Integrative Genome-Wide Analysis of Genomic and Epigenomic Data in Neuropsychiatric Disorders

Doctoral thesis

to obtain a doctorate (PhD)

from the Faculty of Medicine

of the University of Bonn

Sugirtahn Sivalingam

from Manipay, Sri Lanka

2023

Written with authorization of

the Faculty of Medicine of the University of Bonn

First reviewer: Prof. Dr. med. Markus Nöthen

Second reviewer: Prof. Dr. David Ellinghaus

Day of oral examination: 18.01.2023

From the Institute of Human Genetics

Director: Prof. Dr. Markus Nöthen

Table of Contents

Dedicated to my Parents K.P. & L. Sivalingam

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

Die Dissertation ist bisher keiner anderen Fakultät vorgelegt worden.

Ich erkläre, dass ich bisher kein Promotionsverfahren erfolglos beendet habe und dass eine Aberkennung eines bereits erworbenen Doktorgrades nicht vorliegt.

Bonn, den 05.09.2022

Abstract

The inaugural Ph.D. dissertation is divided into two chapters. The first chapter consists of the integrative analysis of genetic and epigenetic data of individuals with familial and environmental risk for affective disorder. The second chapter is focused on the rare variant burden analysis of Niemann-Pick genes in SCZ (Schizophrenia) using smMIP-based targeted sequencing. First, the neurobiological correlates for the development of affective disorder development are largely unknown. There is increasing evidence that epigenetic modifications e.g. DNA methylation play an important role in the development of MDD (Major Depressive Disorder) and BD (Bipolar Disorder). We conducted two separate epigenome-wide methylation studies in whole-blood samples of female individuals with no reported history of psychiatric diseases to identify methylation sites associated with different risk factors for affective disorder: (i) familial risk (at least one 1st-degree relative with a history of affective disorder) or (ii) environmental risk (ranks above the threshold for a minimum of two forms of maltreatment in the CTQ (childhood trauma questionnaire)). Female individuals without any of these risk factors were chosen as controls. The data analysis pipeline included the following steps, probe filtering, functional normalization, and correction for leukocyte subpopulations. After rigorous quality control measurements, 495,406 DNA methylation sites were tested using a linear regression approach and correction for technical and biological covariates. Overall, 22,230 and 21,940 DNA methylation sites were nominally associated (p<0.05) with familial and environmental risk for affective disorder. Not any of the tested methylation sites reached epigenome-wide significance after correction for multiple testing. GO (Gene Ontology) analyses for familial and environmental risk displayed an enrichment of methylation sites in pathways related to neurogenesis and nervous system development. To identify the effects of genetic variants on the DNA methylation level, an integrated analysis of genome-wide genotype and epigenetic data was conducted. At this point, 45 independent single-nucleotide polymorphisms indicated a significant regulatory effect on DNA methylation. These methylation quantitative trait loci may help to understand the functional role of genetic variants in affective disorders. Further, analyses involving post-mortem brain samples, larger cohort sizes, and independent replication cohorts are needed to discover epigenetic mechanisms in affective disorder development. Second, SCZ is a severe neuropsychiatric

disorder with devastating health consequences for the affected individuals. Patients with SCZ display a broad range of symptoms highlighting that it is a clinically heterogeneous and complex mental disorder. The age at onset varies from early childhood to adolescence. The heritability was estimated to be around 60% in families and 80% in twins. Both genetic and environmental risk factors contribute to disease development. As many genes are implicated in SCZ from common variants with small effects to rare variants with large effects, there is currently no valid biomarker available to confirm or rule out the clinical diagnosis of SCZ. Due to the clinical heterogeneity and overlapping symptoms with other neurological and psychiatric disorders, making a correct clinical diagnosis can be challenging. Nieman-Pick type C (NP-C) disease belongs to a group of rare lysosomal storage disorders which are a group of heterogeneous inherited inborn errors of lipid metabolism. This disease is caused by mutations in either one of the genes *NCP1* or *NPC2.* It is a slowly progressing neurodegenerative disease where the clinical spectrum is remarkably heterogeneous and whose manifestations are age-dependent. In young adulthood neuropsychiatric symptoms like major depressive syndromes, sometimes bipolar disorder, or schizophrenia including psychosis start to manifest in the affected individuals. Mimicking psychiatric symptoms such as SCZ might lead to the misdiagnosis of NP-C patients. Correctly diagnosing NP-C is crucial for the affected individuals as NP-C-specific therapies are available. For that, NGS-based targeted sequencing of all coding exons and exon/intron boundaries of the *NPC1* and *NPC2* genes was applied. To test the hypothesis of whether rare functionally relevant *NPC1* and *NPC2* variants are enriched in SCZ patients compared to controls, a rare variant association test using SKAT-O (Optimal Sequence Kernel Association Test) was carried out. This type of test is advantageous as it maximizes the power by adaptively selecting the best linear combination of the burden and non-burden SKAT test. After stringent QC and filtering, 42 and 4 rare functionally relevant variants in *NPC1* and *NPC2* from 1,815 SCZ patients and 1,831 controls served as input for the SKAT-O. None of the tested genes either *NPC1* ($p_{\text{SKAT-O}}<0.929$) or *NPC2* ($p_{\text{SKAT-O}}<0.489$) were significant in the rare variant association test. A lookup in the currently largest meta-analysis of exome sequencing data in SCZ (SCHEMA) revealed similar non-significant findings for *NPC1* (p=0.153) and *NPC2* (p=0.206). Further, to access the effect of rare functionally relevant variants in *NPC1* and *NPC2*, a single marker association test was applied using Pearson’s chi-square test. None of the tested single variants were significant. Together these results suggested that the effects of rare variants in *NPC1* and *NPC2* have no major impact on the development of SCZ in the current study's cohort. However, the enrichment of heterozygous variants is rigorously discussed in the development of late-onset NP-C manifestations and as a potential risk factor for neurodegenerative diseases such as AD (Alzheimer's disease). Overall, the smMIPs assay enabled the screening of a large cohort of patients diagnosed with SCZ (NP-C). This might lead to a first step toward the implementation of a routine clinical diagnostics pipeline for the detection of rare pathogenic variants in *NPC1* and *NPC2* until NGS-based methods such as WES (whole-exome) or WGS (whole-genome) sequencing become more feasible than in the past.

List of Abbreviations

ADCY9	Adenylate Cyclase 9
ANK3	Ankyrin 3
APA	American Psychiatric Association
ASMD	Acid sphingomyelinase deficiency
BD	Bipolar Disorder
BDNF	Brain Derived Neurotrophic Factor
BRCA	Breast And Ovarian Cancer Susceptibility Protein 1
BWA	Burrow-Wheeler Alignment
CACNA1C	Calcium Voltage-Gated Channel Subunit Alpha1 C
CACNB2	Calcium Voltage-Gated Channel Auxiliary Subunit Beta 2
CADD	Combined Annotation Dependent Depletion
CGI	CpG Island
CHR	Chromosome
CLCN3	Chloride Voltage-Gated Channel 3
CNTN4	Contactin 4
CNV	Copy Number Variation
CpG	Cytosine-Guanine Dinucleotide
CTQ	Child Trauma Questionnaire
dbSNP	Single Nucleotide Polymorphism Database
DNA	Deoxyribonucleic Acid
DRD2	Dopamine Receptor D2
DSM	Diagnostic and Statistical Manual of Mental Disorders
EWAS	Epigenome-wide Association Study
FDR	False Discovery Rate
FKBP5	FKBP Prolyl Isomerase 5
FOR2107	DFG Research Group 2107
FOXN3	Forkhead Box N3
GABRP	Gamma-Aminobutyric Acid Receptor Subunit Pi
GATK	Genome Analysis Toolkit
GO	Gene Ontology

GRIA1	Glutamate Ionotropic Receptor AMPA Type Subunit 1
GRIN2A	Glutamate Ionotropic Receptor NMDA Type Subunit 2A
GWAS	Genome-wide Association Study
HeiDE	Heidelberg *Cohort* Study of the Elderly
HEMGN	Hemogen
ICD	International Classification of Diseases
KCTD13	Potassium Channel Tetramerization Domain Containing 13
LD	Linkage Disequilibrium
LIBSVM	Library for Support Vector Machines
MAF	Minor Allele Frequency
MDD	Major Depressive Disorder
MDS	Multidimensional Scaling
MHC	Major Histocompatibility Complex
mQTL	Methylation Quantitative Trait Loci
NGS	Next-Generation Sequencing
NP-B	Type B Niemann-Pick Disease
NP-C	Type C Niemann-Pick Disease
NPC1	Niemann-Pick Intracellular Cholesterol Transporter 1
NPC2	Niemann-Pick Intracellular Cholesterol Transporter 2
NR3C1	Nuclear Receptor Subfamily 3 Group C Member 1
NRXN1	Neurexin 1
PAK6	P21 (RAC1) Activated Kinase 6
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PGC	Psychiatric Genomics Consortium
QQ Plot	Quantile-Quantile Plot
RIMS1	Regulating Synaptic Membrane Exocytosis 1
SCHEMA	Schizophrenia Exome Sequencing Meta-analysis
SCZ	Schizophrenia
SD	Standard Deviation
SETD1A	SET Domain Containing 1A, Histone Lysine Methyltransferase
SKAT	SNP-set (Sequence) Kernel Association Test

SKAT-O	Optimized SNP-set (Sequence) Kernel Association Test
SLC6A4	Solute Carrier Family 6 Member 4
smMIP	Single-molecule molecular inversion probes
SMPD1	Sphingomyelin Phosphodiesterase 1
SNP	Single-Nucleotide Polymorphism
SRR	Serine Racemase
SVD	Singular Value Decomposition
TMEM110	Transmembrane Protein 110
TSS	Transcription Start Site
UTR	Untranslated Region
WES	Whole-Exome Sequencing
WHO	World Health Organization
ZNF197	Zinc Finger Protein 197

CHAPTER 1

EPIGENOME-WIDE METHYLATION IN FEMALE INDIVIDUALS WITH FAMILIAL OR ENVIRONMENTAL RISK FOR AFFECTIVE DISORDER

In this chapter, the findings of the integrative analyses of genetic and epigenetic data in female individuals with different risk factors for affective disorder are presented.

1.1 Introduction

1.1.1 Affective Disorders

Affective disorders including bipolar disorder (BD) and major depressive disorder (MDD) are highly complex and clinically heterogeneous neuropsychiatric disorders [Lohoff et al., 2010; Craddock et al., 2012]. They belong to one of the leading causes of disabilities worldwide with a high economic impact on the global health system [WHO, 2017]. The etiology of BD and MDD involves both genetic and environmental risk factors [Klengel et al., 2013; Aldinger et al., 2017]. The lifetime prevalence among the general population range between 1-2% for BD and 15% for MDD [Merikangas et al., 2007; Kessler et al., 2013]. Up to 5% of people experiencing one of these disorders are likely to commit suicide indicating the fatal consequences of these disorders [Isometsä et al., 2014]. MDD is characterized by the recurrence of depressive episodes which include depressed mood, decreased energy, and lack of interest and joy [McIntoshet al., 2019]. Other characteristics of MDD consist of cognitive symptoms like reduced concentration and behavioral symptoms like suicidal thoughts [McIntoshet al., 2019; Power et al., 2017]. The minimum duration of symptoms should last at least 2 weeks according to the structured diagnostic criteria [American Psychiatric Association, 2013]. Bipolar disorder is defined by the occurrence of manic and hypomanic phases which alternate with depressive periods [Phillips et al., 2013]. The manic phase lasts at least one week and is defined by high energy, increased self-esteem, racing thoughts, and irritable mood [Severus et al, 2013]. The depressive phase includes low energy, sad mood, suicidal thoughts, and psychomotor disturbances [Severus et al, 2013]. The symptoms observed in the depressive periods of bipolar disorder are similar to those observed in major depressive disorder. Two major subtypes of BD are distinguished: i.) Bipolar disorder type I (BD1) involves a severe history of recurring elevation of mania and depression; ii.) Bipolar disorder type II (BD2) consists of milder episodes of mania and depression with at least one-lifetime occurrence [American Psychiatric Association, 2013]. For the diagnosis, two main classification systems called the DSM-V (Diagnostic and Statistical Manual of Mental Disorders) published by the American Psychiatric Association (APA) and the ICD-11 (International Classification of Diseases) manual from the WHO are applied [American Psychiatric Association, 2013; WHO, 2019].

1.1.2 Genetic Factors

Family and twin studies revealed a strong genetic contribution to the etiology of affective disorders. Family studies indicated a five to ten times greater risk of developing an affective disorder among first-degree relatives of BD patients [Song et al., 2015]. However, complex disorders such as BD or MDD are multifactorial and a polygenic contribution of common and rare variants leads to disease susceptibility [Sul et al., 2020, Yu et al., 2018]. Recently, genome-wide association studies (GWAS) have identified susceptibility genes implicated in the disease etiology and generated initial insights into the genetic architecture of BD and MDD [Wray et al., 2018; Stahl et al., 2019; Howard et al., 2019; Mullins et al., 2021]. In GWAS the contribution of common SNPs (single-nucleotide polymorphisms) is assessed in case-control studies. The allele frequency of each SNP in individuals from a given population with and without a given disease trait is compared by statistical testing [Visscher et al., 2017]. The heritability of BD was estimated to be 40% and MDD 80%, respectively [Sullivan et al., 2000; Craddock et al., 2006]. The genetic correlation between bipolar disorder and major depressive disorder was estimated to be 35% using common SNPs [Brainstorm Consortium et al., 2018]. This is in line with other empirical studies that showed that BD and MDDD have not only overlapping clinical symptoms but also shared genetic etiology [Schulze et al., 2014]. However, due to the greater heterogeneity of MDD, more molecular genetic factors (n>100) were identified through GWAS to date than in BD. At the time of writing two of the largest GWAS of BD and MDD were conducted by the Psychiatric Genomics Consortium (PGC). The largest GWAS of BD comprised 41,917 cases and 371,549 controls in which 64 independent genome-wide significant loci were identified from which 33 were novel [Mullins et al., 2021]. BD associations were mainly enriched in genes in synaptic signaling pathways and brain-expressed genes with high specificity of expression in neurons of the prefrontal cortex and hippocampus [Mullins et al., 2021]. In a GWAS of MDD conducted in 2019, a meta-analysis based on 246,363 cases and 561,190 controls, 87 of 102 independent loci were genome-wide significant after correction for multiple testing. These were mainly associated with pathways related to synaptic structure and neurotransmission [Howard et al., 2019]. Although GWAS has identified successfully candidate genes for BD and MDD, a large proportion of the heritability remains unexplained suggesting that rare variants, copy-number variations, and non-genetic or environmental factors influence the risk for affective disorders [Priebe et al., 2012; Goes et al., 2021]. The largest whole-exome study of BD conducted in 2022 comprising 13,933 patients with BD and 14,422 controls revealed a significant contribution of ultra-rare protein-truncating variants (PTVs) to BD risk [Palmer et al., 2022]. In this study overall, 68 candidate gene sets were investigated, among the significantly enriched genes for ultra-rare PTVs were *CHD8* binding targets in the human brain (2,517 genes, OR = 1.09, $P = 5.18 \times 10^{-5}$) and genes from the SCHEMA browser (34 genes, OR = 1.89, $P = 4.81 \times 10^{-5}$). These results highlighted the importance of assessing rare variants for a better understanding of the genetic architecture of BD.

1.1.3 Environmental Factors

A substantial amount of studies identified environmental factors which play an important role in the etiology of affective disorders. Environmental factors can be grouped into three main categories that are neurodevelopmental (e.g. maternal infection during pregnancy or indicators of fetal development), physical or psychological stress (e.g. brain injuries or abuses), and substances (e.g. cannabis or cocaine) [Marangoni et al., 2016]. Most of the attention is focused on early life adversities like childhood maltreatment including physical, emotional, and sexual abuse, stressful life events, socioeconomic status, and substance abuse [Johnson et al., 1995, Barichello et al., 2016; Bortolato et al., 2017, Marangoni et al, 2016]. Childhood maltreatment is a well-studied factor for which there is clear evidence that it increases the risk to develop affective disorders such as bipolar disorder in later life [Bortolato et al., 2017; Schmitt et al., 2014]. However, the neurobiological mechanisms and pathophysiology by which these risk factors influence disease development are hardly understood and remain largely unknown [Forstner et al., 2018]. Several studies revealed that environmental factors partly mediate gene expression through epigenetic modifications such as DNA methylation, histone modifications, and non-coding RNAs [Weder et al., 2014; Cruceanu et al., 2013; Fan et al., 2014]. Such modifications do not alter the genetic code but regulate gene expression in response to environmental factors in a time and cell-type-specific manner. DNA methylation is, in particular, one of the best-studied epigenetic mechanisms which plays a crucial role in normal development, genomic imprinting, X-chromosome inactivation, aging, and carcinogenesis [Smith et al., 2013; Robertsen, 2005; Klutstein et al., 2016]. Increasing evidence suggests that epigenetic modifications such as DNA methylation have important implications for the development of neuropsychiatric disorders including bipolar disorder and major depressive disorder [Januar et al., 2015].

1.1.4 DNA Methylation

DNA methylation occurs at a CpG dinucleotide content where a methyl group is added at the C5 position of a cytosine molecule which is called 5-methylcytosine (5mC). This form of DNA methylation is the most common in the mammalian genome where 70% to 80% of CpG dinucleotides are methylated [Jabbari et al., 1986]. Next to this, DNA methylation can occur at the C4 position in cytosine (4mC) and the C6 position in adenine (6mA) [Ehrlich et al., 1985; Dunn et al., 1958]. Overall, there are more than 28 million CpG sites across the human genome, and these are located in around 30,000 CpG islands (CGI) which are dense areas of CpG dinucleotides [Venter et al., 2001; Lander et al., 2001]. Approximately 60% to 70% of human genes have CpG islands associated with their corresponding promoter regions [Saxonov et al., 2006; Illingworth et al., 2010]. Transcriptional repression or gene silencing occurs through high levels of 5mC methylation in CGIs of promoter regions while methylation in the gene body leads to transcriptional activation or alternative splicing [Bird, 2002; Jones, 2012]. Interestingly, the function of DNA methylation varies with different genomic contexts such as the transcription start site, gene body, regulatory elements, and repeat sequences. Furthermore, DNA methylation of regulatory elements such as enhancers is recognized to be functionally important [Jones et al., 2012]. Another important aspect of DNA methylation is that it stabilizes the genome by suppressing the expression of transposable elements and repeat regions such as centromeres [Moarefi et al., 2011]. Interestingly, DNA methylation is a reversible process in which demethylation occurs during early embryonic cell development orchestrated by methyltransferases [Li et al., 1992]. In general DNA methylation is a highly dynamic and complex process involved in cell differentiation during embryonic and normal cell development [Smith et al., 2013]. Thus, alterations in DNA methylation in conjunction with genetic factors are involved in different types of human diseases i.e. carcinogenesis, imprinting, autoimmune and neuropsychiatric disorders [Counts et al., 1995; Roberston, 2005; Richardson, 2003; Ai et al., 2012].

1.1.5 Epigenome-wide association studies (EWAS)

Over the past years, numerous technologies arose to investigate DNA methylation patterns. Recently, advances in next-generation sequencing (NGS) and microarray technologies e.g. whole-genome bisulfite sequencing and Illumina BeadArray technology (Infinium Methylation EPIC Array) facilitated the investigation of DNA methylation levels in a high-throughput genome-wide way. Thus, these technologies enabled a comprehensive, highly accurate, and reproducible analysis of epigenome-wide DNA methylation patterns. These approaches have been utilized to explore DNA methylation with regard to environmental factors in neuropsychiatric disorders. Up to date, candidate gene studies have focused on promoter methylation as one plausible biological mechanism for reduced gene expression. Several studies have demonstrated that alterations in DNA methylation mediate gene expression in response to the environment. In addition, there is increasing evidence that alterations in DNA methylation at specific candidate gene loci are associated with future risk for developing neuropsychiatric disorders including bipolar disorder and major depressive disorder [Vinkers et al., 2015; Matosin et al., 2017]. In particular, early life adversity, such as childhood maltreatment, can alter DNA methylation in complex neuropsychiatric disorders, such as BD or MDD [Weder et al., 2014; Bustamante et al., 2016]. Promising candidate genes for promoter methylation include *BDNF*, *FKBP5, NR3C1*, and *SLC6A4* [Oberlander et al., 2008; Fuchikami et al., 2011; Sugawara et al., 2011; Roy et al., 2017]. Several studies have investigated epigenetic alterations in the context of affective disorders via epigenome-wide association studies (EWAS) [Córdova-Palomera et al., 2015; Ratanatharathorn et al., 2017; Kuan et al., 2017; Story Jovanova et al., 2018; Starnawska et al., 2019]. For example, the largest currently conducted EWAS meta-analysis on DNA methylation of depressive symptoms included more than 7,900 middle-aged and elderly individuals with depressive symptoms [Story Jovanova et al., 2018]. The EWAS discovered methylation of 3 CpG sites significantly associated with depressive symptoms. Axon guidance was the most common disrupted pathway of the 3 methylated sites [Story Jovanova et al., 2018]. At the same time, epigenetic analyses in psychiatric patients might be affected by several confounding factors, *e.g.*, sex, age, disease course, and/or medication. Recent studies investigated methylation changes in post-mortem brain samples associated with MDD in which cell-type-specific deconvolution was applied to correct for subpopulations of neurons and glial cells [Aberg et al., 2020; Chan et al., 2020; Hüls et al., 2020]. Thus, analyses of the correlation between brain and blood-based methylation signatures are of big interest [Hannon et al., 2015]. Likewise, the integrative analyses of multi-omics data promoted methylation quantitative trait loci (mQTL) analyses which provide a way to functionally annotate genetic variation within the context of DNA methylation and the risk for neuropsychiatric disorders [Starnawska et al., 2021; Villicaña et al., 2021].

1.1.6 Methylation Quantitative Trait Loci (mQTLs)

The majority of GWAS susceptibility loci are located in the non-coding part of the genome and likely affect the phenotype through gene regulation [Ng et al., 2021]. Many genetic associations were identified through GWAS but still many candidate gene loci and their functional role remain unanswered [Howard et al., 2019; Mullins et al., 2021]. Therefore, DNA methylation can provide valuable insights through gene regulatory mechanisms as a potential pathomechanism of GWAS susceptibility variants. Many recent studies pointed out that genetic variants have a strong impact on levels of DNA methylation [Min et al., 2021, Ng et al. 2021]. These studies systematically correlated genetic variation (SNP genotypes) with DNA methylation levels at CpG sites in a genome-wide approach to identify DNA methylation quantitative trait loci (mQTLs). In general, mQTLs can be divided into *cis*-mQTLs, which are mQTLs that have local effects on the methylation from 500 kb to 1 Mb, or trans-mQTLs that have an effect on the long-distance within at least 5 Mb [Nica et al., 2013]. Overall, the genetic regulation of DNA methylation is highly complex and variable across different tissues and cell types [Januar et al., 2015]. The integration of mQTLs in epigenome-wide association studies is crucial and recommended as the identification of non-genetic effects is needed to distinguish between the contribution of environmental and genetic factors in disease progression. Recently, in many studies, genome-wide SNP and methylation array (e.g., Illumina EPIC array) data are analyzed in an integrative way [Januar et al., 2015]. For the majority of quantitative trait analyses, linear regression models are used for association testing between SNP genotypes and molecular traits [Shabalin et al., 2012; Ongen et al., 2016]. Thereby, the methylation level at each CpG site serves as the response variable and SNP genotypes as predictor variables. Additionally, important technical and biological covariates such as amplification plate, array, PC components from population stratification, proportions of cell components, age, gender, and/or medication are included in the statistical model. In a recently conducted GWAS, one of the most replicated risk variants in bipolar disorder, located in the intron of *ANK3* (rs10994336) showed to modulate *ANK3* methylation [Tang et al., 2021]. In another candidate gene study, it was shown that variation in DNA methylation levels at *CACNA1C* was associated with genotypes from bipolar disorder risk SNPs providing further evidence that methylation might mediate genotype-phenotype relationships [Starnawska et al., 2016]. As many EWAS and mQTL studies were performed using whole-blood samples they may not be sufficient to reflect the biological processes in the brain of psychiatric patients [Gamazon et al., 2013; Lin et al., 2018]. As DNA methylation is highly tissue and cell-type-specific, replication of these findings in post-mortem brain samples, larger sample sizes, and independent replication cohorts are warranted to further elucidate epigenetic mechanisms in affective disorders.

1.1.7 Aim of the Study

The aim of the present study was threefold. First, analyses were performed to identify methylation sites (CpGs) associated with: (i) familial risk (a family history of affective disorder); and (ii) environmental risk (childhood maltreatment). Second, we performed pathway enrichment analyses to determine the functional relevance of genes associated with significant methylation sites. Third, we investigated the impact of common genetic variants on DNA methylation via *cis*-mQTL analysis to dissect the complex interplay between genetic and epigenetic factors for affective disorders.

1.2 Material and Methods

1.2.1 Participants

The study participants were all unrelated females with no reported history of psychiatric disease (excluding specific phobias), as characterized by the Structured Clinical Interview (SCID) [First et al., 2004]. All participants were recruited at the Departments of Psychiatry of the Universities of Marburg and Münster, Germany, as part of the German Research Foundation unit FOR2107 (http://for2107.de/) [Kircher et al., 2019]. To avoid any potential confounding by sex and disease status, in this study, the focus was on female participants with no reported history of psychiatric disease only. The phenotypic data were requested in 2016 from the FOR2107 database. Three groups of participants served as input for the epigenome-wide DNA methylation analyses. Group one comprised 22 individuals with a reported familial risk for affective disorder. These individuals had at least one first-degree relative (*i.e.*, parent, sibling, or adult child) with a reported lifetime diagnosis of an affective disorder as assessed by the semi-structured clinical interview. None of the participants had a reported history of childhood maltreatment (environmental risk) in this group. One participant had a familial risk for BD and no additional risk for MDD, whereas 21 reported a familial risk for MDD but no additional risk for BD. Group two comprised 22 individuals with a reported environmental risk for affective disorder. These individuals scored above the cut-off for at least two forms of maltreatment in the childhood trauma questionnaire (CTQ) [Bernstein et al., 1994] and had no reported family history of affective disorder. Due to the longitudinal design, the CTQ data and the disease status were retrieved from the FOR2107 database in 2020, resulting in subsequent changes and the exclusion of three participants in this group. One participant scored above the cutoff for only one form of maltreatment, and two other participants were diagnosed with anorexia nervosa and an adaptation disorder with depressed mood. Group three consists of 22 healthy control samples with no reported familial or environmental risk. The present work was approved by the ethics committee of the Universities of Marburg and Münster, Germany. Written informed consent was provided by all participants before inclusion.

1.2.2 DNA Methylation Microarrays

Whole blood was bisulfite-treated using the EpiTect Bisulfite Kit (Qiagen, Hilden, Germany) to extract genomic DNA (500 ng). For assessing DNA methylation, the Infinium MethylationEPIC BeadChip (Illumina, San Diego, USA) and the Infinium HD Methylation Assay protocol were applied. Batch effects were prevented by random sampling of all samples across the 96-well plate before processing. Amplification, fragmentation, extension, hybridization, staining, and scanning were carried out according to the manufacturer's instructions.

