Aus dem ZentralInstitut für Seelische Gesundheit
der Medizinischen Fakultät Mannheim
(Direktor: Prof. Dr. Med. Andreas Meyer-Lindenberg)

Biologically informed risk scoring in schizophrenia based on genome-wide omics data

Inauguraldissertation
zur Erlangung des Doktor scientiarum humanarum (Dr.sc.hum.)
der
Medizinischen Fakultät Mannheim
der Ruprecht-Karls-Universität
zu
Heidelberg

vorgelegt von
Junfang Chen

aus
Putian, China
2019

Dekan: Prof. Dr. med. Sergij Goerdt
Referent: Dr. Emanuel Schwarz

To my family

TABLE OF CONTENTS

ABBREVIATIONS........ 1

1 INTRODUCTION 4

1.1 Background 4

1.1.1 Clinical features of schizophrenia........ 4

1.1.2 Genetic and environmental hypotheses of schizophrenia 5

1.1.3 Current progress on personalized medicine 7

1.2 Omics........ 8

1.2.1 Genomics........ 8

1.2.2 Transcriptomics 9

1.2.3 Epigenomics........10

1.3 Systems Biology 12

1.3.1 Pathway analysis 12

1.3.2 Developmental specificity........13

1.4 Machine learning........14

1.4.1 Machine learning models........14

1.4.2 The curse of dimensionality and dimensionality reduction........16

1.4.3 Biologically informed machine learning 17

1.5 Imaging genetics 19

1.5.1 From candidate association to GWAS........19

1.5.2 Machine learning models - From PRS to multivariate and machine learning 20

2 EMPIRICAL STUDIES 22

2.1 Study 1 - Identification of a reproducible epigenetic risk profile for schizophrenia linked to brain methylation and function 22

2.1.1 Key points........22

2.1.2 Abstract........22

2.1.3 Introduction........23

2.1.4 Methods........25

2.1.5 Results 26

2.1.6 Discussion........27

2.1.7 Conclusion 30

2.1.8 Supplementary Methods 33
2.1.9 Supplementary Results..41
2.1.10 Supplementary Tables 43
2.1.11 Supplementary Figures..49
2.2 Study 2 - Male increase in brain gene expression variability is linked to genetic risk for schizophrenia 52
2.2.1 Abstract..52
2.2.2 Introduction..52
2.2.3 Materials and Methods 53
2.2.4 Results 56
2.2.5 Discussion..58
2.2.6 Supplementary Tables..63
2.2.7 Supplementary Figures..69

3 DISCUSSION..72

3.1 Epigenetic modulation of risk..72
3.2 Spatiotemporal characterization of risk..73
3.3 Synaptic and immune pathways..74
3.4 Explainable machine learning models 75
3.5 Limitations..76
3.6 Future work..78

4 SUMMARY 80

5 REFERENCE 82

6 PUBLICATION..109

7 CURRICULUM VITAE 111

8 ACKNOWLEDGEMENT..112

ABBREVIATIONS

A1C	Primary auditory cortex
AAL	Automated Anatomical Labeling
AI	Artificial intelligent
AMY	Amygdala
ANOVA	Analysis of variance
ANCOVA	Analysis of covariance
AUC	Area Under the receiver operating characteristic Curve
AUT	Autism
BA	Brodmann area
BDNF	Brain-derived neurotrophic factor,
BioMM	Biologically informed machine learning
BP	Bipolar disorder
CACNA1C	Calcium Voltage-Gated Channel Subunit Alpha1 C
CACNA1I	Calcium Voltage-Gated Channel Subunit Alpha1 I
CACNB2	Calcium Voltage-Gated Channel Auxiliary Subunit Beta 2
CBC	Cerebellar cortex
CNV	Copy number variation
COMT	Catechol-O-methyltransferase
CpG	The " 5'—C—phosphate—G—3' " sequence of nucleotides
CSF	Cerebral spinal fluid
CTNNA3	Catenin Alpha 3
DARPP-32	dopamine- and cyclic-AMP-regulated phosphoprotein of molecular weight 32,000
DARTEL	Diffeomorphic Anatomical Registration Through Exponential Lie algebra
dbGaP	Database of Genotypes and Phenotypes
DFC	Dorsal prefrontal cortex
DISC1	Disrupted-in-Schizophrenia 1
DLPFC	Dorsolateral prefrontal cortex
DLPFC-HC	Dorsolateral-Prefrontal-Cortex – Hippocampal
DNA	Deoxyribonucleic acid
DRD2	Dopamine receptor D2
DRD4	Dopamine receptor D4
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders IV edition
EPI	Echo planar imaging
eQTL	Expression quantitative trait loci
ESTs	Expressed sequence tags
EWAS	Epigenome-Wide Association Study
FDR	False Discovery Rate
fMRI	Functional magnetic resonance imaging
FOV	Field of view
FWHM	Full width at half maximum
GAIN	Genetic Association Information Network
GLM	Generalized linear model

GM	Grey matter
GO	Gene Ontology
GRCh37	Genome Reference Consortium Human Build 37
GRM3	Glutamate Metabotropic Receptor 3
GRIN2A	Glutamate Ionotropic Receptor NMDA Type Subunit 2A
GRIA1	Glutamate Ionotropic Receptor AMPA Type Subunit 1
GSEA	Gene Set Enrichment Analysis
GWAS	Genome-wide association study
HC	Healthy controls
hESC	Human embryonic stem cell
hg19	Human genome version 19
HIP	Hippocampal anlage or hippocampus
hiPSC	Human-induced pluripotent stem cell
ICD10	International Classification of Diseases 10
IPC	posterior inferior parietal cortex
ITC	Inferior temporal cortex
LCLs	Lymphoblastoid cell line
LD	Linkage disequilibrium
M1C	Primary motor cortex
MAF	Minor allele frequency
MD	Mediodorsal nucleus of the thalamus
MDD	Major depressive disorder
MFC	Medial prefrontal cortex
MNI	Montreal Neurological Institute
MOFA	Multi-omics factor analysis
MPRAGE	magnetization-prepared rapid acquisition with gradient echo
NF-kappaB	The regulation of nuclear factor kappaB
NHS	National Health Service
OFC	Orbital prefrontal cortex
OPCRIT	Operational Criteria Checklist
PCW	Post-conceptional week
PGC	Psychiatric Genomics Consortium
PMI	Post-mortem interval
PMS	Poly-methylation-signature
PRS	Polygenic risk score
QC	Quality control
RDoC/RDC	Research Domain Criteria
relAUT	First-degree relatives of patients with autism
relBP	First-degree relatives of patients with bipolar disorder
relMDD	First-degree relatives of patients with major depressive disorder
RELN	Reelin
relSCZ	First-degree relatives of patients with schizophrenia
RGS4	Regulator Of G Protein Signaling 4
RIN	RNA Integrity Number
RNA	Ribonucleic acid
RNA-Seq	RNA-sequencing

RPKM	Reads per kilobase million mapped reads
S1C	Primary somatosensory cortex
SADS-L	Schedule for Affective Disorders and Schizophrenia-Lifetime Version
SBE	Single base extension
SCID	Structured Clinical Interview for DSM-IV
SCZ	Schizophrenia
SNP	Single nucleotide polymorphism
STC	Superior temporal cortex
STR	Striatum
SubCtl	Subset of healthy controls
SVM	Support Vector Machine
V1C	Primary visual cortex
VFC	Ventral prefrontal cortex
ZNF804A	Zinc finger protein 804A

1 INTRODUCTION

1.1 Background

Mental health is an essential component of human wellbeing, and has a profound impact on individual and civil society. With a prevalence of 22.1%, mental illness is one of the predominant global disease burdens (Charlson, van Ommeren et al. 2019). These debilitating conditions often have a young age-of-onset and long term impairments that lead to a substantial reduction in life-expectancy. As a consequence, mental illnesses have an enormous socioeconomic cost that has been estimated to exceed 600 billion annually in the European Union alone (OECD and Union 2018). The clinical management of most mental illnesses is severely hampered by our lack of understanding the underlying biology, clinical tools for objective diagnosis and treatment selection, and personalized therapy for individual patients.

Schizophrenia is a severe, highly heritable, mental health disorder with a lifetime prevalence of 0.5-1% (Saha, Chant et al. 2005, American Psychiatric Association 2013). Despite the massive clinical and socioeconomic burden, process in tackling this devastating illness has been painfully slow with few efficacious treatments and many unanswered questions. Moreover, the pharmaceutical industry has broadly withdrawn from the psychiatric field, due to a lack of suitable targets and difficulties in stratifying patients to address the low response rate in clinical studies. There is an urgent need to move biological research in schizophrenia towards clinical applications that facilitate more accurate early diagnostics for identifying at-risk subjects and novel pharmaceutical targets.

1.1.1 Clinical features of schizophrenia

Schizophrenia is characterized by a broad spectrum of symptoms that include hallucinations and delusions, reduced volition, and disorganized speech and behavior (Andreasen 1995). The typical age of onset is in adolescence or young adulthood (Messias, Chen et al. 2007) (Leung and Chue 2000, Messias, Chen et al. 2007) with meta analytical studies reporting that a higher prevalence in males who often experience earlier onsets and a more severe illness course (odds ratio = 1.4 (Aleman, Kahn et al. 2003, McGrath, Saha et al. 2008). The prognosis is moreover highly heterogeneous with many patients experiencing a poor outcome despite some recent advances in treatment (Häfner and an der Heiden 1997, Hafner 1998, Millan, Goodwin et al. 2015, Millan, Andrieux et al. 2016, Häfner 2019).

Currently, the diagnosis of schizophrenia is based on two major diagnostic systems: the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (Association 2013) and the International Classification of Diseases 10 (ICD-10) (Organization 1992). These systems have heuristic clinical value, high inter-rater reliability and are continuously updated. It is noteworthy that these diagnostic criteria are remarkably similar to the first classification of mental disorders by Emil Kraeplin in 1898. Indeed, the diagnosis of schizophrenia is still purely phenomenological and relies on clinical observation and psychological self-reports with limited consideration of biological measures. Given the heterogeneous presentation of schizophrenia, whose symptoms overlaps with several other mental disorders (Kirkpatrick, Buchanan et al. 2001, Buckley, Miller et al. 2009), there is concern about over reliance on categorical systems that do not index patients according to their underlying neurobiology.

This is particularly pertinent in light of evidence for biological similarities between several neuropsychiatric disorders which blur the line between diagnostically separate entities (McDonald, Bullmore et al. 2004, International Schizophrenia, Purcell et al. 2009, De Peri, Crescini et al. 2012, Consortium 2013, Li, Cai et al. 2016).
In 2010 the National Institute of Mental Health responded to this unmet need by launching the Research Domain Criteria framework (RDoC) (Insel, Cuthbert et al. 2010). RDoC introduced a new nosology that uses multidimensional constructs that are guided by neuroscience and behavioral science instead of descriptive phenomenology. The RDoC framework is built around multidimensional constructs that are better aligned to the underlying neurobiology and can be readily assessed on the genetic, behavioral and neural functional level (Morris and Cuthbert 2012). However, the RDoC concept has thus far not been translated towards clinical application as further effort is needed to characterize the underlying molecular and neurobiological dimensions and turn the resulting insights into predictive algorithms for future clinical use.

1.1.2 Genetic and environmental hypotheses of schizophrenia

Despite substantial progress in the understanding of the biology underlying schizophrenia, the illness' etiology remains elusive. Gaining a better understanding of complex genetic and environmental risk factors is thus considered a top priority (Tsuang 2000, van Os, Rutten et al. 2008). In the next section research strategies for investigating genetic and environmental aspect of schizophrenia are discussed.

Genetics

Over the last few decades deep insights into the genetic basis of schizophrenia have been gained through genetic study of schizophrenia patients and their first and second degree relatives. These studies have shown schizophrenia is substantially aggregated in families, and that genetic factors play a significant role in its development (Aberg, Liu et al. 2013). One meta analyses of twin studies produced an estimated heritability of liability of 80% (Sullivan, Kendler et al. 2003). Despite the identification of high heritability, such epidemiological approaches, however, provide no direct insight into the specific genetic factors contributing to illness risk.
The completion of the Human Genome Project in 2003 ushered in a new wave of genetic research in schizophrenia. This project, sought to map the entire human genetic code, identifying approximately 22,300 protein-coding genes in the process. This maps of the human genome allowed use of the molecular biology technique to establish correlation 'or linkage' between proximally located genetic markers (which tend to be inherited together during meiosis). Although the linkage approach was first used to investigate simple Mendelian traits, these studies have successfully identified a number of genes and chromosomal abnormalities associated with schizophrenia (Kendler and Diehl 1993) (Bassett, Chow et al. 2000). However, sample size limitations and the considerable genetic heterogeneity of schizophrenia have posed a problem, resulting in comparably low replicability of findings (Risch and Merikangas 1996).
The next generation of research took a markedly different approach: In contrast to linkage analysis, which feature an unbiased exploration of potentially relevant genes across the entire genome, candidate gene studies aimed to uncover risk associated genetic variants belonging to genes thought to be particularly relevance for the biology of schizophrenia. However, notwithstanding early successes, this approach was limited to the a priori selection of plausible genes where some information was already known, and did not take into account the complex gene-gene interactions

and regulatory elements so important for elucidating underlying causal mechanisms. These limitations impacted the biological reproducibility of findings and lead to a shift back towards hypothesis free approaches (Gejman, Sanders et al. 2011). However, despite some misgivings, candidate gene approaches have been very useful for estimating the plausibility of identified variant-associations. For instance, the SzGene database resource summarizes findings on over 1,000 genes associated with schizophrenia (Allen, Bagade et al. 2008).
Finally, considerable advances in our understanding of the genetic basis of schizophrenia have been gained through genome-wide association study (GWAS). GWAS are facilitated by simultaneous, chip-based analysis of hundreds of thousands of common genetic variants. These studies have led to profound insight into the genetic architecture of schizophrenia. Findings from International consortia that have allowed pooling of data resources have been especially bountiful (Visscher, Wray et al. 2017). These studies have identified more than 100 risk loci congruent with major hypotheses of neurobiological basis of schizophrenia, including genes related to dopamine signaling (DRD2), glutamatergic neurotransmission (GRM3, GRIN2A, and GRIA1), and voltage-gated calcium signaling (CACNA1C, CACNB2, and CACNA1I) (Schizophrenia Working Group of the Psychiatric Genomics 2014). Importantly, these large, well powered-studies provide support for the hypothesized polygenic nature of schizophrenia (International Schizophrenia, Purcell et al. 2009, Schizophrenia Working Group of the Psychiatric Genomics 2014). A summary of the last 10 years' worth of psychiatric GWAS research is given in the 'GWAS Catalog' website (https://www.ebi.ac.uk/gwas/) (Horwitz, Lam et al. 2019). One notable finding from these studies is the observation that many psychiatric disorders share common genetic risk loci that are with physical traits or somatic illness. For instance, schizophrenia-related GWAS signals have also been identified in studies exploring cholesterol and body mass index (Horwitz, Lam et al. 2019), and blood pressure (Andreassen, Djurovic et al. 2013). These findings may help to point to the importance of developing personalized medicine approaches in schizophrenia interventions.

Environment

Schizophrenia arises from the complex interplay of genetic predisposition and environmental risk factors. These environmental risk factors are categorized into early life, childhood and later life stages based on their assumed relevance for the etiology of schizophrenia (Dean and Murray 2005, Stilo, Di Forti et al. 2011). The boundary between these stages is, however, blurry and specific factors may impact on susceptibility across stages. Studies focusing on environmental risk during early life point to the contribution of factors such as obstetric complications, season of birth, maternal malnutrition and other stress factors. During childhood, maltreatment, trauma and other adversities increase susceptibility to schizophrenia. Relevant environmental risk factors during later life (i.e. adolescence and early adulthood) include stress, lifestyle, urbanicity, social adversity, traumatic life events, and substance abuse. These environmental risk factors can also be categorized according to their specific social, environmental, familial, neurodevelopmental, economic, and other contexts (Iyegbe, Campbell et al. 2014).

Gene and environment interplay

Individual genetic or environmental factors possesses insufficiently large effect sizes to account for a major portion of schizophrenia risk. Therefore the study of gene-environment interactions, which may explain this so-called 'missing heritability,' have therefore received widespread attention (Manolio, Collins et al. 2009). By definition, a gene-environment interaction (G x E) is "a different effect of environmental exposure on disease risk in persons with different genotypes," or "a different

effect of a genotype on disease risk in persons with different environmental exposures." (Ottman 1996). Indeed, a meta-analysis of 12 twin studies has strengthened the view that nature and nurture are both highly relevant in the development of schizophrenia (Sullivan, Kendler et al. 2003). The diverse G x E studies have been systematically reviewed by (Duncan and Keller 2011, Modinos, Iyegbe et al. 2013, Iyegbe, Campbell et al. 2014). Findings of these studies are, however, rarely been replicated in independent samples due to insufficient sample size and methodological heterogeneity. Furthermore, most of the conducted studies have focused on candidate genes. This knowledge gap on genome level effects was first addressed by Børglum et al (Børglum, Demontis et al. 2013), found the CTNNA3 gene may interact with maternal cytomegalovirus infection, reiterating the importance of environmental risk factors in genetic studies of schizophrenia. One way of overcoming some of the challenges of G x E studies is the use of integrated, large-scale investigations that combine different paradigms and considering the illnesses biological architecture (Modinos, Iyegbe et al. 2013, European Network of National Networks studying Gene-Environment Interactions in Schizophrenia, van Os et al. 2014, Iyegbe, Campbell et al. 2014). In particular, the strategies adopting polygenic risk score in conjunction with well documented environmental factors could help to enhance personalized genetic risk stratification and ensure early detection of high-risk profiles.

1.1.3 Current progress on personalized medicine

The personalized medicine approach describes the tailoring of medical treatment and health care towards a given patient depending on their specific clinical and/or biological characteristics (Schork 2015). Advances in personalized medicine have been made in many other diseases including cancers (Kakimi, Karasaki et al. 2017, Krzyszczyk, Acevedo et al. 2018), HIV (Lengauer, Pfeifer et al. 2014, Mu, Kodidela et al. 2018), and cardiovascular disease (Dainis and Ashley 2018, Leopold and Loscalzo 2018). The psychiatric field has been slow to implement personalized medicine, but it is starting to become more commonplace (Dalvie, Koen et al. 2016, Gandal, Leppa et al. 2016). Personalized medicine approaches in schizophrenia center on diagnostic and predictive biomarkers, including genetic variants (Schizophrenia Working Group of the Psychiatric Genomics 2014, Pardiñas, Holmans et al. 2018), omics (Schwarz, Guest et al. 2012, Montano, Taub et al. 2016, Gandal, Zhang et al. 2018) neuroanatomical factors (van Erp, Hibar et al. 2016, Van Erp, Walton et al. 2018), and drug discovery research (Xu and Wang 2015). One representative example is a genetic study that identified 108 risk loci using up to 36,989 schizophrenia cases and 113,075 controls. A polygenic risk score calculated for each subject can be used to quantify the individual risk, albeit with a limited prediction accuracy. Furthermore, statistical and machine learning models for personalized predictions in schizophrenia are being continuously developed (as reviewed by (Chen and Schwarz 2017, Bzdok and Meyer-Lindenberg 2018)). However, despite the development of high-throughput large-scale screening and computational methodologies, personalized medicine for schizophrenia is still in its infancy. It has been broadly recognized, however, that personalized medicine approaches are rooted in an improved understanding of schizophrenias underlying biology, and are critical for improved clinical management of the illness (DeLisi and Fleischhacker 2016, Buckley and Miller 2017, Zhang, Mao et al. 2018).

1.2 Omics

Increasing interest in systems-wide exploration of schizophrenia is evidenced by a recent wave of studies and funding opportunities. The omics field allows researchers to investigate subtle molecular changes within cells or tissues in a holistic manner, enabling researchers to study a complex molecular system as a whole. Omics technologies have shown substantial utility in uncovering biological factors likely involved in the etiology of schizophrenia across multiple levels of biological organization. This thesis focuses on the application of three omics approaches: genomics, transcriptomics, and epigenomics, as detailed below.

1.2.1 Genomics

Genomics is the comprehensive study of a given organism's genetic sequence. The genome is defined as the complete set of DNA inside a given cell, with the human genome containing tens of thousands of genes organized into 23 pairs of chromosomes. There are two primary goals of genomics: 1) to sequence and analyze the structure and function of the genome with the help of DNA sequencing technologies and bioinformatics methods, and 2) to characterize and quantify all genes of a particular organism and the interplay of these genes with each other and with the environment. In comparison to genetics, which describes the study of heredity, genetic variation, and individual genes (Organization 2002), genomics is the study all genes and their interactions. Genomic approaches have thus particular utility for the exploration of biologically complex psychiatric disorders such as schizophrenia, which are hallmarked by substantial polygenicity and epistasis.
Genomic research in psychiatry has been mainly driven by a fluffy of GWAS that identified a number of reproducible risk loci from millions of genomic variants (Schizophrenia Working Group of the Psychiatric Genomics 2014, Wray, Ripke et al. 2018, Stahl, Breen et al. 2019). In contrast to Mendelian disorders where the mutational patterns are predictable, the majority of psychiatric disorders are polygenic or multi-factorial (O'Donovan 2015). Moreover, the presence of pleiotropic effects in psychiatry has been observed for common (O'donovan, Craddock et al. 2008, Purcell, Wray et al. 2009, Lee, Ripke et al. 2013) and rare genetic variants (Malhotra and Sebat 2012, Kirov, Rees et al. 2014, Schizophrenia Working Group of the Psychiatric Genomics 2014), which suggested the need of new approaches for psychiatric diagnostic delineation (Owen 2014, O'Donovan 2015). Moreover, the genetic correlation between schizophrenia, bipolar and major depressive disorder has been shown to converge upon genes implicated in biological processes involving histone methylation, and immune and neuronal pathways (Network, O'Dushlaine et al. 2015). As individual common variants mostly have small effect sizes, a fundamental question is whether the aggregation of variants can explain a larger portion of the heritable variance. The predominant approach for performing this integration is to sum up effect size weighted allele counts across variants, yielding a poly-genic risk score (PRS) (International Schizophrenia, Purcell et al. 2009, Schizophrenia Working Group of the Psychiatric Genomics 2014). When determined from large-scale GWAS data, PRS have been shown to explain up to 18% of schizophrenia-associated variance. It is notable that the comparatively simple PRS have also been shown to outperform more complex machine learning algorithms, such as kernel support vector machines (Vivian-Griffiths, Baker et al. 2019) and are still considered the best aggregate measure of genetic risk for common genetic variants. While the increasing predictive value of PRS scores highlights their potential diagnostic utility (Torkamani, Wineinger et al. 2018), larger genetic samples and more studies are required before schizophrenia PRS can be translated into clinical application.

A major advance made by the psychiatric research community is the increasing public availability of data and summary statistics that can be used for calculating PRS for genotyped samples from large-scale GWAS analyses. For example, the association-analysis results from thousands of phenotypes (diseases and traits) have been produced by UK Biobank's rapid GWAS program (http://www.nealelab.is/uk-biobank/). Remarkably, many identified schizophrenia-linked GWAS signals have been found to be located in non-coding regions that encompass regulatory elements (Ripke, O'Dushlaine et al. 2013, Schizophrenia Working Group of the Psychiatric Genomics 2014). These findings imply a possible role for gene regulation in the pathophysiology of schizophrenia. While the functional impact of most schizophrenia risk loci still remains elusive, a number of large-scale studies (Fromer, Roussos et al. 2016, Gusev, Mancuso et al. 2018, Jaffe, Straub et al. 2018) provide evidence that these chromosomal regions are associated with perturbations in gene expression and highlighted the utility of gene expression profiling for providing more mechanistic interpretations of genetic liability for schizophrenia.

