöthen MM, O'Donovan MC, Oedegaard KJ, Owen MJ, Paciga SA, Pato C, Pato MT, Posthuma D, Ramos-Quiroga JA, Ribasés M, Rietschel M, Rouleau GA, Schalling M, Schofield PR, Schulze TG, Serretti A, Smoller JW, Stefansson H, Stefansson K, Stordal E, Sullivan PF, Turecki G, Vaaler AE, Vieta E, Vincent JB, Werge T, Nurnberger JI, Wray NR, Di Florio A, Edenberg HJ, Cichon S, Ophoff RA, Scott LJ, Andreassen OA, Kelsoe J, Sklar P; Bipolar Disorder Working Group of the Psychiatric Genomics Consortium. Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat Genet. 2019 May;51(5):793-803. doi: 10.1038/s41588-019-0397-8. Epub 2019 May 1. PMID: 31043756; PMCID: PMC6956732. [IF: 27.6]

6. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, Pallesen J, Agerbo E, Andreassen OA, Anney R, **Awashti S,** Belliveau R, Bettella F, Buxbaum JD, Bybjerg-Grauholm J, Bækvad-Hansen M, Cerrato F, Chambert K, Christensen JH, Churchhouse C, Dellenvall K, Demontis D, De Rubeis S, Devlin B, Djurovic S, Dumont AL, Goldstein JI, Hansen CS, Hauberg ME, Hollegaard MV, Hope S, Howrigan DP, Huang H, Hultman CM, Klei L, Maller J, Martin J, Martin AR, Moran JL, Nyegaard M, Nærland T, Palmer DS, Palotie A, Pedersen CB, Pedersen MG, dPoterba T, Poulsen JB, Pourcain BS, Qvist P, Rehnström K, Reichenberg A, Reichert J, Robinson EB, Roeder K, Roussos P, Saemundsen E, Sandin S, Satterstrom FK, Davey Smith G, Stefansson H, Steinberg S, Stevens CR, Sullivan PF, Turley P, Walters GB, Xu X; Autism Spectrum Disorder Working Group of the Psychiatric Genomics Consortium; BUPGEN; Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium; 23andMe Research Team, Stefansson K, Geschwind DH, Nordentoft M, Hougaard DM, Werge T, Mors O, Mortensen PB, Neale BM, Daly MJ, Børglum AD. Identification of common genetic risk variants

for autism spectrum disorder. Nat Genet. 2019 Mar;51(3):431-444. doi: 10.1038/s41588-019-0344-8. Epub 2019 Feb 25. PMID: 30804558; PMCID: PMC6454898. [IF:27.6]

7. Kaminski JA, Schlagenhauf F, Rapp M, **Awasthi S,** Ruggeri B, Deserno L, Banaschewski T, Bokde ALW, Bromberg U, Büchel C, Quinlan EB, Desrivières S, Flor H, Frouin V, Garavan H, Gowland P, Ittermann B, Martinot JL, Martinot MP, Nees F, Orfanos DP, Paus T, Poustka L, Smolka MN, Fröhner JH, Walter H, Whelan R, Ripke S, Schumann G, Heinz A; IMAGEN consortium. Epigenetic variance in dopamine D2 receptor: a marker of IQ malleability? Transl Psychiatry. 2018 Aug 30;8(1):169. doi: 10.1038/s41398-018-0222-7. PMID: 30166545; PMCID: PMC6117339. [IF: 5.18]
8. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, de Leeuw CA, Nagel M, **Awasthi S**, Barr PB, Coleman JRI, Grasby KL, Hammerschlag AR, Kaminski JA, Karlsson R, Krapohl E, Lam M, Nygaard M, Reynolds CA, Trampush JW, Young H, Zabaneh D, Hägg S, Hansell NK, Karlsson IK, Linnarsson S, Montgomery GW, Muñoz-Manchado AB, Quinlan EB, Schumann G, Skene NG, Webb BT, White T, Arking DE, Avramopoulos D, Bilder RM, Bitsios P, Burdick KE, Cannon TD, Chiba-Falek O, Christoforou A, Cirulli ET, Congdon E, Corvin A, Davies G, Deary IJ, DeRosse P, Dickinson D, Djurovic S, Donohoe G, Conley ED, Eriksson JG, Espeseth T, Freimer NA, Giakoumaki S, Giegling I, Gill M, Glahn DC, Hariri AR, Hatzimanolis A, Keller MC, Knowles E, Koltai D, Konte B, Lahti J, Le Hellard S, Lencz T, Liewald DC, London E, Lundervold AJ, Malhotra AK, Melle I, Morris D, Need AC, Ollier W, Palotie A, Payton A, Pendleton N, Poldrack RA, Räikkönen K, Reinvang I, Roussos P, Rujescu D, Sabb FW, Scult MA, Smeland OB, Smyrnis N, Starr JM, Steen VM, Stefanis NC, Straub RE, Sundet K, Tiemeier H, Voineskos AN, Weinberger DR, Widen E, Yu J, Abecasis G, Andreassen OA, Breen G, Christiansen L, Debrabant B, Dick DM, Heinz A, Hjerling-Leffler J, Ikram MA, Kendler KS, Martin NG, Medland SE, Pedersen NL, Plomin R, Polderman TJC, Ripke S, van der Sluis S, Sullivan PF, Vrieze SI, Wright MJ, Posthuma D. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nat Genet. 2018 Jul;50(7):912-919. doi: 10.1038/s41588-018-0152-6. Epub 2018 Jun 25. PMID: 29942086; PMCID: PMC6411041. [IF:27.6]
9. Nenadic, I., Gräger, S., **Awasthi, S.,** Langbein, K., Dietzek, M., Besteher, B., Ripke, S., Noethen, M. M., Sauer, H., Gaser, C., Degenhardt, F., F170. Schizophrenia Polygenic Risk Score Associated With Left Temporal Gyrification, Schizophrenia Bulletin, 2018 April 01, Volume 44, Issue suppl_1, 1, Page S287, doi: 10.1093/schbul/sby017.701. [IF: 7.57]
10. Witt SH, Streit F, Jungkunz M, Frank J, **Awasthi S**, Reinbold CS, Treutlein J, Degenhardt F, Forstner AJ, Heilmann-Heimbach S, Dietl L, Schwarze CE, Schendel D, Strohmaier J,