1.2.3 Data Preprocessing and Normalization

All raw data and downstream analyses were carried out using *R* (3.4.0) and Bioconductor packages [Aryee et al., 2014; Pidsley et al., 2013; Leek et al., 2018; Ritchie et al., 2015]. The median methylated versus unmethylated signal intensities were inspected for sample-level quality control (QC) purposes. By assessing the methylation of X and Y chromosome probe intensities the observed sex was estimated and compared with the reported sex. To identify sample swaps, genotype concordance was captured by using SNP probes (n=65) on the MethylationEPIC BeadChip and genotypes from a genome-wide genotyping array (PsychArray BeadChip, Illumina, San Diego, USA; see also section Genotyping and Imputation). Probe-level QC consisted the following filtering steps: (i) detection *p*-value >0.05; (ii) bead count <3 in >5% of samples; (iii) probes overlapping with a SNP (MAF >5%, dbSNP, v137) [Sherry et al., 2001]; (iv) cross-hybridizing probes; (v) non-autosomal probes [Chen et al., 2013]. Background correction was applied using the Noob (normal-exponent out-of-band) algorithm with dye-bias normalization [Triche et al., 2013]. Functional normalization was performed, to get rid of technical variance between arrays [Fortin et al., 2014]. β-values were calculated by determining the intensity of the methylated and unmethylated allele ratio of fluorescent signals. β-values were logit-transformed into M-values to improve the detection and true positive rate of strongly methylated and unmethylated CpG sites [Du et al., 2010]. Combat was used to apply batch effect correction on the methylation data based on known covariates such as chip and chip row [Johnson et al., 2007]. The estimation of leukocyte subpopulations (B cells, $CD4^+$ T cells, $CD8^+$ T cells, monocytes, natural killer cells, and granulocytes) from heterogenous tissue sources like whole blood was carried out using the regression calibration algorithm [Houseman et al., 2012]. The most significant sources of technical and biological variation were calculated using the SVD (singular value decomposition) approach [Teschendorff et al., 2009].

1.2.4 Statistical Analysis of DNA Methylation

To assess the significance of DNA methylation at single CpG sites with familial or environmental risk, we carried out two separate epigenome-wide methylation analyses: (i) participants with environmental risk versus controls, and (ii) participants with familial risk versus controls. For that, the methylation level of each CpG site served as a response variable and each risk group served as the explanatory variable in a linear regression model. Technical and biological covariates were applied according to their effect on the first three principal components (PCs) of the genome-wide methylation data. The statistical significance threshold was chosen at an FDR (false discovery rate) <0.05 [Benjamini & Hochberg, 1995]. Further, the epigenome-wide significance threshold was proposed at a p-value $<1\times10^{-7}$ corresponding to approximately 495,000 independent tests and an FDR of 0.05. QQ (quantile-quantile) plots were generated to illustrate deviations of the test statistics from the null hypothesis. In the case of deflated or inflated test statistics, a correction was applied using the inflation factor lambda (λ).

1.2.5 Pathway Enrichment Analysis

To assess the functional relevance of significantly associated methylation sites and their corresponding genes, GO (Gene Ontology) pathway enrichment analyses [Ashburner et al., 2000] were carried out using the missMethyl package [Phipson et al., 2015]. This package estimates an enrichment of GO categories based on a hypergeometric test and accounts for the differing number of probes per gene. Correction for the differing number of probes per gene is a crucial step as this can prevent any bias in the GO enrichment analysis. For each test statistic, methylation sites that were unique for each of the risk groups served as input for the GO analyses. All probes that passed the aforementioned quality control steps served as background for these analyses. The Benjamini-Hochberg method was applied for multiple testing correction [Benjamini & Hochberg, 1995]. Statistical significance was considered at an FDR q-value <0.05. To discard redundant GO categories, the web-based tool Revigo was applied which is based on a clustering algorithm for semantic similarity measures [Supek, et al., 2011].

1.2.6 Genotyping and Imputation

Genotyping and imputation were carried out on the larger FOR2107 dataset (http://for2107.de/) and are described elsewhere [Kircher et al., 2019]. The present study's participants were a subset of the larger FOR2107 dataset. Briefly, genotyping was applied using the Illumina Infinium PsychArray BeadChip (Illumina, San Diego, CA, USA) according to standard protocols. Clustering and initial QC were performed in GenomeStudio v.2011.1 (Illumina, San Diego, USA) using the Genotyping Module v.1.9.4. Full QC was conducted in PLINK (v1.90b5) and *R* (v3.3.3) [Purcell et al., 2007; R Core Team, 2020]. Genotype data were imputed to the 1000 Genomes Phase 3 reference panel using SHAPEIT and IMPUTE2 [1000 Genomes Consortium, 2015; Delaneau et al., 2014; Howie et al., 2011]. Finally, the dataset contained 1,673 individuals and 8,578,636 variants after imputation and post-imputation QC. Population stratification was assessed using MDS (multidimensional scaling) as implemented in PLINK [Purcell et al., 2007]. Four individuals were classified as genetic outliers in the present cohort and were discarded from all downstream analyses of the present study (Figure 1).

1.2.7 Statistical Analysis of *cis*-mQTLs

Local genetic effects that influence variation in DNA methylation were identified through *cis*-mQTL analysis using FastQTL v2.184 which is based on a linear regression approach [Ongen et al., 2016]. The methylation level of each CpG site was used as the response variable. The SNP genotype served as the explanatory variable, and risk group, age, smoking history, and estimates of leukocyte subpopulations were used as covariates. For this analysis, only the overlap between methylation sites associated with both risk groups ($p<0.05$) was taken into account. The window size for *cis*-mQTLs was restricted to ±500 kb. FastQTL was performed using an adaptive permutation scheme via beta approximation. The number of permutations was set from 10,000 to 100,000. Beta approximated *p*-values were calculated for each SNP-CpG pair. Multiple testing correction was applied using Benjamini-Hochberg [Benjamini & Hochberg, 1995]. To identify LD-independent *cis*-mQTL associations, LD clumping based on the 1000 Genomes phase 3 release was applied. Further, a pairwise r^2 threshold of 0.5 within 500 kb of the most significant *cis*-mQTLs was considered.

1.2.8 Enrichment of *cis*-mQTLs among GWAS SNPs

To detect if *cis*-mQTLs were enriched among genetic associations of BD and MDD [Stahl et al., 2019; Howard et al., 2019], an enrichment analysis of *cis*-mQTLs and GWAS SNPs was carried out. For this purpose, the *p*-values of *cis*-mQTL SNPs ($q<0.05$) were extracted from the GWAS of MDD and BD conducted by the PGC. A permutation-based approach was applied, to estimate the significance of the enrichment. For that, 1,000 sets of SNPs were randomly selected from the FOR2107 dataset by matching the MAF threshold (MAF>0.05). The overlap for each random set and the *cis*-mQTLs was calculated under the null distribution. An empirical *p*-value was estimated by comparing the cumulative distribution of the control SNPs over the MDD- and BD-associated GWAS SNPs.

1.3 Results

1.3.1 Demographic Characteristics

A re-examination of the phenotype data in 2020 led to the exclusion of three participants due to mismatches in the phenotypic database. One of these participants scored above the cutoff for only one form of maltreatment, and the other two participants were diagnosed with anorexia nervosa and an adaptation disorder with depressed mood. Four additional individuals were classified as genetic outliers, and their data were not included in the genome-wide methylation analysis (Figure 1).

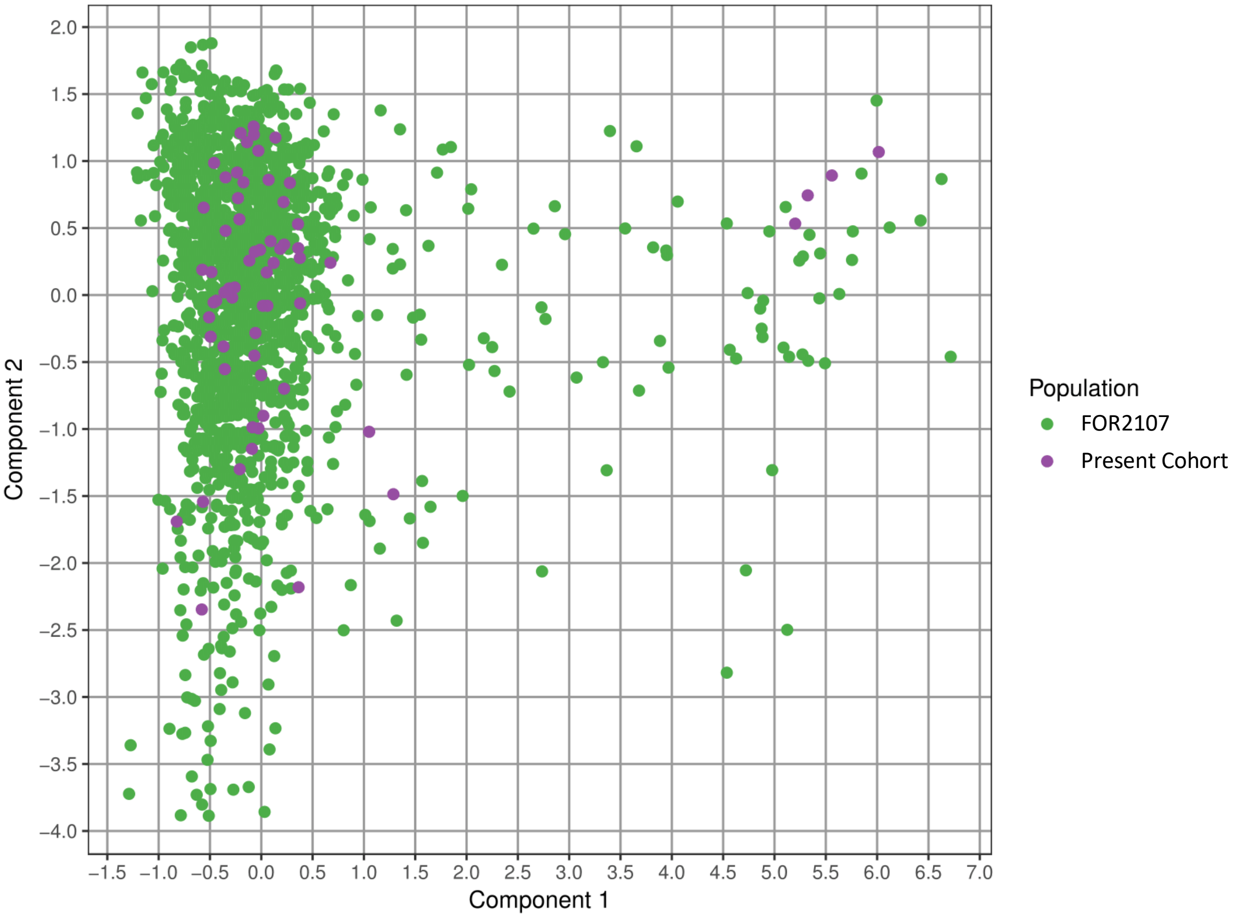

Figure 1 | Multidimensional Scaling (MDS) plot. MDS was performed using PLINK from genome-wide genotype data of the larger FOR2107 dataset for which the present cohort constituted a subset. First and second component of the MDS analysis are plotted using R. The larger FOR2107 cohort is indicated in green and the present cohort of this study is indicted in purple.

The final sample size thus comprised 59 study participants. The demographic characteristics of the three study groups after sample exclusion are shown in Table 1. Kruskal-Wallis test revealed a significant difference between the test groups in terms of the CTQ score (p<0.003). This was as expected as the environmental risk group scored above the cut-off for at least two forms of maltreatment in the childhood trauma questionnaire.

Table 1 | Demographic characteristics of the study cohort

Variable	**Familial Risk (n=20)**	**Environmental Risk (n=17)**	**Controls (n=22)**	χ^2	***df***	***p***
Age (years, mean ±SD)	32.00±11.09	36.06±8.94	26.68±6.88	42.87	42	0.43
Current smoker N (%)	3 (15%)	5 (29%)	5 (23%)	1.12	2	0.57
Former smoker N (%)	6 (30%)	8 (47%)	8 (36%)	1.16	2	0.56
Medication N (%)	4 (20%)	2 (12%)	0 (0%)	4.65	2	0.10
Organic Disease N (%)	4 (20%)	4 (23%)	0 (0%)	4.97	2	0.083
CTQ - total (mean ±SD)	30.95±4.38	51.24±10.44	29.41±4.6	76.51	46	0.003

Legend Table 1 | SD: Standard Deviation; N: Number; CTQ: Childhood Trauma Questionnaire; χ^2: Chi-squared test statistic; df: Degree of freedom; p: *p*-Value.

1.3.2 Quality Control of Methylation Data

Sample-level quality control of the methylation data was assessed by inspecting the median methylated versus unmethylated signal intensities. Figure 2 indicates a distinct clustering of all samples based on median intensities and did not reveal any sample outlier. Usually, sample outliers tend to separate from the main cluster and have lower median intensities. Further, exploring the quality of the samples using M-value densities before and after normalization did not reveal any sample-level quality control issues. As anticipated, the methylation data displayed a bimodal distribution in terms of methylated and unmethylated signal intensities. The mean detection *p*-value summarises the quality of the signal across all the probes and was <0.05 for each sample indicating a good quality of the methylation data on the sample level.

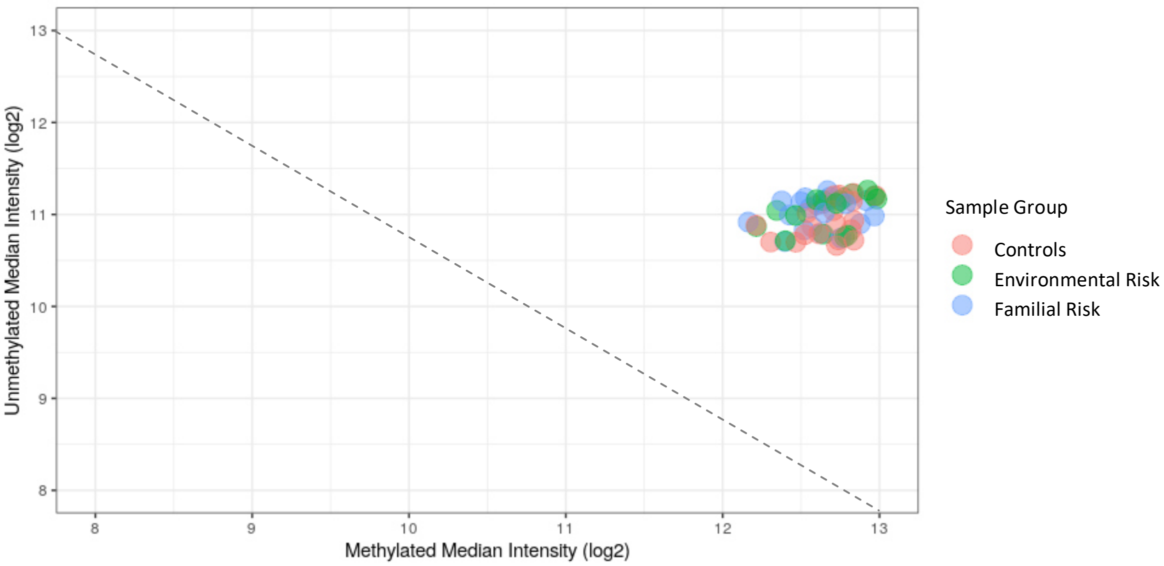

Figure 3 | Sample-level QC plot. In this quality control plot methylated and unmethylated median signal intensities are plotted. The dashed line indicates the regression line. Samples are colored by test group.

No mismatches were found for gender by looking at the median total intensities of the X- and Y-chromosome mapped probes on the EPIC array. The first two principal components explained 13.8% and 4.8% of the variance in the methylation data, respectively (Figure 3). Hence, most of the variance in the methylation data is captured on the first two principal components. In the principal component analysis, no outlier samples were detected. At the same time, no distinct clustering of technical batches or phenotypic data was observed (Figure 3). About 41.72% of the CpG sites on the EPIC array were discarded due to quality control measurements. The largest proportion of these was due to cross-reactive probes (3.44%), probes located in common SNPs (20.26%), and probes with low variance (14.57%). After filtering as specified in Table 2, a total of 495,406 CpG sites (58.28%) remained for all statistical and downstream analyses. Singular Value Decomposition (SVD) analysis revealed significant correlations of the methylation data with biological and technical covariates. The effects of technical covariates (e.g., slide or array) on the methylation data were discarded successfully after batch effect correction (Figure 4). None of the technical covariates such as slide or array were correlated with the first three principal components afterward.

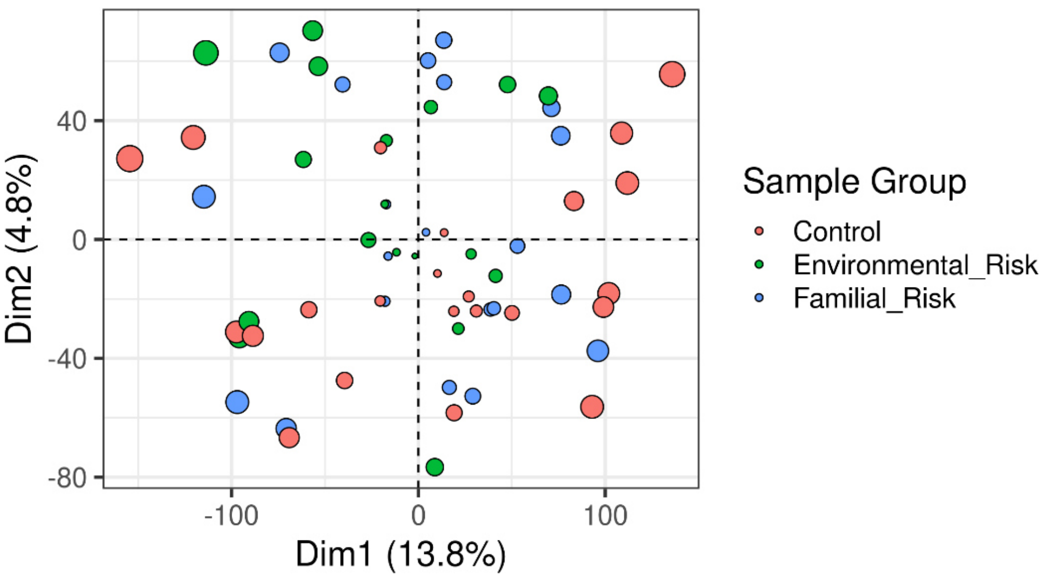

Figure 5 | Principal Component Analysis. The PCA plot shows the percentage of variance explained by each principal component. The x-axis represents the first and the y-axis the second principal component. The different colors indicate the sample group and the size of each data point specifies the contribution of each sample to the overall variance.

Table 2 | Probe Filtering

Feature	Probes
Detection p-Value	4,941 (0.58%)
Bead Count	12,157 (1.43%)
Cross-Reactive Probes	29,205 (3.44%)
SNPs in Probes	172,252 (20.26%)
Non-Specific Probes	6,813 (8.00%)
Low Variance	123,852 (14.57%)
Total	**495,406 (58.28%)**

None of the technical covariates such as slide or array were correlated with the first three PC components afterward. A significant correlation with the first PC component was found for age (p<0.01); former smoking status (p<0.05); CD8+ T cells (p<0.01); CD4+ T cells ($p<1\times10^{-5}$); natural killer cells (p=0.01); monocytes (p<0.05); and granulocytes($p<1\times10^{-10}$) (Figure 4 A/B). Among these CD4+ T cells and granulocytes had the highest effect on the methylation data. Age and three out of six cell-type components were also significantly correlated with the second principal component highlighting their importance and effect on the methylation (Figure 4 A/B).

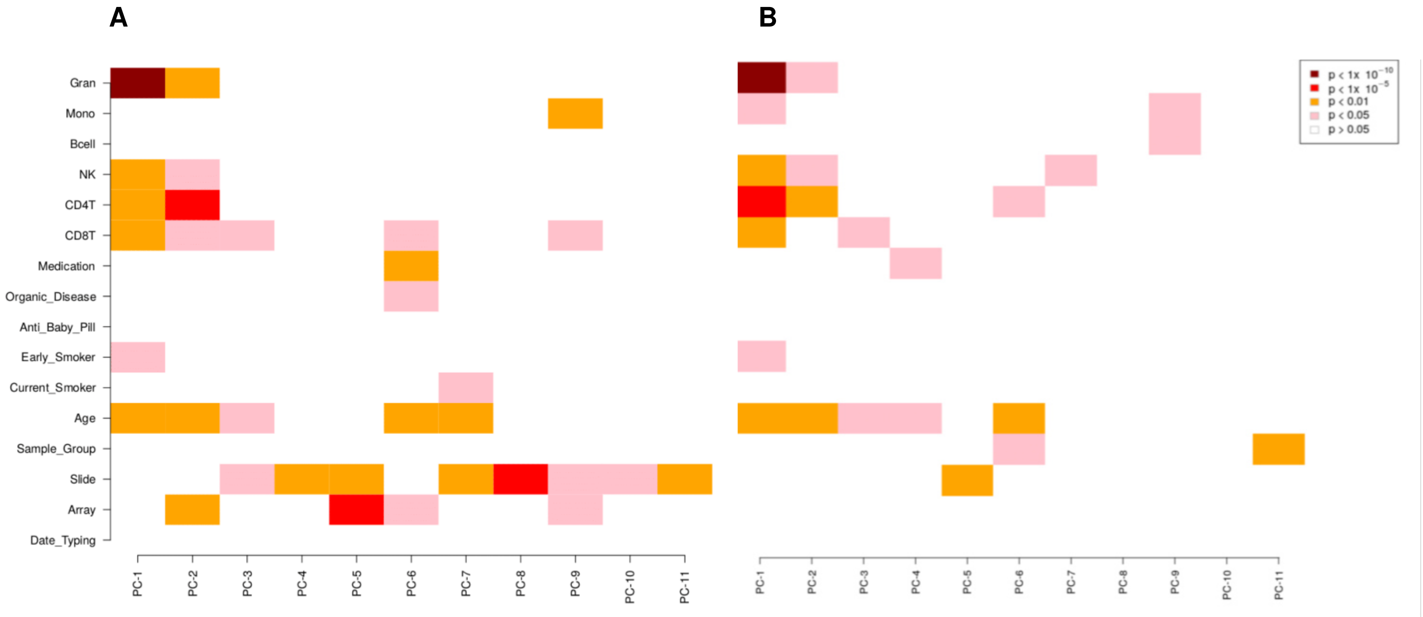

Figure 4 | Singular value decomposition (SVD) plot. Figure A) and B) specify the SVD analysis before and after batch effect correction. The x-axis represents the number of principal components and the y-axis technical and biological covariates. The darker the color the stronger the correlation between principal components and covariates as indicated in the legend.

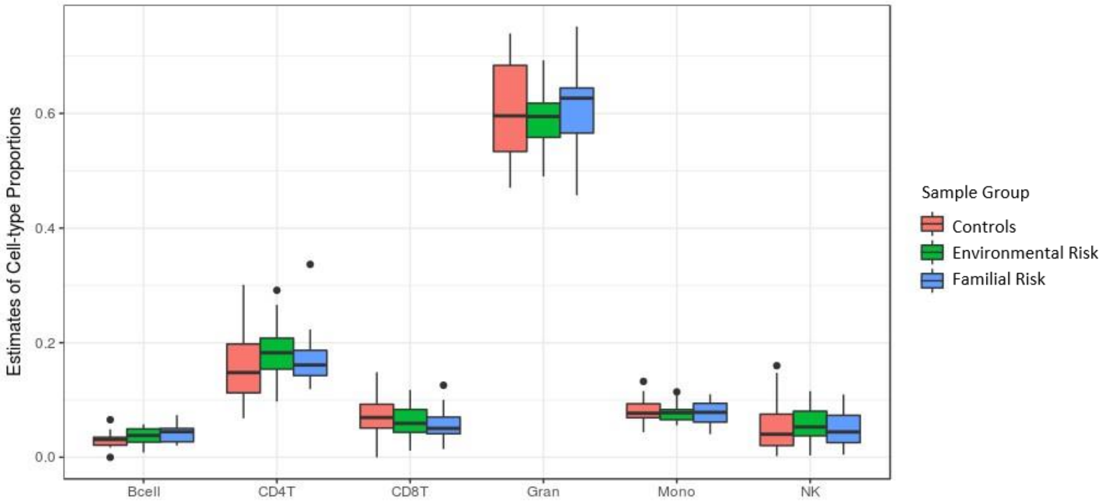

Figure 5 | Estimates of Cell-type Proportions. Relative proportions of CD4+ and CD8+ T-cells, natural killer cells, monocytes, granulocytes, and B-cells in the stud cohort are plotted. The Houseman regression calibration approach [Houseman et al., 2012] for the Illumina EPIC array for deconvoluting heterogeneous tissue sources like whole-blood was applied. Colors indicate the different test groups.

As the EPIC array provides CpG sites associated with cell-type-specific methylation patterns in whole blood, the cell-type composition can be estimated to avoid any confounding. The estimated proportions of leukocyte subpopulations indicate large proportions of granulocytes of approximately 60% (Figure 5). There were no significant differences between the test groups in leukocyte subpopulations. These findings highlight the importance of adjusting for cell-type-specific confounding in methylation data with regard to statistical analyses and models.

1.3.3 Differential DNA Methylation

After rigorous QC, a total of 495,406 CpG-sites were tested using multivariable linear regression models. The genomic inflation factor lambda (λ) for familial and environmental risk was 1.17 and 0.96, respectively (Figure 6 A/B). Due to an inflation of the test statistics in the familial risk group (Figure 6A), a correction was applied by dividing the chi-square test statistics through λ. Strong effects of population stratification could be excluded as the final participants were all of central European origin, as shown by the multidimensional scaling (MDS) plot of the genotype data (Figure 1).

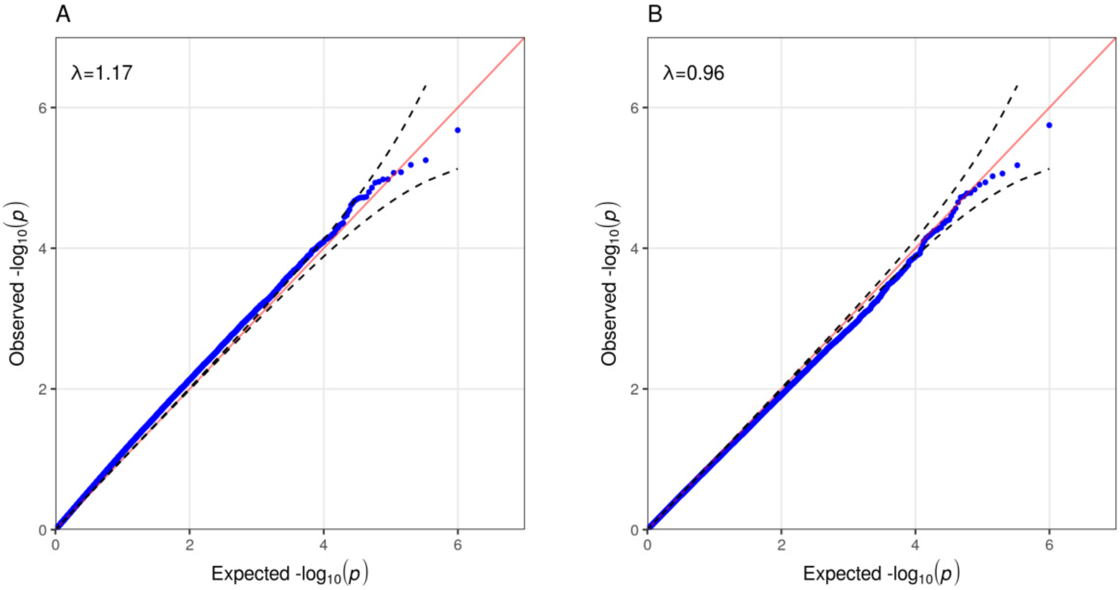

Figure 6 | Quantile-Quantile (QQ) plot. QQ plots for the epigenome-wide methylation of familial and environmental risk for affective disorder. A) Familial risk versus controls. B) Environmental risk versus controls. QQ plots display the distribution of expected and observed p-values according to a logarithmic scale. Deviations from the red line indicate an inflation or deflation of the test statistics.