1.2.2 Transcriptomics

Transcriptomics describes the study of the complete set of RNA transcripts that are generated in a specific cell or a population of cells at a one time. The first human transcriptomics study was conducted in 1991 (Adams, Kelley et al. 1991) and facilitated the discovery of new genes, as well as the functional tagging of genomic elements. There are two main procedures to quantify whole-genome RNA transcripts: Microarrays and RNA-Sequencing (RNA-Seq) (Lowe, Shirley et al. 2017). Microarrays determine the abundance of pre-selected transcripts through nucleic acid hybridization of transcripts to an ordered array of complementary nucleotide probes. Microarrays allow large numbers of transcripts to be measured at the same time, facilitating the generation of transcriptome-wide gene expression data. In RNA-Seq, individual transcripts or expressed sequence tags (ESTs) are sequenced across the genome, and abundance is determined from the number of counts of each transcript. RNA-Seq offers an advantage over microarrays due to its ability to detect alternatively spliced, non-coding, and novel transcripts. Moreover, RNA-Seq has higher accuracy and more reproducibility compared to than microarray-based gene expression analysis (Wang, Gerstein et al. 2009, Martin, Dehler et al. 2016). On the other hand, due to more mature experimental protocols and affordable cost, microarrays are still widely used in the psychiatric field.

Transcriptome profiling has a wide range of applications which include, the identification of differentially expressed genes in a given patient cohort (Sanders, Drigalenko et al. 2017, Wu, Bendriem et al. 2017), interrogation of co-expressed genes (Pacifico and Davis 2017, van Dam, Vosa et al. 2018), and exploration of regulatory processes important for development (Kang, Kawasawa et al. 2011, Shi, Zhang et al. 2016, Semick, Collado-Torres et al. 2018). The investigation of differentially expressed transcripts in schizophrenia has been the focus of numerous studies. For instance, exploration of abnormal gene regulation has been performed in whole blood (de Jong, Boks et al. 2012), lymphoblastoid cell line (LCLs) (Sanders, Goring et al. 2013, Sanders, Drigalenko et al. 2017, Duan, Goring et al. 2018) post-mortem brain tissue (Roussos, Katsel et al. 2012, Fillman, Cloonan et al. 2013, Fromer, Roussos et al. 2016, Ramaker, Bowling et al. 2017), and human-induced pluripotent stem cell (hiPSC) (Maschietto, Tahira et al. 2015, Roussos, Guennewig et al. 2016, Hoffman, Hartley et al. 2017). Despite substantial inconsistencies in findings (particularly in regard to the most differentially expressed genes) some replicable themes have emerged. These are immune system-related dysregulation (Fillman, Cloonan et al. 2013, Gardiner, Cairns et al. 2013, Mistry, Gillis et al. 2013, Sanders, Goring et al. 2013, Bergon, Belzeaux et al. 2015, Hess, Tylee et al. 2016, Sanders,

Drigalenko et al. 2017, Duan, Goring et al. 2018, Kos, Duan et al. 2018, Leirer, Iyegbe et al. 2019), neural or synaptic function (Mistry, Gillis et al. 2013, Sanders, Goring et al. 2013, Fromer, Roussos et al. 2016, Ramaker, Bowling et al. 2017, Sanders, Drigalenko et al. 2017, Duan, Goring et al. 2018, Gusev, Mancuso et al. 2018, Jaffe, Straub et al. 2018, Kos, Duan et al. 2018, Pergola, Di Carlo et al. 2019), and oxidative stress and mitochondrial dysfunction (Torkamani, Dean et al. 2010, Maschietto, Tahira et al. 2015, Kos, Duan et al. 2018, Leirer, Iyegbe et al. 2019). Consequently, it has been argued that schizophrenia may result from alterations across numerous molecular pathways rather than from impairment in a single biological process (Horváth and Mirnics 2015), which is also in line with GWAS findings (Network, O'Dushlaine et al. 2015). Lastly, sex-specific gene expression profiling has suggested the possibility of distinct molecular mechanism in males and females (Qin, Liu et al. 2016, Tiihonen, Koskuvi et al. 2019).
Microarray-based expression profiling initially focused on the analysis of postmortem brain tissue, and later peripheral samples. Despite the relevance of brain tissue for the investigation of psychiatric illness, its analysis is limited by sample size constraints, sample preparation issues and the impact of postmortem effects on downstream analysis. In contrast, peripheral samples are more readily available and the data is thought to be less heterogeneous as it captures only effects which have a systemic manifestation in schizophrenia.

1.2.3 Epigenomics

Epigenomics describes an the use of omics profiling to study all possible biochemical modifications of DNA inside a cell or a population of cells. Epigenetic changes contribute significantly to the regulation of gene expression and activity (Jaenisch and Bird 2003, Gibney and Nolan 2010). Epigenetic modifications can be inherited through mitosis and meiosis without involving the alteration of the genomic sequence. There are three major types of epigenetic modifications: DNA methylation, histone modification, and non-coding RNA interference. In contrast to genetic effects, a notable property of epigenetic modifications is their reversibility (Ramchandani, Bhattacharya et al. 1999, Jia, Fu et al. 2013, Wu and Zhang 2014) meaning that epigenetic alterations have a significant role in regulating genome function during development through non-mutagenic mechanisms. Of note, it allows the engineering of targeted molecular changes and provides the potential for therapeutic development (Kelly, De Carvalho et al. 2010).
DNA methylation describes the addition of a methyl group to DNA with the aid of a family of enzymes termed DNA methyltransferases. DNA methylation is one of the best-characterized epigenetic mechanisms. The methylome refers to the complete set of DNA methylation events in the genome as determined by whole-genome DNA methylation analysis. The first high-resolution DNA methylome was reported for the flowering plant Arabidopsis (Zhang, Yazaki et al. 2006) and the first human methylome studies were carried out in human embryonic stem cell (hESC) (Lister, Pelizzola et al. 2009, Laurent, Wong et al. 2010).
While the epigenetic landscape of schizophrenia has only been studied only in recent years, numerous DNA methylation differences have already been identified (Nishioka, Bundo et al. 2012, Pries, Gülöksüz et al. 2017). Analogous to GWAS, unbiased, hypothesis-free DNA methylome profiling (also termed Epigenome-Wide Association Study (EWAS)) has been employed to investigate aberrant DNA methylation in schizophrenia in blood (Aberg, McClay et al. 2012, Nishioka, Bundo et al. 2013, Aberg, McClay et al. 2014, Hannon, Dempster et al. 2016, Montano, Taub et al. 2016), brain (Pidsley, Viana et al. 2014, Jaffe, Gao et al. 2016) and saliva (Lin, Chen et al. 2018, Braun, Han et al. 2019). The resulting epigenetic signals were compared in terms of CpGs, genes, differentially methylated

regions, and biological pathways. While these studies have identified up to thousands of schizophrenia-associated differentially methylated CpGs, reproducibility of the illness-relevant CpGs across studies has been poor. This is likely due to tissue specificity, as well as methodological and disease-related heterogeneity. Nevertheless, more than ten genes with differentially methylated sites have been replicated in at least four independent studies (as reviewed by Pries and colleagues (Pries, Gülöksüz et al. 2017)). Some of the genes were consistent with those obtained from candidate gene methylation profiling including COMT (Murphy, O'Reilly et al. 2005, Nishioka, Bundo et al. 2013, Chen, Zhang et al. 2014, Wockner, Noble et al. 2014), RELN (Grayson, Jia et al. 2005, Aberg, McClay et al. 2012, Aberg, McClay et al. 2014, Fikri, Norlelawati et al. 2017) and BDNF (Kordi-Tamandani, Sahranavard et al. 2012). Although DNA methylation is cell type-specific (Ziller, Gu et al. 2013, Titus, Gallimore et al. 2017), there is a degree of overlap between blood, saliva, and brain tissue (Smith, Kilaru et al. 2015, Walton, Hass et al. 2016). This facilitates use of the more readily accessible blood or saliva samples for exploration of DNA methylation effects with brain-functional relevance for schizophrenia.

Pathway and gene ontology analyses have also highlight some convergent biological processes relevant to schizophrenia. Consistent with findings from genomic and transcriptomic data, these are primarily immune-related (Liu, Chen et al. 2013, Aberg, McClay et al. 2014, Hannon, Dempster et al. 2016), neurodevelopmental, or related to synaptic functioning (Mill, Tang et al. 2008, Dempster, Pidsley et al. 2011, Aberg, McClay et al. 2014, Pidsley, Viana et al. 2014, Hannon, Dempster et al. 2016). Importantly, the hypothesized neurodevelopmental component of schizophrenia can be probed by mapping DNA methylation changes across brain development and for schizophrenia-associated GWAS loci (Hannon, Spiers et al. 2015, Jaffe, Gao et al. 2016).

A present, there have been few attempts to build machine learning models for classification of schizophrenia from genome-wide DNA methylation data. Two of these (Richfield, Alam et al. 2016, Alam, Lin et al. 2018) which used kernel based machine learning models in conjunction with blood-based DNA methylation and fMRI data for approximately 200 genotyped subjects (92 schizophrenia patients and 116 controls, and 79 patients and 104 controls, respectively), have produced good prediction accuracy. However, the models were built without appropriate confounder adjustment making the final outcome challenging to interpret. Another recent study (Moghadam, Etemadikhah et al. 2019) analyzed post-mortem bulk frontal cortex tissue from 73 schizophrenia cases and 52 controls to build a machine learning classifier that identified methylation patterns that differentiated cases from controls. The classifier was trained using Monte Carlo feature selection and ROSSETA rule-based modeling consisting of a set of minimal IF-THEN rules. However, no significant patterns were identified with these models. Nevertheless, the authors highlighted strategies for improved classification, such as use of single-cell methylation data and whole-genome bisulfite re-sequencing technologies. We have also developed a machine learning framework termed BioMM (Chen and Schwarz 2019) (see also **Figure 3**) for schizophrenia case-control status classification using genome-wide DNA methylation data from two independent cohorts (in total 767 cases and 755 controls). The prediction accuracy of our model was shown to outperform five conventional machine learning models. However, similar to other peripheral studies, brain-functional insight was not gained due to exclusive focus on peripheral DNA methylation differences. The lack of meaningful biological interpretation and clinically sufficient accuracy in the existing studies has highlighted the need for more explainable machine learning models that integrate large-scale datasets (Roscher, Bohn et al. 2019).

1.3 Systems Biology

On a biological level, schizophrenia is hallmarked by small, illness-associated changes that can be observed across many different data modalities. Therefore, and a systems-wide, integrative approach is essential for characterizing the illnesses biology. A core focus of this thesis is use of systems level biological knowledge in combination with data science to identify putative biomarker candidates for schizophrenia. This section describes the application of pathway analysis and the analysis of developmental specificity for the systems-biological exploration of high-dimensional omics resources.

1.3.1 Pathway analysis

A biological pathway refers to set of linked molecular events in a cell or a tissue that lead to alteration of a particular phenotype. The genes that encode the proteins involved in these events can be grouped into different sets and analyzed together to identify important biological aspects of a dataset. Pathway approaches can be particularly helpful in casting light on hidden trends in high-throughput omics data. Over the last two decades a number of pathway resources have been developed with the view of increasing the explanatory power of omics studies (Khatri, Sirota et al. 2012, Garcia-Campos, Espinal-Enriquez et al. 2015). The existing methods are broadly classified into three classes: over-representation analysis, (Capper, Jones et al.)(Capper, Jones et al.)(Capper, Jones et al.)(Capper, Jones et al.)(Capper, Jones et al.)(Capper, Jones et al.)(Capper, Jones et al.)(Capper, Jones et al.)(Capper, Jones et al.)(Capper, Jones et al.)(Capper, Jones et al.)(Capper, Jones et al.)functional class scoring, and pathway topological analysis. All these methods aim to identify a list of pathways significantly associated with a given outcome or disease state. In addition to commercial pathway analysis resources such as Ingenuity Pathways Analysis (https://www.qiagenbioinformatics.com/), there are many public repositories are available for the research community (Bauer-Mehren 2009, Garcia-Campos, Espinal-Enriquez et al. 2015). Of particular note is the Gene Ontology (GO; http://geneontology.org/), which is the world largest resource concerning the functions of genes (Ashburner, Ball et al. 2000, The Gene Ontology Consortium 2018). The GO is organized into a tree like structure with three major branches (or ontologies). These 'biological process', 'molecular function' and 'cellular component' ontologies recursively categorize a given set of genes based on bimolecular functions and cellular location. The GO knowledgebase can be used to investigate gene functions via an enrichment analysis.

Pathway analysis has been extensively applied to explore various complex diseases (Jin, Zuo et al. 2014). Convergent pathways have been identified across schizophrenia, depression and bipolar disorder using five different pathway enrichment analyses of GWAS data from more than 60,000 adult subjects (Network and Pathway Analysis Subgroup of Psychiatric Genomics 2015). These pathways indicate a shared importance for genes linked to immune, neuronal and histone methylation processes in these disorders. This finding suggested that pathway analysis is biologically informative when integrating association signals therefore providing value insight into disease-relevant etiological mechanisms.

1.3.2 Developmental specificity

The influential neurodevelopmental hypothesis of schizophrenia dates back to the 1980s (Weinberger 1987) and is supported by mounting evidence (Fatemi and Folsom 2009, Andreasen 2010, Rapoport, Giedd et al. 2012). The neurodevelopmental hypothesis posits that neurological insults which occur many years before an illness onset can lead to developmental trajectories which sensitize the brain to certain risk factors. These risk factors may, in turn, affect neurodevelopmental processes to the point that the overt symptoms of schizophrenia emergence during adolescence or early adulthood. The hypothesis of an early risk-associated developmental processes having a delayed influence on later-developing brain functions particularly in cortical regions has been supported by various studies employing animal models of the disorder (Saunders, Kolachana et al. 1998, Lipska and Weinberger 2000, Floresco, Geyer et al. 2005, Fung, Webster et al. 2010, Marco, Macri et al. 2011).

The availability of post-mortem brain tissue-derived transcriptomics data, which covers multiple developmental stages and brain regions (Colantuoni, Lipska et al. 2011, Kang, Kawasawa et al. 2011, Hawrylycz, Lein et al. 2012, Miller, Ding et al. 2014) has afforded new opportunities to study schizophrenia-associated developmental trajectories. Despite the complexity of human brain development it has been shown the human brains transcriptome has a highly reproducible organization (Oldham, Konopka et al. 2008, Hawrylycz, Lein et al. 2012). The developmental characterization of schizophrenia risk has mainly investigated by mapping schizophrenia susceptibility genes onto developmental networks. For example, trajectories of gene expression profiles enriched for SNPs associated for schizophrenia are associated with schizophrenia-relevant developmental stages (Clifton, Hannon et al. 2019). Furthermore, relatively higher gene expression during the early mid-fetal period and early infancy are correlated with schizophrenia genetic risk. In contrast, the expression during late childhood showed a negative correlation with genetic risk and then remained steady during adolescence. In addition, a similar gene expression pattern associated with schizophrenia in these two critical periods was more pronounced in the prefrontal cortex than in non-prefrontal and subcortical regions at any developmental stage. Another study (Huckins, Dobbyn et al. 2019) integrated GWAS data with and postmortem brain eQTL information to identify schizophrenia risk genes. These genes were then mapped onto developmental transcriptomic data obtained from the BrainSpan database (Miller, Ding et al. 2014). The identified risk genes were predominantly co-expressed during the early prenatal and postnatal stages. However, the identified brain region with significant co-expression does not have good evidence linking to schizophrenia. In addition to studies involving common genetic variants, studies have been conducted that focus on the 108 risk-associated loci (Birnbaum, Jaffe et al. 2015, Jaffe, Straub et al. 2018), genes harboring de novo mutations in schizophrenia (Xu, Ionita-Laza et al. 2012, Gulsuner, Walsh et al. 2013), and a combination of schizophrenia-related genetic variants (Gilman, Chang et al. 2012). These mappings on the developmental trajectory are predominantly linked to early fetal life through transcriptomic profiling. Lastly, temporal gene expression profiling revealed that schizophrenia risk gene expression patterns during prenatal neurodevelopmental period largely did not overlap with those for bipolar disorder (Clifton, Hannon et al. 2019). While there is substantial genetic overlap between these disorders, this is in line with the lower neurodevelopmental impairment seen in bipolar disorder (Owen and O'Donovan 2017). Overall, the analysis of developmental specificity is beneficial process that may permit a new, more systematic characterization of illness risk.

1.4 Machine learning

1.4.1 Machine learning models

Over the last few years massive amounts of molecular data have been made available to the biomedical research community. To handle this data, contemporary artificial intelligent (AI) techniques have been adopted to assess the probability of being affected by specific diseases. In conjunction with medical expert opinion, these techniques have enhanced clinical diagnosis and decision-making performance (Yu, Beam et al. 2018). Jiang and colleagues, (Jiang, Jiang et al. 2017) who surveyed the current status of AI applications for medical diagnostics reported that a large part of the AI literature focuses on diagnostic imaging, genetic testing and electro-diagnosis. The attention primarily concentrated on three disease areas: cancer, neurology, and cardiology. Due to the increasing availability of large-scale omics data, there has also been substantial adoption of AI applications in psychiatry (Bzdok and Meyer-Lindenberg 2018). These application aim to integrate frequently small biological alterations into classifiers with high predictive value for diagnostic or therapeutic applications.
Machine learning, a sub-field of AI, aims to teach a model to identify patterns, by employing a specific input/output paradigm based on the learned model. Generally speaking machine learning models typically fall into one of two classes: supervised and unsupervised learning. Supervised learning aims to identify a combination of variables (i.e. predictors) that can optimally predict an outcome (e.g. diagnostic status), which needs to be specified in advance. It is then possible to quantify how accurately the model can perform this prediction when applied to unseen data not used for model training. In contrast, unsupervised learning does not require the prior specification of an outcome, instead aiming to learn structural properties of the data from the available variables only. The most prominent application of unsupervised learning is clustering, i.e. the identification of observations that show more similar patterns of variable values compared to observations that are part of a different cluster. In contrast to supervised learning, it is more challenging to evaluate the performance of these models due to a lack of ‘ground truth’. Furthermore, clustering approaches always yield a certain clustering solution, which may not be biologically meaningful and has resulted in clustering approaches being frequently considered ‘hypothesis-generating’. Ideally, clustering solutions should be evaluated in regard to their association with independent variables, such as clinical course or treatment response. This section describes three frequently used supervised machine learning models that are applied in this work.

Generalized Linear model. A general linear model is a simple model that forms a linear combination of variables to predict the dependent variable. The generalized linear model (GLM) is an extension to non-linear forms. Some examples of well-known GLMs are summarized in **Table 1** (Nelder and Wedderburn 1972). For GLMs, the probability distribution of the response variable is allowed to be non-Gaussian which would not be appropriate in a general linear models. For example, when modeling a categorical or multinomial distribution for the dependent variable, a logistic regression model can be a good choice (a graphical presentation of logistic regression is shown in **Figure 1A)**. The objective of logistic regression is to establish the relationship between a binary (in the case of two classes) dependent variable and one or more input variables.

Table 1. Popular GLM models

Model	Random	Link	Systematic
Linear Regression	Normal	Identity	Continuous
ANOVA	Normal	Identity	Categorical
ANCOVA	Normal	Identity	Mixed
Logistic Regression	Binomial	Logit	Mixed
Loglinear	Poisson	Log	Categorical
Poisson	Poisson	Log	Mixed
Multinomial	Multinomial	Generalized Logit	Mixed

Support Vector Machines. The Support Vector Machine (SVM) identifies a hyperplane separating observations into different classes (Cortes and Vapnik 1995). This is represented by single line for two-dimensional data, a plane for three-dimensional data, and a hyperplane for data with more than three dimensions. The distance between this hyperplane and its nearest observations is maximized, leading to a classification 'margin'. In the frequent scenario where classes are not linearly separable, a cost-parameter is optimized to penalize observations falling on the wrong side of the hyperplane. A graphical depiction of SVM is provided in **Figure 1B**. A comprehensive review of the application of SVM in the detection of imaging biomarkers in neurological and psychiatric disorders can be found elsewhere (Orru, Pettersson-Yeo et al. 2012).

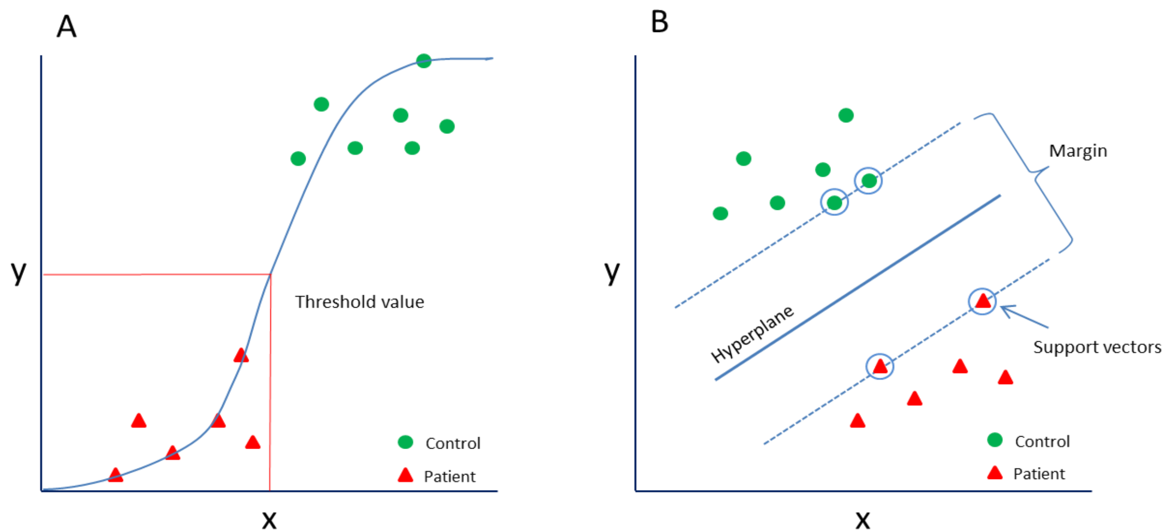


Figure 1 A.) Logistic regression; B.) Support vector machine.

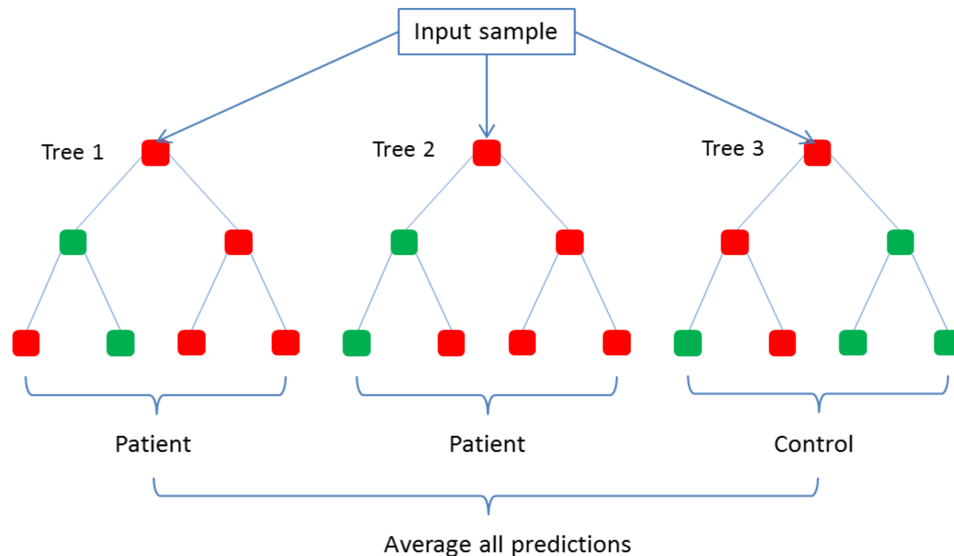


Figure 2. Random forest.