Abdellaoui A, Adolfsson R, Air TM, Akil H, Alda M, Alliey-Rodriguez N, Andreassen OA, Babadjanova G, Bass NJ, Bauer M, Baune BT, Bellivier F, Bergen S, Bethell A, Biernacka JM, Blackwood DHR, Boks MP, Boomsma DI, Børglum AD, Borrmann-Hassenbach M, Brennan P, Budde M, Buttenschøn HN, Byrne EM, Cervantes P, Clarke TK, Craddock N, Cruceanu C, Curtis D, Czerski PM, Dannlowski U, Davis T, de Geus EJC, Di Florio A, Djurovic S, Domenici E, Edenberg HJ, Etain B, Fischer SB, Forty L, Fraser C, Frye MA, Fullerton JM, Gade K, Gershon ES, Giegling I, Gordon SD, Gordon-Smith K, Grabe HJ, Green EK, Greenwood TA, Grigoroiu-Serbanescu M, Guzman-Parra J, Hall LS, Hamshere M, Hauser J, Hautzinger M, Heilbronner U, Herms S, Hitturlingappa S, Hoffmann P, Holmans P, Hottenga JJ, Jamain S, Jones I, Jones LA, Juréus A, Kahn RS, Kammerer-Ciernioch J, Kirov G, Kittel-Schneider S, Kloiber S, Knott SV, Kogevinas M, Landén M, Leber M, Leboyer M, Li QS, Lissowska J, Lucae S, Martin NG, Mayoral-Cleries F, McElroy SL, McIntosh AM, McKay JD, McQuillin A, Medland SE, Middeldorp CM, Milaneschi Y, Mitchell PB, Montgomery GW, Morken G, Mors O, Mühleisen TW, Müller-Myhsok B, Myers RM, Nievergelt CM, Nurnberger JI, O'Donovan MC, Loohuis LMO, Ophoff R, Oruc L, Owen MJ, Paciga SA, Penninx BWJH, Perry A, Pfennig A, Potash JB, Preisig M, Reif A, Rivas F, Rouleau GA, Schofield PR, Schulze TG, Schwarz M, Scott L, Sinnamon GCB, Stahl EA, Strauss J, Turecki G, Van der Auwera S, Vedder H, Vincent JB, Willemsen G, Witt CC, Wray NR, Xi HS; Bipolar Disorders Working Group of the Psychiatric Genomics Consortium; Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium; Schizophrenia Working Group of the Psychiatric Genomics Consortium, Tadic A, Dahmen N, Schott BH, Cichon S, Nöthen MM, Ripke S, Mobascher A, Rujescu D, Lieb K, Roepke S, Schmahl C, Bohus M, Rietschel M. Genome-wide association study of borderline personality disorder reveals genetic overlap with bipolar disorder, major depression and schizophrenia. Transl Psychiatry. 2017 Jun 20;7(6): e1155. doi: 10.1038/tp.2017.115. PMID: 28632202; PMCID: PMC5537640. [IF: 5.18]

11. Sharma, G., **Awasthi, S**., Dixit, A., Sharma, G., Patient safety risk assessment and risk management: A review on Indian hospitals. Chronicles of Young Scientists 2011; 2:186-91. [IF: 0.3]

14) Acknowledgement (Danke)

First, I would like to express my sincere gratitude to my first supervisor, Prof. Dr. Dr. med. Stephan Ripke, for allowing me to conduct the project under his supervision. I am highly thankful for his mentorship, constant motivation, and encouragement, without which this work would not have been possible. I also want to thank Prof. Dr. Henrik Walter for being my second supervisor and providing me with valuable suggestions regarding this project. I thank Prof. Dr. med. Dr. phil. Andreas Heinz as the director of the department which provided me with infrastructure and administrative support.

Further, I want to thank all the collaborators Andreas J. Forstner, Christiane Wolf, Johannes Schumacher, Max Lam, Jakob A Kaminski, Stephanie Witt, Fabian Streit.

A sincere thanks to all the patients and controls that have contributed their genotypic information for this project.

Special thanks to my colleague and friend Vassily Trubetskoy, for helping me with the editing of the dissertation. I also want to thank all of my colleagues Julia Kraft, Alice Braun, Georgia Panagiotaropoulou for the support and their advice.

Thanks to Stanley center at Broad Institute for financially supporting me. Moreover, thanks to the BROAD and LISA supercomputing facilities and IT department for providing me with adequate resources and timely support.

I want to dedicate this dissertation to my grandparents Mr. K.K Awasthi, Mr. Shrikant Dixit, Mrs. Shashi Awasthi and Mrs. Shushila Dixit and to my parents, Mr. A. P. Awasthi, Mrs. Alka Awasthi, Mr. Kislay Dixit, and Mrs. Madhu Dixit. I Heartily thanks to other family members and friends.

Finally, I want to thank my loving wife, Archana Dixit, who has continuously supported me in countless ways throughout this journey.