None of the tested CpG-sites achieved epigenome-wide significance ($p<1\times10^{-7}$) after correction for multiple testing. However, the epigenome-wide methylation analysis of familial and environmental risk revealed that 22,230 and 21,940 methylation sites, respectively, achieved nominal significance ($p<0.05$). Of these, 20 and 40 methylation sites for familial and environmental risk achieved a *p*-value of $<1\times10^{-4}$, respectively. The strongest association for familial risk was found at cg25160593 ($p=1.12\times10^{-5}$) located on chromosome 3 within the 5'UTR of the *ZNF197* gene (Table 3). The strongest association for environmental risk was identified at cg06253966 ($p=1.78\times10^{-6}$) located on chromosome 6 in an intergenic region (Table 4). The overlap of nominally significant methylation sites ($p<0.05$) in both risk groups comprised 5,717 CpG sites.

Table 3 | Top 10 associations between DNA methylation and familial risk for affective disorder

Probe ID	**Chr**	**Position**	**Strand**	**Gene**	**Feature**	**logFC**	**logOdds**	***p* (λ adj.)**	***q* (λ adj.)**
cg25160593	3	44666684	-	*ZNF197*	5'UTR	-0.451	2.415	1.12E-05	0.997
cg14088628	14	71023160	+	-	-	0.343	1.819	2.62E-05	0.997
cg25016544	5	72511412	+	-	-	0.393	1.726	2.99E-05	0.997
cg01612292	8	144809598	-	*FAM83H*	Intron	-1.409	1.579	3.68E-05	0.997
cg16514214	17	28009516	+	*SSH2*	Intron	-0.246	1.567	3.74E-05	0.997
cg05604487	17	77386277	+	*HRNBP3*	5'UTR	-0.361	1.437	4.49E-05	0.997
cg27097386	15	96400610	+	-	-	-0.420	1.436	4.50E-05	0.997
cg06103654	2	27072413	-	*DPYSL5*	5'UTR	0.370	1.389	4.81E-05	0.997
cg25738505	3	101757700	+	-	-	1.753	1.369	4.94E-05	0.997
cg25513853	17	72272298	+	*DNAI2*	5'UTR	0.296	1.269	5.69E-05	0.997

Legend Table 3 | Methylation sites are ranked in descending order of statistical significance. Abbreviations: Probe ID: Unique identifier for each methylation site in the genome; Chr: Chromosome; Strand: Forward (+) or reverse (-) designation of the DNA strand; Feature: Gene regulatory feature; UTR: Untranslated region; logFC: Log-fold change; logOdds: Log-odds ratio; λ adj.: Inflated test statistics corrected by the genomic inflation factor (λ).

Table 4 | Top 10 associations between DNA methylation and environmental risk for affective disorder

Probe ID	Chr	Position	Strand	Gene	Feature	logFC	logOdds	*p*-Value	*q*-Value
cg06253966	chr6	9851903	+	-	-	-0.205	2.230	1.78E-06	0.854
cg14078118	chr3	5047351	+	-	-	-0.339	1.473	6.62E-06	0.854
cg21666801	chr10	77285588	-	C10orf11	Intron	0.253	1.315	8.68E-06	0.854
cg08991615	chr11	75222390	+	GDPD5	5'-UTR	-0.377	1.263	9.49E-06	0.854
cg03696617	chr7	150097762	+	-	-	0.250	1.147	1.16E-05	0.854
cg02331649	chr3	52926322	-	TMEM110	Intron	0.740	1.102	1.25E-05	0.854
cg08134671	chr19	2542837	-	GNG7	5'-UTR	0.218	1.010	1.46E-05	0.854
cg14171448	chr10	1711025	-	ADARB2	Intron	0.186	0.944	1.64E-05	0.854
cg05847183	chr1	90316786	+	LRRC8D	5'-UTR	1.603	0.935	1.66E-05	0.854
cg16322388	chr18	72954277	+	TSHZ1	5'-UTR	0.290	0.879	1.83E-05	0.854

Legend Table 4 | Methylation sites are ranked in descending order of statistical significance. Abbreviations: Probe ID: Unique identifier for each methylation site in the genome; Chr: Chromosome; Strand: Forward (+) or reverse (-) designation of the DNA strand; Feature: Gene regulatory feature; UTR: Untranslated region; logFC: Log-fold change; logOdds: Log-odds ratio.

1.3.4 Pathway Enrichment Analysis

We carried out GO (Gene Ontology) pathway enrichment analyses with methylation sites nominally associated with each risk group uniquely ($p<0.05$). For familial risk, a total of 112 GO pathways were significant at an FDR *q*-value <0.05 (Figure 7). Among the top enriched pathways for familial risk were neurogenesis ($q<1.93\times10^{-8}$) and generation of neurons ($q<2.87\times10^{-7}$). For methylation sites associated with environmental risk, 52 GO pathways remained significant ($q<0.05$) after correction for multiple testing (Figure 8). Nervous system development ($q<1.52\times10^{-6}$) was the top finding in the environmental risk group. GO pathway enrichment analyses indicated an enrichment of brain-derived and nervous system-related categories in the methylation data of both risk groups. This indicates that both risk groups share common pathways. The Revigo analysis revealed redundant GO terms. For example, the parent term for nervous system development and neurogenesis belonged to the GO class tissue development. As a result, 59 of 112 (53%) and 23 of 59 (39%) GO categories were independent of each other for familial and environmental risk.

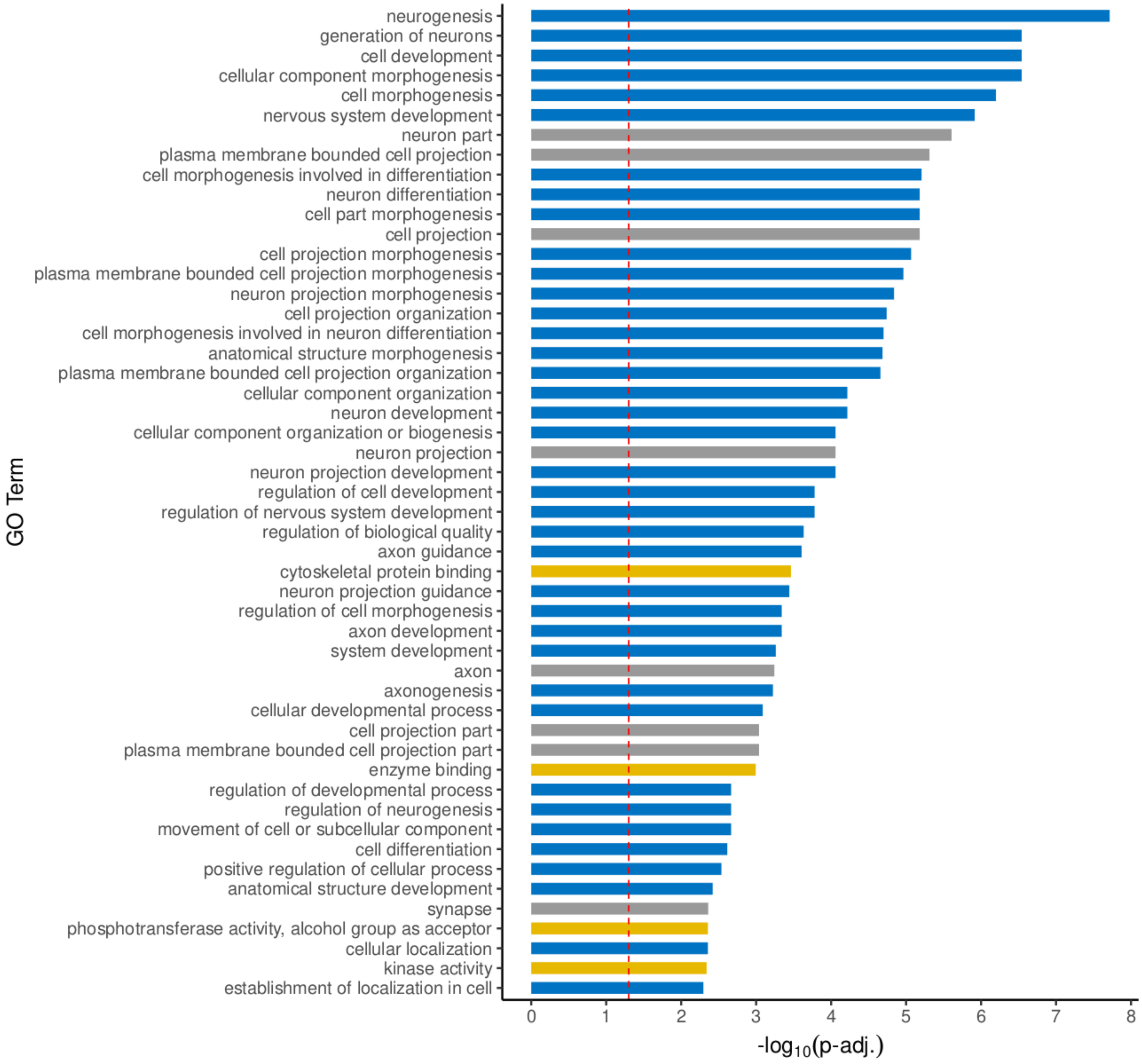

Figure 7 | Gene Ontology pathway enrichment analysis of associations between DNA methylation and familial risk for affective disorder. Top 50 enriched GO categories are shown. GO terms are ranked in descending order of statistical significance and GO ontology categories. Different colors of the bars represent gene ontology classes. Red dashed line indicates an adjusted *p*-value threshold of 0.05.

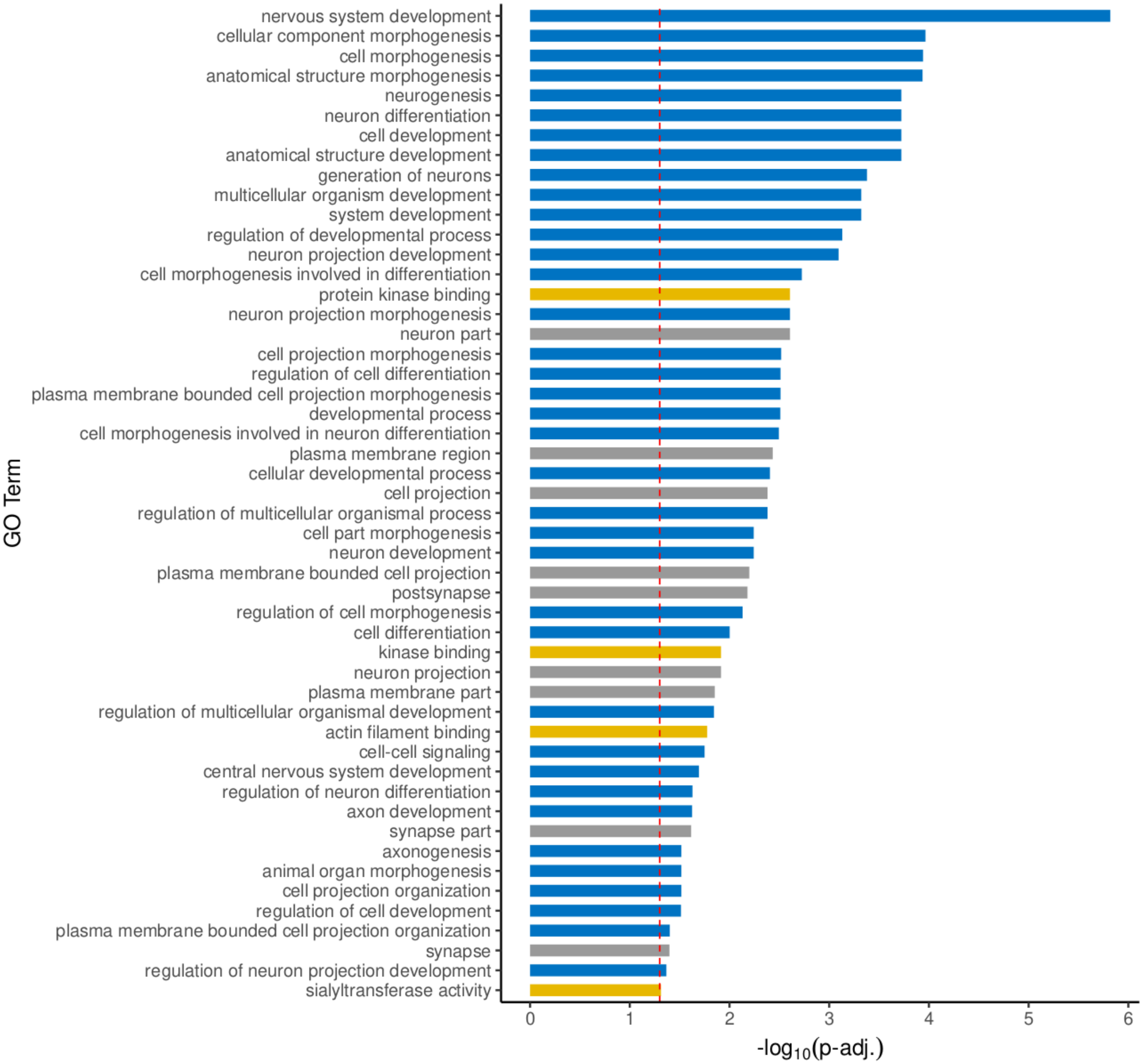

Figure 8 | Gene Ontology pathway enrichment analysis of associations between DNA methylation and environmental risk for affective disorder. Top 50 enriched GO categories are shown. GO terms are ranked in descending order of statistical significance and GO ontology categories. Different colors of the bars represent gene ontology classes. Red dashed line indicates an adjusted *p*-value threshold of 0.05.

1.3.5 Effect and Enrichment of *cis*-mQTLs

To identify SNPs that regulate DNA methylation in sites/genes relevant to environmental and familial risk, the *cis*-mQTL analysis was restricted to CpG sites that were significantly associated with both risk groups (n=5,715, p<0.05). Using FastQTL via the adaptive permutation scheme, we identified a total of 61 SNP-CpG associations (q<0.05) after correction for multiple testing. From these, 45 independent *cis*-mQTLs remained after LD clumping (Table 5). The top-ranked *cis*-mQTL SNP rs3780420 (q_{mQTL}=1.37×10^{-6}) was associated with the methylation at cg11093459, which is located in the 5'-untranslated region (5'UTR) or promoter region (TSS1500) of the *HEMGN* gene (Figure 9A). The second-ranked *cis*-mQTL SNP rs7722394 (q_{mQTL}=2.23×10^{-6}) was associated with the level of methylation at cg08323540 which mapped to the promoter region (TSS1500) of *GABRP* (Figure 9B). Further, the third and fourth-ranked *cis*-mQTL in *FOXN3* and *ADCY9* are depicted in Figures 9C and 9D.

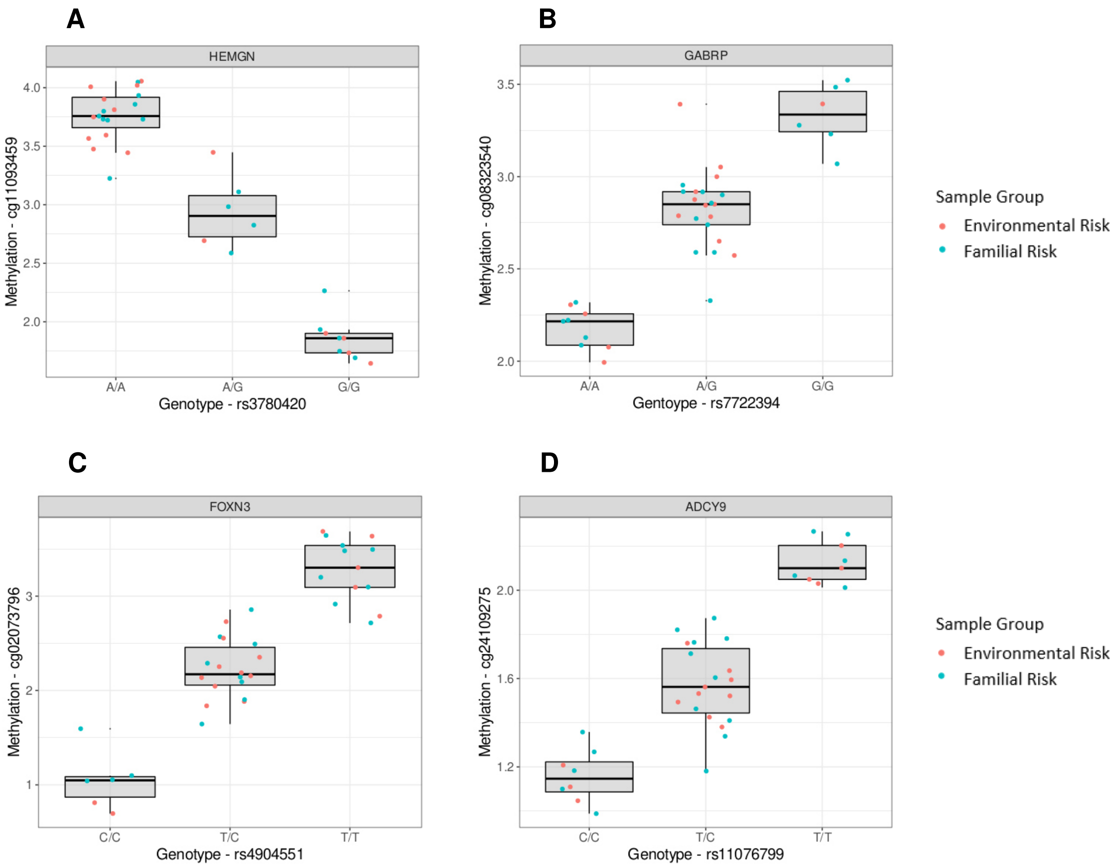

Figure 9 | Boxplot of the four most significant *cis*-mQTLs identified in environmental and genetic risk group (n=37) for affective disorder. For the *cis*-mQTL analysis both risk groups were combined. X-axis represent the genotype and y-axis the methylation level of each individual.

Table 5 | Top 20 associations between SNP genotypes and DNA methylation (*cis*-mQTLs) in the environmental and genetic risk groups combined.

Chr	SNP	Distance	CpG	Effect	*p*-Value	Adj.*p*-Val.	Nearby Gene	Function
9	rs3780420	-23976	cg11093459	-0.9235	2.39E-10	1.37E-06	HEMGN	5'UTR; TSS1500
5	rs7722394	660	cg08323540	0.5653	9.69E-10	2.23E-06	GABRP	TSS1500
14	rs4904551	-1255	cg02073796	-1.1036	1.17E-09	2.23E-06	FOXN3	Intron
16	rs11076799	-357	cg24109275	-0.4485	3.44E-09	4.14E-06	ADCY9	Intron
9	rs2147257	-13613	cg06800115	0.4621	3.67E-09	4.14E-06	HSD17B3	Intron
3	rs7433472	-3418	cg02772928	0.6621	4.35E-09	4.14E-06	-	-
8	rs2006937	3274	cg09039475	0.9399	1.38E-08	1.13E-05	CCDC25	Intron
19	rs10500292	53479	cg14061069	-0.8063	2.10E-08	1.50E-05	DMPK	Intron
2	rs1629979	-5303	cg00532797	-0.4718	3.02E-08	1.92E-05	LOC101929231	Intron
6	rs263184	238301	cg21182457	0.8677	1.92E-07	9.99E-05	ADGRG6	Intron
22	rs5757207	-4663	cg03443888	0.3433	1.25E-06	5.95E-04	FAM227A	TSS1500
14	rs4902360	74	cg15311201	-0.3420	2.22E-06	9.75E-04	MAX	Intron
5	rs11955291	6837	cg18269756	-0.5709	3.19E-06	1.30E-03	-	-
3	rs1398609	-7339	cg25738505	-1.3772	3.50E-06	1.33E-03	-	-
1	rs947367	14899	cg16771827	-0.3962	3.74E-06	1.34E-03	PRELP	5'UTR
17	rs11655504	-19684	cg21657704	-0.3195	4.39E-06	1.45E-03	TBCD	Intron
17	rs56078934	-96531	cg06153925	0.9220	4.83E-06	1.45E-03	RPTOR	Body
3	rs12495073	70284	cg08033130	0.8967	4.96E-06	1.45E-03	CXCR6	TS1500
12	rs6598154	1105	cg04155630	-0.2455	5.34E-06	1.45E-03	-	-

Legend Table 5 | *cis*-mQTLs are ranked in descending order of statistical significance. Abbreviations: Chr: Chromosome; Strand: SNP: Single-nucleotide polymorphism; Distance: Distance between the CpG site and variant in bp; Effect: Slope from the linear regression; Function: Gene regulatory feature; UTR: Untranslated region; TSS: Transcription start site.

Enrichment analysis of the 45 independent *cis*-mQTLs was performed using the summary statistics of the large GWAS of BD and MDD. A permutation-based testing approach was applied to estimate an empirical *p*-value by comparing the cumulative distribution of randomly drawn control SNPs over the MDD- and BD-associated GWAS SNPs. Due to a lack of overlap (missing test statistics) between *cis*-mQTLs and GWAS risk variants for MDD (*p*<0.05), an empirical *p*-value could not be retrieved. Three of the 45 independent *cis*-mQTLs were nominally associated in the GWAS of bipolar disorder (rs7152726: *p*<0.04; rs4904551 *p*<0.02); rs6781560 *p*<0.01). However, the enrichment in the GWAS of BD was not significant (*p*<0.771) as indicated by the permutation-based enrichment test. Further, the three *cis*-mQTLs were only moderately associated with bipolar disorder and thus might not have an impact in terms of the risk or pathophysiology of affective disorder.

1.4 Discussion

1.4.1 Key Findings

In the present study, epigenome-wide methylation analyses of familial and environmental risk for affective disorders were carried out in whole-blood samples from female individuals with no reported history of psychiatric disease. A total of 22,230 methylation sites were associated at nominal significance ($p<0.05$) with familial risk, while 21,940 sites were nominally associated with environmental risk. However, none of the tested methylation sites achieved epigenome-wide significance after correction for multiple testing. The lowest p-value for familial risk was found at cg25150593 ($p=2.10\times10^{-6}$), which is located in the 5'UTR (untranslated region) of *ZNF197* on chromosome 3. However, to our knowledge, no associations between alterations in DNA methylation of this gene and affective disorders has yet been reported. The top-ranked methylation site for environmental risk cg16713962 ($p=1.60\times10^{-6}$) was located in an intergenic region of chromosome 16. Furthermore, one of the top-ranked methylation sites for environmental risk (cg02331649, $p=1.25\times10^{-5}$) was located at the intron of *TMEM110,* which is a previously reported genome-wide significant risk locus for BD, schizophrenia (SCZ) and autism spectrum disorder [Sklar et al., 2011; Ripke et al., 2014; Anney et al., 2017]. *TMEM110* encodes a brain-expressed transmembrane protein that connects the endoplasmic reticulum and plasma membrane and acts as a positive regulator of Ca^{2+} influx in mammalian cells [Quintana et al., 2015; Jing et al., 2015; Song et al., 2017]. Interestingly, research has implicated dysregulation of Ca^{2+} homeostasis in the pathophysiology of several neuropsychiatric disorders, including BD and SCZ [Forstner et al., 2017; Berridge et al., 2014]. However, among the other top ten findings for familial as well as environmental risk, we did not find any gene linked to neuropsychiatric illness. To assess the potential functional relevance of genes associated with significantly associated methylation sites, GO pathway enrichment analyses were performed. In both risk groups, we identified pathways that are related to the brain and neuronal system. In particular, neurogenesis was the most significant finding for familial risk and among the top pathways for environmental risk. Neurogenesis plays an important role in the maintenance and function of the hippocampus circuitry [Lazarov et al., 2016]. It is a process in which new neurons are generated in the subgranular zone of the dentate gyrus regulated by environmental and genetic factors [Samuels et al., 2011; Jesulola et al., 2018]. The neurogenesis theory hypothesizes that MDD is linked to impairments of adult neurogenesis [Jacobs et al., 2000]. Furthermore, the effects and mechanisms of antidepressants are partly mediated through increased neurogenesis [Tunc-Ozcan et al., 2019]. Some research implicated that neurogenesis plays a pivotal role in the treatment and prevention of neurological and psychiatric disorders [DeCarolis et al., 2010]. These findings and the results of the present study suggest that neurogenesis might be of potential relevance in terms of the pathophysiology of affective disorders. Finally, we performed a *cis*-mQTL analysis to identify genetic variants that impact DNA methylation levels. In this analysis, 45 independent SNP-CpG pairs were significant after correction for multiple testing via the permutation scheme. For the most significant SNP-CpG association (rs3780420; cg11093459; $q_{\mathrm{mQTL}}<1.37\times10^{-6}$) located nearby *HEMGN*, we did not find any link to psychiatric diseases. However, the second most significant SNP-CpG association (rs7722394; cg08323540; $q_{\mathrm{mQTL}}<2.23\times10^{-6}$) was located in the promoter of *GABRP* (gamma-aminobutyric acid type A receptor subunit pi) which encodes for a multisubunit chloride channel expressed in the hippocampus and several non-neuronal tissues. This gene plays an important role in the inhibitory synaptic transmission of the central nervous system [Neelands et al., 1999; O'Leary et al., 2016]. However, the effect sizes of the *cis*-mQTLs were moderate and their functional role in the development of affective disorder needs further investigation. In addition, an enrichment of the *cis*-mQTLs in the GWAS of MDD or BD could not be confirmed. On the individual level, only three *cis*-mQTLs were retrieved from the GWAS summary statistics but these were only moderately associated with BD.