Random Forest. The Random Forest algorithm is an ensemble learning method that combines a set of weak learners to reduce the overall variance (Breiman 2001). Here each weak learner is a decision tree. A random forest comprises a number of random decision trees with two kinds of built in randomness. First each tree is built on a random input sample of the training data. Second, at each node of the tree, a subset of features are selected at random to yield the best split (**Figure 2**). In the classification setting, the prediction result is determined by the majority vote cast from a committee of trees and the averaged predicted result from bootstrapping predictions is used for regression. A particular advantage of the random forest is its intrinsic ranking of important predictors. There are two popular types of importance scores provided for each feature: accuracy-based importance and Gini-based importance. The accuracy-based measure is determined from the decrease in accuracy when a given feature is permuted while the Gini-based importance is accessed by the reduction of Gini impurity when the feature is selected for a split at a given node.

1.4.2 The curse of dimensionality and dimensionality reduction

The analysis of high-throughput omics data is challenging due to the problem of 'large p, small n', meaning that the number of variables greatly exceeds that of the observations. This can often result in a phenomenon termed "curse of dimensionality" when machine learning models operate on data in high-dimensional spaces. In this situation, the dimensionality rapidly increases with the addition of input variables with the volume of the space increasing so quickly that the data becomes sparse. Due to this sparsity, it is highly difficult to evaluate statistical significance or find a decision boundary for any learning methods. Consequently, machine learning models then tend to capture effects that are overly specific for a given dataset, and lack of reliability and generalizability, leading to so-called 'overfitting'. A commonly applied strategy to address this issue is the reduction of data dimensionality, which also aids the visualization and interpretation of high-dimensional omics datasets. Data dimensionality reduction techniques applied in this thesis fall into two broad categories: 1) feature selection and II) feature extraction or engineering.

Feature selection aims to identify useful features directly from a given dataset, while feature extraction describes the process of extracting or generating new features from the dataset via feature transformation or mapping to reduce dimensionality. Three commonly used feature selection

techniques used in the context of classification are filters, wrappers and embedded or hybrid methods, which are comprehensively reviewed elsewhere (Saeys, Inza et al. 2007). In the case of feature extraction, there are two general approaches: linear and nonlinear feature extraction. Principal Component Analysis is the most well-known linear feature extraction method and works by projecting the original data into a lower-dimensional feature space such that the variance of the data in this set of features is maximized. Non-linear dimension reduction methods are based on manifold learning in which the high dimensionality of the data may relate to redundant information or noise. Here, relevant information can be condensed into lower dimension manifolds which can be especially useful when the data contains nonlinear dependencies. Some popular methods regarding the nonlinear dimensionality reduction were systematically discussed by (Van Der Maaten, Postma et al. 2009).

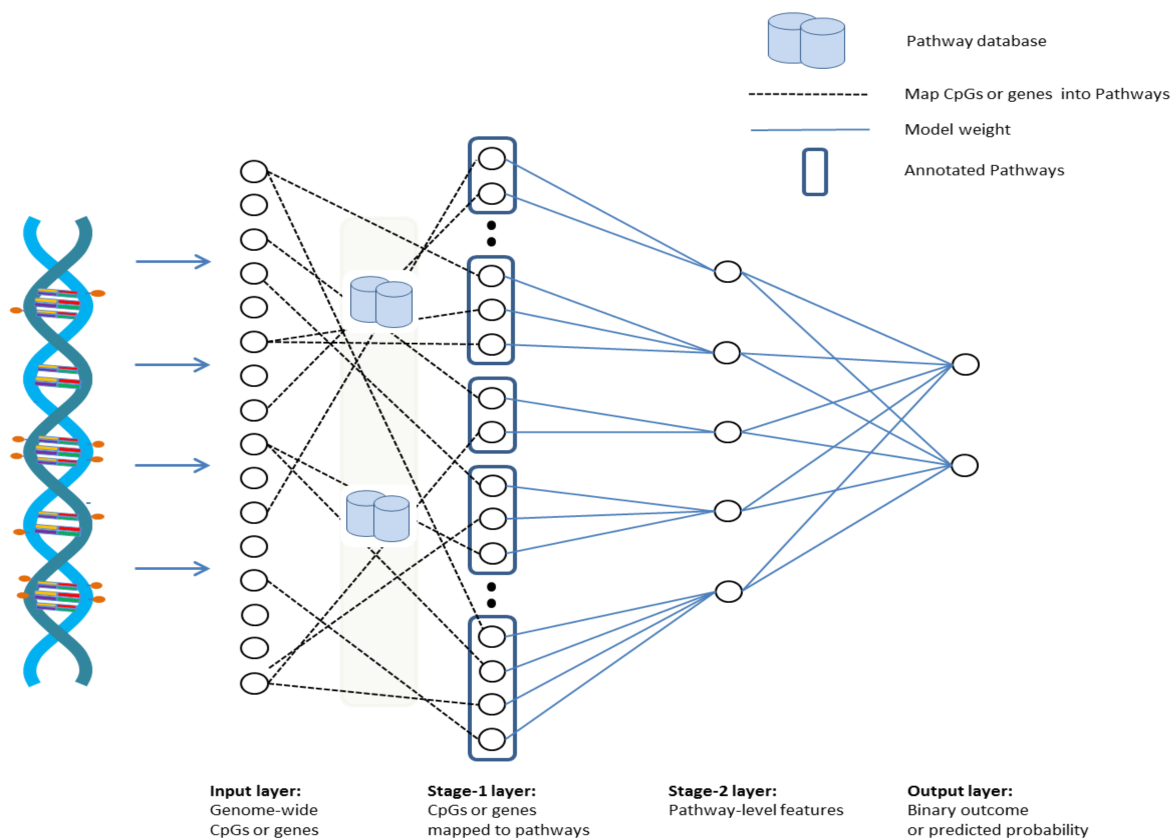


Figure 3. The basic architecture of *BioMM*. From left to right: the input layer consists of genome-wide CpGs; The stage-1 layer comprises individual CpGs mapped into their corresponding pathways by means of the pathway database; The stage-2 layer contains the pathway level features created using machine learning; the output layer contains the predicted outcome, which can be either probabilistic or binary. *BioMM* 1st stage describes the learning process to create stage-2 layer features. *BioMM* 2nd stage integrates information from stage-2 layer features using machine learning and yields the final prediction.

1.4.3 Biologically informed machine learning

Despite high-performing machine learning models holding the promise of revolutionizing the clinical management of schizophrenia, their development is hampered by several unresolved challenges. One issue is that machine learning models trained on high-dimensional data are often regarded black boxes that do not yield an intuitive explanation of their predictive value, limiting biological insight and the possibility for further improvement (Ribeiro, Singh et al. 2016). In terms of the development

of more explainable machine learning models three descriptive terms are currently being distinguished: transparency, interpretability and explainability (Roscher, Bohn et al. 2019). By and large, transparency refers to the machine learning model itself, interpretability refers to the model together with data, and explainability refers to the model, the data and human involvement. An essential part of explainability is the explicit integration of domain knowledge into the machine learning framework, which is called informed machine learning (von Rueden, Mayer et al. 2019). On the other hand, scientific consistency or reproducibility is another fundamental element for creating explainable machine learning approaches (Reichstein, Camps-Valls et al. 2019, Roscher, Bohn et al. 2019). To this end, there has been a number of biologically informed machine learning models developed for different purposes in the context of omics data (Cun and Fröhlich 2012, Chen and Schwarz 2017, Kang, Ding et al. 2017, Zarringhalam, Degras et al. 2018).

An example of these methods are the so-called 'synthetic feature random forest' [SF-RF] (Pan, Hu et al. 2014), first developed to detect phenotype-associated pathways by taking the gene-gene and pathway-pathway interactions into account. Genotyping data from bladder cancer patients and healthy controls consisting of 1,303 preselected SNPs were used in this study. It was hypothesized that each disease-associated pathway could be represented as a single synthetic feature, which is learned by the random forest classifier. High-level interactions between such features that are captured by random forest, as well as statistical epistasis networks may better explain illness-associated genetic mechanisms. A similar strategy was developed by Liu and colleagues who created a pathway-based machine learning framework that identified a subtle overlap of schizophrenia-associated pathways between patients of three different ancestries using GWAS data (Liu, Bousman et al. 2017).

Pathway-informed machine learning has also been reported for gene-level associations. A Markov Random Field approach that incorporated a pathway's topological information was used to identify a set of disease-related genes using both simulated and real SNP data for Crohn's disease (Chen, Cho et al. 2011). The genetic interactions encoded by the topology of a pathway can be explicitly modeled so that disease-associated genes can be identified with increased statistical power. In addition to genetic association data, these methods have been applied to the analysis of gene expression data. For instance, a two-stage machine learning approach has been developed to identify pathways associated with traits of interest using three gene expression datasets (Zhang, Emrich et al. 2010). The significance of gene sets was investigated using four different machine learning classifiers, combined with four different feature selection methods and compared with results from Gene Set Enrichment Analysis (GSEA) (Subramanian, Tamayo et al. 2005). This demonstrated that the two-stage machine learning approach to integrating pathway information outperformed GSEA by detecting larger number of active pathways with more statistical power.

A recent study incorporating gene ontology into deep learning models has shown the improved clustering of single-cell RNA-seq data (Peng, Wang et al. 2019). In this study the authors employed both supervised and unsupervised models to reduce the dimensionality for the clustering of mammalian cell types from two different RNA-seq datasets, and demonstrated better performance over eight other dimension reduction methods. The approaches illustrate the potential application of explainable machine learning model in omics analyses of schizophrenia. The present thesis describes such an application in the context of biologically-informed multi-stage machine learning (BioMM), which was used to identify predictive signatures from epigenetic data.

1.5 Imaging genetics

Imaging genetics is concerned with the study of how genetic variation impacts on brain structural and functional phenotypes that are captured using neuroimaging. This rapidly evolving field has existed for nearly two decades and aims to better characterize risk mechanisms contributing to brain disorders. During this time it has provided a fascinating window into potential neurobiological processes underlying brain images (Mufford, Stein et al. 2017). Neuroimaging techniques allows characterization of brain structure and function in-vivo and have hugely advanced our knowledge of brain disorders (Bandettini 2009) such as schizophrenia (Pfefferbaum and Zipursky 1991, Kotrla, J. et al. 1995), by providing a map of candidate structural, functional and network alterations in patients and at-risk subjects (Meyer-Lindenberg 2010).
The exploration of genetic effects on imaging phenotypes has progressed from investigation of variants in candidate genes to GWAS and polygenic scores with more recent multivariate machine learning approaches being discussed in the next section.

1.5.1 From candidate association to GWAS

Candidate gene approaches offer a more direct path towards mechanistic interpretation compared to genome-wide analyses and have been the primary focus of imaging genetics research. As reviewed elsewhere (Meyer-Lindenberg 2010, Meyer-Lindenberg 2010), genetic variation in a number of well-known candidate genes has shown associations with imaging phenotypes, with genes including COMT (Egan, Goldberg et al. 2001), NRG1 (Hall, Whalley et al. 2006), DISC1 (Callicott, Straub et al. 2005), DARPP-32 (Meyer-Lindenberg, Straub et al. 2007), RGS4 (Buckholtz, Meyer-Lindenberg et al. 2007), DRD2 (Bertolino, Fazio et al. 2008), and BDNF (Ho, Milev et al. 2006). However, candidate gene studies frequently produce inconsistent associations across studies (Nickl-Jockschat, Janouschek et al. 2015, Bogdan, Salmeron et al. 2017). Beyond variants harbored by candidate genes, several SNPs supported by genome-wide significant illness-association have been explored for neural associations in a comparatively more hypothesis-free manner. Representative examples include ZNF804A (Esslinger, Walter et al. 2009, Rasetti, Sambataro et al. 2011) and CACNA1C (Paulus, Bedenbender et al. 2014).
The intermediate phenotype or endophenotype concept of Gottesman and Shields (Gottesman and Shields 1973) refers to a latent biological feature that lies somewhere between gene effects and the overt clinical phenotype. Endophenotypes are thought to be closer to underlying molecular processes that contribute to an illnesses development and progression and thus show greater genetic penetrance than conventional phenotypes. This is particularly relevant for imaging genetics which often focus on analysis of heritable aspects of brain related function or structure (Gottesman and Gould 2003, Meyer-Lindenberg and Weinberger 2006, Greenwood, Braff et al. 2007, Rasetti and Weinberger 2011, Cao, Dixson et al. 2016, Dixson, Tost et al. 2018). The aforementioned, genome-wide supported variants in ZNF804A and CACNA1C were associated with one of the most well-established imaging intermediate phenotypes for schizophrenia. This phenotype quantifies the decoupling of the dorsolateral prefrontal cortex and the hippocampus (DLPFC-HC) during a working memory paradigm (Meyer-Lindenberg, Olsen et al. 2005, Schneider, Walter et al. 2017), and shows how schizophrenia patients possess a significantly different connectivity pattern compared to healthy controls and unaffected first degree relatives. Although single variant focus study in imaging genetics focusing have substantial utility in characterizing the genetic architecture of neural functioning,

these studies are limited to small effect sizes and further do not capture complex epistastic effects that likely contribute to the polygenic architecture of schizophrenia.

1.5.2 Machine learning models - From PRS to multivariate and machine learning

Due to the low explanatory power of individual common variants, recent imaging genetics research has increasingly focused on the analysis of PRS scores in order to aggregate genetic effects (Dima and Breen 2015, Bogdan, Salmeron et al. 2017). For example, the association of a schizophrenia PRS with neural activity during working memory performance has been explored in a cohort of 79 schizophrenia cases and 99 healthy controls (Walton, Turner et al. 2012). This PRS summarizes the combined effect of variants harbored by 34 schizophrenia risk genes for schizophrenia which were correlated with decreased activation in the left DLPFC during working memory. Several other studies have observed similar patterns linking higher PRS scores to lower DLPFC activity during working memory (Kauppi, Westlye et al. 2014, Whalley, Hall et al. 2015, Miller, Scult et al. 2017). Moreover, PRS for schizophrenia has been associated with reduced hippocampal activity during memory encoding (Chen, Ursini et al. 2018), increased global cortical thickness (Neilson, Bois et al. 2017), and decrease cortical gyrification (Liu, Zhang et al. 2017). Interestingly, a recent study (Alnæs, Kaufmann et al. 2019) reported that there is substantially increased brain structural heterogeneity in cortical thickness, and cortical and hippocampal volumes in schizophrenia. In this study, the PRS was associated with mean changes in schizophrenia implicated regions, but could not capture the brain heterogeneity warranting longitudinal investigations to disentangle the hidden factors underlying inter-individual variability. Taken together these findings further supported the utility of polygenic score analysis for characterizing the genetic architecture of schizophrenia.

It should be noted that the PRS approach has several limitations which are reviewed in detail elsewhere. Limitation include the lack of consideration of potential epistatic effects, the restricted biological insight into underlying mechanisms (Bogdan, Salmeron et al. 2017), constrained diversity in the populations under study (Bogdan, Salmeron et al. 2017, De La Vega and Bustamante 2018), and insufficient generalizability (Torkamani, Wineinger et al. 2018).Multivariate or machine learning approaches are an alternative to the PRS approaches that can substantially profit from big data collected in large-scale collaborations such as the ENIGMA consortium (Thompson, Stein et al. 2014, Thompson, Andreassen et al. 2017). However, these approaches have not yet been widely applied in the schizophrenia field, likely due to the scarcity of the multimodal data and the lack of well-developed methodological frameworks. While there are several studies investigate brain disorders using multivariate or machine learning techniques these do not primarily focus on schizophrenia (Liu and Calhoun 2014, Bogdan, Salmeron et al. 2017, Mufford, Stein et al. 2017). One of these few studies using a semi-blind multivariate approach termed parallel ICA with reference (Chen, Calhoun et al. 2013). This study integrated imaging and SNP data with prior knowledge from 140 patients with schizophrenia and 160 healthy controls to identify genomic risk variants involved in neurotransmission and neural signaling pathways that were significantly associated with schizophrenia-linked grey matter reduction in prefrontal and temporal regions .

Multivariate or machine learning frameworks should preferably feature under researched factors such as epistatic effects, gene-environment interactions, with transcriptomic, epigenetic, clinical and environmental data (Bogdan, Salmeron et al. 2017, Mufford, Stein et al. 2017). However, there is currently insufficient data to effectively cover this broad spectrum of modalities. That said, the ongoing efforts in international collaborations provide opportunities for increasingly sophisticated

data analysis that integrate small-effect changes across modalities to better characterize the risk architecture of schizophrenia and its impact on neural functioning.

In summary, the application of machine learning offers promising approaches to characterize the molecular risk mechanisms contributing to the polygenetic nature of schizophrenia. The characterized effect of molecular perturbations can then be mapped onto the brain imaging space to infer systems-level consequences. It is hoped such a strategy it would allow a more precise and comprehensive understanding of risk mechanisms and aid in the personalization of precognitive therapy.

2 EMPIRICAL STUDIES

2.1 Study 1 - Identification of a reproducible epigenetic risk profile for schizophrenia with brain methylation and function[1]

2.1.1 Key points

Question Can a blood marker of epigenetic risk for schizophrenia be derived that is specific for the disease and predicts epigenetic changes in brain and disease-associated brain function?

Findings In this case-control study, machine learning was used to identify a reproducible schizophrenia blood DNA-methylation signature that was associated with functional dorsolateral prefrontal cortex hippocampal connectivity, mapped to methylation differences found in dorsolateral prefrontal cortex hippocampal connectivity postmortem samples, and indexed biological pathways associated with synaptic function. No interactions with polygenic schizophrenia risk were found.

Meaning These findings support the presence of a systemic methylation profile in schizophrenia that is associated with established intermediate functional phenotypes as well as with epigenetic signatures in brain and should thus be useful to capture the biological effects of gene-environment interactions.

2.1.2 Abstract

Importance: Schizophrenia is a severe mental disorder in which epigenetic mechanisms may contribute to illness risk. Epigenetic profiles can be derived from blood cells, but to our knowledge, it is unknown whether these predict established brain alterations associated with schizophrenia.

Objective: To identify an epigenetic signature (quantified as polymethylation score [PMS]) of schizophrenia using machine learning applied to genome-wide blood DNA-methylation data; evaluate whether differences in blood-derived PMS are mirrored in data from postmortem brain samples; test whether the PMS is associated with alterations of dorsolateral prefrontal cortex hippocampal (DLPFC-HC) connectivity during working memory in healthy controls (HC); explore the association between interactions between polygenic and epigenetic risk with DLPFC-HC connectivity; and test the specificity of the signature compared with other serious psychiatric disorders.

Design, setting, and participants: In this case-control study conducted from 2008 to 2018 in sites in Germany, the United Kingdom, the United States, and Australia, blood DNA-methylation data from 2230 whole-blood samples from 6 independent cohorts comprising HC (1238 [55.5%]) and participants with schizophrenia (803 [36.0%]), bipolar disorder (39 [1.7%]), major depressive disorder 35 [1.6%]), and autism (27 [1.2%]), and first-degree relatives of all patient groups (88 [3.9%]) were analyzed. DNA-methylation data were further explored from 244 postmortem DLPFC samples from

[1] **Published as**: Chen, J., Zang, Z., Braun, U., Schwarz, K., Harneit, A., Kremer, T., ... & Schwarz, E. (2020). Association of a Reproducible Epigenetic Risk Profile for Schizophrenia With Brain Methylation and Function. JAMA psychiatry.

136 HC and 108 patients with schizophrenia. Neuroimaging and genome-wide association data were available for 393 HC. The latter data was used to calculate a polygenic risk score (PRS) for schizophrenia. The data were analyzed in 2019.

Main outcomes and measures: The accuracy of schizophrenia control classification based on machine learning using epigenetic data; association of schizophrenia PMS scores with DLPFC-HC connectivity; and association of the interaction between PRS and PMS with DLPFC-HC connectivity.

Results: This study included 7488 participants (4395 men [58.7%]), of whom 3158 (2230 men [70.6%]) received a diagnosis of schizophrenia. The PMS signature was associated with schizophrenia across 3 independent data sets (area under the curve [AUC] from 0.69 to 0.78; P value from 0.049 to 1.24 × 10-7) and data from postmortem DLPFC samples (AUC = 0.63; P = 1.42 × 10-4), but not with major depressive disorder (AUC = 0.51; P = .16), autism (AUC = 0.53; P = .66), or bipolar disorder (AUC = 0.58; P = .21). Pathways contributing most to the classification included synaptic processes. Healthy controls with schizophrenia-like PMS showed significantly altered DLPFC-HC connectivity (validation methylation/magnetic resonance imaging, $t < -3.81$; P for familywise error, $<.04$; validation magnetic resonance imaging, $t < -3.54$; P for familywise error, $<.02$), mirroring the lack of functional decoupling in schizophrenia. There was no significant association of the interaction between PMS and PRS with DLPFC-HC connectivity ($P > .19$).

Conclusions and relevance: We identified a reproducible blood DNA-methylation signature specific for schizophrenia that was correlated with altered functional DLPFC-HC coupling during working memory and mapped to methylation differences found in DLPFC postmortem samples. This indicates a possible epigenetic contribution to a schizophrenia intermediate phenotype and suggests that PMS could be of interest to be studied in the context of multimodal biomarkers for disease stratification and treatment personalization.

2.1.3 Introduction

Schizophrenia is a severe brain disorder thought to be caused by a complex interplay of genetic predisposition and environmental exposure (Weinberger 1987, Cannon 1998, Sullivan, Kendler et al. 2003, van Os, Rutten et al. 2008, Bergen, O'Dushlaine et al. 2012). In the context of gene-environment may account for heritability not captured by other current methods, such as polygenic interplay and developmental programming (van Os, Kenis et al. 2010, Brown 2011), epigenetic mechanisms, such as DNA methylation, have received substantial attention in schizophrenia(Nishioka 2012, Aberg, McClay et al. 2014) and other neuropsychiatric disorders (Mill, Tang et al. 2008, Cecil, Walton et al. 2015, Teroganova, Girshkin et al. 2016, Zhang and Gelernter 2017). As epigenetic mechanisms risk scores (PRS), potential interactions between genetic susceptibility and epigenetic changes are of particular interest. To date, much of this work has studied single genes. For example, methylation at the Single Nucleotide Polymorphism (SNP) rs6265 within the neurodevelopmentally important brain-derived neurotrophic factor (*BDNF)* gene shows a genotype-dependent association with WM, hypoxia-related early life events, and a WM-related schizophrenia intermediate phenotype (Ursini, Cavalleri et al. 2016). Similarly, methylation differences in dopamine receptor D4 (*DRD4)* predict WM, suggesting that the dopaminergic methylation status could affect cognitive functions in a dissociable manner (Lewis, Henderson-Smith et al. 2019). Likewise, membrane-bound catechol-O-

methyltransferase (MB-COMT) promoter methylation is associated with DLPFC activity during WM (Walton, Liu et al. 2014) and methylation of the COMT Val(158) allele with lifetime stress, WM performance and prefrontal activity during WM (Ursini, Bollati et al. 2011).