1.4.2 Strengths and Limitations

The present study had several limitations. First, whole-blood samples were used for the assessment of DNA methylation levels. Although the heterogeneity of whole-blood samples was considered via correction for leukocyte subpopulations, it is currently unclear whether blood is a relevant tissue for investigating the pathophysiology of affective disorders. Previous studies have demonstrated only a limited correlation between DNA methylation in the whole blood of psychiatric patients and brain tissue [Walton et al., 2016; Farré et al., 2015]. The tissue specificity of neuropsychiatric disorders might explain the moderate number of findings on the epigenome-wide scale. Investigations of methylation alterations in whole-blood and post-mortem brain samples of psychiatric patients are warranted to identify tissue-dependent epigenetic signatures of relevance to psychiatric diagnosis [Aberg et al., 2020; Chan et al., 2020; Boström et al., 2017]. Second, we only examined individuals without a reported history of psychiatric disease to avoid any potential confounding by disease status or course. Therefore, it remains unclear whether the methylation sites identified in the present study also show an association in patients with affective disorder, which should be investigated in future studies. The third important limitation of this study was the sample size which limited the power to detect associations on the epigenome-wide scale. The largest epigenome-wide association study of depressive symptoms to date was performed by the CHARGE consortium (discovery cohort, n= 7,948; replication cohort, n= 3,308) [Story Jovanova et al., 2018]. In that meta-analysis, three methylation sites showed an epigenome-wide significant association ($p<1\times10^{-7}$) to depressive symptoms with moderate effect sizes. This demonstrates that the association between methylation sites and depressive symptoms may be subtle and that very large sample sizes are needed. Interestingly, the three methylated sites were targeting the axonal guidance pathway which is one of the commonly affected pathways in depressive symptoms. The respective three methylation sites were not significant in either of the risk groups in the present study. Future studies involving larger sample sizes are therefore warranted and should include case/control analyses to elucidate epigenetic mechanisms in affective disorders.

1.4.3 Conclusions

To conclude, the present work is one of the first epigenome-wide methylation studies in which environmental and familial risk for affective disorders were analyzed in an integrative way using the MethylationEPIC BeadChip. No epigenome-wide significant methylation site was found after correction for multiple testing. However, the analyses implicated >20,000 nominally significant CpG sites for familial or environmental risk that converged into pathways associated with neurogenesis and nervous system development.

1.5 List of Figures

1.6 List of Tables

CHAPTER 2

RARE VARIANT BURDEN ANALYSIS OF NIEMANN-PICK GENES IN SCHIZOPHRENIA USING TARGETED SEQUENCING

In this chapter, the findings of the rare variant burden analysis of Niemann-Pick genes *NPC1* and *NPC2* in SCZ patients and controls using smMIP-based targeted sequencing are presented.

2.1 Introduction

2.1.1 Schizophrenia

Schizophrenia (SCZ) is a severe neuropsychiatric disorder that affects about 1% of the population worldwide [Leucht et al., 2007]. Patients with SCZ display a broad range of symptoms that may include delusions, hallucinations, disorganized speech, and grossly disorganized or catatonic behavior. The latter include affective flattening, alogia, avolition, and social withdrawal [Crow et al., 1985; Andreasen 1999; Sass et al., 2003; American Psychiatric Association, 1994]. All of these symptoms highlight that SCZ is a clinically heterogeneous and complex mental disorder with devastating consequences on the physical and mental health of the affected individuals [Millier et al., 2014]. The age at onset of SCZ is in early childhood or adolescence [Kessler et al., 2007]. The disorder is highly heritable, exceeding 60% in family studies and 80% in twin studies [Wray et al., 2012; Sullivan et al., 2003]. However, environmental factors such as childhood traumatic life events, migration, substance use, and psychosocial factors also contribute to the development of SCZ and these may interact with genetic factors [Vilain et al., 2013]. Diagnosis is based on psychiatric evaluations and an extensive assessment by an experienced mental health professional [American Psychiatric Association, 1994]. The advent of genome-wide association studies (GWAS) and next-generation sequencing (NGS) technologies allowed dissecting the genetic architecture of SCZ. In GWAS it was found that SCZ is highly polygenic which means that many genes with relatively low effect sizes contribute to the disease development. One of the largest GWAS of SCZ was conducted by the Psychiatric Genomics Consortium (PGC) in 2014 which included 36,989 cases and 113,075 controls and revealed 108 independent risk loci from which 83 were newly implicated with SCZ. Multiple associations such as *DRD2*, genes involved in glutamatergic neurotransmission (*GRIN2A*, *SRR*, *CLCN3*, and *GRIA1*), neuronal calcium signaling (*CACNA1C*, *CACNA1I*, *CACNB2*, *RIMS1*) and broader synaptic function (*KCTD13*, *CNTN4*, *PAK6*) were identified [Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014]. Interestingly, one of the top new loci was associated with the major histocompatibility complex (MHC) locus on chromosome 6. This led to the conclusion that there might be an important component of immune mechanisms involved in the pathophysiology of SCZ [Schizophrenia Working Group of the Psychiatric

Genomics Consortium, 2014]. However, common variants explain only partially the heritability and phenotypic variance. In addition, most of these variants are located in the noncoding part of the genome and do not affect protein-coding genes with known functions. There is increasing evidence that copy number variations (CNVs) and rare single-nucleotide variants (SNVs) also play an important role in the development of SCZ. Over many years multiple rare chromosomal deletions and duplications were identified by large consortia-based studies [International Schizophrenia Consortium, 2008; CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium, 2017]. It was found that there was a global as well as gene-specific increase in the burden of rare CNVs in SCZ patients [Walsh et al., 2008; International Schizophrenia Consortium, 2008]. The Psychiatric Genomics Consortium (PGC) provided an analysis in which it pinpointed more than eight genome-wide significant CNVs such as 1q21.1, 2p16.3 (*NRXN1*), 3q29, 7q11.2, 15q13.3, distal 16p11.2, proximal 16p11.2 and 22q11.2 [CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium, 2017]. Most of the findings unveiled that the identified CNVs overlapped with genes associated with neurodevelopmental disorders [Grayton et al., 2012]. In sequencing studies such as exome studies, it was found that rare loss-of-function (LoF) variants with a minor allele frequency (MAF) ≤0.1% were associated with the risk of schizophrenia and developmental disorders [Singh et al., 2016]. In particular, in this study, rare variants in *SETD1A* contributed significantly to the risk of SCZ. This gene is a component of a histone methyltransferase (HMT) that catalyzes the trimethylation of histone H3 at lysine 4 (H3K4me4) chromatin modification. This modification is generally known to play a role as a regulator of gene transcription [Wang et al., 2021]. The SCZ Exome Sequencing Meta-analysis (SCHEMA) consortium is a large collaboration in which currently exomes from 24,248 cases and 97,322 controls, and de novo mutations from 3,402 parent-proband trios were analyzed [Singh et al., 2022]. This large collaborative study results revealed that ultra-rare coding variants (URVs) in ten genes impact the risk of developing schizophrenia substantially (odds ratios 3 - 50, $p<2.14 \times 10^{-6}$), and 32 genes at an FDR < 5% [Singh et al., 2022]. The majority of the genes indicated the greatest expression in the central nervous system neurons. This large-scale analysis of exome sequencing data highlighted the importance of assessing the role of rare variants in the risk of developing SCZ. As many genes are implicated in SCZ from common variants with small effects to rare single mutations or CNVs of large effect sizes, so far there is no valid biomarker available to confirm or rule out the clinical diagnosis of SCZ. And due to its clinical heterogeneity and overlapping symptoms with other neurological and psychiatric disorders, making a correct clinical diagnosis shortly after the first symptom presentation can be challenging [Doherty et al., 2016].

2.1.2 Niemann-Pick Type C Disease

NP-C (Niemann-Pick type C) disease belongs to a group of rare lysosomal storage disorders which are a group of heterogeneous inherited inborn errors of lipid metabolism and were first described by Albert Niemann (1880-1921) and Ludwig Pick (1868-1944) [Niemann, 1914; Pick, 1926]. In principle, there are 3 subtypes A, B, and C which are inherited in an autosomal recessive way and are classified based on the genetic cause and clinical symptoms [Crocker et al., 1961; Vanier et al., 2013]. Niemann-Pick Type A (NP-A) and Type B (NP-B), also called acid sphingomyelinase deficiency (ASMD) are caused by deficiency of the enzyme acid sphingomyelinase (ASM). Loss of function mutations in the *SMPD1* gene encoding for the acid sphingomyelinase (ASM) on chromosome 11 lead to a toxic accumulation of sphingomyelin and other lipids within the cell causing malfunctions of multiple organ systems [Schuchman et al., 2007]. However, Type A and Type B differ in the presented clinical symptoms and severity based on the mutations in the *SMPD1* gene. Based on the mutations, the amount of the gene product varies in Type A and Type B [Schultz et al., 2016]. In Type A (classic infantile form) very low to no functional ASM (~1%) is retained and is associated with hepatosplenomegaly (abnormal growth of liver and spleen) and severe progressive neurological symptoms such as hypotonia, ataxia, spasticity which leads to the early death of the affected individuals by an age of 2 to 3 years [Vanier et al., 2013]. In Type B (visceral form) approximately ~10% of ASM is available, this type does not show any neurological pathology so the affected individuals survive until late adulthood with health complications arising from chronic hepatosplenomegaly, respiratory and cardiovascular diseases [Schuchman et al., 2015, Schultz et al., 2016]. Niemann-Pick Type C is caused by mutations in either one of the genes *NPC1* or *NPC2* in the affected individuals. The *NPC1* gene encodes for a lysosomal transmembrane protein and *NPC2* for a soluble lysosomal protein, both of these gene products work collaboratively by regulating the transport of intracellular cholesterol and thus play an important role in cholesterol homeostasis [Carstea et al., 1997; Loftus et al., 1997]. When this function is impaired it comes to an accumulation of unesterified cholesterol, sphingolipids, and other lipids in multiple organ systems. Thus, Niemann-Pick Type C is a slowly progressing neurodegenerative disease belonging to the aforementioned group of lysosomal storage diseases. *NPC1* is affected in 95% of the cases whereas *NPC2* is involved in 5% of the affected individuals, respectively [Patterson et al., 2012; Vanier et al., 2016]. The estimated lifetime prevalence varies on the causing gene which is 1:92,000 for *NPC1* and 1:2,860,000 for *NPC2* [Wassif et al., 2016]. The clinical spectrum is remarkably heterogeneous with manifestations being age-dependent. Most of the symptoms in early infancy are visceral with hepatosplenomegaly, jaundice, and pulmonary infiltrates. In later infancy, neurological manifestations start to dominate such as ataxia, dysarthria, seizures, dementia, or developmental delay [Pattersen et al., 2000]. In young adulthood (age > 15 years) neuropsychiatric symptoms like major depressive syndromes, sometimes bipolar disorder, or schizophrenia including psychosis overshade the neurological manifestations [Pattersen et al., 2000; Vanier et al., 2013; Kawazoe et al., 2018]. Due to clinically heterogeneous symptoms, the diagnostics of NP-C are challenging and long-lasting. Since the 90s, molecular diagnostic analyses based on the polymerase chain reaction method (PCR) and Sanger sequencing were applied for the diagnostics of NP-C. This was, in particular, used when biochemical biomarkers such as filipin testing were not convincing. However, in the end, genetic testing is crucial for high-precision diagnostics of NP-C to rule out any false-positive diagnoses. Nowadays targeted next-generation sequencing (NGS) or whole-exome sequencing of *NPC1* and *NPC2* may be applied [Bounford et al., 2914; Sitarska et al., 2019]. This saves costs because at the same time many other genes in a panel or exome can provide valuable insights into disease pathology. The *NPC1* and *NPC2* gene consists of 25 and 5 exons and are located on chromosome 18q11.2 and 14q24.3 [Carstea et al., 1997; Naureckiene et al., 2000]. At the time of writing, there were 185 and 30 pathogenic mutations for *NPC1* and *NPC2* listed in ClinVar [Landrum et al., 2018]. But rare private family-specific mutations including indels can cause the development of the disease as well. Further, mimicking psychiatric symptoms such as SCZ might lead to a misdiagnosis of NP-C patients [Kawazoe et al., 2018; Sandu et al., 2009]. Correctly diagnosing NP-C is crucial for the affected individuals as NP-C-specific therapies are available [Alavi et al., 2013]. Therefore, NGS-based targeted or whole-exome sequencing of all exons and exon/intron boundaries of the *NPC1* and *NPC2* genes is warranted and might become the method of choice in the diagnostics of Niemann-Pick disease [Málaga et al., 2019].

2.1.3 smMIP-based Targeted Sequencing

Single-molecule Molecular Inversion Probes (smMIPs) enable massive parallel targeted resequencing of genes by increasing throughput and reducing costs. This technology combines single-molecule tagging with a multiplex targeted capture that allows the detection of rare and low-frequency variations [O'Roak et al., 2012; Hiatt et al., 2013]. The smMIP technology provides many advantages including its flexible design meaning that genes can be easily integrated into an existing panel, the workflow is highly automatable, variant calling is highly reproducible due to high coverage, and its low costs compared to other commercial kits. The accurate genotyping of variants in subclonal frequencies is of importance in the detection of somatic mutations for example in cancer diagnostics e.g. *BRCA* testing [Neveling et al., 2017]. The smMIP technology is applied in the diagnostics of many other diseases such as male infertility, congenital disorders, hypopituitarism, dystonia, and many more indicating that the technology has already been established so far in clinical practice [Oud et al., 2017; Bakar et al., 2022; Millán et al., 2018; Pogoda et al., 2019]. However, the design of the smMIPs is a complex process in which the capture uniformity and specificity are evaluated using statistical models [Boyle et al., 2014]. The following steps are carried out (i) sequences corresponding to the targeted regions are extracted from the reference sequence, (ii) then it is checked whether any SNPs are located in the target region to place the probe arm in non-polymorphic sites, (iii) all targeting arms and insert sequences are checked for overlapping copy number variations using an alignment tool called Burrows-Wheeler Aligner (BWA, (iv) finally a machine learning approach called Library for Support Vector Machines (LIBSVM) is applied for scoring the best fitting combinations of the targeting arms. Then the MIP selection is guided by the scoring approach until an optimal MIP tiling that covers all targeted bases has been reached. Targets that cannot be tiled due to low complexity or specificity are accessed separately [Boyle et al., 2014]. Although the technology is highly scalable and demonstrated comprehensive multiplexing there are some important limitations such as the non-uniformity of the capture efficiency across long exons and GC-rich target regions. To address this, a re-pooling to balance the dropouts is necessary. This step can lower the turnaround time and increase a certain amount of the costs and resources for sequencing [Boyle et al., 2014]. While rare variants and *de novo* mutations contribute to the development of complex genetic disorders including neuropsychiatric disorders like bipolar disorder [Forstner et al., 2020; Toma et al., 2021], major depressive disorder [Zhou et al., 2021; Tombácz et al., 2019] or schizophrenia [Singh et al., 2016; Zoghbi et al., 2021; Purcell et al., 2014], implicated genes need to be resequenced and analyzed in large quantities of cases and control samples. The smMIP technology has tremendous potential to influence the diagnostics and therapy of different diseases on a large scale. Further, this technology might also replace conventional gold-standard Sanger sequencing by inaugurating a new era of NGS-based diagnostics [Diekstra et al., 2015; Beck et al., 2016].

2.1.4 Rare Variant Association Test

The missing heritability problem highlighted that common variants identified by GWAS only account partially for the heritability of complex genetic diseases [Eichler et al., 2010]. It is well known that numerous Mendelian disorders are caused by rare highly penetrant variants. There is increasing evidence that rare variants with a minor allele frequency (MAF) <1% in the general population contribute to the missing heritability of complex genetic disorders such as SCZ [Purcell et al., 2014]. The detection of rare variants became possible through the rapid advent of NGS technologies that were capable of identifying rare single nucleotide variants (SNVs) and copy number variations (CNVs). In particular, targeted resequencing and whole-exome sequencing (WES) are applied to investigate the role of putative rare functionally relevant variants in the coding region of the genome. The assessment of the pathogenicity (*e.g.* CADD score) of these variants is of major interest as they can be critical for correct diagnosis and therapy intervention [Kircher et al., 2014]. However, due to their rare allelic spectrum much larger sample sizes are needed than in GWAS for rare variant association testing. Therefore, typically rare variants are collapsed into genes, gene sets, or genomic regions for a particular disease of interest. The statistical testing is then based on a gene or genomic region rather than on a single variant level [Li et al., 2008; Madsen et al., 2009; Morgenthaler et al., 2007]. This integration of the effects of rare variants in units defined by gene annotations or genomic regions with functional impact improves the power to detect any associations. Because of the grouping of variants into genes, rare variant tests are also called gene-based tests. In principle, the frequencies of the individuals carrying rare variants in a gene or genomic region of interest are calculated in cases and controls and the differences in the frequencies of both of these groups are tested for quantitative or binary traits, respectively. Mainly, linear regression-based approaches are applied which enable the adjustment of the underlying models using covariates [Lee et al., 2014]. For gene-based tests, the genome-wide significance threshold is set to 2.5×10^{-6} assuming 20,000 genes in the human genome. In general, there exist two types of gene-based tests that are either burden or variance-component tests including methods that combine both of these approaches [Bansal et al., 2010]. These methods differ in the varying assumptions they make in the underlying genetic model and power. The classical Burden test is powerful when a large number of variants are causal and the effect sizes have the same direction. Whereas it is not of advantage when variants have different directionalities or weak effect sizes. In burden tests, usually, variants are collapsed into a single genetic score which tests for association under a dominant rather additive genetic model [Li et al., 2008; Madsen et al., 2009; Morgenthaler et al., 2007]. The Variance-Component test takes into account a random-effect model. Instead of collapsing all variants into a single gene-based score, the distribution of aggregated score statistics of individual variants is assessed. This method is powerful when it comes to variants with mixed effects that are pathogenic or protective at the same time, but is less powerful when most of the variants are causal and the effects are in the same direction [Wu et al., 2011; Neale et al., 2011]. One of the most applied methods of the Variance-Component test is the Sequence Kernel Association Test (SKAT) [Wu et al., 2011]. Its frequently applied as it is computationally efficient and enables modeling epistatic effects (SNP-SNP interactions). Several other methods have been introduced to unify burden and variance-component tests as implemented in SKAT-O. This approach maximizes the power by adaptively selecting the best linear combination of the burden and non-burden SKAT [Lee et al., 2012]. Hence, it is more robust in terms of causal variants that can have different directions. Usually, this type of approach is of advantage as the information on the underlying genetic model is often unknown. Thus, for these reasons, SKAT-O was the tool of choice for the rare variant association test of rare functionally relevant variants in *NPC1* and *NPC2* of SCZ patients and controls.

2.1.5 Aim of the Study

The present study had three main aims: (i) to test the hypothesis that functionally relevant variants in *NPC1* and *NPC2* are enriched in patients with SCZ compared to controls; (ii) to screen a large cohort of patients diagnosed with SCZ (NP-C) and (iii) to implement an NGS-based routine diagnostics pipeline for the targeted resequencing of *NPC1* and *NPC2*.

2.2 MATERIALS & METHODS

2.2.1 Cohort Description

The study was approved by the respective ethics committee. All individuals provided written informed consent before inclusion. All study procedures were performed following the Code of Ethics of the World Medical Association [World Medical Association, 2013]. All participants were of German descent according to self-reported ancestry. DNA was extracted from whole venous blood. In total, 1,947 patients with SCZ and 1,921 controls were included in this study. All patients were recruited from departments of psychiatry across Germany. They were assessed by an experienced psychiatrist. The minimum assessment included medical records, family history, and performance of the Structured Clinical Interview [Spitzer et al., 1992]. A lifetime "best estimate" diagnosis of SCZ was assigned following the International Statistical Classification of Diseases and Related Health Problems ICD-10 criteria [Leckman et al., 1982]. The controls were recruited at different sites in Germany. They included individuals from the Heidelberg Cohort Study of the Elderly [HeiDE; Amelang et al., 2004] and the Heinz Nixdorf recall study [HNR; Schmermund et al., 2006].

2.2.1 Design of single molecule Molecular Inversion Probes (smMIPs)

The genes *NPC1* and *NPC2* comprised of 25 and 5 coding exons, respectively (target region). In addition, 2 intronic regions in *NPC1* were included for known pathogenic variants. The target region was resequenced using *smMIPs* [Hiatt et al., 2013]. The *smMIPs* were designed using an in-house pipeline based on the program MIPgen [Boyle et al., 2014] with the following modifications: (i) without untranslated regions (UTR); (ii) target size 190 bp (170-210 bp); (iii) target regions: all exons ±5 bp. The logistic priority score was 0.5. In total, 48 and 6 smMIPs were designed, respectively. All 4,936 bp of the 30 exons, 2 introns in *NPC1, NPC2,* and the corresponding splice sites were *in silico* sufficiently covered by the designed probes.

2.2.3 Library Preparation and Sequencing

Library preparation was performed as described elsewhere [Eijkelenboom et al., 2016]. Briefly, a total of 100ng of genomic DNA served as input material. After denaturation, incubation of the probe mix for hybridization, extension, and ligation is carried out before exonuclease treatment. Exonuclease-treated capture is then used for PCR with common forward and barcoded reverse primers. After pooling and purification of the PCR products to ensure a homogeneous read depth across all target regions, the concentration of the different *smMIPs* was adjusted according to their performance in a test run using Illumina's MiSeq. Purified and diluted libraries were then sequenced using Illumina's HiSeq 2500 resulting in 2 × 150 bp paired-end reads [Hiatt et al., 2013]. Equivalent amount of patient and control samples were pooled prior sequencing of each flowcell.

2.2.4 Sample-level QC - SNP Genotyping Data

For all individuals presented in this study, data from Illumina SNP genotyping arrays were available. The genotyping information was used to detect sex mismatches, duplicates, relatedness, and population outliers. The latter sample QC steps were performed using PLINK [Purcell et al., 2007] and KING [Manichaikul et al., 2010]. Standard QC parameters as defined in PLINK were applied to the SNP dataset. Due to the low number of overlapping variants, we were unable to assess sequencing-array genotype concordance. Samples from our targeted sequencing cohort were excluded if they: (i) showed a mismatch between X-chromosomally inferred and phenotypic sex; (ii) had a kinship coefficient >0.0884; were identified as population outliers from the mean distance of > 3 standard deviations (SD) in the first two components according to multidimensional scaling (MDS) analysis using 1000 Genomes Project Phase 3 [The 1000 Genomes Project Consortium; Auton et al., 2015].

2.2.5 Sequencing Data - Variant Calling and Annotation

Raw sequence reads were processed according to a previously published pipeline [Hiatt et al., 2013]. Briefly, the pre-processing of the raw data includes merging overlapping regions of read-pairs using PEAR [Zhang et al., 2013]. The extension and ligation arm (smMIP tags) are inserted into the header of the read-pairs applying a custom python script from the MIPgen repository (https://github.com/shendurelab) called "mipgen_fq_cutter_pe.py". Mapping of the read-pairs to the human genome assembly GRCh37 (hg19) is carried out using the Burrow-Wheeler Alignment tool (BWA) [Li et al., 2009]. Collapsing the smMIP tags and removing extension and ligation arms is performed using a custom python script called "mipgen_smmip_collapser.py" which was also available from the MIPgen repository. Variant calling was performed using the Unified Genotyper from the Genome Analysis Tool Kit (GATK) [Van der Auwera et al., 2013]. Variants were annotated using Annovar [Wang et al., 2010].

2.2.6 Sequencing Data - Sample Level Quality Control

Samples were excluded from downstream analyses if they fulfilled at least one of the following criteria (i) <90% of the target sequence covered at $\geq$10x as calculated using Picard [Picard Tool Kit, 2018]; and (ii) $\geq$3 SD in Het/Hom and/or Ti/Tv Ratios as calculated by BCFTools [Li et al., 2011].

2.2.7 Variant-Level QC and Filtering

Multi-allelic variants were left normalized according to BCFTools [Li et al., 2011] Then, all variants were filtered using the VariantFiltration tool in GATK applying hard-filtering thresholds [Van der Auwera et al., 2013]. Sequencing-based variants in the VCF file were converted to PLINK's binary format using PLINK [Purcell et al., 2007]. Variants were excluded if the missing genotype rate was >10% and/or fail the Hardy-Weinberg test ($p<1\times10^{-6}$) in the study's control cohort using PLINK [Purcell et al., 2007]. All synonymous variants were excluded. Subsequently, the non-synonymous variants were filtered according to their minor allele frequency (MAF) and their putative *in silico* functional effect. Variants were excluded from our downstream analyses if they had a MAF > 0.001 in either the present study's cohort and/or the Non-Finish European (nonPsych) subset of the Exome Aggregation Consortium [ExAC; Lek et al., 2016] (n=45,376). Variants with a Combined Annotation Dependent Depletion [CADD; Kircher et al., 2014] score < 20 were discarded. No CADD scores were available for frameshift indels but these were included in downstream analyses due to their protein-truncating effects.

2.2.8 Statistical Analysis

We applied the optimized SKAT-O which is a gene-based kernel-regression association test for rare variants [Lee et al., 2012]. To get model parameters and residuals for the SKAT-O, the SKAT null model for a binary trait without any corrections for covariates was assigned. SKAT-O default parameters were used: (i) exclusion of non-polymeric variants and (ii) variants with a missing rate >0.15 [Lee et al., 2012]. Using a linear-weighted kernel we carried out association tests between SCZ patients and controls for each gene separately. Correction for multiple testing was performed using the Benjamini-Hochberg method [Benjamini et al. 1995]. A *p*-value significance threshold for gene-based testing (0.05/2) of *p*<0.025 was considered. Additionally, a classical burden-like test statistic by counting heterozygous, compound heterozygous, and homozygous carriers in patients and controls were generated to run a one-sided Fisher`s exact test on a 2 x 2 contingency table. First, an SNP file with qualifying variants that are potentially functional according to the aforementioned filter criteria was created in which variants are assigned to genes. Then the allele counts of carriers in the cases and controls are assessed using a custom script separately. Finally, a table is generated in which individual carriers of heterozygous, compound heterozygous, and homozygous variants in cases and controls are listed. This table serves as input for the one-sided Fisher`s exact test which is run under a dominant or recessive model. Furthermore, to access the significance of single markers in SCZ patients versus controls, an association test based on allele frequency distributions using a Pearson's Chi-square test as implemented in PLINK [Purcell et al., 2007] was carried out. All analyses were based on MAF<0.001 and CADD score >20 including frameshift variants without any CADD score. All statistical analyses were performed in R v3.6.3 [R Core Team, 2017] and PLINK [Purcell et al., 2007].

2.3 RESULTS

2.3.1 Integration of SNP Array Data

We included SNP genotyping array data for checking sex mismatches, duplicates, relatedness, and population stratification. No sex mismatches and duplicate samples were found. 115 related pairs of individuals up to 2nd degree (kinship coefficient >0.0844) were identified by estimating the kinship coefficients (Figure 1). These 115 individuals were discarded from all downstream analyses.

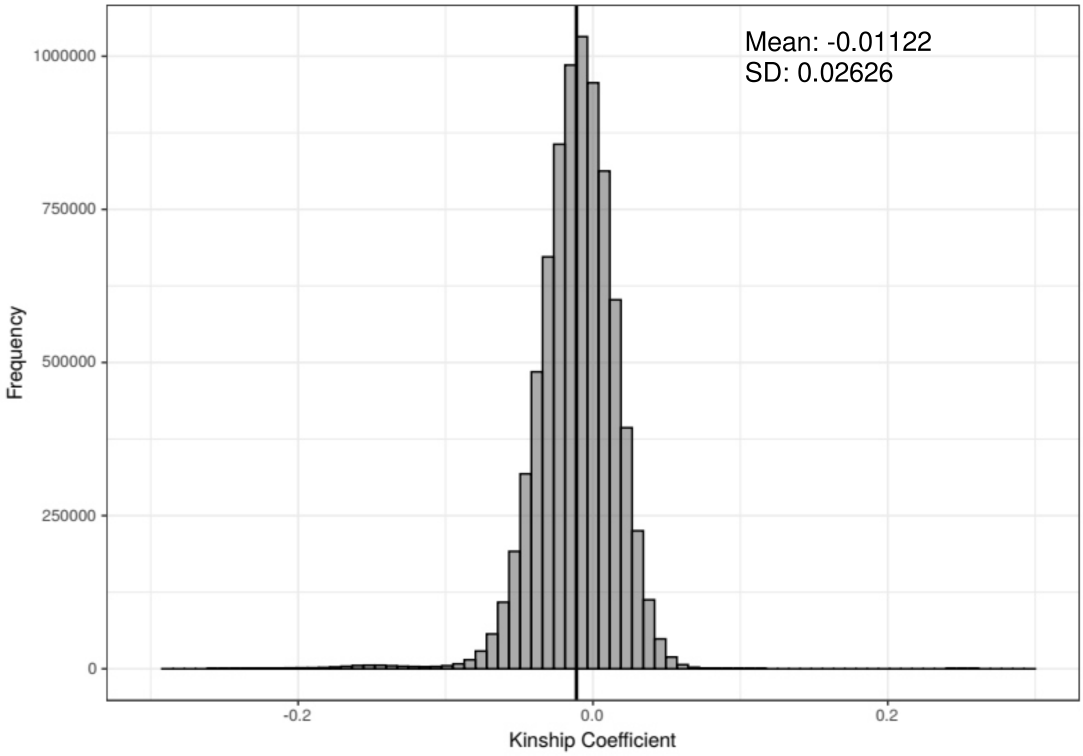

Figure 1 | Kinship Coefficient. The histogramm shows the distribution of kinship coefficient estimates as calculated by the relationship inference tool KING. The mean and standard deviation from the distribution are specified. The straight black line indicates the mean of the distribution.

Additionally, 70 population outliers (>3 SD) were identified which were also excluded from all downstream analyses (Figure 1). For that 3 standard deviations were applied to the first and second components of the MDS analysis and population outliers were detected and visualized as presented in Figure 2.

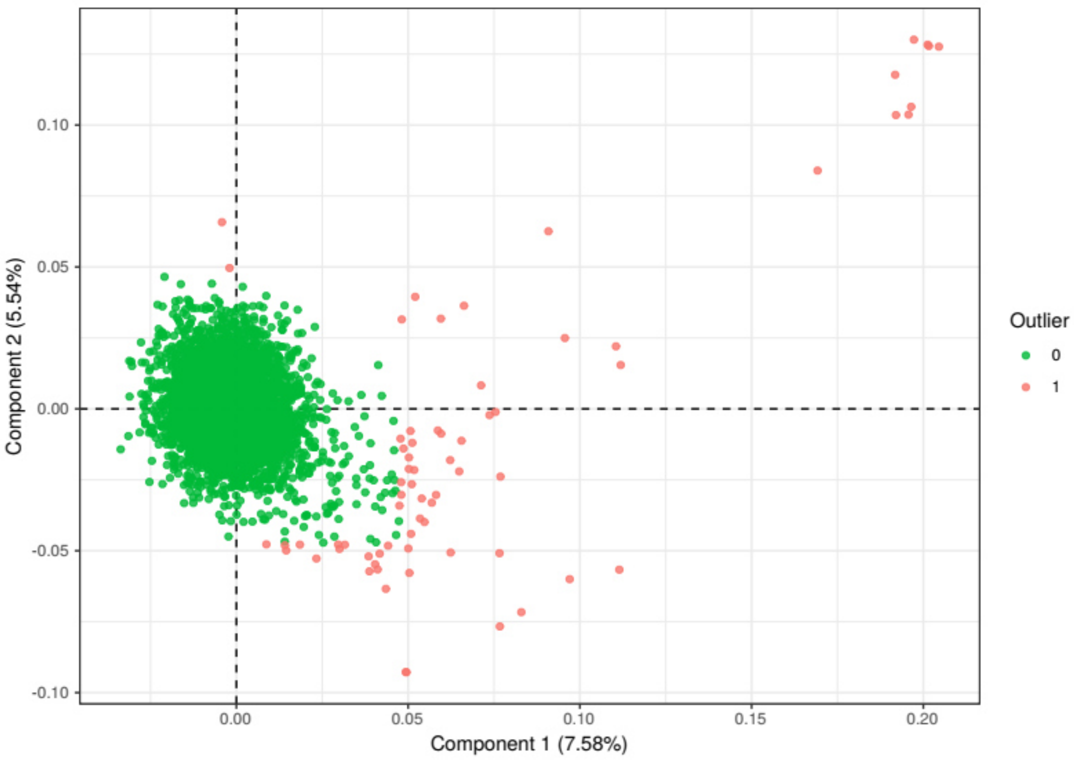

Figure 2 | Multidimensional Scaling Plot. MDS was performed using PLINK from genome-wide genotype data of the present cohort. First and second component of the MDS analysis are plotted. Population outlier as defined by >3 SD in the cohort are indicated in red. The green cluster represents a distinct cohort and is within <3 SD on the first and second component. The proportion of variance explained by each of the components is specified in brackets.

To capture the population structure along with other ancestries, we performed an MDS analysis using the 1000 genomes reference data. Only variants that overlap with the 1000 genomes reference and the present cohort were taken into account for MDS analysis. The combined analysis with the 1000 genomes data indicated for the rest of the study cohort a central European origin (blue, Figure 3) and a homogenous population (green, Figure 3) which is important in genetic association studies to reduce false-positive findings. After stringent GWAS-like quality control steps, we identified 115 related individuals and 70 population outliers. By integrating SNP array genome-wide genotype data population stratification using common variants could be carried out. Overall 185 from 3,868 (4.8%) samples were discarded from the smMIPs data for the rare variant association analysis in *NPC1* and *NPC2*. These QC steps would not have been possible without the use of genome-wide SNP array data.

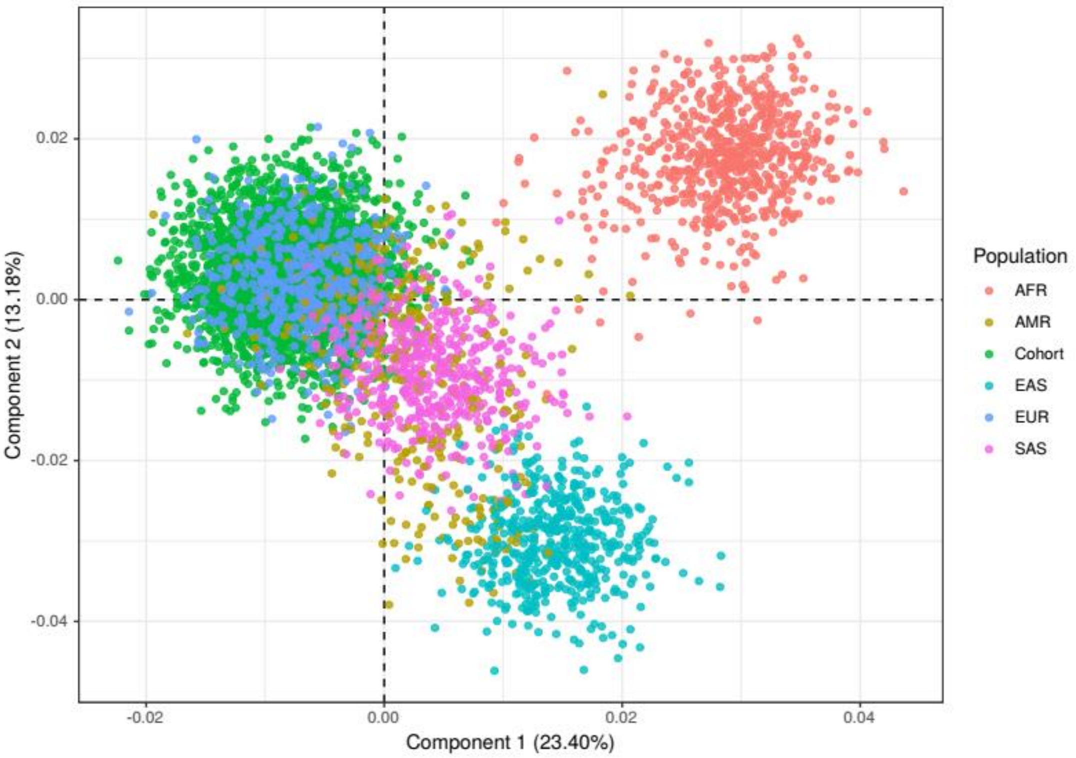

Figure 3 | Multidimensional Scaling Plot using 1000 Genomes data. MDS was performed using PLINK from genome-wide genotype data of the present cohort and the 1000 Genomes reference data. First and second component of the MDS analysis are plotted. The proportion of variance explained on each of the components is indicated in brackets.

2.3.2 Targeted Sequencing using smMIPs

We selected two genes *NPC1* and *NPC2* for the targeted sequencing of 1,947 schizophrenia patients and 1,921 controls (pre-QC). 48 and 6 smMIPs were designed to sequence coding (25 and 5 exons), splicing (exons ±5 bp), and 2 intronic regions in *NPC1*, totaling in 4,936 target bases. The mean coverage of *NPC1* and *NPC2* for all samples was 269.45 (sd=79.68) and 214.37 (sd=63.06) (Figures 4 and 5). When comparing the mean coverage stratified by patients and controls for *NPC1* and *NPC2* there were differences in the overall coverage distribution (Figures 4 and 5). In both genes, there was a small proportion of control samples with less coverage than the patients, indicating cohort-specific effects of the target region captured by the smMIPs technology. The patients had an overall better coverage than the control samples.

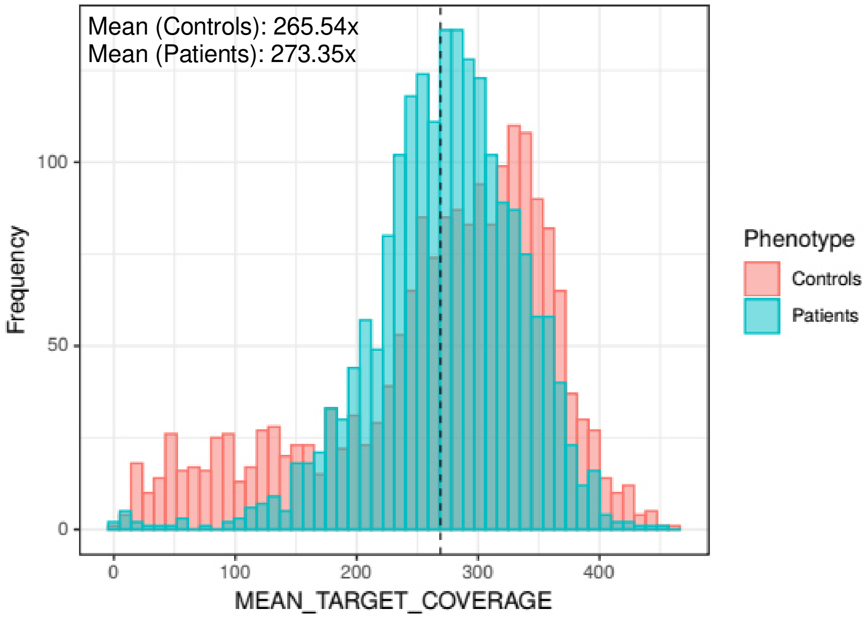

Figure 4 | Mean Coverage Plot of *NPC1*. The mean coverage of the target regions of the *NPC1* gene is depicted as a histogram. The red and light blue distributions represent the coverage stratified by controls and patients. The black dashed line indicates the overall mean of controls and patients. The control and patients-specific means are shown in the upper left corner of this chart.

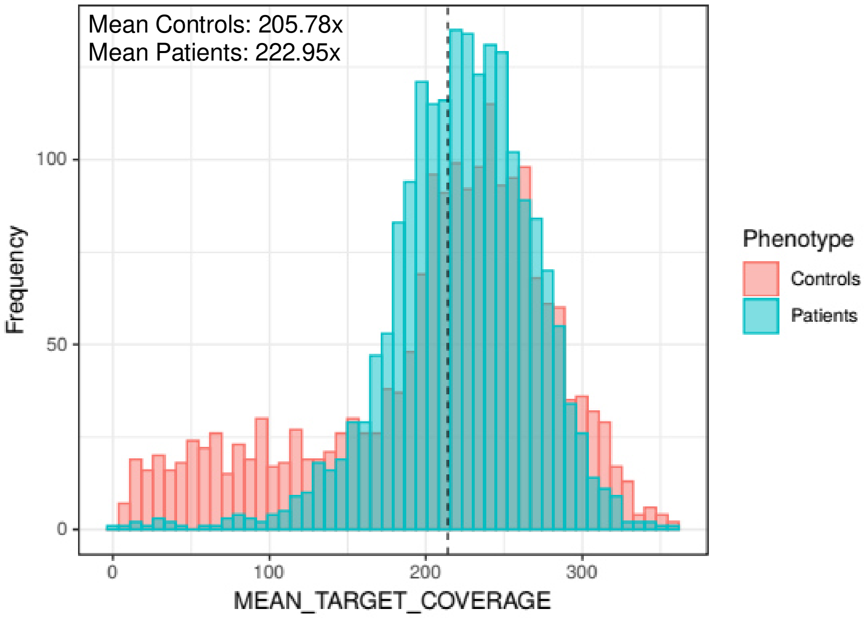

Figure 5 | Mean Coverage Plot of *NPC2*. The mean coverage of the target regions of the *NPC2* gene is depicted as a histogram. The red and light blue distributions represent the coverage stratified by controls and patients. The black dashed line indicates the overall mean of controls and patients. The control and patients-specific means are shown in the upper left corner of this chart.

For sample-based quality control of the targeted sequencing data, the threshold of 90% of the target region covered at 10x was checked for both genes together. Overall, 46 from 3,683 (1.25%) samples with a target region coverage <90% of 10x were identified and discarded from downstream analyses (Figure 5). In *NPC1*, 133 bp from 4,430 bp (3%) from the targeted bases were not covered at >10x in the entire cohort, and in *NPC2*, 91 bp from 506 bp (17.98%) from the targeted bases were not covered at >10x. Meaning the overall target region capture efficiency of the smMIPs was better in *NPC1* than *NPC2*. Despite, the differences the majority of the samples (98.75%) attain 90% of the 10x target region-based coverage threshold indicating an overall good performance of the smMIP sequencing on the sample level (Figure 5).

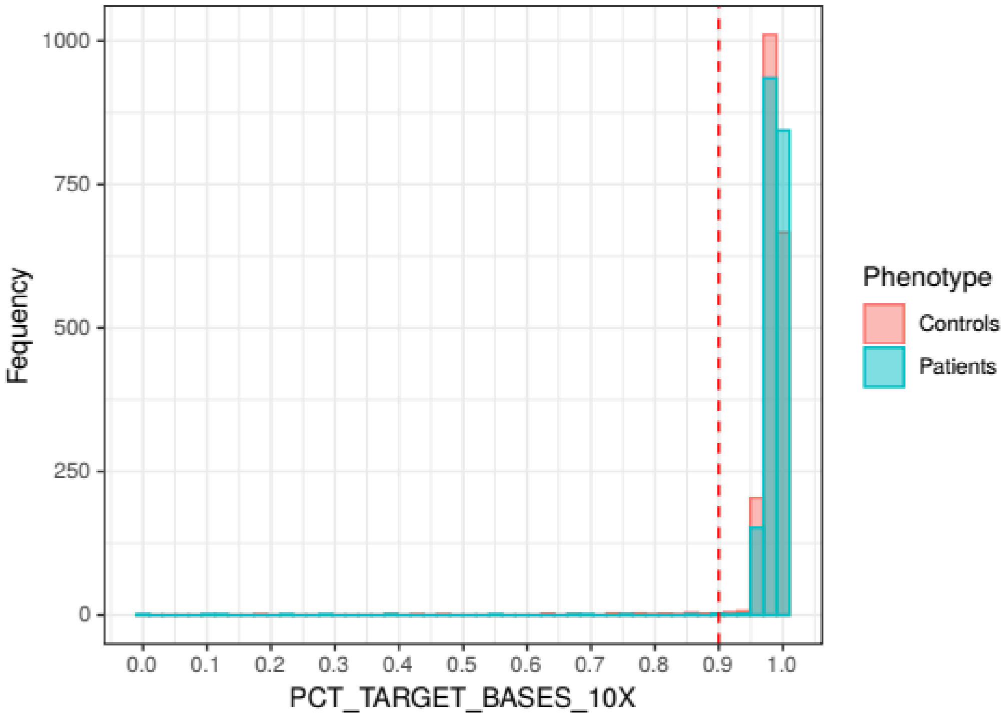

Figure 6 | 10x Target Region Coverage. In this histogram the distribution of 90% of the target region covered at 10x are plotted. The red and light blue distributions represent the controls and patients from the study cohort. The red dashed line indicates the 90% cutoff threshold.

The exon-based mean coverage analysis revealed that exon 8 in *NPC1* was fully affected by the loss of coverage meaning all targeted bases were covered <10x. Further, partially affected by the coverage loss were exon 11 (0.8% <10x), exon 24 (39% <10x) and one of the intronic regions (26% <10x) (Figure 7). In *NPC2* exon 4 (90% <10x) was fully affected by the loss of coverage (Figure 8). This revealed that some of the targeted regions such as exon 8 in *NPC1* and exon 4 in *NPC2* were not sufficiently covered for downstream analysis in particular for variant calling thus the contribution of these regions to the rare variant burden analysis could not be assessed. The loss of coverage of these particular exons and regions needs to be further investigated and evaluated.

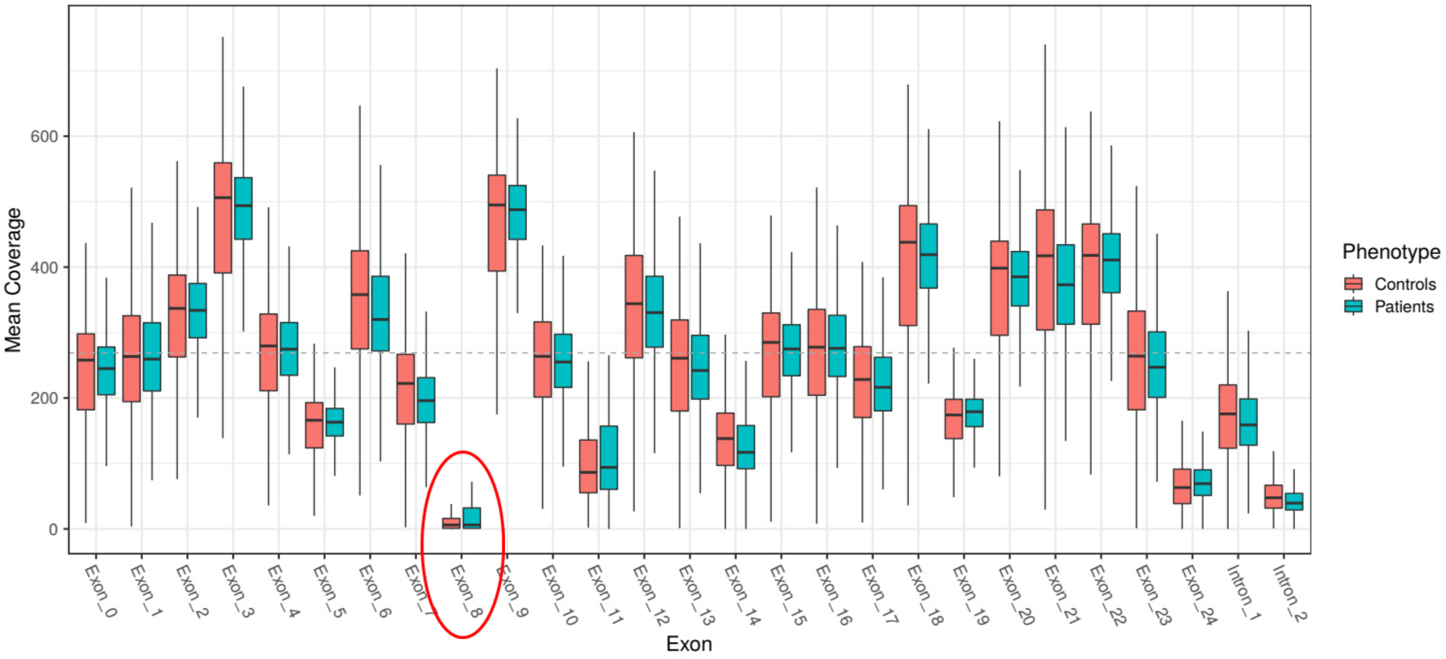

Figure 7 | Exon-based mean coverage of *NPC1*. The mean coverage depth of all coding exons and 2 intronic regions in *NPC1* are plotted by stratifiying controls and patients. The light grey dashed line indicates the overall mean coverage. The red circle outlines the failed exon 8 in *NPC1*.

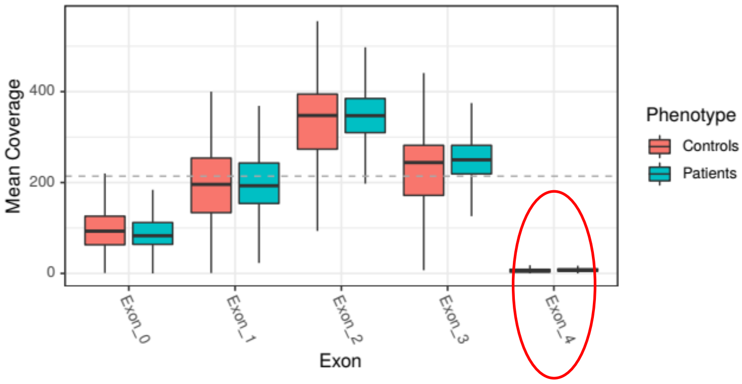

Figure 8 | Exon-based mean coverage of *NPC2*. The mean coverage depth of all exons in *NPC2* are plotted by stratifiying controls and patients. The light grey dashed line indicates the overall mean of the coverage. The red circle outlines the failed exon 4 in *NPC2*.

None of the other important quality control parameters on the sample level such as individual missing genotype data >15%, Het-/Hom- or Ti-/Tv-ratio (>3 SD) were exceeded. The mean Het-/Hom- and Ti/Tv-ratios were 1.403 and 2.09. After vigorous QC, data from a total of 1,815 patients and 1,831 controls were included in downstream analyses. Overall 222 (5.74%) from 3,868 samples were discarded due to quality control issues.

2.3.3 Variant Prioritization

In total, 164 variants in *NPC1* (91.5%) and *NPC2* (8.5%) were identified. Among the 164 variants, 2 MNPs (multiallelic SNP sites), 155 SNVs (single nucleotide variants), and 7 Indels (insertion-deletions) were found (Table 1). Figure 9 highlights the main variant prioritization steps in a schematic overview. The overall genotype concordance with dbSNP (v144) within the sufficiently covered target regions (>10x) was 100%. After applying left normalization, GATK hard-filter, filtering for non-polymorphic sites, missing genotype rate (>15%), and MAF >0.001, 94 SNVs and 6 indels were left in *NPC1* (94%) and *NPC2* (6%) for the identification of rare potentially functional variants (Table 1). In particular, we included variants that had a MAF <0.001 in the study's cohort or the Non-Finish European (nonPsych) cohort of ExAC which were nonsynonymous with a PHRED-scaled CADD score >20, and frameshift mutations for which no CADD scores were available. Overall, in *NPC1* three frameshift deletions (6.5%), two frameshift insertions (4.3%), thirty-five non-synonymous (76.1%), and two stopgain (4.3%) variants were detected. The thirty-five non-synonymous and two stopgain variants had a CADD score >20 (Table 2). For two of the frameshift insertions, no CADD score was available. However, nineteen non-synonymous variants in *NPC1* had a CADD score <20 and were not included in downstream analyses. In *NPC2*, we distinguished between one frameshift deletion (2.2%), two non-synonymous (4.3%), and one stopgain (2.2%) mutation. The two non-synonymous and one stopgain mutation had a CADD score >20 (Table 2) and for the frameshift deletion, no CADD score was accessible. For the statistical analyses, overall 42 and 4 rare variants with a high impact on the protein-level and potentially functionally relevant in *NPC1* and *NPC2* were chosen (Tables 2 and 3). In total, 46 (28%) from 164 variants were kept for all downstream analyses.

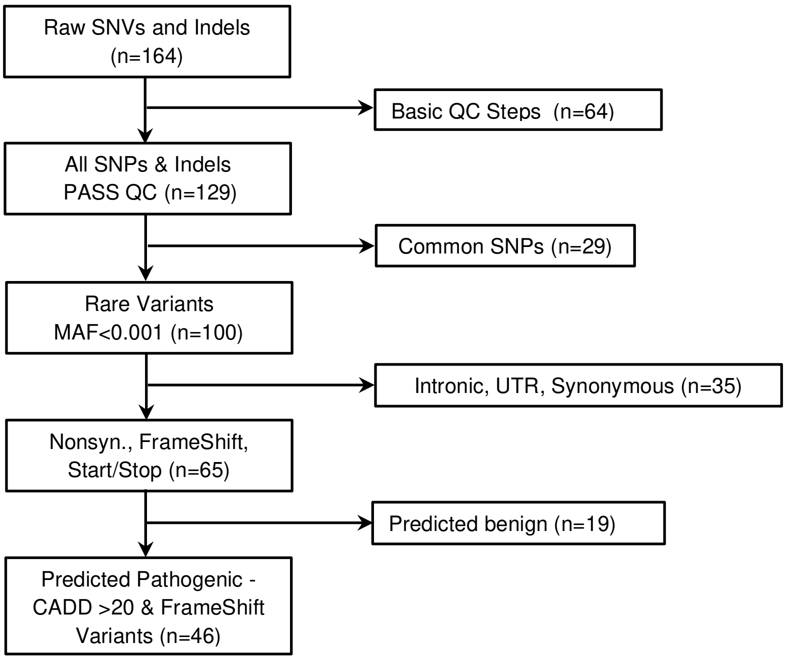
Raw SNVs and Indels (n=164)
Basic QC Steps (n=64)
All SNPs & Indels PASS QC (n=129)
Common SNPs (n=29)
Rare Variants MAF<0.001 (n=100)
Intronic, UTR, Synonymous (n=35)
Nonsyn., FrameShift, Start/Stop (n=65)
Predicted benign (n=19)
Predicted Pathogenic - CADD >20 & FrameShift Variants (n=46)

Figure 9 | Workflow overview for rare variant prioritization in *NPC1* and *NPC2*. The flowchart highlights the prioritization of rare variants for the SKAT-O and Burden Test. The numbers in the brackets represent SNPs and Indels in *NPC1* and *NPC2* together.