While these studies have explored targeted hypotheses, methylation differences in single genes (Ursini, Bollati et al. 2011, Ursini, Cavalleri et al. 2016) are only weakly associated with schizophrenia. To capture systems-level effects, methylation differences have been explored across the genome in whole-blood (Aberg, McClay et al. 2014, Hannon, Dempster et al. 2016) and post-mortem brain tissue samples (Pidsley, Viana et al. 2014, Jaffe, Gao et al. 2016). Joint analysis of genome-wide methylation and genotyping data provided evidence that the methylation changes found in schizophrenia differ from those associated with polygenic risk, but overlap with previously identified genetic susceptibility loci (Schizophrenia Working Group of the Psychiatric Genomics 2014, Hannon, Dempster et al. 2016, Jaffe, Gao et al. 2016). This supports the notion that the methylation differences are indeed associated with schizophrenia and not merely a result of disease-unrelated environmental factors (Lim and Song 2012, Zakhari 2013, Voisin 2015). An important question is whether, similar to polygenic scores, there is a combined contribution of these methylation differences on illness risk. A machine learning study provided evidence for a schizophrenia poly-methylation signature (PMS) that could be validated in independent data (Chen and Schwarz 2017). We do not, however, yet understand whether this PMS is relevant in the brain and how it is associated with genomic risk with regard to neural effects.

Components of the risk architecture of schizophrenia have been successfully interrogated using a strategy termed *imaging genetics* (Meyer-Lindenberg 2010), an approach that has facilitated the identification of so called neural 'intermediate phenotypes', illness-associated, heritable traits that reflect a manifestation of illness liability(Lenzenweger 2013). One of the best established intermediate phenotypes of schizophrenia is Dorsolateral-Prefrontal-Cortex–Hippocampus (DLPFC-HC) connectivity during working memory (Malki, Tosto et al.) performance (Meyer-Lindenberg, Olsen et al. 2005, Malki, Tosto et al. , Schneider, Walter et al. 2017). DLPFC-HC connectivity is altered in healthy first-degree relatives of patients with schizophrenia and is linked to risk alleles of established genome-wide significant schizophrenia gene variants, such as ZNF804A (Rasetti, Sambataro et al. 2011). WM is impaired in schizophrenia, linked to genetic risk (Schwarz, Tost et al. 2016) and affected by environmental factors contributing to illness risk, including childhood trauma and socioeconomic status (Evans and Schamberg 2009, Vargas, Lam et al. 2018), strengthening the argument that this intermediate phenotype reflects risk-related pathophysiological processes.

Using this approach, we investigated genome-wide DNA methylation data from 2041 whole-blood samples from four independent cohorts comprising 1238 healthy controls (HC) and 803 patients with schizophrenia (SCZ). We aimed to identify and validate a PMS differentiating SCZ from HC. Subsequently, we assessed whether the PMS reproducibly predicted altered DLPFC-HC connectivity in HC and explored potential interactions with a schizophrenia PRS. Finally, we used genome-wide DNA methylation data from post mortem DLPFC samples from 136 healthy donors and 108 donors with schizophrenia to assess whether the peripheral PMS was mirrored by analogous changes in the brain. This investigation aimed at characterizing the systems-level relationship between genetic and epigenetic risk for schizophrenia, and to test the effects of these parameters on schizophrenia-relevant neural functioning.

2.1.4 Methods

Cohorts

Genome-wide DNA methylation data from 7 cohorts, denoted as discovery methylation, validation methylation, validation methylation/magnetic resonance imaging (MRI), validation MRI, validation post mortem, specificity methylation, and relatives methylation, were analyzed in this work (demographic characteristics are summarized in **Table 1** and eTable 1 in the Supplement). All participants (or their legal next-of-kin in case of brain tissue donors) gave written or audiotaped informed consent and all studies were approved by the local ethics committees (eMethods in the Supplement). Discovery methylation and validation methylation consisted of 2 independent cohorts 767 patients with schizophrenia [95.5%] and 755 HC [61.0%]). These data sets were used to identify and validate a PMS using machine learning. Validation methylation/MRI (36 patients with schizophrenia [4.5%] and 331 HC [26.7%]) was used for additional validation of the PMS, while MRI data from a subset of the HC (241 [19.5%]; n-back WM functional MRI paradigm (Rasetti, Sambataro et al. 2011, Schneider, Walter et al. 2017)) were used to test associations with DLPFC-HC connectivity. The MRI data from validation MRI (n = 152 HC) acquired during the Sternberg WM task (Geiger, Moessnang et al. 2018) were used to validate the identified DLPFC-HC connectivity associations. The functional MRI face-matching task (Cao, Bertolino et al. 2016) was used in validation MRI to explore the specificity of findings for WM. Validation post mortem comprised genome-wide methylation data from postmortem DLPFC samples 108 schizophrenia [44.3%]; 136 HC [55.7%]) and was used to assess the overlap of the PMS with brain methylation changes. Genome-wide association study (GWAS) data in validation methylation/MRI and validation MRI were used to compute a schizophrenia PRS to test associations with the PMS and interactions with PMS on brain functional connectivity. For specificity testing, we explored DNA methylation data from a cohort (specificity methylation) of patients with bipolar disorder (BD; 39 [1.7%]), major depressive disorder (MDD; 35 [1.6%]) and autism (27 [1.2%]). To further characterize the effect of genetic schizophrenia risk, a cohort (relatives methylation) of unaffected first-degree relatives of participants with schizophrenia (27 [1.2%]), BD (15 [0.7%]), MDD (29 [1.3%]), and autism (17 [0.8%]) were analyzed. The machine-learning procedure used here was further applied to GWAS data from GWAS molecular genetics of schizophrenia (GWAS MGS) (n = 2718 HC and n = 2296 schizophrenia) to identify risk components not captured by the PRS that could potentially explain PMS-associated findings.

The data, methods, and confounding correction are detailed in the eMethods in the Supplement. Most patients were taking medication. Neuroimaging analyses focused on HCs not taking medication to demonstrate that PMS associations were not associated with medication. Additionally, we explored the association of the PMS and DLPFC-HC connectivity with chlorpromazine equivalents.

Gene and pathway assignment of genome-wide DNA methylation data

For each gene, CpGs harbored by the gene with an extended window size of 20 kb downstream and upstream were used for analysis. CpG locations and gene annotations used for mapping were obtained from the R library *IlluminaHumanMethylationEPICanno.ilm10b2.hg19 (R Foundation)*. Genes were binned into biological process categories (denoted as pathways) using the *GeneOntology* database (data freeze in Dec. 2018) (Gene Ontology, Blake et al. 2013). 2846 pathways comprising between 10 and 200 genes with at least 10 CpGs per pathway were used for analysis.

Machine learning, replication and validation

An updated biologically informed machine learning (BioMM) approach was used (eMethods in the Supplement). (Chen and Schwarz 2017). BioMM is a 2-stage machine-learning approach that first builds separate machine learning models for methylation sites mapping to each of the 2846 pathways, yielding 1 machine-learning model per pathway (first-stage). This procedure compresses data from individual methylation sites into a pathway-level feature. Then, a second-stage algorithm integrates these pathway-level features into a systems-level classifier. BioMM was trained on discovery methylation and the algorithm then applied to all other data sets. In each data set, the output of BioMM was a score (PMS) that quantified the likelihood of a given participant being in the schizophrenia group. To assess predictive accuracy, we determined the area under the receiver operating characteristic curve (AUC) as well as Nagelkerke R^2.

Statistical analysis

Associations between PMS and DLPFC-HC functional connectivity were assessed using a linear regression in SPM, version 12, using PMS as covariate of interest and age and sex as covariates of noninterest. Statistical significance was set at $P < .05$. The PMS and DLPFC-HC connectivity associations in the imaging space were corrected using a familywise small-volume correction in the hippocampus (eFigures 1 and 2 in the Supplement) based on the automated anatomical labeling template. Associations between PMS and the schizophrenia PRS, as well as the association of the PMS by PRS interaction with DLPFC-HC connectivity, were tested using a multiple linear regression that accounted for the effects of sex, age, and (for PRS-associated analyses) 10 genetic principal components. For details and analysis of potential confounding, see the eMethods in the Supplement.

2.1.5 Results

Determination and validation of a PMS in blood samples

Genome-wide DNA methylation data from discovery methylation (675 [30.3%]) were used for model training. In this data set the model showed a cross-validation accuracy of an AUC of 0.78 ($P = 2.95 \times 10{-}6$, corrected for 20 potential confounders; $R^2 = 21\%$). The model was then predicted into validation methylation (847 [38.0%]), yielding an AUC of 0.69 ($P = 1.24 \times 10{-}7$; $R^2 = 10.5\%$). For additional validation, the model was predicted into validation methylation/MRI (367 [16.5%]), yielding an AUC of 0.74 ($P = .049$; $R^2 = 21.8\%$). This demonstrates the biological reproducibility of the PMS (**Table 2** and **Figure 1**).

Identification of implicated biological pathways

To identify pathways with methylation changes contributing strongly to the epigenetic signature, pathway-level (second-stage) data from BioMM in discovery methylation were used. The 10 pathways most associated with schizophrenia are shown in **Figure 2** (eTable 2 in the Supplement). Of the 2846 pathways, 917 (32.2%) were positively associated with diagnosis and 57 (6.2%) of these showed significance at P for familywise error <.05. They did not differ significantly from the remaining pathways regarding size (determined as the number of CpGs per pathway; $\beta = 2.46 \times 10{-}5$; $P = .61$). Individual genes within the 10 pathways harboring methylation differences have been previously implicated in schizophrenia (eTables 3 and 4 in the Supplement).

Identifying and validating the association between PMS and functional DLPFC-HC coupling

The schizophrenia PMS predicted in data from HCs not taking medication from validation methylation/MRI (241 [10.8%]) was negatively associated with DLPFC-HC functional connectivity during WM (P for familywise error, <.04; $F_{1,237}$ = 14.55; t_{237} = −3.81; bilateral hippocampus-corrected, peak voxel at 33, −22, −13) in the right posterior hippocampus (Figure 3; eFigure 3 in the Supplement). The negative association between PMS and DLPFC-HC connectivity was replicated in validation MRI (n = 152), with the Sternberg WM paradigm within right posterior hippocampus (P for familywise error = .02; t148 = −3.54; peak voxel at 33, −37, −7) corrected for right posterior hippocampus (Figure 3; eFigure 3 in the Supplement). Post hoc 1-sample t tests revealed negative connectivity between DLPFC and HC in the validation methylation/MRI (t240 = −14.09; P < .001) and the validationMRI (t_{151} = −6.94; P < .001) sample. Associations between PMS and DLPFC-HC connectivity were specific for the WM tasks compared with the faces tasks and were not confounded by brain-structural effects (eResults in the Supplement). Association between PMS and PRS and interactions on DLPFC-HC coupling
The DLPFC-HC connectivity was not associated with the schizophrenia PRS in validation methylation/MRI (β = 126.6; P = .79) or validation MRI (β = 401.3; P = .67). Similarly, no significant interactions between PMS and PRS on DLPFC-HC connectivity were found in validation methylation/MRI (β = −7280; P = .19) or validation MRI (β = −10 540; P = .41). The BioMM model was used to identify a risk score from GWAS MGS data using the same pathways assignment as used for DNA methylation data. The resulting associations with PMS and DLPFC-HC connectivity are described in the eResults and eTable 5 in the Supplement. Analysis of relatives methylation demonstrated that none of the groups of relatives showed significant PMS differences compared with HC (eFigure 4 in the Supplement).

Prediction of PMS in DLPFC post-mortem brain samples
Predicted PMS in validation post mortem was significantly higher in schizophrenia compared with HC (AUC = 0.63; P = 1.42 × 10−4; R^2 = 8.3%). Notably, the reverse prediction (i.e., building a PMS from postmortem brain data and testing this score in blood data) did not allow case-control differentiation (eTable 6 in the Supplement) and showed no association with DLPFC-HC connectivity (validation methylation/MRI: P = .89; validation MRI: P = .92).

Assessment of the robustness, diagnostic specificity and residual confounding
Permutation of diagnostic labels, as well as the random selection of pathways used by BioMM, supported the significance and robustness of the PMS finding (eResults and eTables 7 and 8 in the Supplement). Analysis of the specificity methylation cohort further showed that the PMS increase was specific for schizophrenia (eTable 9 and eFigure 5 in the Supplement). Because despite the confounding correction the PMS was associated with some of all potential variables, additional analyses regarding confounding can be found in the eResults and eTables 10 and 11 in the Supplement.

2.1.6 Discussion

In this article, we identified a blood DNA methylation signature that reproducibly differentiated schizophrenia from HC and several other major neuropsychiatric disorders. The underlying biological pathways implicated several synaptic processes as contributing most to the classification. A more schizophrenia-like epigenetic profile was associated with an established intermediate phenotype for the disorder,

persistent DLPFC-HC connectivity, in HC during 2 different WM tasks. Notably, the epigenetic signature could also differentiate schizophrenia from HC in data from DLPFC postmortem samples, supporting the relevance of the identified blood-derived epigenetic signature for brain-associated phenotypes associated with schizophrenia in vivo and ex vivo.

The pathways contributing most to the schizophrenia classification comprised synaptic and neurodevelopmental processes. This is consistent with previous results showing a co-localization of epigenetic changes with genetic susceptibility variants of schizophrenia (Hannon, Dempster et al. 2016), which are over-represented in synaptic pathways (Network and Pathway Analysis Subgroup of Psychiatric Genomics 2015). The important role of synaptic processes in schizophrenia is supported by findings from induced pluripotent stem cells (Brennand, Simone et al. 2011, Wen, Nguyen et al. 2014), and changes in different omics modalities pointing towards a synaptic pathology [e.g.(Osimo, Beck et al. 2019)]. Furthermore, schizophrenia-related neurodevelopmental processes show an over-representation of methylation changes in the schizophrenia prefrontal cortex, and these are enriched for sites undergoing epigenetic changes during fetal neocortex brain development (Pidsley, Viana et al. 2014). Notably, post-mortem expression and methylation studies support that the neural effects of epigenetic and genetic risk factors for schizophrenia already occur during early brain development, rather than the typical onset-age of the illness (Birnbaum and Weinberger 2017), highlighting their importance for altered neurodevelopment in schizophrenia.

We show that during two WM tasks, higher, more schizophrenia-like scores of the identified DNA methylation signature were associated with stronger negative DLPFC-HC connectivity. DLPFC-HC connectivity is altered in the same way in SCZ, unaffected first-degree relatives, as well as HC carrying specific risk genetic variants (Meyer-Lindenberg, Olsen et al. 2005, Esslinger, Walter et al. 2009, Rasetti, Sambataro et al. 2011, Schneider, Walter et al. 2017), with more negative DLPFC-HC connectivity being related to higher risk for schizophrenia. Furthermore, functional and structural abnormalities in the hippocampus and prefrontal cortex are at the core of the schizophrenia pathophysiology (Weinberger 1987, Bahner, Demanuele et al. 2015) and disease processes of both areas are highly interassociated (Lipska and Weinberger 2000, Bahner, Demanuele et al. 2015). For example, altered DLPFC-HC microcircuits in post-mortem brains of SCZ influence both excitatory and inhibitory cells as well as interneurons (Harrison and Weinberger 2005). Interestingly, while the general liability for schizophrenia and specific genetic risk variants do impact this phenotype, common polygenic risk for schizophrenia has been repeatedly demonstrated not to (Erk, Mohnke et al. 2017, Miller, Scult et al. 2018), suggesting either a restricted set of genes responsible and/or more complex gene-environment interactions to shape DLPFC-HC connectivity. We extend these findings by demonstrating that an epigenetic risk signature for schizophrenia correlates with DLPFC-HC coupling and we did not find sufficient evidence for a direct polygenic and interactive association, suggesting that epigenetic analysis provides pathophysiologically relevant information not captured by PRS. Also, a BioMM-derived PRS was not associated with PMS or DLPFC-HC and showed no interactions with PMS on DLPFC-HC connectivity. This supports that the PMS outperformed classifiers identified from genetic association data and the observed PMS effects were not primarily driven by underlying genetics. The lack of PMS differences in relatives of patients with schizophrenia, BD, MDD and autism supported this finding. Notably, the accuracy a classifier can achieve is limited by the clinical and biological heterogeneity of schizophrenia. Applying multimodal subgroup identification strategies may aid in deconstructing this heterogeneity and ultimately contribute to an alternative disease classification.

Epigenetic modifications including DNA methylation are strongly influenced by life-span environmental exposures, such as postnatal mother-infant interactions (Fagiolini, Jensen et al. 2009) and experiencing of stress-related events (Klengel, Pape et al. 2014), making the poly-epigenetic signature a potential proxy on which cumulative environmental risk factors could converge. Moreover, studies in animals and humans indicated that several neurobiological processes at different stages of development can modify DLPFC-HC connectivity across the life span. This includes early neuronal formation (Lee, Dvorak et al. 2012, Phillips, Bartsch et al. 2012) and synaptic plasticity related processes (Fagiolini, Jensen et al. 2009, Meadows, Guzman-Karlsson et al. 2015), raising the possibility that the neural impact of epigenetic and genetic schizophrenia risk on synaptic processes during early development may have a lasting impact on DLPFC-HC connectivity. Such an interpretation is also consistent with results from animal studies showing that lesions in the hippocampus lead to delayed maturation and dysfunction of the DLPFC (Lipska and Weinberger 2000, Bahner, Demanuele et al. 2015). Taken together, the present results may suggest that DLPFC-HC connectivity is influenced by environmental risk accumulation mediated by the association of altered DNA methylation with synaptic plasticity (Fagiolini, Jensen et al. 2009, Meadows, Guzman-Karlsson et al. 2015).

In this article, the PMS was found to differentiate schizophrenia from HC when predicted in DLPFC DNA methylation data, indicating that elements of the signature were represented in the brain and may have mediated the observed DLPFC-HC connectivity association. This finding is consistent with the previously observed correlation between postmortem brain and blood methylation. (Walton, Hass et al. 2016, Edgar, Jones et al. 2017, Braun, Han et al. 2019). Such cross-tissue correlation has been hypothesized to result from genetic influences, casting doubt on the added value of peripheral epigenome-wide association studies in brain disorders (Hannon, Lunnon et al. 2015). However, the association of the PMS with schizophrenia and schizophrenia-relevant neural phenotypes based on epigenetic changes in pathophysiologically relevant pathways, as well as the lack of evidence for an association with polygenic susceptibility, supports their use for integrative, multimodal approaches toward disease stratification and potentially treatment personalization. Notably, a schizophrenia PMS derived from DLPFC samples did not predict case-control status in blood samples and was not correlated with brain functional connectivity. This may have been due to the larger biological and methodologic variability in brain samples (such as cell heterogeneity or postmortem effects) or the comparatively smaller sample size.

The present study has several limitations. First, most investigated patients were taking medication and the association of medication with the PMS could not be excluded. However, the association between the PMS and DLPFC-HC connectivity in HC not taking medication contradicts the idea that the identified methylation signature is a consequence of medication. Second, because of limited available data, the associations between DLPFC-HC connectivity and the PMS could not be explored in patients. This could have identified a stronger contribution of genetic schizophrenia risk toward the PMS associations, which may have been affected by limited statistical power. Third, while the PMS replicated in data from postmortem DLPFC samples, the explained variance in the brain data was comparatively low. This necessitates further studies to isolate the brain-specific associations of methylation differences with brain function. Fourth, despite extensive efforts to correct findings for confounding effects, we cannot exclude the potential presence of residual confounding. However, the analysis of patients with BD and MDD as well as autism suggested that the PMS increase in schizophrenia was not driven by potential confounders with transdiagnostic relevance. These findings, and PMS effects in relatives, require validation in larger cohorts. Finally, the machine learning approach used was based on the annotation of CpGs to genes and biological pathways,

which may be biased and not suitably index biological function and can lead to an overlap of annotated genes between pathways. As machine learning isolates the most predictive features mapped to a given pathway, this can lead to a loss of biological specificity if the most predictive CpGs are shared among different pathways. This necessitates follow-up experiments to more precisely characterize the role of the identified pathways for mediating epigenetic risk in schizophrenia.

2.1.7 Conclusion

This study shows that a reproducible and specific blood DNA-methylation signature of schizophrenia was correlated with functional DLPFC-HC coupling as an intermediate phenotype for schizophrenia, mapped to methylation differences found in DLPFC postmortem samples of schizophrenia, and indexed biological pathways associated with synaptic function. These results help to characterize the systems-level association between genetic, epigenetic, and environmental risk for schizophrenia. They further support the use of PMS for multimodal biomarker discovery strategies aimed at disease stratification and the development of novel, personalized therapeutic approaches.

Tables

Table 1. Overview of sample characteristics. The rightmost three columns indicate the sample numbers for which the respective data were available (Methylation: DNA-Methylation; Genotypes: whole-genome genetic association data; Imaging: functional MRI data [validation$_{Meth/MRI}$: n-back task; validation$_{MRI}$: Sternberg task]; Case: schizophrenia patients; SubCtl: Subset of healthy controls; AUT: Autism; BP: Bipolar disorder; MDD: Major depressive disorder; relSCZ: First-degree relatives of patients with schizophrenia; relAUT: First-degree relatives of patients with autism; relBP: First-degree relatives of patients with bipolar disorder; relMDD: First-degree relatives of patients with major depressive disorder).*40 subjects with missing age information.

Cohort	Tissue	Status	Sex (m/f)	Age	Methylation	Genotype	Imaging
Discovery Methylation	Whole Blood	Control	142/180	37.7±15.2	322		
		Case	254/99	43.7±14.7	353		
Validation Methylation		Control	319/114	45.0±12.1	433		
		Case	283/131	46.6±13.6	414		
Validation Methylation/MRI		Control	168/163	27.8±10.2	331		
		SubCtl	126/115	28±10.7	241	241	241
		Case	28/8	33.8±10.4	36		
specificity Methylation		AUT	18/9	33.8±9.6	27		
		BP	16/23	36.4±10.5	39		
		MDD	11/24	37.2±12.2	35		
Relatives Methylation		relSCZ	8/19	37.2±14.4	27		
		relAUT	8/9	44.9±9.1	17		
		relBP	8/7	36.9±12.3	15		

		relMDD	11/18	30.3±10.4	29		
validation$_{MRI}$		Control	66/86	26.9±8.8	152	152	152
validation$_{postmortem}$	Brain	Control	90/46	46.5±16.1	136		
		Case	59/49	52.7±14.5	108		
GWAS$_{MGS}$	Whole Blood	Control	1301/1417	50.6±16.4		2718	
		Case	1606/690	43.4±11.7*		2296	

Table 2. Prediction performance of BioMM on different cohorts using discovery$_{Meth}$ as the discovery set

Data sets	AUC	R^2
discovery$_{Methylation}$	0.78	0.21
validation$_{Methylation}$	0.69	0.105
validation$_{Methylation/MRI}$	0.74	0.218
validation$_{postmortem}$	0.63	0.083

Figures

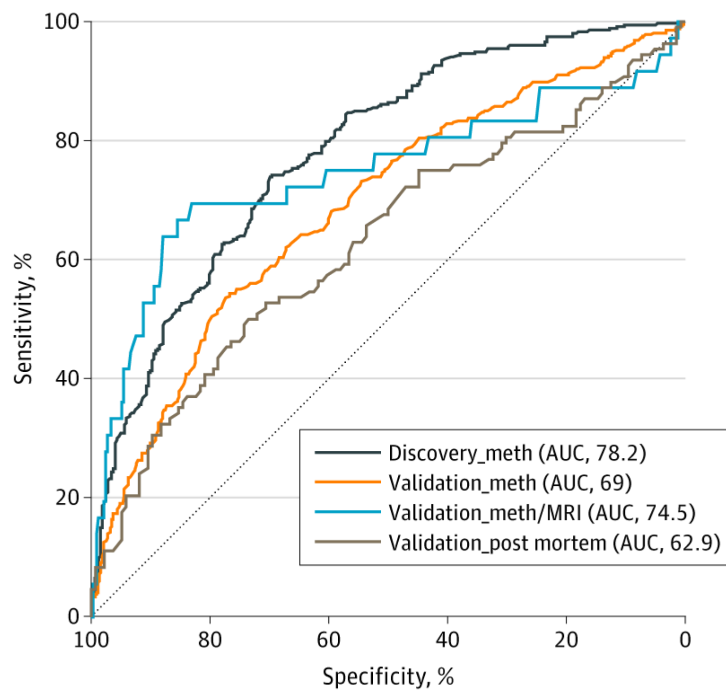


Figure 1. Prediction performance of Biologically Informed Machine Learning (BioMM) on Different Cohorts Using Discovery Methylation as the Training Data

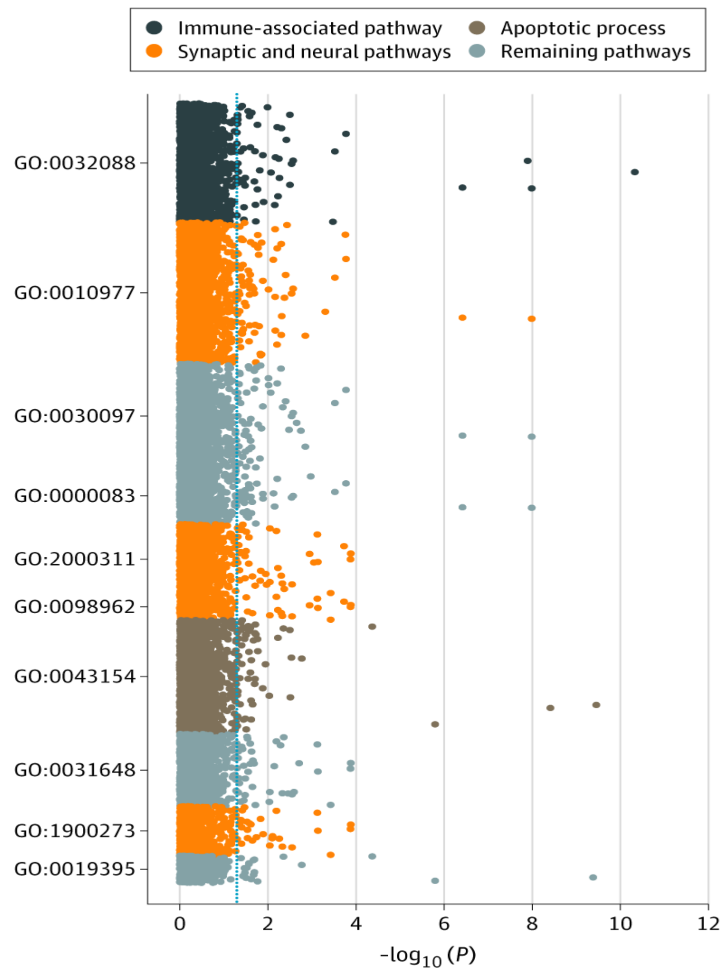


Figure 2. Biological pathways contributing most to the PMS and significance of the individual methylation sites. GO:0032088: negative regulation of NF-kappaB transcription factor activity; GO:0010977: negative regulation of neuron projection development; GO:0030097: hemopoiesis; GO:0000083: regulation of transcription involved in G1/S transition of mitotic cell cycle; GO:2000311: regulation of AMPA receptor activity; GO:0098962: regulation of postsynaptic neurotransmitter receptor activity; GO:0043154: negative regulation of cysteine-type endopeptidase activity involved in apoptotic process; GO:1903427: negative regulation of reactive oxygen species biosynthetic process; GO:0031648 protein destabilization; GO:1900273: positive regulation of long-term synaptic potentiation; GO:0019395: fatty acid oxidation. The immune related pathway is illustrated in yellow, an apoptotic process in purple, synaptic and neural pathways in green and the remaining pathways in gray. The dots indicate the significance of individual CpG within these pathways. The blue line marks the uncorrected significance level of 0.05.

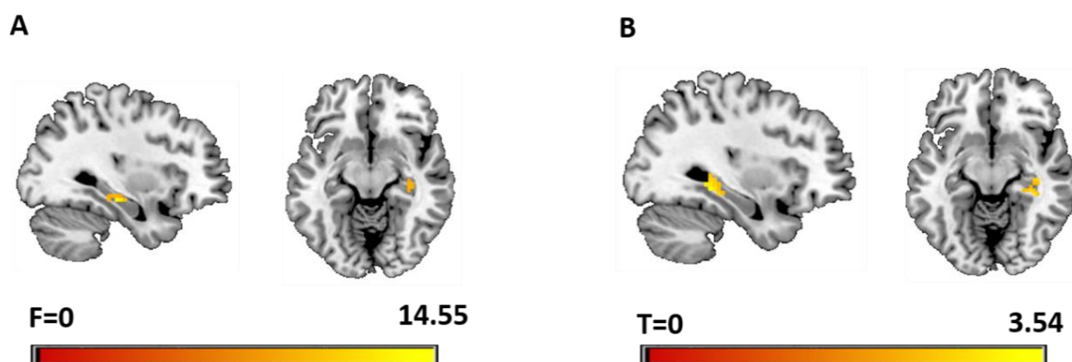


Figure 3. Association between the predicted PMS and DLPFC-HC connectivity in un-medicated, healthy controls. A) validation$_{Methylation/MRI}$ (n-back task): uncorrected results ($p < 0.05$, shown in hippocampus) with

peak voxel ($F_{1,237}$=14.55, T_{237}=-3.81, MNI: [33 -22 -13]) in the right hippocampus significant after bilateral hippocampus correction in healthy subjects (n = 241) (P_{FWE}=0.040). **B)** validation$_{MRI}$ (Sternberg task): uncorrected results (p < 0.05, shown in hippocampus) with peak voxel (T_{148}=-3.54, peak voxel at [33 -37 -7]) in the right hippocampus showing a significantly negative association after right posterior hippocampus correction in healthy subjects (n = 152) (P_{FWE}=0.016). Age and sex are controlled as covariates of non-interests. For presentation purpose, imaging results are shown in p<0.05 uncorrected threshold at similar spatial plane (MNI: X=32, Y=-33, Z=-10).

2.1.8 Supplementary Methods

eMethods: Genome-wide DNA methylation data
eMethods: Clinical characteristics of investigated cohorts
eMethods: Preprocessing of genome-wide DNA methylation data
eMethods: Genotyping QC and imputation
eMethods: Population structure and relatedness testing
eMethods: Polygenic risk score (PRS) determination on genotype data
eMethods: BioMM framework
eMethods: Permutation test procedure
eMethods: Analysis of schizophrenia specificity
eMethods: Machine learning for case-control prediction based on GWAS data
eMethods: Working memory and faces matching task - imaging data acquisition and preprocessing
eMethods: First-level DLPFC-HC functional connectivity
eMethods: Structural MRI data acquisition, preprocessing and influence.
eMethods: Task specificity and structural influences
eMethods: Medication influences on PMS and DLPFC-HC connectivity
eMethods: Statistical analysis
eResults: Task specificity and structural confounding influences on the identified associations between PMS and DLPFC-HC coupling
eResults: Analysis of residual confounding effects on the PMS and DLPFC-HC connectivity
eResults: Associations between PMS, DLPFC-HC connectivity and medication in schizophrenia patients
eResults: Association between BioMM-derived polygenic risk signature, PMS and DLPFC-HC connectivity
eResults: Permutation test analysis
eFigure1: PMS comparision in validationmeth/MRI and relativesMeth
eFigure2: PMS comparision in validationMeth/MRI and specificityMeth
eFigure3: Post-hoc partial regression plots of the association between PMS and DLPFC-HC connectivity
eFigure4: DLPFC ROI and hippocampus mask in the validationMeth/MRI sample
eFigure5: DLPFC ROI and hippocampus mask in the validationMRI sample
eTable1: Differences of subject demographics between patients and controls
eTable2: Overall prediction performance of BioMM on different cohorts and specificity analysis
eTable3: The association between PMS and DLPFC-HC, PMS and PRS, PMS and PRSBioMM
eTable4: Permutation of diagnostic label for both machine learning prediction and the subsequent testing of imaging associations
eTable5: Permutation of pathway level features for both machine learning prediction and subsequent testing of imaging associations
eTable6: The association between predicted PMS and the confounding variables in controls
eTable7: Prediction performance of BioMM on different cohorts and specificity analysis following residualization of predicted PMS scores against all potential covariates using linear regression
eTable8: Top 10 schizophrenia-associated pathways in the discovery sample (discoveryMeth)
eTable9: 30 most significant CpGs derived from top 10 pathways
eTable10: The existing evidence for the top genes harboring top CpGs from eTable 9
eTable11: Prediction performance of BioMM on different cohorts using validationpostmortem as the discovery set

eMethods

Genome-wide DNA methylation data

DNA methylation data from discovery$_{Meth}$ and validation$_{Meth}$ were downloaded from the GEO database (GSE80417 and GSE84727, respectively)(Edgar, Domrachev et al. 2002). All samples were analyzed using Illumina Infinium HumanMethylation450 BeadChip (450k) arrays. Detailed descriptions of cohorts and data acquisition can be found elsewhere(Hannon, Dempster et al. 2016). Subjects part of validation$_{Meth/MRI}$ and validation$_{MRI}$ were recruited at the Central Institute of Mental Health, Mannheim, Germany and data from these cohorts has partially been described previously(Schneider, Walter et al. 2017). Validation$_{postmortem}$ was obtained from the GEO database under accession GSE74193(Jaffe, Gao et al. 2016). A total of 675 samples were scanned on 534 unique subjects, but after exclusion (as suggested by the authors of the original study) of a problematic processing plate and subjects with an age of less than 16, data from 244 subjects were used for analysis. Data processing is described in detail below.

Clinical characteristics of investigated cohorts

Dicovery$_{meth}$

Dicovery$_{meth}$ case-control samples were obtained from the UK and consist of unrelated ancestrally matched cases and controls as described in the original reference (Datta, McQuillin et al. 2010). Briefly, National Health Service (NHS) multicentre and local research ethics approval was obtained and all subjects signed an approved consent form. Patients with schizophrenia were selected based on the International Classification of Diseases 10 (ICD10) criteria and interviewed with the Schedule for Affective Disorders and Schizophrenia-Lifetime Version (SADS-L) schedule(Mannuzza, Fyer et al. 1986). All cases were further interviewed by a second psychiatrist at the probable level of schizophrenia with Research Diagnostic Criteria (RDC). Subjects with brain damage were not included. The control subjects were also interviewed with the initial clinical screening questions of the SADS-L and selected on the basis of not having a family history of schizophrenia, alcoholism or bipolar disorder and for having no past or present personal history of any RDC-defined mental disorder.

Validation$_{meth}$

Validation$_{meth}$ case-control samples were self-identified as born in the British Isles (95% in Scotland) as originally described in (Stone, O'Donovan et al. 2008). All cases met the Diagnostic and Statistical Manual for Mental Disorders-IV edition (DSM-IV) (Sharp, Mefford et al. 2008) and International Classification of Diseases 10th edition (ICD-10) (Janca, Ustun et al. 1993) criteria for schizophrenia. Diagnosis was made by Operational Criteria Checklist (OPCRIT) (McGuffin, Farmer et al. 1984). All case participants were outpatients or stable inpatients. Detailed medical and psychiatric histories were collected. A clinical interview using the Structured Clinical Interview for DSM-IV (SCID) was also performed on schizophrenia cases. Controls were volunteers recruited through general practices in Scotland. Practice lists were screened for potentially volunteers matched by age and sex and by excluding subjects with major mental illness or use of neuroleptic medication. Volunteers who replied to a written invitation were interviewed using a short questionnaire to exclude major mental illness in individual themselves and first degree relatives. All cases and controls gave informed consent. The study was approved by both local and multiregional academic ethics committees.

$Validation_{meth/MRI}$

We included 367 subjects comprising of 331 healthy volunteers and 36 patients with schizophrenia. None of the healthy volunteers had a first-degree relative with a history of mental illness. All subjects were recruited at the Central Institute of Mental Health in Mannheim, Germany (Cao, Bertolino et al. 2016, Schneider, Walter et al. 2017). Patients were recruited from inpatient and outpatient treatment facilities and psychiatric diagnoses were confirmed based on DSM-IV criteria(Schwarz, Moessnang et al. 2019). The association between PMS and DLPFC-HC connectivity was assessed in 241 healthy subjects for which n-back fMRI data was acquired. The same 241 subjects were also used for genotyping. All participants provided written informed consent that has been approved by the ethical committees of the Universities of Heidelberg.

$Specificity_{Meth}$

We have acquired DNA methylation data from patients with bipolar disorder (n=39), major depressive disorder (n=35) and autism (n=27). All patients were obtained at the Central Institute of Mental Health in Mannheim, Germany (Cao, Bertolino et al. 2016, Schneider, Walter et al. 2017). Patients were recruited from inpatient and outpatient treatment facilities and psychiatric diagnoses were confirmed based on DSM-IV and ADOS-G26 (for autism) criteria(Schwarz, Moessnang et al. 2019). All participants provided written informed consent that has been approved by the ethical committees of the Universities of Heidelberg.

$Relatives_{Meth}$

We included 88 unaffected first-degree relatives of patients with schizophrenia (n=27), bipolar disorder (n=15), major depressive disorder (n=29) and autism (n=17). All subjects were recruited at the Central Institute of Mental Health in Mannheim, Germany (Cao, Bertolino et al. 2016, Schneider, Walter et al. 2017). For the first-degree relatives, diagnostic assessments were based on ICD-10 and DSM-IV criteria. All participants provided written informed consent that has been approved by the ethical committees of the Universities of Heidelberg.

$Validation_{MRI}$

In the $Validation_{MRI}$ sample, we investigated 152 healthy participants. All subjects had no history of psychiatric and neurological illness, prior head trauma, or current alcohol or drug abuse. The samples were collected at Central Institute of Mental Health in Mannheim, Germany (Geiger, Moessnang et al. 2018, Zang, Geiger et al. 2018). All participants provided written informed consent that had been approved by ethics committee of the Universitiy of Heidelberg.

$Validation_{postmortem}$

$Validation_{postmortem}$ postmortem brain samples were donated through the Offices of the Chief Medical Examiners of the District of Columbia and of the Commonwealth of Virginia, Northern District to the NIMH Brain Tissue Collection at the National Institutes of Health in Bethesda, MD, according to NIH Institutional Review Board guidelines (Protocol #90-M-0142), which is originally described in (Jaffe, Gao et al. 2016). All postnatal non-psychiatric control donors (N=300) were free from psychiatric and/or neurologic diagnoses and substance abuse according to DSM-IV. All control donors had toxicology screening for the exclusion of acute drug and alcohol intoxication/use at time of death.

$GWAS_{MGS}$

$GWAS_{MGS}$ comprised genome-wide association data from 2718 healthy controls and 2296 patients with schizophrenia. All subjects were of European ancestry. The genotype data was obtained from Database of Genotypes and Phenotypes (dbGaP) (http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap)(Mailman, Feolo et al. 2007). The samples that included subjects with Caucasian origin were collected from (a) GAIN (Genetic Association Information Network) dataset [dbGAP accession number: phs000021.v2.p1] genotyped for 906,600 SNPs(Shi, Levinson et al. 2009); (b) non-GAIN dataset [dbGAP accession number: phs000167.v1.p1] genotyped for 909,622 SNPs(Shi, Levinson et al. 2009). Diagnostic assessments were based on the Diagnostic and Statistical Manual of Mental Disorders criteria. Approximately 10% of the cohorts consisted of patients with schizoaffective disorder who had schizophrenia-like symptoms for at least six months. Detailed information on these cohorts has previously been described in (Shi, Levinson et al. 2009).

Preprocessing of genome-wide DNA methylation data

The Infinium MethylationEPIC BeadChip was used to obtain genome-wide DNA methylation profiles for $validation_{Meth/MRI}$ and $validation_{MRI}$ from whole blood samples. Raw signal intensities were obtained from IDAT files using the *minfi* Bioconductor R package (Aryee, Jaffe et al. 2014). Both background noise subtraction and dye-bias normalization are performed using the function *preprocessNoob* for each sample individually. Red and green intensities were mapped to the M(Cao, Bertolino et al.) and U(nmeth) channels, and the log median intensity in both channels was used to check for low quality samples (the cutoff for low quality samples was defined as 10.5 for both channels, one subject was removed). Intensities from the sex chromosomes were used to predict sex, and we removed 10 samples that had predicted sex distinct from the phenotypic sex. Probes with a detection P-value > 0.05 in at least 1% of samples and annotated with SNPs at the target CpG or single base extension (SBE) site with minor allele frequency > 5% were filtered out. DNA methylation data from the $validation_{postmortem}$ set were assessed using the Illumina HumanMethylation450 microarray. Raw signal intensities from IDAT files were preprocessed and normalized same as above using the *minfi* R package.

The data of four cohorts from the whole blood samples were corrected to account for the influence of potential confounders, which comprised cigarette smoking (de Leon and Diaz 2005), population structure (Liu, Hutchison et al. 2010), cellular composition, gender and age at the time of recruitment. Smoking was quantified from DNA methylation levels as described previously (Zeilinger, Kuhnel et al. 2013, Elliott, Tillin et al. 2014). Population structure was determined from methylation data via Principal Components Analysis. Specifically, the first 10 principal components were considered as covariates. Cellular composition was quantified using the Epigenetic Clock tool (https://dnamage.genetics.ucla.edu/) (Horvath 2013) and included the seven recommended cell types: CD8.naive, CD8pCD28nCD45RAn, PlasmaBlast, CD4T, NK, Mono, Gran. The $validation_{postmortem}$ set was corrected to adjust for gender, age, ethnic background and the first four PCs of the negative control probes on the microarrays, as well as brain cell types provided by (Jaffe, Gao et al. 2016). All covariates were used in a linear model to residualize each given methylation probe. This was performed separately for the each cohort and the resulting residuals were used for downstream analysis. For all cohorts, we focused on the overlapping set of autosomal methylation sites to limit the potential influence of sex on machine learning due to the phenomenon of X chromosome inactivation or the existence of an additional X chromosome in female subjects. In total, 389,228 CpGs were retained for subsequent analysis.

Genotyping QC and imputation

Infinium PsychArray BeadChip by Illumina ("PsychChip") was used for genotyping samples from $Validation_{Meth}/_{MRI}$ and $Validation_{MRI}$. The initial number of SNPs was 577,832 without considering chromosome Y and the mitochondrial DNA. For all given samples, standard quality control (QC) and imputation are performed using Gimpute pipeline (Chen, Lippold et al. 2018). The following QC steps were applied: 1.) Remove male subjects with more than 10 haploid heterozygous SNPs on chromosome X; 2.) Remove SNPs with missing genotyping rate > 5%; 3.) Exclude samples with missingness >= 0.02; 4.) Exclude samples with autosomal heterozygosity deviation |Fhet| >= 0.2; 5.) Remove SNPs with the proportion of missing genotyping > 2%; 6.) If controls existed in the dataset, remove SNPs with difference in SNP missingness between cases and controls >= 0.02; 7.) Remove SNPs if the Hardy-Weinberg equilibrium exact test P-value was < 1 x 10-6 in controls. Imputation was carried out using IMPUTE2/SHAPEIT (Howie, Donnelly et al. 2009, Howie, Fuchsberger et al. 2012, Delaneau, Zagury et al. 2013), which chooses a European reference panel for each study sample in each 3 Mb segment of the genome. This imputation reference set is from the full 1000 Genome Project dataset (August 2012, 30,069,288 variants, release "v3.macGT1"). The length of buffer region is set to be 500 kb on either side of each segment. All other parameters were set to default values implemented in IMPUTE2.

Genome-Wide Human SNP Array 6.0 by Affymetrix was used for $GWAS_{MGS}$ genotyping. The same genotyping QC and imputation procedure as above was applied for $GWAS_{MGS}$ data.

Population structure and relatedness testing

After imputation, SNPs with high imputation quality (INFO >= 0.6) and successfully imputed in >= 20 samples were retained. The proportion of alleles shared identity-by-decent estimated using PLINK was used to identify relatedness for all pairs of samples (Stevens, Heckenberg et al. 2011). The following criteria were used to select a subset of autosomal SNPs for relatedness testing: 1.) SNPs from the MHC region were excluded (chr6:28,477,797-33,448,354); 2.) SNPs were pruned based on linkage disequilibrium (r2 > 0.02 within 50 variant windows); 3.) SNPs with minor allele frequency (MAF) < 0.05 were removed. A threshold of $\hat{\pi}$ > 0.2 was used to identify related pairs of samples and exclude one member of each pair at random. Using the same set of autosomal SNPs, we determined principal components to be used as covariates during downstream analyses. The final imputed dataset comprised of N=7,660,409 autosomal SNPs for $Validation_{Meth}/_{MRI}$ and $Validation_{MRI}$, and N= 11,798,966 autosomal SNPs for $GWAS_{MGS}$.

Polygenic risk score (PRS) determination on genotype data

The schizophrenia PRS was computed using Psychiatric Genomics Consortium (PGC) summary statistics taken from (Schizophrenia Working Group of the Psychiatric Genomics 2014) following the method developed by Purcell and colleagues (International Schizophrenia, Purcell et al. 2009) and using the PRSice software (Euesden, Lewis et al. 2015). Briefly, PRSs were calculated by summing schizophrenia-associated alleles, weighted by the natural log of the odds ratio. To ensure that SNPs were not in high linkage disequilibrium (LD) with one another, clumping was applied on the genotype data using an LD r^2 threshold of 0.1 and a genomic distance threshold of 250 kb. PRSs were constructed based on the *P*-value threshold 0.05. The PRS score was then transformed into z-scores separately for $validation_{Meth/MRI}$ and $validation_{MRI}$ and utilized for subsequent analyses.

BioMM framework

BioMM framework consisted of two major stages which is an updated version of the BioMM (Chen and Schwarz 2017) and is publically available as a Bioconductor R package (vs 1.1.11) (Chen and Schwarz 2019). Parameters of BioMM were optimized through 100-fold bootstrapping in the training data. Feature selection and random forest models with 1000 trees were employed at both stages. At the first stage, methylation sites with a case-control difference of $P < 0.05$, 0.1, 0.5 and 1 (no feature selection) using the Wilcoxon test were selected for a given pathway to reduce the impact of non-predictive features. The best p value threshold was determined by 100 times bootstrapping. When no methylation sites passed the significance threshold, the 10% most significant sites determined using the Wilcoxon test were selected. At the second stage, the association between pathway-level features and diagnosis was tested using Wilcoxon tests and corrected for multiple hypothesis testing according to the method of Bonferroni. Then only features that showed a positive association with diagnosis were used for building the second-level classifier. This was due to the fact that in random data, the machine learning-based compression of methylation sites into pathway-level features can yield features that are strongly, but negatively associated with diagnoses (Perlich and Swirszcz 2011). The above procedure ensured such features were not used for prediction. Prediction performance for the training data was evaluated using 10-fold cross validation. Second-stage analyses were repeated 20 times and the predictions were averaged, to reduce the effect of variability intrinsic to random forest prediction.

Permutation test procedure

Two permutation strategies were employed to explore the predictive value of the DNA methylation signatures identified by the BioMM procedure:

1.) Permutation of diagnostic label

During this procedure, diagnostic labels were permuted prior to application of the BioMM procedure. It was expected that BioMM will identify a random DNA methylation 'signature' that would lead to an AUC of approximately 0.5 when applied in independent test data. Due to the computational complexity of the procedure, only 30 permutations were performed.

2.) Permutation of pathway level features

This procedure explores the specificity of the DNA methylation signature for the set of pathways selected at the second stage of the BioMM procedure. For this, the same number (here 57) of pathways were selected randomly and used for building the second stage classifier. It was expected that a high biological specificity for the actual pathways would lead to a substantial drop in AUC values when applying the classifier that is based on randomly selected pathways to independent data. 500 permutations were performed. An empirical P-value was determined as (Good 2013):

$$p = \frac{\left|\left\{\mathrm{D}' \in \check{\mathrm{D}}: \mathrm{e}\left(\mathrm{f}, \mathrm{D}'\right)\right\} \leq e(f, D)\right| + 1}{\mathrm{k} + 1}$$

Where Ď is a set of k permuted versions D' of the original data D sampled from a given null distribution.