Table 1 | Overview of variant prioritization in *NPC1* and *NPC2*. Basic quality control steps on the sample- and variant-level are shown. The total amount of discarded samples and variants are summarized. The numbers in the brackets represent variants in *NPC1* and *NPC2* separately.

Filtering Steps	Variants	Variants Removal	Samples	Sample Removal
Initial Raw Data	164 (150/14)	0	3,868	0
Samples - Array QC	164 (150/14)	0	3,683	185
Samples - MIPs QC	164 (150/14)	0	3,646	37
Samples - Het-/Hom (3 SD)	164 (150/14)	0	3,646	0
Samples - Ti-/Tv (3 SD)	164 (150/14)	0	3,646	0
Samples - Missingness 15%	164 (150/14)	0	3,646	0
Variants - GATK Hardfilter	142 (128/14)	22	3,646	0
Variants - Non-Polymorphic	132 (120/12)	10	3,646	0
Variants - Missingness 15%	129 (117/12)	3	3,646	0
Variants - MAF >0.001	100 (92/8)	29	3,646	0
Variants - Intronic, UTR, Synonymous	65 (61/4)	35	3,646	0
Variants - Benign	46 (42/4)	19	3,646	0
Sum	**46 (28%)**	**118 (72%)**	**3,646 (94%)**	**222 (6%)**

Table 2 | List of rare variants in *NPC1* and *NPC2*. All variants prioritized for the rare variant association test are listed in this table.

Gene	Position	Change	Exon	Exonic Function	Impact	Clinical Significance	Internal MAF	External MAF	CADD	MAC$_{Pat./Cont.}$
NPC1	18:21113384	c.T3689C	exon24	nonsynonymous SNV	MODERATE	0	0.000274273	0.000094990	26.3	1/1
NPC1	18:21113400	c.3666_3672del	exon24	frameshift deletion	HIGH	Likely_pathogenic	0.000137137	0	0	1/0
NPC1	18:21113459	c.C3614A	exon24	nonsynonymous SNV	MODERATE	Pathogenic/Likely_pathogenic	0.000137137	0.000023770	28.8	0/1
NPC1	18:21114441	c.C3560T	exon23	nonsynonymous SNV	MODERATE	Uncertain_significance	0.000137137	0.000100000	23.7	1/0
NPC1	18:21114444	c.G3557A	exon23	nonsynonymous SNV	MODERATE	Pathogenic/Likely_pathogenic	0.000137137	0.000047800	25.8	0/1
NPC1	18:21114451	c.G3550A	exon23	nonsynonymous SNV	MODERATE	Uncertain_significance	0.000137137	0.000071690	23.8	1/0
NPC1	18:21114508	c.G3493A	exon23	nonsynonymous SNV	MODERATE	Conflicting_interpretations	0.000137137	0.000025010	31	1/0
NPC1	18:21115468	c.A3442G	exon22	nonsynonymous SNV	MODERATE	0	0.000137137	0	24.1	1/0
NPC1	18:21115582	c.T3328C	exon22	nonsynonymous SNV	MODERATE	Uncertain_significance	0.000137137	0.000023750	25.7	1/0
NPC1	18:21115621	c.G3289A	exon22	nonsynonymous SNV	MODERATE	Likely_pathogenic	0.000137137	0	29.8	1/0
NPC1	18:21116679	c.C3203T	exon21	nonsynonymous SNV	MODERATE	Uncertain_significance	0.000274273	0.000095410	23.7	0/2
NPC1	18:21116700	c.T3182C	exon21	nonsynonymous SNV	MODERATE	Pathogenic	0.000137137	0.000500000	23.9	0/1
NPC1	18:21116722	c.G3160A	exon21	nonsynonymous SNV	MODERATE	Pathogenic	0.000137137	0	26.8	1/0
NPC1	18:21118573	c.2972_2973del	exon20	frameshift deletion	HIGH	Conflicting_interpretations	0.000274273	0	0	1/1
NPC1	18:21119357	c.G2873A	exon19	nonsynonymous SNV	MODERATE	Uncertain_significance	0.000137137	0.000048640	24.8	0/1
NPC1	18:21119358	c.C2872T	exon19	stopgain	HIGH	Pathogenic	0.000137137	0	33	1/0
NPC1	18:21119369	c.C2861T	exon19	nonsynonymous SNV	MODERATE	Pathogenic	0.000137137	0.000073210	24	0/1
NPC1	18:21119411	c.C2819T	exon19	nonsynonymous SNV	MODERATE	Pathogenic/Likely_pathogenic	0.000137137	0.000025420	31	0/1
NPC1	18:21119429	c.G2801A	exon19	nonsynonymous SNV	MODERATE	Pathogenic/Likely_pathogenic	0.000137137	0	23.5	1/0
NPC1	18:21119430	c.C2800T	exon19	stopgain	HIGH	Pathogenic	0.000137137	0.000026480	39	0/1
NPC1	18:21119793	c.C2777T	exon18	nonsynonymous SNV	MODERATE	Pathogenic/Likely_pathogenic	0.000137137	0.000023790	28.4	1/0
NPC1	18:21119842	c.G2728A	exon18	nonsynonymous SNV	MODERATE	Pathogenic/Likely_pathogenic	0.000137137	0.000023750	25.6	1/0
NPC1	18:21119949	c.A2621T	exon18	nonsynonymous SNV	MODERATE	Pathogenic/Likely_pathogenic	0.000137137	0.000047670	25.2	1/0
NPC1	18:21121091	c.C2455T	exon16	nonsynonymous SNV	MODERATE	0	0.000137137	0.000023730	22.9	1/0
NPC1	18:21121147	c.G2399A	exon16	nonsynonymous SNV	MODERATE	0	0.000137137	0	24.5	0/1
NPC1	18:21123467	c.2196dupT	exon14	frameshift insertion	HIGH	Pathogenic	0.000137137	0.000023810	0	1/0

Legend Table 2: Impact: Severity of variant consequence as assessed by VEP (Variant Effect Predictor); Clinical Significance: ACMG terms provided by ClinVar; Internal MAF: Minor allele frequency in the study's cohort; External MAF: ExAC nonpsych non-finish European (NFE) MAF; CADD: Phred-based CADD score; MAC: Minor allele count in SCZ patients versus controls.

Table 2 cont. | List of rare variants in *NPC1* and *NPC2*. All variants prioritized for the rare variant association test are listed in this table.

Gene	Position	Change	Exon	Exonic Function	Impact	Clinical Significance	Internal MAF	External MAF	CADD	MAC$_{Pat./Cont.}$
NPC1	18:21124366	c.C2072T	exon13	nonsynonymous SNV	MODERATE	Pathogenic/Likely_pathogenic	0.000137137	0	29.1	1/0
NPC1	18:21124441	c.G1997A	exon13	nonsynonymous SNV	MODERATE	0	0.000137137	0	23	1/0
NPC1	18:21124448	c.G1990A	exon13	nonsynonymous SNV	MODERATE	Pathogenic/Likely_pathogenic	0.000137137	0.000023780	25.6	0/1
NPC1	18:21124988	c.A1883C	exon12	nonsynonymous SNV	MODERATE	Likely_pathogenic	0.000137137	0	24.9	0/1
NPC1	18:21125112	c.T1759C	exon12	nonsynonymous SNV	MODERATE	0	0.000137137	0	24.4	1/0
NPC1	18:21134845	c.C1430T	exon9	nonsynonymous SNV	MODERATE	Conflicting_interpretations	0.000137137	0.000071200	21.7	1/0
NPC1	18:21134854	c.C1421T	exon9	nonsynonymous SNV	MODERATE	Pathogenic	0.000137137	0.000023730	23.7	0/1
NPC1	18:21134927	c.A1348G	exon9	nonsynonymous SNV	MODERATE	Conflicting_interpretations	0.000274273	0.000095620	20.6	1/1
NPC1	18:21137120	c.G916A	exon7	nonsynonymous SNV	MODERATE	Uncertain_significance	0.000137137	0	23.6	0/1
NPC1	18:21137150	c.C886T	exon7	nonsynonymous SNV	MODERATE	0	0.000137137	0.000048520	31	0/1
NPC1	18:21140387	c.C689A	exon6	nonsynonymous SNV	MODERATE	0	0.000274273	0	24.2	0/2
NPC1	18:21141335	c.616_619del	exon5	frameshift deletion	HIGH	0	0.000137137	0	0	1/0
NPC1	18:21141414	c.G541A	exon5	nonsynonymous SNV	MODERATE	Uncertain_significance	0.00041141	0.000071190	23.5	2/1
NPC1	18:21141488	c.T467C	exon5	nonsynonymous SNV	MODERATE	Uncertain_significance	0.000137137	0.000100000	21.5	1/0
NPC1	18:21148814	c.435dupA	exon4	frameshift insertion	HIGH	0	0.000137137	0	0	0/1
NPC1	18:21153427	c.G169A	exon2	nonsynonymous SNV	MODERATE	0	0.000137137	0.000047460	23.3	1/0
NPC2	14:74947469	c.376delG	exon4	frameshift deletion	HIGH	0	0.000137137	0	0	1/0
NPC2	14:74951129	c.G352T	exon3	stopgain	HIGH	Pathogenic	0.000137137	0.000047460	41	1/0
NPC2	14:74951148	c.T333G	exon3	nonsynonymous SNV	MODERATE	Uncertain_significance	0.000137137	0.000023730	24.1	0/1
NPC2	14:74953113	c.G109A	exon2	nonsynonymous SNV	MODERATE	0	0.000137137	0	23.6	1/0

Legend Table 2: Impact: Severity of variant consequence as assessed by VEP (Variant Effect Predictor); Clinical Significance: ACMG terms provided by ClinVar; Internal MAF: Minor allele frequency in the study's cohort; External MAF: ExAC nonpsych non-finish European (NFE) MAF; CADD: Phred-based CADD score; MAC: Minor allele count in SCZ patients versus controls.

2.3.4 Rare Variant Association Test

To examine whether rare functionally relevant variants in *NPC1* and *NPC2* may cause schizophrenia, we carried out a SKAT-O and a classical burden (Fisher's exact) test. In total, 42 and 4 potentially functional variants in *NPC1* and *NPC2* from 1,815 schizophrenia patients and 1,831 controls were included in these analyses. In *NPC1*, 49 heterozygote carriers were identified in schizophrenia patients (51%) and controls (49%). In *NPC2*, overall 4 heterozygote carriers were detected in patients (75%) and controls (25%). Except for 3 variants (c.C3203T; c.C689A; c.G541A), all were singletons and observed only once in patients or controls (Tables 2 and 3). None of the samples in the cohort neither in patients nor in controls were enriched for homozygous variants. According to the SKAT-O and one-sided Fisher's exact test, both genes *NPC1*, as well as NPC2, were not significantly associated with schizophrenia. For *NPC1* the SKAT-O and Fisher's exact test revealed a *p*-value of 0.9289 and 0.4885. For *NPC2,* both tests revealed a *p*-value of 0.2487 and 0.3095 (Table 4). A lookup in the currently largest exome sequencing meta-analysis of SCZ (SCHEMA) revealed non-significant findings for *NPC1* (*p*=0.153) and *NPC2* (*p*=0.206) as well.

Table 3 | SKAT-O and Fisher exact test for rare variant association test in *NPC1* and *NPC2*.

Gene	Marker All	Marker Test	$\text{MAC}_{\text{Pat./Cont.}}$	$\text{M}_{\text{Pat./Cont.}}$	P_{SKATO}	P_{Burden}
NPC1	42	42	49 (25/24)	49 (25/24)	0.9289	0.4885
NPC2	4	4	4 (3/1)	4 (3/1)	0.2487	0.3095

Legend Table 4: Marker All: Number of the SNPs in the gene; Marker Test: Number of SNPs used for the test; MAC: Minor allele count; M: Number of individuals with minor allele;

2.3.5 Single Marker Association Test

Finally, to access the effect of rare functionally relevant variants in *NPC1* and *NPC2* associated with schizophrenia, a single marker association test was applied using Pearson's chi-square test (Table 5). An adaptive permutation-based testing approach was applied to derive empirical *p*-values. The single marker association analysis was applied to the same set of variants and samples as in the rare variant association test. None of the 46 single ultra-rare variants were significantly associated with schizophrenia (Table 5).

Table 4 | Single marker association test in *NPC1* and *NPC2*.

Gene	Variant	F_A	F_U	CHISQ	*P*	$P_{emp.}$
NPC1	18:21113384:G:A	0.0002755	0.0002731	0.00003853	0.9950	0.6429
NPC1	18:21113400:A:AGAATATC	0.0002755	0	1.009	0.3152	0.2895
NPC1	18:21113459:T:G	0	0.0002731	0.9914	0.3194	0.8125
NPC1	18:21114441:A:G	0.0002755	0	1.009	0.3152	0.2365
NPC1	18:21114444:T:C	0	0.0002731	0.9914	0.3194	0.8125
NPC1	18:21114451:T:C	0.0002755	0	1.009	0.3152	0.2615
NPC1	18:21114508:T:C	0.0002755	0	1.009	0.3152	0.2742
NPC1	18:21115468:C:T	0.0002755	0	1.009	0.3152	0.3265
NPC1	18:21115582:G:A	0.0002755	0	1.009	0.3152	0.2742
NPC1	18:21115621:T:C	0.0002755	0	1.009	0.3152	0.2576
NPC1	18:21116679:A:G	0	0.0005461	1.983	0.1591	0.2742
NPC1	18:21116700:G:A	0	0.0002731	0.9914	0.3194	0.7083
NPC1	18:21116722:T:C	0.0002755	0	1.009	0.3152	0.2797
NPC1	18:21118573:C:CCT	0.0002755	0.0002731	0.00003853	0.9950	0.8571
NPC1	18:21119357:T:C	0	0.0002731	0.9914	0.3194	0.7778
NPC1	18:21119358:A:G	0.0002755	0	1.009	0.3152	0.2615
NPC1	18:21119369:A:G	0	0.0002731	0.9914	0.3194	0.8571
NPC1	18:21119411:A:G	0	0.0002731	0.9914	0.3194	0.7273
NPC1	18:21119429:T:C	0.0002755	0	1.009	0.3152	0.2946
NPC1	18:21119430:A:G	0	0.0002731	0.9914	0.3194	0.7273
NPC1	18:21119793:A:G	0.0002755	0	1.009	0.3152	0.2797
NPC1	18:21119842:T:C	0.0002755	0	1.009	0.3152	0.2308
NPC1	18:21119949:A:T	0	0.0002731	0.9914	0.3194	0.8571
NPC1	18:21121091:A:G	0	0.0002731	0.9914	0.3194	0.7083
NPC1	18:21121147:T:C	0.0002755	0	1.009	0.3152	0.3444
NPC1	18:21123467:GA:G	0	0.0002731	0.9914	0.3194	0.8125
NPC1	18:21124366:A:G	0.0002755	0	1.009	0.3152	0.2056
NPC1	18:21124441:T:C	0.0002755	0	1.009	0.3152	0.3077
NPC1	18:21124448:T:C	0	0.0002731	0.9914	0.3194	0.8571
NPC1	18:21124988:G:T	0	0.0002731	0.9914	0.3194	0.8571
NPC1	18:21125112:G:A	0.0002755	0	1.009	0.3152	0.2576
NPC1	18:21134845:A:G	0.0002755	0	1.009	0.3152	0.2615
NPC1	18:21134854:A:G	0	0.0002731	0.9914	0.3194	0.7778
NPC1	18:21134927:C:T	0.0002755	0.0002731	0.00003853	0.9950	0.7273
NPC1	18:21137120:T:C	0	0.0002731	0.9914	0.3194	0.7273
NPC1	18:21137150:A:G	0	0.0002731	0.9914	0.3194	0.9999
NPC1	18:21140387:T:G	0	0.0005461	1.983	0.1591	0.3780
NPC1	18:21141335:G:GGAGT	0.0002755	0	1.009	0.3152	0.2333
NPC1	18:21141414:T:C	0.000551	0.0002731	0.3423	0.5585	0.4375
NPC1	18:21141488:G:A	0.0002755	0	1.009	0.3152	0.2397
NPC1	18:21148814:AT:A	0	0.0002731	0.9914	0.3194	0.8571
NPC1	18:21153427:T:C	0.0002755	0	1.009	0.3152	0.2576
NPC2	14:74947469:A:AC	0.0002755	0	1.009	0.3152	0.2365
NPC2	14:74951129:A:C	0.0002755	0	1.009	0.3152	0.2333
NPC2	14:74951148:C:A	0	0.0002731	0.9914	0.3194	0.9286
NPC2	14:74953113:T:C	0.0002755	0	1.009	0.3152	0.3000

Legend Table 6: F_A: Frequency in cases; F_U: Frequency in controls; CHISQ: Basic allelic chi-square test statistics; *P*: Asymptotic *p*-value; $P_{emp.}$: Empirical *p*-value derived by adaptive permutation approach.

2.4 DISCUSSION

2.4.1 Key Findings

To explore the role of *NPC1* and *NPC2* in schizophrenia, we have carried out rare variant association tests of smMIP-based targeted sequencing data. A total of 46 and 4 rare functionally relevant variants (MAF <0.01) were identified in *NPC1* and *NPC2* from 1,815 schizophrenia patients and 1,831 controls (post-QC). The major goal was to test the hypothesis that rare functionally relevant variants in *NPC1* and *NPC2* are enriched in patients with schizophrenia compared to controls. The second aim was to screen a large cohort of patients diagnosed with schizophrenia (NP-C). The negative results of the SKAT-O analyses indicated that rare heterozygous variants in *NPC1* or *NPC2* do not play an important role in the current schizophrenia cohort of this study. Further, we did not find any rare homozygous or compound heterozygote variant which might have confirmed an autosomal recessive form of NP-C in the current schizophrenia cohort. Thus, we conclude that in the current cohort a misdiagnosis of NP-C patients with psychiatric symptoms such as schizophrenia can be ruled out. However, we identified 42 and 4 rare heterozygous variants (internal MAF or ExAC non-psych NFE <0.001) in *NPC1* and *NPC2* with moderate to high impact on the protein level (Tables 2 and 3). All of these variants had a CADD score >20 except the frameshift mutations for which no CADD scores were available. Interestingly, three stop gain mutations in *NPC1* (rs759826138, R958X; rs370721218, R934X) and NPC2 (rs80358266, E118X) had the highest CADD scores (33, 39, and 41) and were predicted to be pathogenic in Clinvar. All three variants were singletons in the current cohort, and rs759826138 had no allele frequency available in the ExAC (non-psych NFE) database. Interestingly, in an international genetic screening study called ZOOM, an enrichment of heterozygous variants as a dominant condition with reduced penetrance was discussed in the development of late-onset NP-C manifestations (Bauer et al., 2013). However, they point out that there is not much evidence in the literature and clinic that heterozygous *NPC1* and *NPC2* variants lead to NP-C-based psychiatric symptoms (Bauer et al., 2013). Nonetheless, heterozygous mutations in *NPC1* and *NPC2* are rigorously discussed as potential risk factors for neurodegenerative diseases such as Alzheimer's disease (AD) (Kresojević et al., 2014). Notably, in GWAS as well as targeted sequencing studies, heterozygous carriers of *NPC1* loss-of-function mutations were associated with an increased risk of obesity due to an incomplete transport of cholesterol (Meyre et al., 2009; Liu et al., 2017). However, due to missing BMI-based data, a follow-up on the obesity trait of heterozygous *NPC1* carriers could not be performed. A revisiting of pathogenic heterozygous variant carriers in *NPC1* or *NPC2* might be warranted to rule out any of the aforementioned phenotypic features or incidental findings. Notably, the main results were in concordance with the findings of the SCHEMA browser [Schizophrenia exome meta-analysis consortium, Singh et al., 2020] where exomes from 24,248 cases and 97,322 controls, and *de novo* mutations from 3,402 parent-proband were meta-analyzed. A lookup in the SCHEMA browser displayed gene-based *p*-values of 0.153 and 0.206 for *NPC1* and *NPC2*. Likewise, the single marker association test revealed no significant association of rare variants in *NPC1* or *NPC2* with schizophrenia. Together these results indicate that the effects of rare variants in *NPC1* and *NPC2* have no major contribution to the development of SCZ. Thus, rare functionally relevant variants in *NPC1* and *NPC2* in the current cohort and the newly established SCHEMA consortium might not play an important role for SCZ. However, further detailed investigation of clinically pathogenic variants in *NPC1* and *NPC2* is warranted to confirm or rule out the role of *NPC1* and *NPC2* in SCZ.

2.4.2 Strengths and Limitations

This study had several strengths, first, a smMIPs-based targeted next-generation sequencing approach was successfully implemented for reliably identifying rare variants in the target genes of *NPC1* and NPC2. This technology enabled the processing of large quantities of samples with high coverage up to 200x for both target genes in a highly automatable and reproducible way compared to commercial kits. This setup can now be used for NGS-based diagnostics of recessive NP-C cases to rule out any misdiagnosis of NP-C patients who might mimic psychiatric symptoms. Therefore, this technology might change the diagnostics and therapy of NP-C patients as NP-C specific therapies such as Miglustat are available. Importantly, targeted sequencing provides a cost-effective solution for screening candidate genes for rare variants in a large cohort. Second, different approaches to test the effect of rare variants in *NPC1* and *NPC2* associated with schizophrenia were deployed in this study. Mainly, gene-based tests such as the SKAT-O and conservative burden test were conducted. But also single marker association tests were carried out to access the significance on a single variant level. Although several other methods have been proposed to study the effects of rare variants, the SKAT-O test was chosen as it maximizes the power by adaptively selecting the best linear combination of the burden and non-burden SKAT. Further, the SKAT-O analysis is highly automatable and computationally efficient so that a large cohort of patients and controls can be analyzed very quickly. Third, we integrated SNP array-based genome-wide genotype data for sophisticated sample-based quality control steps. Common variants from genome-wide genotyping arrays enabled an improved analysis for estimating population stratification as it was shown that including only rare variants was not effective in controlling population stratification [Ma et al., 2020]. These steps are important in general to avoid any confounding and false-positive associations. This study had several limitations. First, although the mean coverage was higher than >200x for both genes, it was observed that in *NPC1* 3% and *NPC2* 17.98% from the targeted bases were not covered at >10x. Thus, some exons such as exon 8 in *NPC1* or exon 4 in *NPC2* failed to be covered fully. Due to the size of the genes, *NPC1* (25 exons) and *NPC2* (5 exons) the loss of coverage looks more dramatic in *NPC2* than in *NPC1*. However, the chance to detect any pathogenic variant in these target regions is therefore prevented. A ClinVar lookup revealed that overall 9 (4/5) pathogenic variants in exon 8 of *NPC1* and exon 4 of *NPC2* would have been missed to be detected due to the loss of coverage. Some of the important limitations of the smMIPs technology are the non-uniformity of the capture efficiency across long exons and GC-rich target regions. Whether these factors influenced the efficacy of the sequencing of the failed target regions needs to be further investigated and optimized on the smMIPs assay level. Second, the limited sample size to detect any associations implicated less power on the rare variant level. However, the findings from the SCHEMA consortium revealed similar not significant findings for *NPC1* and *NPC2*. This might suggest that rare variants in *NPC1* and *NPC2* might not play an important role or that small effects contribute to the development of SCZ. Third, we applied a basic statistical model for rare variant association testing without any correction for covariates such as age, sex, BMI, or ancestry. This leads to the fact that the effects of covariates on the rare variant level could not be estimated.

2.4.3 Conclusions

In the present work, the targeted sequencing of *NPC1* and *NPC2* of a relatively large cohort of 1,815 schizophrenia cases and 1,831 controls was conducted. For that, an NGS-based targeted sequencing of all exons and exon/intron boundaries of the *NPC1* and *NPC2* genes was established. To test the hypothesis of whether pathogenic *NPC1* and *NPC2* variants are enriched in SCZ patients compared to controls, a rare variant association test using SKAT-O was applied. According to the SKAT-O test, both genes *NPC1*, as well as NPC2, were not significantly associated with schizophrenia. However, the correct diagnosis of NP-C patients who might mimic psychiatric symptoms is crucial for the correct clinical diagnosis. Therefore, the smMIP technology has the potential to improve the diagnostics and therapy of the affected individuals. It can be run in parallel to conventional sequencing methods until NGS-based methods such as whole-exome (WES) or whole-genome sequencing (WGS) become more feasible than in the past. To conclude, this is a first step toward the setup of a routine clinical diagnostics pipeline for the identification of rare pathogenic variants in *NPC1* and *NPC2* which might lead to personalized individual care and treatment of NP-C or SCZ patients.