Analysis of schizophrenia specificity

To test the specificity of the identified PMS for schizophrenia, predictions were analyzed in other related psychiatric conditions including bipolar disorder, major depressive disorder and autism. We applied the same BioMM framework on genome-wide DNA methylation data from these conditions.

Machine learning for case-control prediction based on GWAS data

We applied the same BioMM framework on $GWAS_{MGS}$ but with 50-fold bootstrapping to reduce computational running time. Feature selection and random forest models with 1000 trees were carried out at both stages. At the first stage, SNPs with a case-control difference of $P < 0.05$ using the Wilcoxon test were picked for each pathway and if no SNPs passed this significance threshold, the 10% most significant SNPs determined using the Wilcoxon test were selected. At the second stage, the association between pathway-level features and diagnosis was tested using Wilcoxon tests and corrected for multiple hypothesis testing according to the method of Bonferroni. Only features that showed a positive association with diagnosis were used for building the second-level classifier. The final prediction performance for the training data was evaluated using 10-fold cross validation.

Working memory and faces matching task - imaging data acquisition and preprocessing

Whole-brain fMRI was performed on two 3T MR systems (Siemens Trio, Erlangen, Germany) in Mannheim, Germany. In the $validation_{Meth/MRI}$ sample, we studied brain function during working memory using a well-established n-back fMRI paradigm as previously described (Rasetti, Sambataro et al. 2011, Schneider, Walter et al. 2017). In addition, we tested the specificity of the working memory imaging phenotype by using the 'faces matching' task(Cao, Bertolino et al. 2016). Briefly, for the n-back working memory task, subjects were instructed to press the button when a stimulus (number 1-4) was presented. For the sensorimotor control condition (0-back), subjects pressed the button that was linked to the current number while during the working memory condition (2-back), subjects were asked to press the button corresponding to the number presented two stimuli before. For the faces matching task, subjects were instructed to respond to the fearful or angry face expressions during emotional conditions and simple geometric shapes during the control condition. Both the n-back and faces matching tasks were block design and each of the block was 30 seconds. Here, all functional data were acquired using echo planar imaging (EPI) sequence with the following specifications: 28 axial slices, 4 mm slice thickness, 1 mm gap, TR/TE = 2000/30 ms, 80° flip angle, 192 mm × 192 mm field of view (FOV), 64 × 64 matrix. In the $validation_{MRI}$ sample, we studied brain function during working memory using a well-established Sternberg paradigm as previously described(Geiger, Moessnang et al. 2018). Briefly, for the Sternberg task, subjects were trained to memorize five letters (e.g. FGMPT) and were asked to press the button if the upcoming presented letter belonged to the five trained letters during training condition. For novel conditions, five novel letters were presented (e.g. DCKWX) first and subjects were asked to respond to the upcoming presented letter if it was shown before. The Sternberg paradigm comprised a sensorimotor control condition (where five 'A' were presented) and a resting baseline condition (where subjects were instructed to only look at the screen). Here, functional images were acquired with an EPI sequence with the following specifications: TR = 1790 ms, TE = 28 ms, flip angle = 76°, 34 axial slices, 3 mm slice thickness, matrix size: 64 × 64, FOV: 192 × 192 mm. Images were preprocessed using standard processing routines in SPM12 (http://www.fil.ion.ucl.ac.uk/spm/). Briefly, preprocessing procedures in included realignment to the first image of the time series, then registered to the mean of the images, slice time correction. Functional images were then co-registered to 3D T1 weighted anatomical images. The 3D T1-weighted anatomical images were segmented into grey matter, white matter and cerebrospinal fluid and other non-brain tissues. The computed warps from the

segmentation steps then were applied to the functional images for nonlinear normalization to the template in Montreal Neurological Institute (MNI) space, resampling to 3 mm isotropic voxels, and smoothed with an 8 mm full-width at half-maximum (FWHM) Gaussian Kernel.

First-level DLPFC-HC functional connectivity

DLPFC ROI definition and calculation of DLPFC-HC functional connectivity were performed consistent with our previous study (Schneider, Walter et al. 2017). In short, for each individual, we extracted the first eigenvariate of the seed time series from a 6 mm sphere centered on the individual activation maximum in the "2-back > 0-back" (n-back task, $validation_{Meth/MRI}$) or "novel > practice" (Sternberg task, $validation_{MRI}$) contrasts in the right DLPFC (defined by anatomical masks covering Brodmann area (BA) 46 and BA 9). Then, individual first-level models were defined with the subject-specific DLPFC time series as covariate of interest, and the following covariates of non-interest: (1) movement parameters from the realignment step, (2) first eigenvariates derived from cerebral spinal fluid and white matter masks, and regressors encoding for the experimental conditions. The model estimation step included a high pass filter of 128 seconds and an adjustment for the global brain signal. For an illustration of the DLPFC ROI and hippocampus mask used in $validation_{Meth/MRI}$ and $validation_{MRI}$, please see **Supplementary Figures 4** and **5**, respectively.
Functional connectivity for the 'faces matching' task were computed following the same procedures and parameters as the working memory fMRI data. We used individual DLPFC ROIs that were defined for the working memory tasks.

Structural MRI data acquisition, preprocessing and influence.
The high resolution structural MRI data were acquired using the magnetization-prepared rapid gradient echo sequence (3D-MPRAGE) and the following parameters: for $validation_{Meth/MRI}$ sample, TR 1570 ms, TE 2.75 ms, TI 800 ms, 176 slices, 256 mm FOV, and 15° flip angle and 1 mm^3 spatial resolution (Cao, Bertolino et al. 2016). For $validation_{MRI}$ sample, TR = 2530 ms, TE = 3.8 ms, TI = 1100 ms, 176 slices, 256 x 256 mm field of view, 7° flip angle, and 1 mm^3 spatial resolution (Zang, Geiger et al. 2018). Briefly, the grey matter (GM), white mater (Malki, Tosto et al.) and cerebral spinal fluid (CSF) were segmented and spatially normalized linearly and nonlinearly to the standard MNI template using the Diffeomorphic Anatomical Registration Through Exponential Lie algebra (DARTEL) template and approach. Then the images were modulated with the Jacobian determinants to correct for differences in head size. Additionally, gray matter maps were corrected for bias-field inhomogeneities and were cleaned up for gray matter partitions. Next, we applied a classical Markov random field model and spatial adaptive nonlocal means denoising. We then smoothed the preprocessed structural data using a 8mm FWHM Gaussian kernel.

Task specificity and structural influences

To investigate whether the identified association between PMS and DLPFC-HC connectivity, we tested the association between PMS and DLPFC-HC connectivity using emotion processing 'faces matching' task in the sample $validation_{Meth}/_{MRI}$ (Cao, Bertolino et al. 2016). In addition to evaluate the structural influences of DLPFC and hippocampus grey matter volume, we extracted mean grey matter volume from the individual DLPFC ROIs and the right hippocampus (6 mm sphere ROI MNI [33 -22 -12] for $validation_{Meth/MRI}$ and MNI [30 -37 -7] for $validation_{MRI}$). We assessed the task specificity of the association between PMS and DLPFC-HC connectivity using post-hoc analyses, treating age and sex as

covariates of non-interests. For analyses of evaluating the structural influences on PMS and DLPFC-HC connectivity association, we further treated grey matter volume of the individual DLPFC ROIs and the hippocampus as covariates of non-interests.

Previous publications have indicated that during working memory processing, the connectivity of DLPFC to parietal regions were significantly increased in a population with high genetic risk for schizophrenia(Whalley, Simonotto et al. 2005) and interhemispheric DLPFC connectivity was reduced in risk gene carriers (ZNF804a rs1344706)(Rasetti, Sambataro et al. 2011). We in addition tested the association between PMS and DLPFC-parietal regions (MNI [±42 -48 48], 6mm sphere ROI) and interhemispheric DLPFC (left DLPFR ROI at MNI [-48 33 30]) connectivity.

Medication influences on PMS and DLPFC-HC connectivity

To clarify the potential influence of medication on the PMS as well as DLPFC-HC connectivity, we acquired and calculated chlorpromazine equivalents (for details please refer to our previous study(Schwarz, Moessnang et al. 2019)) from 33 out of 36 patients in our validation$_{Meth}$/$_{MRI}$. We used linear regression model to test the association between chlorpromazine equivalents and PMS as well as DLPFC-HC connectivity.

Statistical analysis

Association between PMS and DLPFC-HC functional connectivity were assessed using linear regression in SPM12 (https://www.fil.ion.ucl.ac.uk/spm/software/spm12/), using PMS as covariate of interest and age and sex as covariates of non-interest. An F-test was applied for the discovery sample (validation$_{Meth}$/$_{MRI}$) and a one-tailed T-test for the replication sample (validation$_{MRI}$). PMS and DLPFC-HC connectivity associations in the imaging space were corrected using family-wise small-volume correction in the hippocampus based on the Automated Anatomical Labeling (AAL) template. Using small-volume correction approach based on the hippocampus mask could allow us to detect location within the hippocampus that shows strongest association between DLPFC-HC connectivity and PMS. For validation$_{Meth}$/$_{MRI}$, we used bilateral hippocampus for small-volume correction. For validation$_{MRI}$, we focused on the anterior or posterior section of the hippocampus (see(Erickson, Voss et al. 2011), determined by y coordinate in MNI space of the hippocampus center of gravity) that showed the highest correlation between PMS and DLPFC-HC connectivity in validation$_{Meth}$/$_{MRI}$. Other post-hoc statistics were performed in R using mean DLPFC-HC connectivity extracted from a 6mm sphere ROI located at the peak voxel identified by SPM group statistics outlined above. Associations between PMS and the schizophrenia PRS, as well as the PMS by PRS interaction on DLPFC-HC connectivity were tested using multiple linear regression, accounting for the effects of sex and age. All analyses involving PRS additionally incorporated 10 genetic principal components as covariates to account for the potential confounding effect of genetic population structure.

2.1.9 Supplementary Results

Task specificity and structural confounding influences on the identified associations between PMS and DLPFC-HC coupling

In the validation$_{Meth/MRI}$ sample, we found no significant association between the PMS and DLPFC-HC connectivity during emotion processing ‘faces matching’ task (P=0.60). In the validation$_{Meth/MRI}$ sample, we did not find significant association between PMS and DLPFC-parietal region and interhemispheric DLPFC connectivity (P_{FDR} values > 0.14, Hochberg correction). In the validation$_{MRI}$ sample, the association between PMS and DLPFC-parietal region and interhemispheric DLPFC

connectivity were not significant (P_{FDR} values > 0.42, Hochberg correction). In addition, we found the association between PMS and DLPFC-HC connectivity were still significant after controlling for grey matter volume of DLPFC and the hippocampus (t = -3.70, P<0.001 in validation$_{Meth/MRI}$ sample; t = -3.48, P=0.001 in validation$_{MRI}$ sample).

Analysis of residual confounding effects on the PMS and DLPFC-HC connectivity

Linear regression was used to explore the association between the PMS and potential confounding variables in HC across studies (see **Supplementary Table** 6). We found that despite the employed correction procedure for confounding, the predicted PMS were significantly associated with some of all available potential confounders. Since this may have downstream impact on classification accuracy, we additionally determined AUC-values for case-control differentiation using PMS residualized for the effects of all potential variables (see **Supplementary Table 7**). The residualized predictions showed AUC values of approximately 0.60 across cohorts, which supports the reproducibility of the PMS-effects despite potentially present residual confounding.

We additionally performed linear regression analyses on the association of DLPFC-HC connectivity and covariates of non-interests including age, sex, cigarette smoking score, 10 principal components determined via principle component analyses of the methylation data and seven cell types: CD8.naive, CD8pCD28nCD45RAn, PlasmaBlast, CD4T, NK, Mono, Gran. None of the 20 covariates were associated with the DLPFC-HC connectivity (P_{FDR} values > 0.32 for validation$_{Meth/MRI}$ sample; P_{FDR} values > 0.30 for validation$_{MRI}$ sample).

Associations between PMS, DLPFC-HC connectivity and medication in schizophrenia patients

We acquired and calculated chlorpromazine equivalents from 33 out of 36 patients in our validation$_{Meth/MRI}$ sample and found no correlation between the predicted PMS and chlorpromazine equivalents (p = 0.30). Neither did we observed significant correlation between DLPFC-HC connectivity and chlorpromazine equivalents (p = 0.28).

Association between BioMM-derived polygenic risk signature, PMS and DLPFC-HC connectivity

We used the BioMM procedure to identify a polygenic risk score from GWAS data using the same pathways assignment as used for DNA methylation data (PRS$_{BioMM}$). PRS$_{BioMM}$ was significantly predictive of schizophrenia in GWAS$_{MGS}$ cohort with a p value of 2.88×10^{-15} accounting for sex, age and 10 PCs (AUC=0.58, R^2=0.015). However, DLPFC-HC connectivity was not associated with the schizophrenia PRS$_{BioMM}$ in validation$_{Meth/MRI}$ (P=0.299). It was significantly associated in validation$_{MRI}$ (P=0.032) but in a wrong direction (t=2.168). Similarly, no significant interactions between PMS and PRS$_{BioMM}$ on DLPFC-HC connectivity were found in validation$_{Meth/MRI}$ (P=0.104) or validation$_{MRI}$ (P=0.443). Detailed comparative information is shown in **Supplementary Table 3**.

Permutation test analysis

Two permutation strategies were employed to characterize the predictive value of the DNA methylation signatures identified by the BioMM procedure: I) Permutation of the diagnostic label where diagnostic labels were permuted prior to application of the BioMM procedure; II) Permutation of pathway level features, where a random set of 57 pathways was selected from stage 2 data to build a stage-2 classifier. The results of the permutation tests are shown in **Supplementary Tables 4 and 5**. Permutation of the diagnostic label yielded AUC values close to 0.5 for cross-validation in discovery$_{Meth}$ as well as independent prediction in the remaining cohorts. These AUC values were significantly lower (empirical P < 0.032, based on 30 permutations due to the high computational

demand of the procedure) than those observed in the non-permuted data, with the exception of validation$_{postmortem}$ (P=0.23). The subsequent imaging associations were also significantly weaker compared to non-permuted data ($P < 0.032$).
For the permutation of pathway level features, the predictions of all datasets but validation$_{postmortem}$ showed an empirical P-value < 0.002 based on 500 permutations, with the exception of validation$_{postmortem}$ (P=0.10), supporting a degree of specificity of the predictive signal for the 57 originally identified pathways.

2.1.10 Supplementary Tables

Supplementary Table 1. Differences of subject demographics between patients and controls (P values are based on the Wilcoxon signed-rank test (continuous variables) or logistic regression (categorical variables), respectively)

	discovery$_{Meth}$*	**validation$_{Meth}$#**	**validation$_{Meth/MRI}$**		**validation$_{postmortem}$**
Covariates		whole blood		**Covariates**	post-mortem
Sex	$7.44x10^{-14}$	0.088	$3.27x10^{-3}$	Sex	0.067
Age	1.50×10^{-9}	0.199	1.59×10^{-4}	Age	7.63×10^{-3}
PC1	2.82×10^{-5}	8.09×10^{-13}	0.713	negControl_PC1	0.347
PC2	9.21×10^{-12}	1.85×10^{-15}	0.868	negControl_PC2	0.107
PC3	5.71×10^{-2}	0.181	0.613	negControl_PC3	0.969
PC4	0.357	8.58×10^{-4}	0.623	negControl_PC4	0.496
PC5	2.18×10^{-2}	0.615	0.470	race	0.102
PC6	1.30×10^{-2}	2.20×10^{-7}	0.031		
PC7	0.701	0.589	0.331		
PC8	6.42×10^{-3}	4.29×10^{-2}	0.903		
PC9	3.45×10^{-2}	1.24×10^{-3}	0.574		
PC10	4.65×10^{-2}	5.45×10^{-2}	0.760		
smokeScore	3.31×10^{-42}	1.32×10^{-26}	9.750×10^{-6}		
Cell types					
CD8.naive	2.86×10^{-2}	0.221	0.034	ES	0.286
CD8pCD28nCD45RAn	0.963	3.80×10^{-2}	0.401	NPC	0.412
PlasmaBlast	5.05×10^{-3}	9.32×10^{-7}	0.964	DA_NEURON	0.359
CD4T	4.68×10^{-3}	3.47×10^{-8}	0.410	NeuN_pos	0.190
NK	2.56×10^{-8}	1.57×10^{-8}	0.818	NeuN_neg	0.265
Mono	0.698	2.50×10^{-2}	0.956		
Gran	2.34×10^{-10}	6.60×10^{-20}	0.531		

*age information was missing for 37 subjects and estimated using the Epigenetic Clock tool as described in the methods.

age information was missing for 182 subjects and estimated using the Epigenetic Clock tool as described in the methods.

Supplementary Table 2. Overall prediction performance of BioMM on different cohorts and specificity analysis. * low statistical power likely due to small sample size in these cohorts. *P*-value is adjusted for all potential confounding variables.

Cohorts	Status	Control/Case	PMS prediction		
			AUC	**R2**	**P value**
Discovery$_{Meth}$	SCZ	322/353	0.78	0.21	2.95×10^{-6}
Validation$_{Meth}$	SCZ	433/414	0.69	0.10	1.24×10^{-7}
***Validation$_{Meth/MRI}$**	SCZ	331/36	0.74	0.22	4.95×10^{-2}
	AUT	331/27	0.526	0.006	0.658
***Specificity$_{Meth}$**	BP	331/39	0.578	0.015	0.211
	MDD	331/35	0.509	0.004	0.164
Validation$_{postmortem}$	SCZ	136/108	0.63	0.08	4.20×10^{-4}

Supplementary Table 3. The association between PMS and DLPFC-HC, PMS and PRS, PMS and PRS$_{BioMM}$, as well as the interaction of these two PRSs and PMS on DLPFC-HC in two different cohorts. ** Significant result was observed in the imaging space after family-wise error correction.

Association	**Validation$_{Meth/MRI}$** (N=241)		**Validation$_{MRI}$** (N=152)	
	T value	**P value**	**T value**	**P value**
****PMS vs DLPFC-HC**	-3.81	0.04	-3.54	0.016
PRS$_{BioMM}$ vs DLPFC-HC	1.042	0.299	2.168	0.032
PRS vs DLPFC-HC	0.263	0.793	0.432	0.667
PRS$_{BioMM}$ vs PMS	0.389	0.698	0.031	0.975
PRS vs PMS	0.002	0.999	0.252	0.802
PMSxPRS$_{BioMM}$ vs DLPFC-HC	1.633	0.104	0.769	0.443
PMSxPRS vs DLPFC-HC	-1.312	0.191	-0.835	0.405

Supplementary Table 4. Permutation of diagnostic label for both machine learning prediction and the subsequent testing of imaging associations. Due to computational complexity, 30 permutations were performed.

Cohorts	Prediction/Association	Measure	Mean/SD	P value
discovery$_{Meth}$	Cross validation	AUC	0.532±0.026	< 0.032
validation$_{Meth}$	Independent prediction	AUC	0.517±0.017	< 0.032
validation$_{Meth/MRI}$	Independent prediction	AUC	0.539±0.03	< 0.032
validation$_{postmortem}$	Independent prediction	AUC	0.58±0.059	=0.226
validation$_{Meth/MRI}$	Imaging association	T value	-0.348±0.852	< 0.032
validation$_{MRI}$	Imaging association	T value	-0.243±0.999	< 0.032

Supplementary Table 5. Permutation of pathway level features for both machine learning prediction and subsequent testing of imaging associations. 500 permutations were performed.

Cohorts	Prediction/Association	Measure	Mean/SD	P value
discovery$_{Meth}$	Cross validation	AUC	0.707±0.017	< 0.002
validation$_{Meth}$	Independent prediction	AUC	0.563±0.041	< 0.002
validation$_{Meth/MRI}$	Independent prediction	AUC	0.572±0.058	< 0.002
validation$_{postmortem}$	Independent prediction	AUC	0.57±0.045	= 0.1018
validation$_{Meth/MRI}$	Imaging association	T value	-1.203±0.982	< 0.002
validation$_{MRI}$	Imaging association	T value	-0.486±1.024	< 0.002

Supplementary Table 6. The association between predicted PMS and the confounding variables in controls (*P* values are based on generalized linear regression)

	discovery$_{Meth}$*	**validation$_{Meth}$#**	**validation$_{Meth/MRI}$**		**validation$_{postmortem}$**
Covariates		whole blood		**Covariates**	post-mortem
Sex	0.928	0.679	2.81×10^{-7}	Sex	0.460
Age	0.038	0.506	1.11×10^{-4}	Age	0.179
PC1	2.08×10^{-13}	0.011	8.26×10^{-5}	negControl_PC1	0.052
PC2	0.134	0.487	3.99×10^{-3}	negControl_PC2	0.224
PC3	0.160	0.340	0.040	negControl_PC3	0.234
PC4	0.282	0.121	1.99×10^{-5}	negControl_PC4	0.006
PC5	0.068	0.244	1.27×10^{-8}	race	0.376
PC6	0.041	0.243	8.13×10^{-4}		
PC7	0.091	0.304	1.83×10^{-7}		
PC8	2.81×10^{-5}	0.610	0.012		
PC9	0.119	0.981	0.164		
PC10	0.084	0.791	0.691		
smokeScore	4.78×10^{-16}	4.89×10^{-10}	2.64×10^{-7}		
Cell types					
CD8.naive	0.339	0.446	1.98×10^{-4}	ES	0.316
CD8pCD28nCD45RAn	0.855	0.649	1.53×10^{-6}	NPC	0.089
PlasmaBlast	0.804	6.59×10^{-3}	0.022	DA_NEURON	0.088
CD4T	0.361	8.39×10^{-3}	1.64×10^{-3}	NeuN_pos	0.925
NK	0.988	0.027	0.086	NeuN_neg	0.964
Mono	2.54×10^{-4}	0.029	0.427		
Gran	0.072	0.014	0.062		

Supplementary Table 7. Prediction performance of BioMM on different cohorts and specificity analysis following residualization of predicted PMS scores against all potential covariates using linear regression. * low statistical power likely due to small sample size in these cohorts. *P*-value is adjusted for all potential confounding variables.

Cohorts	Status	Control/Case	PMS prediction	
			AUC	**P value**
$\text{Discovery}_{\text{Meth}}$	SCZ	322/353	0.60	$2.95\ \text{x}10^{-6}$
$\text{Validation}_{\text{Meth}}$	SCZ	433/414	0.59	$1.24\ \text{x}10^{-7}$
$^*\text{Validation}_{\text{Meth/MRI}}$	SCZ	331/36	0.62	$4.95\text{x}10^{-2}$
	AUT	331/27	0.524	0.658
$^*\text{Specificity}_{\text{Meth}}$	BP	331/39	0.578	0.211
	MDD	331/35	0.565	0.164
$\text{Validation}_{\text{postmortem}}$	SCZ	136/108	0.62	$4.20\text{x}10^{-4}$

Supplementary Table 8. Top 10 schizophrenia-associated pathways in the discovery sample ($\text{discovery}_{\text{Meth}}$). (*P* values are based on the Wilcoxon signed-rank test)

ID	Description	Z score	Size	P value
GO:0032088	negative regulation of NF-kappaB transcription factor activity	9.188	1623	$7.72\text{x}10^{-23}$
GO:0010977	negative regulation of neuron projection development	8.905	1929	$2.41\text{x}10^{-21}$
GO:0030097	hemopoiesis	8.667	1423	$9.01\text{x}10^{-20}$
GO:0000083	regulation of transcription involved in G1/S transition of mitotic cell cycle	8.557	765	$1.01\text{x}10^{-19}$
GO:2000311	regulation of AMPA receptor activity	7.037	955	$1.25\ \text{x}10^{-13}$
GO:0098962	regulation of postsynaptic neurotransmitter receptor activity	6.948	353	$3.13\text{x}10^{-13}$
GO:0043154	negative regulation of cysteine-type endopeptidase activity involved in apoptotic process	7.081	1554	$3.33\text{x}10^{-13}$
GO:0031648	protein destabilization	6.747	986	$2.08\text{x}10^{-12}$
GO:1900273	positive regulation of long-term synaptic potentiation	6.637	675	$3.94\text{x}10^{-12}$
GO:0019395	fatty acid oxidation	6.753	372	$4.96\text{x}10^{-12}$

Supplementary Table 9. 30 most significant CpGs derived from top 10 pathways. (*P* values are based on the Wilcoxon signed-rank test)

ID	CHR	pos	RefGene	island	P value	Rho
cg14385231	13	76112274	COMMD6	S_Shore	$4.70x10^{-11}$	-0,023
cg20227766	1	27998703	IFI6	S_Shelf	$3.52x10^{-10}$	-0,002
cg22062597	12	1905547	CACNA2D4	Island	$4.15x10^{-10}$	-0,03
cg21012647	2	201981814	CFLAR	N_Shore	$3.87x10^{-9}$	0,008
cg18316974	1	92947035	GFI1	Island	$1.03x10^{-8}$	0,079
cg18316974	1	92947035	GFI1	Island	$1.03x10^{-8}$	0,079
cg18316974	1	92947035	GFI1	Island	$1.03x10^{-8}$	0,079
cg18316974	1	92947035	GFI1	Island	$1.03x10^{-8}$	0,079
cg12017057	20	377006	TRIB3	Island	$1.27x10^{-8}$	-0,003
cg18146737	1	92946700	GFI1	Island	$3.84x10^{-7}$	0,077
cg18146737	1	92946700	GFI1	Island	$3.84x10^{-7}$	0,077
cg18146737	1	92946700	GFI1	Island	$3.84x10^{-7}$	0,077
cg18146737	1	92946700	GFI1	Island	$3.84x10^{-7}$	0,077
cg25286967	7	75623934	TMEM120A	Island	$1.60x10^{-6}$	-0,034
cg25286967	7	75623934	TMEM120A	Island	$1.60x10^{-6}$	-0,034
cg01415275	7	75624096	TMEM120A	Island	$4.28x10^{-5}$	-0,03
cg01415275	7	75624096	TMEM120A	Island	$4.28x10^{-5}$	-0,03
cg09846458	19	55972646	ISOC2	N_Shore	$1.31 x10^{-4}$	0,02
cg09846458	19	55972646	ISOC2	N_Shore	$1.31 x10^{-4}$	0,02
cg09846458	19	55972646	ISOC2	N_Shore	$1.31 x10^{-4}$	0,02
cg09846458	19	55972646	ISOC2	N_Shore	$1.31 x10^{-4}$	0,02
cg12195369	19	55972957	ISOC2	Island	$1.33 x10^{-4}$	0,018
cg12195369	19	55972957	ISOC2	Island	$1.33 x10^{-4}$	0,018
cg12195369	19	55972957	ISOC2	Island	$1.33 x10^{-4}$	0,018
cg12195369	19	55972957	ISOC2	Island	$1.33 x10^{-4}$	0,018
cg06338710	1	92946187	GFI1	Island	$1.69 x10^{-4}$	0,047
cg06338710	1	92946187	GFI1	Island	$1.69 x10^{-4}$	0,047
cg06338710	1	92946187	GFI1	Island	$1.69 x10^{-4}$	0,047
cg06338710	1	92946187	GFI1	Island	$1.69 x10^{-4}$	0,047
cg02105261	17	1945138	DPH1	Island	$1.74 x10^{-4}$	-0,108

Supplementary Table 10. The existing evidence for the top genes harboring top CpGs from Supplementary Table 9.

Gene names	Definition	Reference
COMMD6	COMMD6 belongs to a family of NF-kappa-B	(Roussos, Katsel et al. 2013)
CACNA2D4	Calcium channel, voltage-dependent, alpha 2/delta subunit 4	(Cross-Disorder Group of the Psychiatric Genomics 2013, Purcell, Moran et al. 2014)
GFI1	Growth Factor Independent Protein 1	(Hannon, Dempster et al. 2016)
TRIB3	Tribbles Pseudokinase 3	(Duan, Sanders et al. 2015)
TMEM120A	Transmembrane protein 120A	(Glatt 2009)

Supplementary Table 11. Prediction performance of BioMM on different cohorts using $\text{validation}_{\text{postmortem}}$ as the discovery set.

Data sets	AUC	R2
$\text{validation}_{\text{postmortem}}$	0.738	0.17
$\text{discovery}_{\text{Meth}}$	0.572	0.006
$\text{validation}_{\text{Meth}}$	0.566	0
$\text{validation}_{\text{Meth/MRI}}$	0.52	0.001

2.1.11 Supplementary Figures

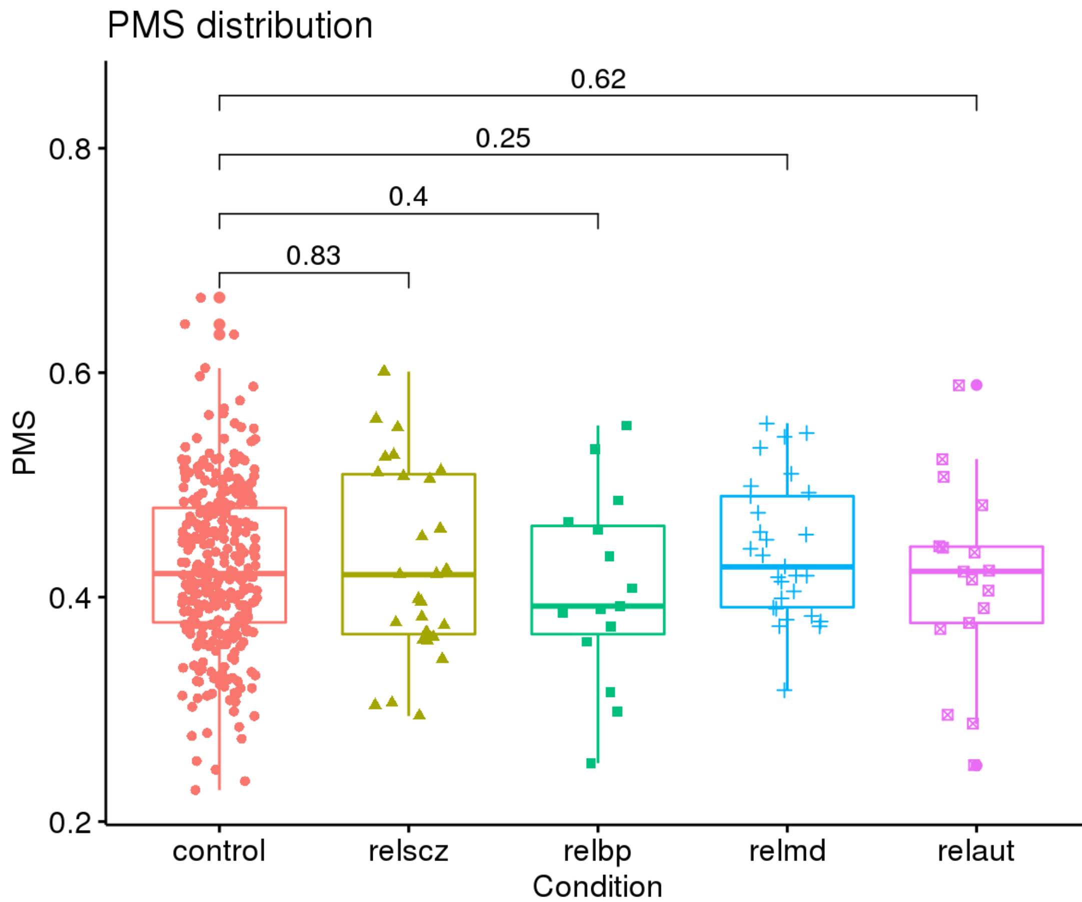


Figure S1. PMS comparision in validation$_{meth/MRI}$ and relatives$_{Meth}$. The control group is compared with the relscz: first-degree relatives of schizophrenia patients; relbp: first-degree relatives of patients with bipolar disorder; relmdd: first-degree relatives of patients with major depressive disorder; relaut: first-degree relatives of patients with autism. *P*-values are denoted between comparison groups based on the t-test.

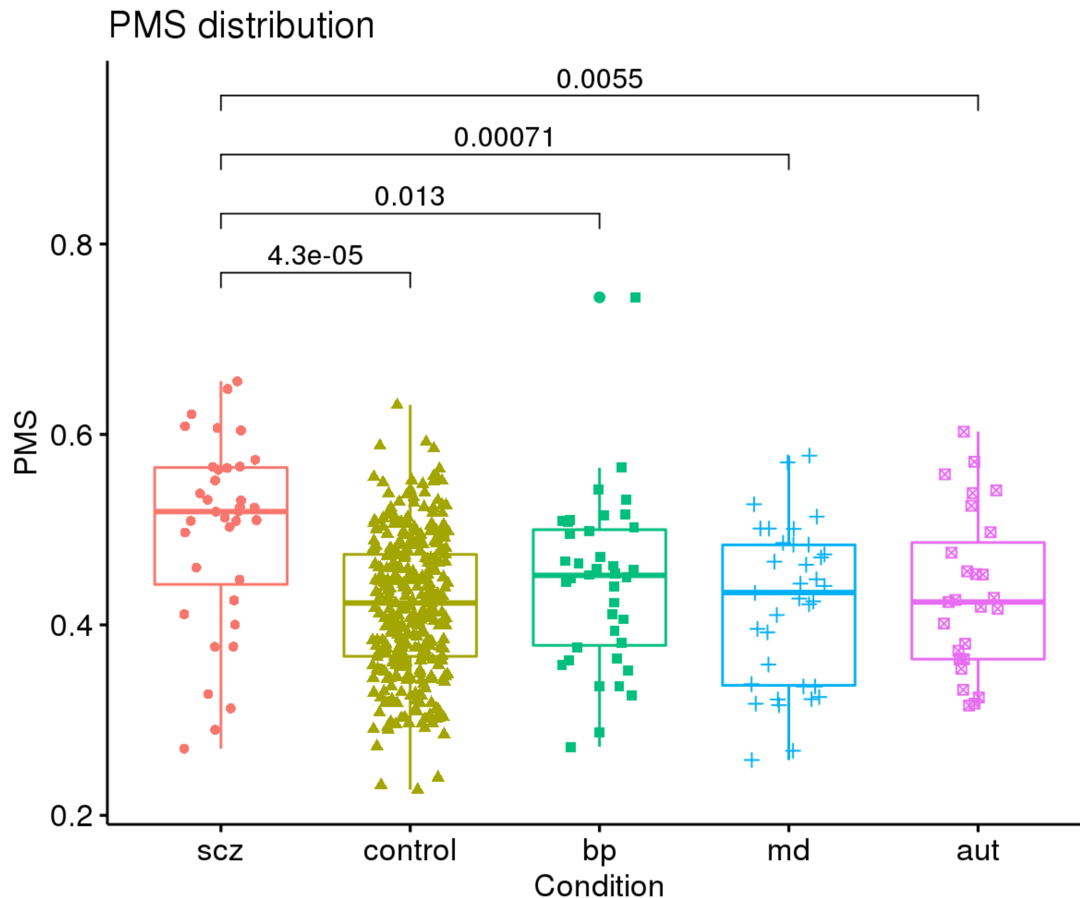


Figure S2. PMS comparision in validation$_{Meth/MRI}$ and specificity$_{Meth}$. The schizophrenia group is compared with the healthy control group; bp: bipolar disorder; md: major depressive disorder; aut: autism. *P*-values are denoted between comparison groups based on the t-test.

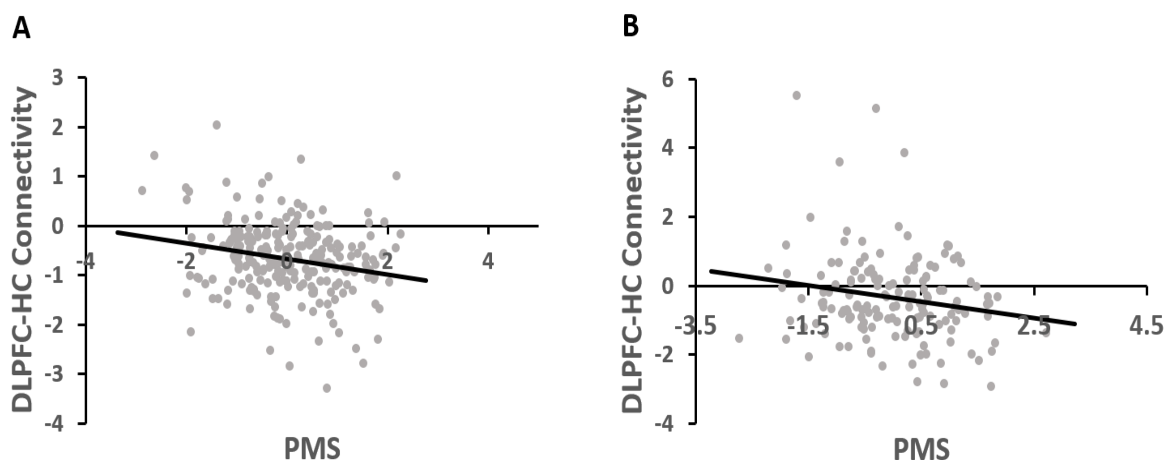


Figure S3. Post-hoc partial regression plots of the association between PMS and DLPFC-HC connectivity in the n-back (validation$_{Meth/MRI}$, **panel A**) and Sternberg (validation$_{MRI}$, **panel B**) working memory fMRI data. The post-hoc partial regression plot from the identified 6mm sphere ROI centered at the peak voxel (**Panel A**: T = 3.81, p_{FWE} = 0.040, MNI [33 -22 -13], bilateral hippocampus corrected; **Panel B**: T = 3.54, p_{FWE} = 0.016, MNI [33 -37 -7], right posterior hippocampus corrected). The partial regression plots were adjusted for age and sex and are only shown for illustration purpose. No statistical interference was further made based on these plots.

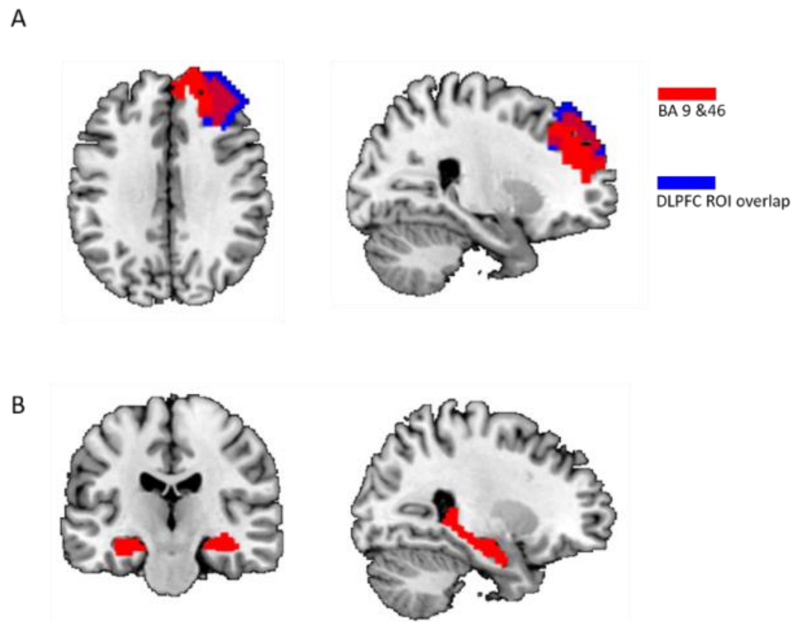


Figure S4 DLPFC ROI and hippocampus mask in the validation$_{Meth/MRI}$ sample. Panel A shows the right BA 9 and 46 area and an overlap of all subjects' 6mm DLPFC sphere ROIs. Panel B shows the bilateral hippocampus masks from AAL.

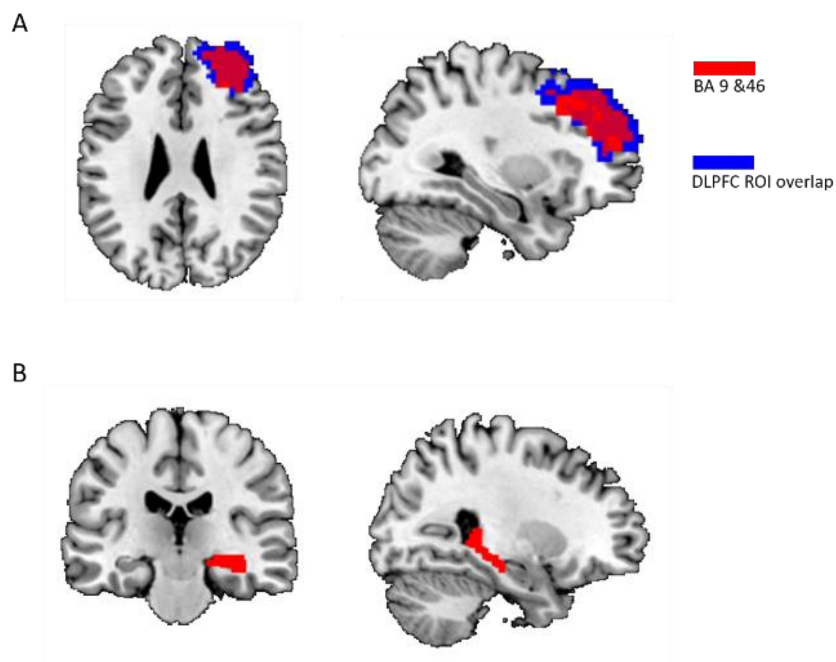


Figure S5 DLPFC ROI and hippocampus masks in validation$_{MRI}$ sample. Panel A shows the right BA 9 and 46 area and an overlap of all subjects' 6mm DLPFC sphere ROIs. Panel B shows the right posterior hippocampus masks (determined by the y MNI coordinates of COG and we choose the posterior overlap mask from the AAL template) that we have applied for small volume correction analysis.

2.2 Study 2 - Male increase in brain gene expression variability is linked to genetic risk for schizophrenia[2]

2.2.1 Abstract

Schizophrenia shows substantial sex differences in age of onset, course and treatment response but the biological basis of these effects is incompletely understood. Here we show that during human development, males show a regionally specific decrease in brain expression similarity compared to females. The genes modulating this effect were significantly co-expressed with schizophrenia risk genes during prefrontal cortex brain development in the fetal period as well as during early adolescence. This suggests a genetic contribution to a mechanism through which developmental abnormalities manifest with psychosis during adolescence. It further supports sex differences in brain expression variability as a factor underlying the well-established sex differences in schizophrenia.

2.2.2 Introduction

Schizophrenia is a severe developmental mental illness with an incidence approximately 1.4 times higher in men compared to women(Aleman, Kahn et al. 2003). The disorder is substantially heritable and a large number of common and rare variants have been associated with illness risk(International Schizophrenia, Purcell et al. 2009, Sullivan, Daly et al. 2012, Ripke, O'Dushlaine et al. 2013, Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014). A widely accepted neurodevelopmental hypothesis posits that genetically determined alterations in early brain development interact with developmental changes during adolescence in the prefrontal cortex to lead to the manifestation of psychosis(Weinberger 1987, Birnbaum R 2017). Consistent with this, developmentally changing prefrontal cortex expression has been found to be linked to neuronal differentiation and maturation, as well as genetic schizophrenia risk(Jaffe, Shin et al. 2015).
In men, the illness has a more severe course characterized by more pronounced negative symptoms as well as cognitive impairment(Leung and Chue 2000, Maric, Krabbendam et al. 2003), although evidence has been reported that substance abuse in men may confound such clinical differences(Abel, Drake et al. 2010). Males with schizophrenia have also, albeit inconsistently, been reported to have a lower age of onset, show more pronounced alterations of brain morphology and poorer response to antipsychotic medication(Pinals, Malhotra et al. 1996, Leung and Chue 2000, Morgan VA 2008, Abel, Drake et al. 2010). Genetic risk associations, as well as molecular profiles, contain sex-dependent factors(Goldstein, Cherkerzian et al. 2013, Ramsey, Schwarz et al. 2013) and sex hormones are thought to play an important role for illness course(Leung and Chue 2000, Markham 2012), but again little is known about the underlying neurobiological mechanisms.
We pursued a novel strategy to explore how biological sex differences may impact on the manifestation of genetic risk and the clinical sex differences of schizophrenia. Inspired by a recent study on the human brain connectome(Kaufmann, Alnaes et al. 2017), we tested whether during development human brain gene expression is more variable in males than females. We hypothesized that such increased expression variability might contribute to a predisposition of males for heritable neurodevelopmental disorders. A similar hypothesis has previously been explored for HIV, where gene expression variability has been suggested as a modulator for susceptibility to infection(Li, Liu et

[2] **Published as**: Chen, J., Cao, H., Meyer-Lindenberg, A. and Schwarz, E., 2018. Male increase in brain gene expression variability is linked to genetic risk for schizophrenia. Translational psychiatry, 8(1), p.140.

al. 2010). Our study is further motivated by previous identification of sexual dimorphisms of brain expression(Kang, Kawasawa et al. 2011, Trabzuni, Ramasamy et al. 2013, Werling, Parikshak et al. 2016), protein abundance(Raser and O'Shea 2005), as well as genetic and epigenetic factors modulating gene expression noise(Raser and O'Shea 2004, Alemu, Carl et al. 2014), supporting the possibility of links between polygenic risk and expression variance. The longitudinal exploration of variability differences is further motivated by previous identification of differential variance of transcriptional regulators during human embryonic development(Hasegawa, Taylor et al. 2015). Analysis of gene expression variability has also been successfully applied to identify genes and pathways implicated in several illnesses and highlighted such variability as an informative biological signal(Ho, Stefani et al. 2008, Ran and Daye 2017).

Expression variability as genetic risk mediator can capture polygenic effects beyond sex differences of expression. To investigate this, we identified genes driving brain-region and age specific variability differences between sexes and tested whether these were associated with expression of schizophrenia risk genes.

2.2.3 Materials and Methods

Data preprocessing. To characterize brain expression throughout the human lifespan, we used data from the BrainSpan: Atlas of the Developing Human Brain (funded by ARRA Awards 1RC2MH089921-01, 1RC2MH090047-01, and 1RC2MH089929-01 and available from: http://developinghumanbrain.org), as well as Braincloud microarray data (GSE30272(Colantuoni, Lipska et al. 2011), available from the GEO database(Edgar, Domrachev et al. 2002)).

The primary analysis was performed on BrainSpan exon microarray data (GSE25219, preprocessed as described in(Goyal, Hawrylycz et al. 2014)) due to availability of a larger sample number. BrainSpan RNA sequencing (RNAseq) data was used for replication and Braincloud data for validation of findings. BrainSpan data comprised transcriptome-wide expression information on subjects between the 6th post-conceptional week and 40 years of age (**Table 1**, **supplementary Tables 2, 8 and 9**). We did not consider older subjects, as sex effects on risk are not likely to manifest beyond the typical age of onset that ranges between late adolescence and early adulthood. As performed by Willsey et al.(Willsey, Sanders et al. 2013), subjects were grouped in age-bins by a windowing approach that joins three consecutive age periods into a single group.