2.5 List of Figures

2.6 List of Tables

3 References

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

Abdolmaleky, H. M., Smith, C. L., Faraone, S. V., Shafa, R., Stone, W., Glatt, S. J., & Tsuang, M. T. (2004). Methylomics in psychiatry: Modulation of gene-environment interactions may be through DNA methylation. American journal of medical genetics. Part B, Neuropsychiatric genetics: the official publication of the International Society of Psychiatric Genetics, 127B(1), 51–59. https://doi.org/10.1002/ajmg.b.20142

Aberg, K. A., Dean, B., Shabalin, A. A., Chan, R. F., Han, L., Zhao, M., van Grootheest, G., Xie, L. Y., Milaneschi, Y., Clark, S. L., Turecki, G., Penninx, B., & van den Oord, E. (2020). Methylome-wide association findings for major depressive disorder overlap in blood and brain and replicate in independent brain samples. Molecular psychiatry, 25(6), 1344–1354. https://doi.org/10.1038/s41380-018-0247-6

Ai, S., Shen, L., Guo, J., Feng, X., & Tang, B. (2012). DNA methylation as a biomarker for neuropsychiatric diseases. The International journal of neuroscience, 122(4), 165–176. https://doi.org/10.3109/00207454.2011.637654

Alavi, A., Nafissi, S., Shamshiri, H., Nejad, M. M., & Elahi, E. (2013). Identification of mutation in NPC2 by exome sequencing results in diagnosis of Niemann-Pick disease type C. Molecular genetics and metabolism, 110(1-2), 139–144. https://doi.org/10.1016-/j.ymgme.2013.05.019

Aldinger, F., & Schulze, T. G. (2017). Environmental factors, life events, and trauma in the course of bipolar disorder. Psychiatry and clinical neurosciences, 71(1), 6–17. https://doi.org/10.1111/pcn.12433

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Andreasen N. C. (1999). A unitary model of schizophrenia: Bleuler's "fragmented phrene" as schizencephaly. Archives of general psychiatry, 56(9), 781–787. https://doi.org/10.1001/archpsyc.56.9.781

Anney RJL, Ripke S, et al. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. Molecular Autism. 2017;8:21. doi:10.1186/s13229-017-0137-9

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., & Irizarry, R. A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics, 30(10),1363–1369. doi.org/10.1093/bioinformatics/btu049

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology tool for the unification of biology. The Gene Ontology Consortium. Nature genetics, 25(1), 25–29. https://doi.org/10.1038/75556

Bakar, N. A., Ashikov, A., Brum, J. M., Smeets, R., Kersten, M., Huijben, K., Keng, W. T., de Carvalho, D. R., de Oliveira Rizzo, I., de Mello, W. D., Heiner-Fokkema, M. R., Gorman, K., Grunewald, S., Michelakakis, H., Moraitou, M., Martinelli, D., van Scherpenzeel, M., Janssen, M., de Boer, L., van den Heuvel, L. P., … Lefeber, D. J. (2022). Synergistic use of glycomics and single-molecule molecular inversion probes (smMIPs) for IDENTIFICATION of congenital disorders of glycosylation type-1. Journal of inherited metabolic disease, 10.1002/jimd.12496. Advance online publication. https://doi.org/10.1002/jimd.12496

Bansal, V., Libiger, O., Torkamani, A., & Schork, N. J. (2010). Statistical analysis strategies for association studies involving rare variants. Nature reviews. Genetics, 11(11), 773–785. https://doi.org/10.1038/nrg2867

Barichello, T., Badawy, M., Pitcher, M. R., Saigal, P., Generoso, J. S., Goularte, J. A., Simões, L. R., Quevedo, J., & Carvalho, A. F. (2016). Exposure to Perinatal Infections and Bipolar Disorder: A Systematic Review. Current molecular medicine, 16(2), 106–118. https://doi.org/10.2174/1566524016666160126143741

Beck, T. F., Mullikin, J. C., NISC Comparative Sequencing Program, & Biesecker, L. G. (2016). Systematic Evaluation of Sanger Validation of Next-Generation Sequencing Variants. Clinical chemistry, 62(4), 647–654. https://doi.org/10.1373/clinchem.2015.249623

Benjamini, Y. and Hochberg, Y. (1995), Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological), 57: 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bernstein, D. P., Fink, L., Handelsman, L., Foote, J., Lovejoy, M., Wenzel, K., Sapareto, E., & Ruggiero, J. (1994). Initial reliability and validity of a new retrospective measure of child abuse and neglect. The American journal of psychiatry, 151(8), 1132–1136. https://doi.org/10.1176/ajp.151.8.1132

Berrettini W. H. (2001). Molecular linkage studies of bipolar disorders. Bipolar disorders, 3(6), 276–283. https://doi.org/10.1034/j.1399-5618.2001.30603.x

Berridge, M.J., Calcium signaling and psychiatric disease: bipolar disorder and schizophrenia. Cell Tissue Res. 2014 Aug;357(2):477-92. https://doi:10.1007/s00441-014-1806-z

Bird A. (2002). DNA methylation patterns and epigenetic memory. Genes & development, 16(1), 6–21. https://doi.org/10.1101/gad.947102

Bortolato, B., Köhler, C. A., Evangelou, E., León-Caballero, J., Solmi, M., Stubbs, B., Belbasis, L., Pacchiarotti, I., Kessing, L. V., Berk, M., Vieta, E., & Carvalho, A. F. (2017). Systematic assessment of environmental risk factors for bipolar disorder: an umbrella review of systematic reviews and meta-analyses. Bipolar disorders, 19(2), 84–96. https://doi.org/10.1111/bdi.12490

Boyle, E. A., O'Roak, B. J., Martin, B. K., Kumar, A., & Shendure, J. (2014). MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. Bioinformatics (Oxford, England), 30(18), 2670–2672. https://doi.org/10.1093/bioinformatics/btu353

Brainstorm Consortium, Anttila, V., Bulik-Sullivan, B., Finucane, H. K., Walters, R. K., Bras, J., Duncan, L., Escott-Price, V., Falcone, G. J., Gormley, P., Malik, R., Patsopoulos, N. A., Ripke, S., Wei, Z., Yu, D., Lee, P. H., Turley, P., Grenier-Boley, B., Chouraki, V., Kamatani, Y., … Murray, R. (2018). Analysis of shared heritability in common disorders of the brain. Science, 360(6395), eaap8757. https://doi.org/10.1126/science.aap8757

Bustamante, A. C., Aiello, A. E., Galea, S., Ratanatharathorn, A., Noronha, C., Wildman, D. E., & Uddin, M. (2016). Glucocorticoid receptor DNA methylation, childhood maltreatment and major depression. Journal of affective disorders, 206, 181–188. https://doi.org/10.1016/j.jad.2016.07.038

Cantor, Rita M. (2013), "Analysis of Genetic Linkage", in Rimoin, David; Pyeritz, Reed; Korf, Bruce (eds.), Emery and Rimoin's Principles and Practice of Medical Genetics (6th ed.), Academic Press, pp. 1–9, doi:10.1016/b978-0-12-383834-6.00010-0

Carstea, E. D., Morris, J. A., Coleman, K. G., Loftus, S. K., Zhang, D., Cummings, C., Gu, J., Rosenfeld, M. A., Pavan, W. J., Krizman, D. B., Nagle, J., Polymeropoulos, M. H., Sturley, S. L., Ioannou, Y. A., Higgins, M. E., Comly, M., Cooney, A., Brown, A., Kaneski, C. R., Blanchette-Mackie, E. J., … Tagle, D. A. (1997). Niemann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis. Science (New York, N.Y.), 277(5323), 228–231. https://doi.org/10.1126/science.277.5323.228

Chan, R. F., Turecki, G., Shabalin, A. A., Guintivano, J., Zhao, M., Xie, L. Y., van Grootheest, G., Kaminsky, Z. A., Dean, B., Penninx, B., Aberg, K. A., & van den Oord, E. (2020). Cell Type-Specific Methylome-wide Association Studies Implicate Neurotrophin and Innate Immune Signaling in Major Depressive Disorder. Biological psychiatry, 87(5), 431–442. https://doi.org/10.1016/j.biopsych.2019.10.014

Chen, Y. A., Lemire, M., Choufani, S., Butcher, D. T., Grafodatskaya, D., Zanke, B. W., Gallinger, S., Hudson, T. J., & Weksberg, R. (2013). Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics, 8(2), 203–209. https://doi.org/10.4161/epi.23470

Córdova-Palomera, A., Fatjó-Vilas, M., Gastó, C., Navarro, V., Krebs, M., & Fañanás, L. (2015). Genome-wide methylation study on depression: differential methylation and variable methylation in monozygotic twins. Translational Psychiatry, 5(4), e557-e557. doi:10.1038/tp.2015.49

Counts, J. L., & Goodman, J. I. (1995). Alterations in DNA methylation may play a variety of roles in carcinogenesis. Cell, 83(1), 13–15. https://doi.org/10.1016/0092-8674(95)90228-7

Craddock, N., & Forty, L. (2006). Genetics of affective (mood) disorders. European journal of human genetics: EJHG, 14(6), 660–668. doi.org/10.1038/sj.ejhg.5201549

Craddock, N., & Sklar, P. (2013). Genetics of bipolar disorder. Lancet (London, England), 381(9878), 1654–1662. https://doi.org/10.1016/S0140-6736(13)60855-7

Craddock, N., Owen, M., Burge, S., Kurian, B., Thomas, P., & McGuffin, P. (1994). Familial cosegregation of major affective disorder and Darier's disease (keratosis follicularis). The British journal of psychiatry: the journal of mental science, 164(3), 355–358. https://doi.org/10.1192/bjp.164.3.355

Crocker A. C. (1961). The cerebral defect in Tay-Sachs disease and Niemann-Pick disease. Journal of neurochemistry, 7, 69–80. doi.org/10.1111/j.14714159.1961.tb13499

Cross-Disorder Group of the Psychiatric Genomics Consortium, 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., … 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

Crow T. J. (1985). The two-syndrome concept: origins and current status. Schizophrenia bulletin, 11(3), 471–486. https://doi.org/10.1093/schbul/11.3.471

Cruceanu, C., Alda, M., Nagy, C., Freemantle, E., Rouleau, G. A., & Turecki, G. (2013). H3K4 tri-methylation in synapsin genes leads to different expression patterns in bipolar disorder and major depression. The international journal of neuropsychopharmacology, 16(2), 289–299. https://doi.org/10.1017/S1461145712000363

DeCarolis, N. A., & Eisch, A. J. (2010). Hippocampal neurogenesis as a target for the treatment of mental illness: a critical evaluation. Neuropharmacology, 58(6), 884–893. https://doi.org/10.1016/j.neuropharm.2009.12.013

Delaneau, O., Marchini, J. & The 1000 Genomes Project Consortium. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. Nat Commun 5, 3934 (2014). https://doi.org/10.1038/ncomms4934

Derkach, A., Lawless, J. F., & Sun, L. (2013). Robust and powerful tests for rare variants using Fisher's method to combine evidence of association from two or more complementary tests. Genetic epidemiology, 37(1), 110–121. https://doi.org/10.1002/gepi.21689

Diekstra, A., Bosgoed, E., Rikken, A., van Lier, B., Kamsteeg, E. J., Tychon, M., Derks, R. C., van Soest, R. A., Mensenkamp, A. R., Scheffer, H., Neveling, K., & Nelen, M. R. (2015). Translating sanger-based routine DNA diagnostics into generic massive parallel ion semiconductor sequencing. Clinical chemistry, 61(1), 154–162. https://doi.org/10.1373/clinchem.2014.225250

Docherty, A. R., Moscati, A. A., & Fanous, A. H. (2016). Cross-Disorder Psychiatric Genomics. Current behavioral neuroscience reports, 3(3), 256–263. https://doi.org/-10.1007/s40473-016-0084-3

Du, P., Zhang, X., Huang, C.-C., Jafari, N., Kibbe, W. A., Hou, L., & Lin, S. M. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics, 11, 587. http://doi.org/10.1186/1471-2105-11-587

Dunn, D. B., & SMITH, J. D. (1958). The occurrence of 6-methylaminopurine in deoxyribonucleic acids. The Biochemical journal, 68(4), 627–636. https://doi.org/10.1042/bj0680627

Ehrlich, M., Gama-Sosa, M. A., Carreira, L. H., Ljungdahl, L. G., Kuo, K. C., & Gehrke, C. W. (1985). DNA methylation in thermophilic bacteria: N4-methylcytosine, 5-methylcytosine, and N6-methyladenine. Nucleic acids research, 13(4), 1399–1412. https://doi.org/10.1093/nar/13.4.1399

Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H., & Nadeau, J. H. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. Nature reviews. Genetics, 11(6), 446–450. https://doi.org/10.1038/nrg2809

Eijkelenboom, A., Kamping, E. J., Kastner-van Raaij, A. W., Hendriks-Cornelissen, S. J., Neveling, K., Kuiper, R. P., Hoischen, A., Nelen, M. R., Ligtenberg, M. J., & Tops, B. B. (2016). Reliable Next-Generation Sequencing of Formalin-Fixed, Paraffin-Embedded Tissue Using Single Molecule Tags. The Journal of molecular diagnostics : JMD, 18(6), 851–863. https://doi.org/10.1016/j.jmoldx.2016.06.010

Fan, H. M., Sun, X. Y., Guo, W., Zhong, A. F., Niu, W., Zhao, L., Dai, Y. H., Guo, Z. M., Zhang, L. Y., & Lu, J. (2014). Differential expression of microRNA in peripheral blood mononuclear cells as specific biomarker for major depressive disorder patients. Journal of psychiatric research, 59, 45–52. doi.org/10.1016/j.jpsychires.2014.08.007

Farré, P., Jones, M. J., Meaney, M. J., Emberly, E., Turecki, G., & Kobor, M. S. (2015). Concordant and discordant DNA methylation signatures of aging in human blood and brain. Epigenetics & chromatin, 8, 19. https://doi.org/10.1186/s13072-015-0011-y

First, M. B., & Gibbon, M. (2004). The Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) and the Structured Clinical Interview for DSM-IV Axis II Disorders

(SCID-II). In M. J. Hilsenroth & D. L. Segal (Eds.), Comprehensive handbook of psychological assessment, Vol. 2. Personality assessment (pp. 134–143). John Wiley & Sons, Inc..

Forstner, A. J., Fischer, S. B., Schenk, L. M., Strohmaier, J., Maaser-Hecker, A., Reinbold, C. S., Sivalingam, S., Hecker, J., Streit, F., Degenhardt, F., Witt, S. H., Schumacher, J., Thiele, H., Nürnberg, P., Guzman-Parra, J., Orozco Diaz, G., Auburger, G., Albus, M., Borrmann-Hassenbach, M., González, M. J., … Cichon, S. (2020). Whole-exome sequencing of 81 individuals from 27 multiply affected bipolar disorder families. Translational psychiatry, 10(1), 57. https://doi.org/10.1038/s41398-020-0732-y

Forstner, A. J., Hecker, J., Hofmann, A., Maaser, A., Reinbold, C. S., Mühleisen, T. W., … Nöthen, M. M. (2017). Identification of shared risk loci and pathways for bipolar disorder and schizophrenia. PLoS ONE, 12(2), e0171595. http://doi.org/10.1371/journal.pone.0171595

Fortin, J.-P., Labbe, A., Lemire, M., Zanke, B. W., Hudson, T. J., Fertig, E. J., … Hansen, K. D. (2014). Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biology, 15(11), 503. http://doi.org/10.1186/s13059-014-0503-2

Fuchikami, M., Morinobu, S., Segawa, M., Okamoto, Y., Yamawaki, S., Ozaki, N., … Terao, T. (2011). DNA Methylation Profiles of the Brain-Derived Neurotrophic Factor (BDNF) Gene as a Potent Diagnostic Biomarker in Major Depression. PLoS ONE, 6(8), e23881. doi:10.1371/journal.pone.0023881

Gamazon, E. R., Badner, J. A., Cheng, L., Zhang, C., Zhang, D., Cox, N. J., Gershon, E. S., Kelsoe, J. R., Greenwood, T. A., Nievergelt, C. M., Chen, C., McKinney, R., Shilling, P. D., Schork, N. J., Smith, E. N., Bloss, C. S., Nurnberger, J. I., Edenberg, H. J., Foroud, T., Koller, D. L., … Liu, C. (2013). Enrichment of cis-regulatory gene expression SNPs and methylation quantitative trait loci among bipolar disorder susceptibility variants. Molecular psychiatry, 18(3), 340–346. https://doi.org/10.1038/mp.2011.174

Goes, F. S., Pirooznia, M., Tehan, M., Zandi, P. P., McGrath, J., Wolyniec, P., Nestadt, G., & Pulver, A. E. (2021). De novo variation in bipolar disorder. Molecular psychiatry, 26(8), 4127–4136. https://doi.org/10.1038/s41380-019-0611-1

Grayton, H. M., Fernandes, C., Rujescu, D., & Collier, D. A. (2012). Copy number variations in neurodevelopmental disorders. Progress in neurobiology, 99(1), 81–91. https://doi.org/10.1016/j.pneurobio.2012.07.005

Hannon, E., Lunnon, K., Schalkwyk, L., & Mill, J. (2015). Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes. Epigenetics, 10(11), 1024–1032. https://doi.org/10.1080/15592294.2015.1100786

Hiatt, J. B., Pritchard, C. C., Salipante, S. J., O'Roak, B. J., & Shendure, J. (2013). Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. Genome research, 23(5), 843–854. https://doi.org/10.1101/-gr.147686.112

Houseman, E. A., Accomando, W. P., Koestler, D. C., Christensen, B. C., Marsit, C. J., Nelson, H. H., … Kelsey, K. T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics, 13, 86. doi.org/10.1186/1471-2105-13-86

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

Howie, B., Marchini, J., & Stephens, M. (2011). Genotype imputation with thousands of genomes. G3 (Bethesda, Md.), 1(6), 457–470. doi.org/10.1534/g3.111.001198

Hüls, A., Robins, C., Conneely, K. N., De Jager, P. L., Bennett, D. A., Epstein, M. P., Wingo, T. S., & Wingo, A. P. (2020). Association between DNA methylation levels in brain tissue and late-life depression in community-based participants. Translational psychiatry, 10(1), 262. https://doi.org/10.1038/s41398-020-00948-6

Illingworth, R. S., Gruenewald-Schneider, U., Webb, S., Kerr, A. R., James, K. D., Turner, D. J., Smith, C., Harrison, D. J., Andrews, R., & Bird, A. P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. PLoS genetics, 6(9), e1001134. https://doi.org/10.1371/journal.pgen.1001134

International Schizophrenia Consortium (2008). Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature, 455(7210), 237–241. https://doi.org/10.1038/nature07239

Isometsä E. (2014). Suicidal behavior in mood disorders--who, when, and why?. Canadian journal of psychiatry. Revue canadienne de psychiatrie, 59(3), 120–130. https://doi.org/10.1177/070674371405900303

Jabbari, K., & Bernardi, G. (2004). Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. Gene, 333, 143–149. https://doi.org/10.1016/j.gene.2004.02.043

Jacobs, B. L., van Praag, H., & Gage, F. H. (2000). Adult brain neurogenesis and psychiatry: a novel theory of depression. Molecular psychiatry, 5(3), 262–269. https://doi.org/10.1038/sj.mp.4000712

Januar, V., Saffery, R., & Ryan, J. (2015). Epigenetics and depressive disorders: a review of current progress and future directions. International journal of epidemiology, 44(4), 1364–1387. https://doi.org/10.1093/ije/dyu273

Jesulola, E., Micalos, P., & Baguley, I. J. (2018). Understanding the pathophysiology of depression: From monoamines to the neurogenesis hypothesis model - are we there yet?. Behavioral brain research, 341, 79–90. https://doi.org/10.1016/j.bbr.2017.12.025

Jing, J., He, L., Sun, A., Quintana, A., Ding, Y., Ma, G., … Zhou, Y. (2015). Proteomic mapping of ER-PM junctions identifies STIMATE as regulator of Ca2+ influx. Nature Cell Biology, 17(10), 1339–1347. http://doi.org/10.1038/ncb3234

Johnson, S. L., & Roberts, J. E. (1995). Life events and bipolar disorder: implications from biological theories. Psychological bulletin, 117(3), 434–449. https://doi.org/10.1037/0033-2909.117.3.434

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics (Oxford, England), 8(1), 118–127. https://doi.org/10.1093/biostatistics/kxj037

Jones P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature reviews. Genetics, 13(7), 484–492. doi.org/10.1038/nrg3230

Kawazoe, T., Yamamoto, T., Narita, A., Ohno, K., Adachi, K., Nanba, E., Noguchi, A., Takahashi, T., Maekawa, M., Eto, Y., Ogawa, M., Murata, M., & Takahashi, Y. (2018). Phenotypic variability of Niemann-Pick disease type C including a case with clinically pure schizophrenia: a case report. BMC neurology, 18(1), 117. https://doi.org/10.1186/s12883-018-1124-2

Kennedy S. H. (2008). Core symptoms of major depressive disorder: relevance to diagnosis and treatment. Dialogues in clinical neuroscience, 10(3), 271–277. https://doi.org/10.31887/DCNS.2008.10.3/shkennedy

Kessler, R. C., & Bromet, E. J. (2013). The epidemiology of depression across cultures. Annual review of public health, 34, 119–138. https://doi.org/10.1146/annurev-publhealth-031912-114409

Kessler, R. C., Amminger, G. P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., & Ustün, T. B. (2007). Age of onset of mental disorders: a review of recent literature. Current opinion in psychiatry, 20(4), 359–364. doi.org/10.1097/YCO.0b013e32816ebc8c

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nature genetics, 46(3), 310–315. https://doi.org/10.1038/ng.2892

Kircher, T., Wöhr, M., Nenadic, I., Schwarting, R., Schratt, G., Alferink, J., Culmsee, C., Garn, H., Hahn, T., Müller-Myhsok, B., Dempfle, A., Hahmann, M., Jansen, A., Pfefferle, P., Renz, H., Rietschel, M., Witt, S. H., Nöthen, M., Krug, A., & Dannlowski, U. (2019). Neurobiology of the major psychoses: a translational perspective on brain structure and function-the FOR2107 consortium. European archives of psychiatry and clinical neuroscience, 269(8), 949–962. https://doi.org/10.1007/s00406-018-0943-x

Klengel, T., & Binder, E. B. (2013). Gene-environment interactions in major depressive disorder. Canadian journal of psychiatry. Revue canadienne de psychiatrie, 58(2), 76–83. https://doi.org/10.1177/070674371305800203

Klutstein, M., Nejman, D., Greenfield, R., & Cedar, H. (2016). DNA Methylation in Cancer and Aging. Cancer research, 76(12), 3446–3450. https://doi.org/10.1158/0008-5472.CAN-15-3278

Kuan, P. F., Waszczuk, M., Kotov, R., Marsit, C., Guffanti, G., Yang, X., … Luft, B. (2017). 904. DNA Methylation Associated with PTSD and Depression in World Trade Center Responders: An Epigenome-Wide Study. Biological Psychiatry, 81(10), S365. doi:10.1016/j.biopsych.2017.02.629

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., … International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. Nature, 409(6822), 860–921. https://doi.org/10.1038/35057062

Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., Karapetyan, K., Katz, K., Liu, C., Maddipatla, Z., Malheiro, A., McDaniel, K., Ovetsky, M., Riley, G., Zhou, G., Holmes, J. B., … Maglott, D. R. (2018). ClinVar: improving access to variant interpretations and supporting evidence. Nucleic acids research, 46(D1), D1062–D1067. https://doi.org/10.1093/nar/gkx1153

Lazarov, O., & Hollands, C. (2016). Hippocampal neurogenesis: Learning to remember. Progress in neurobiology, 138-140, 1–18. https://doi.org/10.1016/j.pneurobio.2015.12.00

Leckman, J. F., Sholomskas, D., Thompson, D., Belanger, A., & Weissman, M. M. (1982). Best estimate of lifetime psychiatric diagnosis: a methodological study. Archives of general psychiatry, 39(8), 879-883.

Lee, S., Abecasis, G. R., Boehnke, M., & Lin, X. (2014). Rare-variant association analysis: study designs and statistical tests. American journal of human genetics, 95(1), 5–23. https://doi.org/10.1016/j.ajhg.2014.06.009

Lee, S., Wu, M. C., & Lin, X. (2012). Optimal tests for rare variant effects in sequencing association studies. Biostatistics (Oxford, England), 13(4), 762–775. https://doi.org/10.1093/biostatistics/kxs014

Leek JT, Johnson WE, Parker HS, Fertig EJ, Jaffe AE, Storey JD, Zhang Y, Torres LC (2018). sva: Surrogate Variable Analysis. R package version 3.28.0.

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B., Tukiainen, T., Birnbaum, D. P., Kosmicki, J. A., Duncan, L. E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D. N., … Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature, 536(7616), 285–291. https://doi.org/10.1038/nature19057

Leucht, S., Burkard, T., Henderson, J., Maj, M., & Sartorius, N. (2007). Physical illness and schizophrenia: a review of the literature. Acta psychiatrica Scandinavica, 116(5), 317–333. https://doi.org/10.1111/j.1600-0447.2007.01095.x

Li H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics (Oxford, England), 27(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, B., & Leal, S. M. (2008). Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. American journal of human genetics, 83(3), 311–321. https://doi.org/10.1016/j.ajhg.2008.06.024

Li, E., Bestor, T. H., & Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell, 69(6), 915–926. https://doi.org/10.1016/0092-8674(92)90611-f

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England), 25(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Lin, D., Chen, J., Perrone-Bizzozero, N., Bustillo, J. R., Du, Y., Calhoun, V. D., & Liu, J. (2018). Characterization of cross-tissue genetic-epigenetic effects and their patterns in schizophrenia. Genome medicine, 10(1), 13. https://doi.org/10.1186/s13073-018-0519-4

Loftus, S. K., Morris, J. A., Carstea, E. D., Gu, J. Z., Cummings, C., Brown, A., Ellison, J., Ohno, K., Rosenfeld, M. A., Tagle, D. A., Pentchev, P. G., & Pavan, W. J. (1997). Murine model of Niemann-Pick C disease: mutation in a cholesterol homeostasis gene. Science (New York, N.Y.), 277(5323), 232–235. https://doi.org/10.1126/science.277.5323.232

Lohoff F. W. (2010). Overview of the genetics of major depressive disorder. Current psychiatry reports, 12(6), 539–546. https://doi.org/10.1007/s11920-010-0150-6

Ma, S., & Shi, G. (2020). On rare variants in principal component analysis of population stratification. BMC genetics, 21(1), 34. https://doi.org/10.1186/s12863-020-0833-x

Madsen, B. E., & Browning, S. R. (2009). A groupwise association test for rare mutations using a weighted sum statistic. PLoS genetics, 5(2), e1000384. https://doi.org/10.1371/journal.pgen.1000384

Málaga, D. R., Brusius-Facchin, A. C., Siebert, M., Pasqualim, G., Saraiva-Pereira, M. L., Souza, C., Schwartz, I., Matte, U., & Giugliani, R. (2019). Sensitivity, advantages, limitations, and clinical utility of targeted next-generation sequencing panels for the diagnosis of selected lysosomal storage disorders. Genetics and molecular biology, 42(1 suppl 1), 197–206. https://doi.org/10.1590/1678-4685-GMB-2018-0092

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics (Oxford, England), 26(22), 2867–2873. https://doi.org/10.1093/bioinformatics/btq559

Marangoni, C., Hernandez, M., & Faedda, G. L. (2016). The role of environmental exposures as risk factors for bipolar disorder: A systematic review of longitudinal studies. Journal of affective disorders, 193, 165–174. https://doi.org/10.1016/j.jad.2015.12.055

Marshall, C. R., Howrigan, D. P., Merico, D., Thiruvahindrapuram, B., Wu, W., Greer, D. S., Antaki, D., Shetty, A., Holmans, P. A., Pinto, D., Gujral, M., Brandler, W. M., Malhotra, D., Wang, Z., Fajarado, K., Maile, M. S., Ripke, S., Agartz, I., Albus, M., Alexander, M., … CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium (2017). Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nature genetics, 49(1), 27–35. https://doi.org/10.1038/ng.3725

Matosin, N., Cruceanu, C., & Binder, E. B. (2017). Preclinical and Clinical Evidence of DNA Methylation Changes in Response to Trauma and Chronic Stress. Chronic stress (Thousand Oaks, Calif.), 1, 2470547017710764. doi.org/10.1177/2470547017710764

McKay Bounford, K., & Gissen, P. (2014). Genetic and laboratory diagnostic approach in Niemann Pick disease type C. Journal of neurology, 261 Suppl 2(Suppl 2), S569–S575. https://doi.org/10.1007/s00415-014-7386-8

Merikangas, K. R., Akiskal, H. S., Angst, J., Greenberg, P. E., Hirschfeld, R. M., Petukhova, M., & Kessler, R. C. (2007). Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. Archives of general psychiatry, 64(5), 543–552. https://doi.org/10.1001/archpsyc.64.5.543

Millier, A., Schmidt, U., Angermeyer, M. C., Chauhan, D., Murthy, V., Toumi, M., & Cadi-Soussi, N. (2014). Humanistic burden in schizophrenia: a literature review. Journal of psychiatric research, 54, 85–93. doi.org/10.1016/j.jpsychires.2014.03.021

Min, J. L., Hemani, G., Hannon, E., Dekkers, K. F., Castillo-Fernandez, J., Luijk, R., Carnero-Montoro, E., Lawson, D. J., Burrows, K., Suderman, M., Bretherick, A. D., Richardson, T. G., Klughammer, J., Iotchkova, V., Sharp, G., Al Khleifat, A., Shatunov, A., Iacoangeli, A., McArdle, W. L., Ho, K. M., … Relton, C. L. (2021). Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. Nature genetics, 53(9), 1311–1321. https://doi.org/10.1038/s41588-021-00923-x

Moarefi, A. H., & Chédin, F. (2011). ICF syndrome mutations cause a broad spectrum of biochemical defects in DNMT3B-mediated de novo DNA methylation. Journal of molecular biology, 409(5), 758–772. https://doi.org/10.1016/j.jmb.2011.04.050

Morgenthaler, S., & Thilly, W. G. (2007). A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). Mutation research, 615(1-2), 28–56. https://doi.org/10.1016/j.mrfmmm.2006.09.003

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

Naureckiene, S., Sleat, D. E., Lackland, H., Fensom, A., Vanier, M. T., Wattiaux, R., Jadot, M., & Lobel, P. (2000). Identification of HE1 as the second gene of Niemann-Pick C disease. Science (New York, N.Y.), 290(5500), 2298–2301. https://doi.org/10.1126/science.290.5500.2298

Neale, B. M., Rivas, M. A., Voight, B. F., Altshuler, D., Devlin, B., Orho-Melander, M., Kathiresan, S., Purcell, S. M., Roeder, K., & Daly, M. J. (2011). Testing for an unusual distribution of rare variants. PLoS genetics, 7(3), e1001322. https://doi.org/10.1371/journal.pgen.1001322

Neelands, T. R., & Macdonald, R. L. (1999). Incorporation of the pi subunit into functional gamma-aminobutyric Acid(A) receptors. Molecular pharmacology, 56(3), 598–610. https://doi.org/10.1124/mol.56.3.598

Neveling, K., Mensenkamp, A. R., Derks, R., Kwint, M., Ouchene, H., Steehouwer, M., van Lier, B., Bosgoed, E., Rikken, A., Tychon, M., Zafeiropoulou, D., Castelein, S., Hehir-Kwa, J., Tjwan Thung, D., Hofste, T., Lelieveld, S. H., Bertens, S. M., Adan, I. B., Eijkelenboom, A., Tops, B. B., … Hoischen, A. (2017). BRCA Testing by Single-Molecule Molecular Inversion Probes. Clinical chemistry, 63(2), 503–512. https://doi.org/10.1373/clinchem.2016.263897

Ng, B., Casazza, W., Kim, N. H., Wang, C., Farhadi, F., Tasaki, S., Bennett, D. A., De Jager, P. L., Gaiteri, C., & Mostafavi, S. (2021). Cascading epigenomic analysis for identifying disease genes from the regulatory landscape of GWAS variants. PLoS genetics, 17(11), e1009918. https://doi.org/10.1371/journal.pgen.1009918

Nica, A. C., & Dermitzakis, E. T. (2013). Expression quantitative trait loci: present and future. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 368(1620), 20120362. https://doi.org/10.1098/rstb.2012.0362

Niemann, A. (1914). "Ein unbekanntes Krankheitsbild" [An unknown disease picture]. Jahrbuch für Kinderheilkunde. Neue Folge (in German). 79: 1–10.

Oberlander, T. F., Weinberg, J., Papsdorf, M., Grunau, R., Misri, S., & Devlin, A. M. (2008). Prenatal exposure to maternal depression, neonatal methylation of human glucocorticoid receptor gene (NR3C1) and infant cortisol stress responses. Epigenetics, 3(2), 97-106. doi:10.4161/epi.3.2.6034

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C. M.,

… Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic acids research, 44(D1), D733–D745. https://doi.org/10.1093/nar/gkv1189

Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T., & Delaneau, O. (2016). Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics (Oxford, England), 32(10), 1479–1485. https://doi.org/10.1093/bioinformatics/btv722

O'Roak, B. J., Vives, L., Fu, W., Egertson, J. D., Stanaway, I. B., Phelps, I. G., Carvill, G., Kumar, A., Lee, C., Ankenman, K., Munson, J., Hiatt, J. B., Turner, E. H., Levy, R., O'Day, D. R., Krumm, N., Coe, B. P., Martin, B. K., Borenstein, E., Nickerson, D. A., … Shendure, J. (2012). Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science (New York, N.Y.), 338(6114), 1619–1622. https://doi.org/10.1126/science.1227764

Oud, M. S., Ramos, L., O'Bryan, M. K., McLachlan, R. I., Okutman, Ö., Viville, S., de Vries, P. F., Smeets, D., Lugtenberg, D., Hehir-Kwa, J. Y., Gilissen, C., van de Vorst, M., Vissers, L., Hoischen, A., Meijerink, A. M., Fleischer, K., Veltman, J. A., & Noordam, M. J. (2017). Validation and application of a novel integrated genetic screening method to a cohort of 1,112 men with idiopathic azoospermia or severe oligozoospermia. Human mutation, 38(11), 1592–1605. https://doi.org/10.1002/humu.23312

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

Patterson, M. (2000). Niemann-Pick Disease Type C. In M. P. Adam (Eds.) et. al., GeneReviews®. University of Washington, Seattle.

Patterson, M. C., Hendriksz, C. J., Walterfang, M., Sedel, F., Vanier, M. T., Wijburg, F., & NP-C Guidelines Working Group (2012). Recommendations for the diagnosis and management of Niemann-Pick disease type C: an update. Molecular genetics and metabolism, 106(3), 330–344. https://doi.org/10.1016/j.ymgme.2012.03.012

Pérez Millán, M. I., Vishnopolska, S. A., Daly, A. Z., Bustamante, J. P., Seilicovich, A., Bergadá, I., Braslavsky, D., Keselman, A. C., Lemons, R. M., Mortensen, A. H., Marti, M. A., Camper, S. A., & Kitzman, J. O. (2018). Next generation sequencing panel based on single molecule molecular inversion probes for detecting genetic variants in children with hypopituitarism. Molecular genetics & genomic medicine, 6(4), 514–525. Advance online publication. https://doi.org/10.1002/mgg3.395

Personality, Cardiovascular Disease, and Cancer: Amelang, Manfred, Hasselbach, Petra, and Stürmer, Til Zeitschrift für Gesundheitspsychologie 2004 12:3, 102-115

Phillips, M. L., & Kupfer, D. J. (2013). Bipolar disorder diagnosis: challenges and future directions. Lancet (London, England), 381(9878), 1663–1671. https://doi.org/10.1016/S0140-6736(13)60989-7

Phipson B, Maksimovic J, Oshlack A (2015). “missMethyl: an R package for analysing methylation data from Illuminas HumanMethylation450 platform.” Bioinformatics, btv560

Picard Toolkit. 2019. Broad Institute, https://broadinstitute.github.io/picard/; Broad Institute

Pick, L. (1926). "Der Morbus Gaucher und die ihm ähnlichen Krankheiten (die lipoidzellige Splenohepatomegalie Typus Niemann und die diabetische Lipoidzellenhypoplasie der Milz)" [Gaucher's disease and similar diseases (type Niemann lipoid cell splenohepatomegaly and spleen diabetic lipoid cell hypoplasia)]. Ergebnisse der Inneren Medizin und Kinderheilkunde (in German). 29: 519–627.

Pidsley, Ruth, Wong Y, C C, Volta, Manuela, Lunnon, Katie, Mill, Jonathan, Schalkwyk, C L (2013). “A data-driven approach to preprocessing Illumina 450K methylation array data.” BMC Genomics, 14, 293. doi: 10.1186/1471-2164-14-293.

Pogoda, M., Hilke, F. J., Lohmann, E., Sturm, M., Lenz, F., Matthes, J., Muyas, F., Ossowski, S., Hoischen, A., Faust, U., Sepahi, I., Casadei, N., Poths, S., Riess, O., Schroeder, C., & Grundmann, K. (2019). Single Molecule Molecular Inversion Probes for High Throughput Germline Screenings in Dystonia. Frontiers in neurology, 10, 1332. https://doi.org/10.3389/fneur.2019.01332

Priebe, L., Degenhardt, F. A., Herms, S., Haenisch, B., Mattheisen, M., Nieratschker, V., Weingarten, M., Witt, S., Breuer, R., Paul, T., Alblas, M., Moebus, S., Lathrop, M., Leboyer, M., Schreiber, S., Grigoroiu-Serbanescu, M., Maier, W., Propping, P., Rietschel, M., Nöthen, M. M., … Mühleisen, T. W. (2012). Genome-wide survey implicates the influence of copy number variants (CNVs) in the development of early-onset bipolar disorder. Molecular psychiatry, 17(4), 421–432. https://doi.org/10.1038/mp.2011.8

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

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics, 81(3), 559–575. https://doi.org/10.1086/519795

Quintana, A., Rajanikanth, V., Farber-Katz, S., Gudlur, A., Zhang, C., Jing, J., … Hogan, P. G. (2015). TMEM110 regulates the maintenance and remodeling of mammalian ER–plasma membrane junctions competent for STIM–ORAI signaling. Proceedings of the

National Academy of Sciences of the United States of America, 112(51), E7083–E7092. http://doi.org/10.1073/pnas.1521924112

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Ratanatharathorn, A., Boks, M. P., Maihofer, A. X., Aiello, A. E., Amstadter, A. B., Ashley-Koch, A. E., Baker, D. G., Beckham, J. C., Bromet, E., Dennis, M., Garrett, M. E., Geuze, E., Guffanti, G., Hauser, M. A., Kilaru, V., Kimbrel, N. A., Koenen, K. C., Kuan, P. F., Logue, M. W., Luft, B. J., … Smith, A. K. (2017). Epigenome-wide association of PTSD from heterogeneous cohorts with a common multi-site analysis pipeline. American journal of medical genetics. Part B, Neuropsychiatric genetics: the official publication of the International Society of Psychiatric Genetics, 174(6), 619–630. https://doi.org/10.1002/ajmg.b.32568

Richardson B. (2003). DNA methylation and autoimmune disease. Clinical immunology (Orlando, Fla.), 109(1), 72–79. https://doi.org/10.1016/s1521-6616(03)00206-7

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research, 43(7), e47. http://doi.org/10.1093/nar/gkv007

Robertson K. D. (2005). DNA methylation and human disease. Nature reviews. Genetics, 6(8), 597–610. https://doi.org/10.1038/nrg1655

Roy, B., Shelton, R. C., & Dwivedi, Y. (2017). DNA methylation and expression of stress related genes in PBMC of MDD patients with and without serious suicidal ideation. Journal of psychiatric research, 89, 115–124. https://doi.org/10.1016/j.jpsychires.2017.02.005

Samuels, B. A., & Hen, R. (2011). Neurogenesis and affective disorders. The European journal of neuroscience, 33(6), 1152–1159. doi.org/10.1111/j.1460-9568.2011.07614.x

Sandu, S., Jackowski-Dohrmann, S., Ladner, A., Haberhausen, M., & Bachmann, C. (2009). Niemann-Pick disease type C1 presenting with psychosis in an adolescent male. European child & adolescent psychiatry, 18(9), 583–585. https://doi.org/10.1007/s00787-009-0010-2

Sass, L. A., & Parnas, J. (2003). Schizophrenia, consciousness, and the self. Schizophrenia bulletin, 29(3), 427–444. doi.org/10.1093/oxfordjournals.schbul.a007017

Saxonov, S., Berg, P., & Brutlag, D. L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proceedings of the National Academy of Sciences of the United States of America, 103(5), 1412–1417. https://doi.org/10.1073/pnas.0510310103

Schizophrenia Working Group of the Psychiatric Genomics Consortium, Ripke, S., Neale, B. M., Corvin, A., Walters, J. T., Farh, K.-H., … O'Donovan, M. C. (2014). Biological

Insights From 108 Schizophrenia-Associated Genetic Loci. Nature, 511(7510), 421–427. http://doi.org/10.1038/nature13595

Schmermund, A., Möhlenkamp, S., Berenbein, S., Pump, H., Moebus, S., Roggenbuck, U., Stang, A., Seibel, R., Grönemeyer, D., Jöckel, K. H., & Erbel, R. (2006). Population-based assessment of subclinical coronary atherosclerosis using electron-beam computed tomography. Atherosclerosis, 185(1), 177–182. https://doi.org/10.1016/j.atherosclerosis.2005.06.003

Schmitt, A., Malchow, B., Hasan, A., & Falkai, P. (2014). The impact of environmental factors in severe psychiatric disorders. Frontiers in neuroscience, 8, 19. https://doi.org/10.3389/fnins.2014.00019

Schuchman E. H. (2007). The pathogenesis and treatment of acid sphingomyelinase-deficient Niemann-Pick disease. Journal of inherited metabolic disease, 30(5), 654–663. https://doi.org/10.1007/s10545-007-0632-9

Schuchman, E. H., & Wasserstein, M. P. (2015). Types A and B Niemann-Pick disease. Best practice & research. Clinical endocrinology & metabolism, 29(2), 237–247. https://doi.org/10.1016/j.beem.2014.10.002

Schultz, M. L., Krus, K. L., & Lieberman, A. P. (2016). Lysosome and endoplasmic reticulum quality control pathways in Niemann-Pick type C disease. Brain research, 1649(Pt B), 181–188. https://doi.org/10.1016/j.brainres.2016.03.035

Schulze, T. G., Akula, N., Breuer, R., Steele, J., Nalls, M. A., Singleton, A. B., Degenhardt, F. A., Nöthen, M. M., Cichon, S., Rietschel, M., Bipolar Genome Study, & McMahon, F. J. (2014). Molecular genetic overlap in bipolar disorder, schizophrenia, and major depressive disorder. The world journal of biological psychiatry: the official journal of the World Federation of Societies of Biological Psychiatry, 15(3), 200–208. https://doi.org/10.3109/15622975.2012.662282

Severus, E., & Bauer, M. (2013). Diagnosing bipolar disorders in DSM-5. International journal of bipolar disorders, 1, 14. https://doi.org/10.1186/2194-7511-1-14

Shabalin A. A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics (Oxford, England), 28(10), 1353–1358. https://doi.org/10.1093/bioinformatics/bts163

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic acids research, 29(1), 308–311. https://doi.org/10.1093/nar/29.1.308

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

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

Sitarska, D., & Ługowska, A. (2019). Laboratory diagnosis of the Niemann-Pick type C disease: an inherited neurodegenerative disorder of cholesterol metabolism. Metabolic brain disease, 34(5), 1253–1260. https://doi.org/10.1007/s11011-019-00445-w

Sklar, P., Ripke, S., Scott, L. J., Andreassen, O. A., Cichon, S., Craddock, N., … Purcell, S. M. (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nature Genetics, 43(10), 977–983. http://doi.org/10.1038/ng.943

Smith, Z. D., & Meissner, A. (2013). DNA methylation: roles in mammalian development. Nature reviews. Genetics, 14(3), 204–220. doi.org/10.1038/nrg3354

Song, J., Bergen, S. E., Kuja-Halkola, R., Larsson, H., Landén, M., & Lichtenstein, P. (2015). Bipolar disorder and its relation to major psychiatric disorders: a family-based study in the Swedish population. Bipolar disorders, 17(2), 184–193. https://doi.org/10.1111/bdi.12242

Song, J., Zheng, S., Nguyen, N., Wang, Y., Zhou, Y., & Lin, K. (2017). Integrated pipeline for inferring the evolutionary history of a gene family embedded in the species tree: a case study on the STIMATE gene family. BMC Bioinformatics, 18, 439. http://doi.org/10.1186/s12859-017-1850-2

Spitzer RL, Williams JB, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R (SCID). I: History, rationale, and description. Arch Gen Psychiatry. 1992 Aug;49(8):624-9. doi: 10.1001/archpsyc.1992.01820080032005. PMID: 1637252.

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., … 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

Starnawska, A., & Demontis, D. (2021). Role of DNA Methylation in Mediating Genetic Risk of Psychiatric Disorders. Frontiers in psychiatry, 12, 596821. https://doi.org/10.3389/fpsyt.2021.596821

Starnawska, A., Demontis, D., Pen, A., Hedemand, A., Nielsen, A. L., Staunstrup, N. H., Grove, J., Als, T. D., Jarram, A., O'Brien, N. L., Mors, O., McQuillin, A., Børglum, A. D., & Nyegaard, M. (2016). CACNA1C hypermethylation is associated with bipolar disorder. Translational psychiatry, 6(6), e831. https://doi.org/10.1038/tp.2016.99

Starnawska, A., Tan, Q., Soerensen, M., McGue, M., Mors, O., Børglum, A. D., Christensen, K., Nyegaard, M., & Christiansen, L. (2019). Epigenome-wide association study of depression symptomatology in elderly monozygotic twins. Translational psychiatry, 9(1), 214. https://doi.org/10.1038/s41398-019-0548-9

Story Jovanova, O., Nedeljkovic, I., Derek, S., Walker, R. M., Liu, C., Luciano, M., … Amin, N. (2018). DNA Methylation Signatures of Depressive Symptoms in Middle-aged and Elderly Persons. JAMA Psychiatry. doi:10.1001/jamapsychiatry.2018.1725

Sugawara, H., Iwamoto, K., Bundo, M., Ueda, J., Miyauchi, T., Komori, A., … Kato, T. (2011). Hypermethylation of serotonin transporter gene in bipolar disorder detected by epigenome analysis of discordant monozygotic twins. Translational Psychiatry, 1(7), e24-e24. doi:10.1038/tp.2011.26

Sul, J. H., Service, S. K., Huang, A. Y., Ramensky, V., Hwang, S. G., Teshiba, T. M., Park, Y., Ori, A., Zhang, Z., Mullins, N., Olde Loohuis, L. M., Fears, S. C., Araya, C., Araya, X., Spesny, M., Bejarano, J., Ramirez, M., Castrillón, G., Gomez-Makhinson, J., Lopez, M. C., … Freimer, N. B. (2020). Contribution of common and rare variants to bipolar disorder susceptibility in extended pedigrees from population isolates. Translational psychiatry, 10(1), 74. https://doi.org/10.1038/s41398-020-0758-1

Sullivan, P. F., Kendler, K. S., & Neale, M. C. (2003). Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Archives of general psychiatry, 60(12), 1187–1192. https://doi.org/10.1001/archpsyc.60.12.1187

Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: review and meta-analysis. The American journal of psychiatry, 157(10), 1552–1562. https://doi.org/10.1176/appi.ajp.157.10.1552

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. PloS one, 6(7), e21800. https://doi.org/10.1371/journal.pone.0021800

Tang, L., Liu, J., Zhu, Y., Duan, J., Chen, Y., Wei, Y., Gong, X., Wang, F., & Tang, Y. (2021). ANK3 Gene Polymorphism Rs10994336 Influences Executive Functions by Modulating Methylation in Patients With Bipolar Disorder. Frontiers in neuroscience, 15, 682873. https://doi.org/10.3389/fnins.2021.682873

Tarjinder Singh, Benjamin M. Neale, Mark J. Daly. Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia. medRxiv 2020.09.18.20192815; doi: https://doi.org/10.1101/2020.09.18.20192815

Teschendorff, A. E., Menon, U., Gentry-Maharaj, A., Ramus, S. J., Gayther, S. A., Apostolidou, S., Jones, A., Lechner, M., Beck, S., Jacobs, I. J., & Widschwendter, M. (2009). An epigenetic signature in peripheral blood predicts active ovarian cancer. PloS one, 4(12), e8274. https://doi.org/10.1371/journal.pone.0008274

The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. Nature, 526(7571), 68–74. http://doi.org/10.1038/nature15393

Toma, C., Shaw, A. D., Heath, A., Pierce, K. D., Mitchell, P. B., Schofield, P. R., & Fullerton, J. M. (2021). A linkage and exome study of multiplex families with bipolar disorder implicates rare coding variants of ANK3 and additional rare alleles at 10q11-q21. Journal of psychiatry & neuroscience: JPN, 46(2), E247–E257. https://doi.org/10.1503/jpn.200083

Tombácz, D., Maróti, Z., Kalmár, T., Palkovits, M., Snyder, M., & Boldogkői, Z. (2019). Whole-exome sequencing data of suicide victims who had suffered from major depressive disorder. Scientific data, 6, 190010. https://doi.org/10.1038/sdata.2019.10

Triche, T. J., Weisenberger, D. J., Van Den Berg, D., Laird, P. W., & Siegmund, K. D. (2013). Low-level processing of Illumina Infinium DNA Methylation BeadArrays. Nucleic Acids Research, 41(7), e90. http://doi.org/10.1093/nar/gkt090

Tunc-Ozcan, E., Peng, C. Y., Zhu, Y., Dunlop, S. R., Contractor, A., & Kessler, J. A. (2019). Activating newborn neurons suppresses depression and anxiety-like behaviors. Nature communications, 10(1), 3768. https://doi.org/10.1038/s41467-019-11641-8

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Current protocols in bioinformatics, 43(1110), 11.10.1–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

Vanier M. T. (2013). Niemann-Pick diseases. Handbook of clinical neurology, 113, 1717–1721. https://doi.org/10.1016/B978-0-444-59565-2.00041-1

Vanier, M. T., Gissen, P., Bauer, P., Coll, M. J., Burlina, A., Hendriksz, C. J., Latour, P., Goizet, C., Welford, R. W., Marquardt, T., & Kolb, S. A. (2016). Diagnostic tests for Niemann-Pick disease type C (NP-C): A critical review. Molecular genetics and metabolism, 118(4), 244–254. https://doi.org/10.1016/j.ymgme.2016.06.004

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., … Zhu, X. (2001). The sequence of the human genome. Science (New York, N.Y.), 291(5507), 1304–1351. https://doi.org/10.1126/science.1058040

Vilain, J., Galliot, A. M., Durand-Roger, J., Leboyer, M., Llorca, P. M., Schürhoff, F., & Szöke, A. (2013). Les facteurs de risque environnementaux de la schizophrénie [Environmental risk factors for schizophrenia: a review]. L'Encephale, 39(1), 19–28.

Villicaña, S., & Bell, J. T. (2021). Genetic impacts on DNA methylation: research findings and future perspectives. Genome biology, 22(1), 127. https://doi.org/10.1186/s13059-021-02347-6

Vinkers, C. H., Kalafateli, A. L., Rutten, B. P., Kas, M. J., Kaminsky, Z., Turner, J. D., & Boks, M. P. (2015). Traumatic stress and human DNA methylation: a critical review. Epigenomics, 7(4), 593–608. https://doi.org/10.2217/epi.15.11

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

Walsh, T., McClellan, J. M., McCarthy, S. E., Addington, A. M., Pierce, S. B., Cooper, G. M., Nord, A. S., Kusenda, M., Malhotra, D., Bhandari, A., Stray, S. M., Rippey, C. F., Roccanova, P., Makarov, V., Lakshmi, B., Findling, R. L., Sikich, L., Stromberg, T., Merriman, B., Gogtay, N., … Sebat, J. (2008). Rare structural variats disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science (New York, N.Y.), 320(5875), 539–543. https://doi.org/10.1126/science.1155174

Walton, E., Hass, J., Liu, J., Roffman, J. L., Bernardoni, F., Roessner, V., Kirsch, M., Schackert, G., Calhoun, V., & Ehrlich, S. (2016). Correspondence of DNA Methylation Between Blood and Brain Tissue and Its Application to Schizophrenia Research. Schizophrenia bulletin, 42(2), 406–414. https://doi.org/10.1093/schbul/sbv074

Wassif, C. A., Cross, J. L., Iben, J., Sanchez-Pulido, L., Cougnoux, A., Platt, F. M., Ory, D. S., Ponting, C. P., Bailey-Wilson, J. E., Biesecker, L. G., & Porter, F. D. (2016). High incidence of unrecognized visceral/neurological late-onset Niemann-Pick disease, type C1, predicted by analysis of massively parallel sequencing data sets. Genetics in medicine: official journal of the American College of Medical Genetics, 18(1), 41–48. https://doi.org/10.1038/gim.2015.25

Weder, N., Zhang, H., Jensen, K., Yang, B. Z., Simen, A., Jackowski, A., Lipschitz, D., Douglas-Palumberi, H., Ge, M., Perepletchikova, F., O'Loughlin, K., Hudziak, J. J., Gelernter, J., & Kaufman, J. (2014). Child abuse, depression, and methylation in genes involved with stress, neural plasticity, and brain circuitry. Journal of the American Academy of Child and Adolescent Psychiatry, 53(4), 417–24.e5. https://doi.org/10.1016/j.jaac.2013.12.025

Wittchen, H.-U., Zaudig, M. & Fydrich, T. (1997). SKID Strukturiertes Klinisches Interview für DSM-IV. Achse I und II. Göttingen: Hogrefe. Hiller, W., Zaudig, M. & Mombour, W. (1997). IDCL Internationale Diagnosen Checklisten für DSM-IV und ICD-10. Göttingen: Hogrefe. https://doi.org/10.1026//0084-5345.28.1.68

World Health Organization. (2017). Depression and other common mental disorders: global health estimates (No. WHO/MSD/MER/2017.2). World Health Organization.

World Health Organization. (2019). ICD-11: International classification of diseases (11th revision). Retrieved from https://icd.who.int/

Wray, N. R., & Gottesman, I. I. (2012). Using summary data from the danish national registers to estimate heritabilities for schizophrenia, bipolar disorder, and major depressive disorder. Frontiers in genetics, 3, 118. https://doi.org/10.3389/fgene.2012.00118

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

Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M., & Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. American journal of human genetics, 89(1), 82–93. https://doi.org/10.1016/j.ajhg.2011.05.029

Yoav Benjamini, Yosef Hochberg. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 57, No. 1. (1995), pp. 289-300, doi:10.2307/2346101.

Yu, C., Arcos-Burgos, M., Baune, B. T., Arolt, V., Dannlowski, U., Wong, M. L., & Licinio, J. (2018). Low-frequency and rare variants may contribute to elucidate the genetics of major depressive disorder. Translational psychiatry, 8(1), 70. doi.org/10.1038/s41398-018-0117-7

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics (Oxford, England), 30(5), 614–620. https://doi.org/10.1093/bioinformatics/btt593

Zhou, W., Chen, L., Jiang, B., Sun, Y., Li, M., Wu, H., Zhang, N., Sun, X., & Qin, S. (2021). Large-scale whole-exome sequencing association study identifies FOXH1 gene and sphingolipid metabolism pathway influencing major depressive disorder. CNS neuroscience & therapeutics, 27(11), 1425–1428. https://doi.org/10.1111/cns.13733

Zoghbi, A. W., Dhindsa, R. S., Goldberg, T. E., Mehralizade, A., Motelow, J. E., Wang, X., Alkelai, A., Harms, M. B., Lieberman, J. A., Markx, S., & Goldstein, D. B. (2021). High-impact rare genetic variants in severe schizophrenia. Proceedings of the National Academy of Sciences of the United States of America, 118(51), e2112560118. https://doi.org/10.1073/pnas.2112560118

4 Acknowledgements

At this point, I would like to express my great gratitude to all persons who were involved and have supported me in the preparation of this doctoral thesis. I would like to express my special thanks to Jun. Prof. Dr. med. Andreas Forstner and Prof. Dr. med. Markus Nöthen for their excellent support during the execution of the entire thesis. Also, I would like to thank Prof. Krawitz who accompanied me with suggestions and conversations during the preparation of the thesis. To my mother and sisters, I thank them for their encouragement while I was working on this doctoral thesis. On this occasion, I must express my deepest gratitude to my son Lukas and my wife Jasi, who were my ever-driving and motivational force at any given time and place.