Preprocessing of all datasets followed a similar sequence of steps (**Supplementary Figure 1**). Procedures performed on all datasets comprised: RNA Integrity Number (RIN) filtering (for BrainSpan exon microarray data, all donors were removed that had more than 25% of microarray samples with RIN < 7.5, as in (Goyal, Hawrylycz et al. 2014); for BrainSpan RNAseq data and Braicloud data, a more stringent filtering was performed by removing all samples with RIN <= 7.5); removal of subjects >40 years; log2 transformation of data; extraction of autosomal genes (without minimum expression filter); quantile normalization; surrogate variable determination; covariate adjustment and outlier detection. This data contained the respective median values if multiple replicates per subject were present. Following a previously described pipeline(Werling, Parikshak et al. 2016), processing of RNAseq data included two additional steps: gene-level reads per kilobase million mapped reads (RPKM) were normalized for GC content using conditional quantile normalization based on the R library *cqn*(Hansen, Irizarry et al. 2012) and all genes with less than 1 RPKM in more than 50% of male or female samples were removed. Surrogate variable analysis was performed to account for the potential effects of unobserved confounders(Leek and Storey 2007). The number of surrogate

variables were automatically determined using the *num.sv* function of the R package *sva*(Leek and Storey 2007), using the approximation method by Leek(Leek and Storey 2007, Leek 2011). The underlying full model matrix contained gender, whose effects on expression variability should be preserved, as well as age, PMI, RIN and brain pH (as well as an array indicator for Braincloud data). The null-model matrix contained all covariates but gender. Age was used as a covariate, to prevent artifactual correlations between genes due to their joint association with age. This is particularly important for age-bins covering a broader range of ages, where significant correlations between age and expression can be expected. The number of surrogate variables determined for BrainSpan exon microarray was 0, 2 for BrainSpan RNAseq and 0 for the Braincloud data. Covariate adjustment was performed via residualization against all covariates described above (except for gender) using linear models. Missing brain pH values were replaced by the mean of non-missing values.

Outlier detection. After preprocessing, principal component analysis was used to exclude outliers (**Supplementary Figure 2**). For this, we identified separately for males and females observations that deviated more than three standard deviations from the mean of the respective first two principal components. This removed 7 samples in the BrainSpan exon microarray data (6 from male donors), 11 observations in the BrainSpan RNAseq data (6 from male donors), and 1 outlier (from a female donor) in the Braincloud data.

Schizophrenia risk genes. Schizophrenia risk variants, loci, and associated genes were taken from (Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014) (**Supplementary Table 3**). Previous analyses have pursued different approaches to identify genes linked to genetic schizophrenia risk. Among these approaches is the selection of all genes or those within a certain distance from a given locus(Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014), or genes affected by index variant eQTLs(Gamazon, Wheeler et al. 2015). For the present study, we aimed to identify a single gene per locus. This was due to the risk of introducing statistical bias from including multiple genes per locus, caused by (1) the undue influence of loci harboring a larger number of genes and (2) the gene-gene correlation of genes in close chromosomal proximity. Therefore, for loci harboring multiple genes, we here used the gene in closest chromosomal proximity to the genome-wide significant index variant. If a locus contained more than one index variant, we selected the gene in closest chromosomal proximity to the most significant index variant. Chromosomal locations were determined from the R library org.Hs.eg.db., vs. 3.1.2 (genome build hg19, assembly GRCh37). Genes within the MHC region were not considered due to their significant linkage disequilibrium pattern. Two loci mapped to the genes IMMP2L and TCF4, and these were considered only once for subsequent analyses. C10orf32, C12orf79 and VPS14C were not annotated by the library org.Hs.eg.db. and not considered for further analysis. The final set of schizophrenia risk genes contained 100 genes, of which 97 were autosomal. Of these, 87 were part of the BrainSpan dataset (see **Supplementary Table 3**).

Analysis of expression similarity. First, all samples were identified for a given brain-regional-cluster and age-bin. Based on such data subset, we performed a three stage resampling approach separately for males and females. The objective of this resampling was to quantify the expression similarity (and its confidence interval) between subjects while accounting for the non-independence of multiple samples taken from the same donor:

1. First, we randomly selected a single sample per subject to prevent an impact of sample non-independence on results.
2. Second, we took a bootstrap sample of subjects by sampling with replacement and chose the unique set of subjects. This was performed to prevent the perfect correlation between multiply selected samples.

3. Finally, we subsampled the selected subjects, such that the same number of subjects was chosen for males and females. This was aimed at preventing an influence of unequal sample numbers on results.

Then separately for males and females, we determined the pairwise Pearson correlation coefficients between all subject pairs using expression values from all genes. The mean of these estimates was used as an estimate of expression similarity between subjects for a given regional-cluster age-bin combination. Only the upper triangular matrix of a given correlation matrix was used for estimation. This entire resampling was repeated 100 times and the mean value (for confidence intervals the upper and lower 2.5% percentile) of obtained estimates used to quantify expression similarity.

The difference between males and females was then quantified as the mean difference between the point estimates of each regional-cluster age-bin combination. To assess significance, the resampling procedure was repeated 1,000 times. During each repetition, gender information was permuted for a given regional-cluster age-bin combination, such that different samples of the same subject were always assigned the same gender. The frequency of bootstrapping point estimates at least as high as the one obtained from non-permuted data was used as empirical *P*-value and corrected for multiple comparisons according to the method of Bonferroni. To perform two-sided tests, absolute values were used for this calculation.

Identification of genes driving expression similarity differences. We anticipated that genes driving the difference of expression similarity between males and females would likely show strong differences in expression variance between sexes. For each regional-cluster age-bin combination, we therefore performed the same resampling strategy as described above. For a given set of subjects (males and females separately), we then determined the standard deviation of expression for a given gene. These estimates were averaged over 100 resampling repetitions. We then determined the ratio of these averages between males and females and used the 100 genes (arbitrary cut-off) with the highest ratio as 'variability genes'. To test whether these gene sets were also 'variability genes' in replication (BrainSpan RNAseq data) and validation (Braincloud) data, we determined the difference of expression similarity estimates (using the resampling strategy described above) between males and females. An empirical P-value was then determined by comparing this estimate against those derived from random 'variability genes' identified as described below (1000-fold resampling, one-sided test).

Testing associations with schizophrenia risk genes. To explore associations between variability genes and schizophrenia susceptibility genes, the co-expression between the two gene sets was determined for a given regional-cluster age-bin combination, by calculating a matrix of all pairwise Pearson correlation coefficients using expression values from both gene sets. The median value of this correlation matrix was then used as a measure of co-expression. Again, these calculations were determined as part of the resampling procedure described above, with the exception of the third step (undersampling to obtain equal numbers of male and female subjects), since calculations were performed using males only.

Significance was determined using 1,000 fold resampling. During each repetition and for each regional-cluster age-bin combination, the low number of donors prevented meaningful permutation of gender information. Therefore, random 'variability genes' were selected such that for each real variability gene, one gene with a standard deviation of expression within 5% of the original gene was randomly chosen. The resulting co-expression values were then used to form null-distributions. Empirical *P*-values were determined as the frequency of co-expression values at least as high as that observed from real data (one-sided test). Since a total of 22 sets of variability genes were tested, *P*-values were corrected for the Family Wise Error Rate according to the method of Bonferroni.

Analysis of schizophrenia specificity. To test the specificity of the co-expression between variability genes and schizophrenia susceptibility genes, five additional analyses were performed, using different selections of “susceptibility genes”: (I) random selection of schizophrenia susceptibility genes for a given locus (instead of based on physical proximity to the index SNP). (II) Random selection of genes from loci with comparable DNA sequence variability compared to the schizophrenia loci. For this analysis, the number of common (MAF >=1%) variants recorded in dbSNP (GRCh37, available from https://genome.ucsc.edu/) was used as a proxy for DNA sequence variability. For each schizophrenia locus, a locus of the same size was selected from the same chromosome and retained if the DNA sequence variability was within 10% of the original locus. A random gene was then selected from the locus, extended by 20kbp, using the R library *biomaRt*(Durinck, Spellman et al. 2009). (III) Random selection of genes from the same chromosome as a given schizophrenia gene, irrespective of DNA sequence variability. (IV) Selection of genes in proximity to SNPs associated with major depressive disorder (35 genes; closest gene selected to a given index SNP, as described in (Wray, Ripke et al. 2018)). (V) Selection of genes in proximity to SNPs associated with a non-psychiatric phenotype (coronary artery disease; 35 genes; random gene selected from a given susceptibility locus, as described in (Schunkert, Konig et al. 2011)).

Exploratory age-windowing. To perform a ‘fine-mapping’ of effects within a set of age-bins, we performed separate analyses for subjects within a given age-window (**supplementary Table 7**). The width of the window was determined as four consecutive age-entries among the recorded ages in weeks. Differences of expression similarity and co-expression with schizophrenia susceptibility genes were determined separately for each age-window as described above. Genes identified as ‘variability genes’ of the investigated age-bins were combined and used for this analysis.

Functional analysis. To explore biological functions of genes contributing to differences of expression similarity between sexes, we used the DAVID functional annotation tool using default settings (https://david.ncifcrf.gov/home.jsp)(Huang da, Sherman et al. 2009). In this tool, enrichment is quantified based on a modified Fisher’s exact test. The 14702 autosomal genes part of the BrainSpan exon microarray data were used as background for functional analysis. We retained all functional annotation clusters with at least one annotation term passing the False-Discovery-Rate (FDR) corrected *P*-value threshold of 0.05.

Code availability. Code is available from the corresponding author upon request.

2.2.4 Results

Expression similarity differences in BrainSpan exon microarray data. The filtered dataset contained autosomal, transcriptome-wide expression data on healthy subjects between the 6th post-conceptional week (PCW) and 40 years of age(Kang, Kawasawa et al. 2011) (42 donors, 23 males, 14702 autosomal genes; **Figure 1**). We tested whether gender was confounded by ethnicity, but found no association (P = 0.77, Chi-squared test). Subjects were binned into 11 age groups and the 16 brain areas were aggregated into 4 regional-clusters with similar expression values (**Supplementary Tables 1 and 2**, regional-clustering was taken from(Willsey, Sanders et al. 2013) and based on hierarchical clustering of fetal transcriptome profiles; for abbreviations, see **Figure 1**): (1) the V1C-STC cluster; (2) the prefrontal and primary motor-somatosensory cortex or PFC-MSC cluster; (3) the STR-HIP-AMY cluster; and (4) the MD-CBC cluster.

Figure 2a shows that despite substantial variability, males had significantly lower expression similarity compared to females in three of the four brain regional-clusters ($P_{V1C\text{-}STC}$<0.004, $P_{PFC\text{-}}$

$_{MSC}$<0.004, $P_{STR\text{-}HIP\text{-}AMY}$=0.003, $P_{MD\text{-}CBC}$=0.080; FWER corrected). Due to the more pronounced differences in the regional clusters V1C-STC and PFC-MSC, subsequent analyses focused on these areas. **Figure 2a** further shows that in females, expression similarity tended to decrease across developmental time-points, suggesting that inter-subject similarity was lower in adulthood compared to younger age. We aimed to explore whether sex differences in expression similarity were associated with genetic schizophrenia risk, to pinpoint a potential biological mechanism for the well-known sex differences of the disorder.

Identification of genes driving sex differences in expression similarity. For each age-bin-regional-cluster combination we identified the 100 'variability genes' with the greatest ratio (male divided by female) of standard deviations of expression (see **Supplementary Figure 3** and **Supplementary Dataset** for a list of all 'variability genes'). **Figure 2b** shows that expression similarity determined from these genes differed strongly between sexes.

Co-expression between variability and schizophrenia susceptibility genes. Next, we investigated potential relationships between these variability genes and genes harbored by the 108 well-established schizophrenia susceptibility loci(Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014). This analysis was performed in males, since the lack of variance in female expression levels would prevent meaningful association analyses. Across 22 sets of variability genes (11 age-bins in the 2 regional-clusters V1C-STC and PFC-MSC), we found that variability genes derived from both clusters were significantly co-expressed with schizophrenia susceptibility genes in age-bins 8 (4 months – 4 years, $rho_{V1C\text{-}STC}$=0.05, $rho_{PFC\text{-}MSC}$=0.05), 9 (10 months – 11 years , $rho_{V1C\text{-}STC}$=0.10, $rho_{PFC\text{-}MSC}$=0.12) and 10 (2 years – 19 years , $rho_{V1C\text{-}STC}$=0.07, $rho_{PFC\text{-}MSC}$=0.13; all P_{FWER}<0.022, **Figures 3a** and **b**). Significant co-expression was additionally observed for the PFC-MSC in age-bin 1 (6 PCW – 13 PCW, rho=0.11) and 2 (9 PCW – 16 PCW, rho=0.08; all P_{FWER}<0.022).

Age-bin specificity and pathway analysis. Next, we explored whether differences in expression similarity were age-bin specific. **Figure 2c** shows that PFC-MSC variability genes of age-bin blocks 1-2 and 8-9-10 were also associated, albeit to a lesser extent, with decreased male expression similarity in the respectively other age-bin blocks.

In this brain-regional-cluster, the 257 genes of age-bins 8-9-10 were significantly linked to synaptic processes and (calcium-) ion signaling (**Supplementary Table 4**). Notably, the 138 variability genes from age-bins 1 and 2 in the PFC-MSC cluster were associated with similar ontological categories, including 'post-synaptic membrane' and 'synapse' (**Supplementary Table 5**). Interestingly, the genes from age-bins 1-2 and age-bins 8-9-10 showed only a minimal overlap (8 genes shared). These ontological associations showed regional specificity for the PFC-MSC cluster, as the V1C-STC variability genes (age-bins 8-9-10) that also showed significant co-expression with susceptibility genes, were not associated with similar ontological categories (**Supplementary Table 6**). Furthermore, the ontological overlap between age-bins 1-2 and age-bins 8-9-10 in the PFC-MSC cluster is consistent with the correlation of the male expression similarity profiles (**Figure 2c**).

Schizophrenia specificity. To explore the specificity of co-expression results for schizophrenia, analysis was repeated using (I) schizophrenia susceptibility genes randomly selected for a given locus (instead of based on physical proximity to the index SNP), (II) genes randomly selected from loci with comparable DNA sequence variability compared to the schizophrenia loci, (III) genes randomly selected from the same chromosome as a given schizophrenia gene, irrespective of DNA sequence variability, (IV) genes in proximity to SNPs associated with major depressive disorder, (V) genes in proximity to SNPs associated with a non-psychiatric phenotype (coronary artery disease). **Figure 3c** shows that random and proximity-based selection of genes from schizophrenia loci yielded similar results. Despite a similar co-expression profile across age-bins, schizophrenia gene co-expression

showed specificity against DNA sequence variability-stratified gene selection in age-bins 8 and 9 (*P*=0.05) and a trend towards specificity in age-bins 2 and 10 (*P*=0.06). Randomly selected genes (procedure III) showed substantially lower mean co-expression, leading to specificity of schizophrenia results (age-bins 8-9-10, P<=0.05; age-bin 2, *P*=0.06). For both random selection procedures, schizophrenia specificity could not be observed in age bin 1 (*P*=0.12 and *P*=0.11, for procedures II and III, respectively). Genes in the proximity of SNPs linked to major depression or grip strength led to lower co-expression values in age bins 2 and 8-9-10; in age-bin 1, major depression genes showed higher co-expression than the schizophrenia genes.

Age-windowing. Finally, since age-bins 8-9-10 covered a broad age range (4 months – 19 years), we performed an exploratory 'fine-mapping' of PFC-MSC effects using an age-windowing approach. While based on small sample numbers, this analysis suggested that co-expression had a broad plateau from a mean age of 4.8 years to 10.9 years (**Figure 3d**). Differences in expression similarity between sexes were consistent across all age windows (**Figure 3d**).

Replication in BrainSpan RNAseq data. Preprocessed BrainSpan RNAseq data comprised expression information on 11514 autosomal genes in 400 samples (37 subjects, 20 males). The transcriptome-wide expression similarity showed similar profiles as observed for exon microarray data (**Supplementary Figure 4a**). Similarly, the variability genes identified from exon microarray data were also variability genes in RNAseq data (**Supplementary Figure 4b**, *P*<0.001). These genes were significantly correlated with schizophrenia susceptibility genes in the PFC-MSC regional-cluster for age-bins 2 (rho=0.01, *P*<0.001), 9 (rho=0.05, *P*<0.001) and 10 (rho=0.03, *P*<0.001), validating exon microarray observations. For the V1C-STC cluster, we found significant associations for age-bins 3 (rho=0.03, *P*<0.001), and a trend towards nominal significance in age-bins 8 (*P*=0.07), and 10 (*P*=0.05).

Validation in Braincloud data. Filtered Braincloud data contained dorsolateral prefrontal cortex (DLPFC) expression information on 14773 autosomal genes from 112 subjects (75 males). In covariate-corrected data, expression similarity is dependent on expression variance. Therefore, we compared the standard deviation of expression across all genes overlapping with BrainSpan exon microarray data. We found these estimates to be strongly correlated across datasets (rho=0.40, $P<2.2\cdot10^{-16}$, Spearman correlation), suggesting that preprocessing resulted in high cross-dataset comparability. Since the Braincloud data contained no subjects in age groups 1 and 2 (i.e. age-bin 1 only consisted of subjects in age group 3), age-bin 1 was not used for further analysis. Assessment of expression similarity differences using BrainSpan exon microarray PFC-MSC variability genes validated the decreased similarity in males (*P*<0.001, **Supplementary Figure 5**), which was less pronounced in Braincloud data and driven by genes from age-bins 8 and 9. Consistent with BrainSpan results, co-expression with schizophrenia susceptibility genes was significant in age-bin 2 (rho=0.03, *P*<0.001), age-bin 9, (rho=0.07, *P*<0.001) and age-bin 10 (rho=0.03, *P*<0.001) and showed a trend towards significance in age-bin 8 (rho=0.01, *P*=0.08).

2.2.5 Discussion

The present results demonstrate that the similarity of gene expression profiles in males shows a brain-region specific decrease compared to females. Some of the genes driving this effect were co-expressed with schizophrenia susceptibility genes, in a regionally specific and age-dependent manner. Importantly, co-expression was found in the brain regional-cluster encompassing the prefrontal cortex during fetal brain development, confirming a core prediction of the

neurodevelopmental hypothesis of schizophrenia(Weinberger 1987). Additionally, and again as predicted by this hypothesis, significant co-expression was further found during adolescence. Similar differences of expression similarity were found in RNAseq data acquired on a subset of the same samples. In this dataset, we further replicated associations between variability and susceptibility genes in data from adolescent donors, but found no associations during the fetal period. Expression similarity differences were further validated in the independent Braincloud data and significant co-expression was found in samples from fetal, as well as adolescent donors.

Co-expression did not depend on how genes were selected from a given susceptibility locus and exceeded that observed for major depression (in age-bin 1 by a small margin) and coronary artery disease in the early fetal phase, as well as during adolescence. We observed that genes selected from randomly chosen loci stratified for DNA sequence variability showed a broadly similar, although less pronounced, co-expression trend compared to schizophrenia genes. In contrast, genes selected randomly without consideration of DNA sequence variability were not co-expressed with variability genes, on average. This may suggest that sequence variability associated with schizophrenia susceptibility loci impacted on diversification of gene expression and the sex-differences observed in the present study.

Genes from the fetal and adolescent periods were involved in synaptic processes, which have been implicated in schizophrenia by a range of genetic, histopathological, neuroimaging, pharmacological and neurotransmitter studies(McGlashan and Hoffman 2000, Tsai and Coyle 2002, Glausier and Lewis 2013, Network and Pathway Analysis Subgroup of Psychiatric Genomics 2015, Schwarz, Izmailov et al. 2016). They are affected by genetic and environmental risk in particular during early life, leading to subsequent impairments in synaptic plasticity and connectivity(Lewis and Levitt 2002). The lack of overlap between variability genes from the fetal period and adolescence may hint at biologically divergent risk processes that converge on the same synaptic pathways.

The main limitation of the present study is sample size. The primary analysis of the BrainSpan data reported that brain-region and age are stronger modulators of gene expression compared to sex or inter-individual variation(Kang, Kawasawa et al. 2011). Therefore, the present study focused on analyses that are stratified by regional clusters and age-bins, with significant impact on sample numbers available for a given analysis. In the BrainSpan dataset, data from multiple brain regions was available for most donors. We performed a donor-wise bootstrapping procedure during all resampling analyses, to account for the non-independence of the samples. This procedure further accounted for potential effects arising from differences in donor numbers between sexes, further reducing the effective sample size. The low donor number per regional-cluster age-bin combination prevented meaningful permutation of gender. Therefore, random 'variability genes' were created by randomly sampling genes, stratified by expression variance. This may have led to bias, due to the potential correlation among the actual variability genes that is not captured by the procedure employed here. The low sample number in all three investigated datasets also limits the power to identify and validate significant associations, including expression similarity differences and co-expression between variability and susceptibility genes. This may have contributed to the partial non-replication of findings across datasets.

Another limitation is that we selected a single susceptibility gene per locus to prevent statistical bias, but this selection may not accurately reflect genetic schizophrenia risk. By comparison, other studies have previously selected susceptibility genes by extracting all genes within a given locus(Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014) or by focusing on effectors of index variant eQTLs(Gamazon, Wheeler et al. 2015). Another interesting aspect is that the present findings may relate to underlying, variable phenotypes, such as personality traits and

comorbid psychiatric conditions. Furthermore, we aimed to account for the effects of known and unknown confounders during all analyses, but this may not have comprehensively captured experimental artefacts that may have influenced between-subject or gene-gene correlations. Finally, we did not use genetic association data to correct for potential subject relatedness or population structure, due to data availability and sample size limitations.

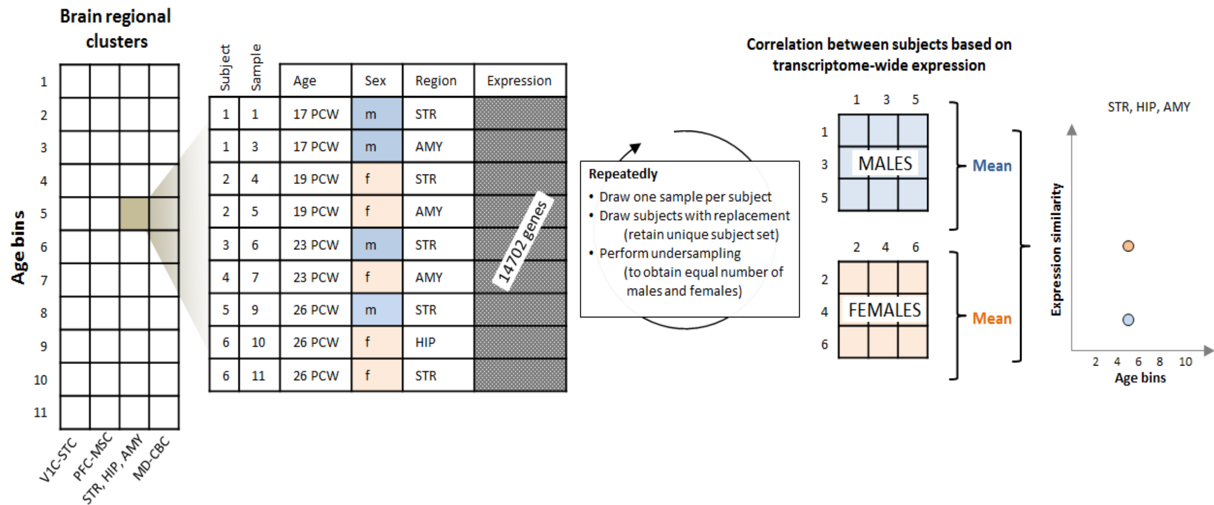
In conclusion, this study indicates sex specific genetic mechanisms operating during fetal brain development linked to the variability of prefrontal brain gene expression during adolescence, as predicted by the neurodevelopmental hypothesis of schizophrenia. These effects may contribute to the well-established clinical sex differences of schizophrenia and underlying gene sets may be valuable for biologically stratified exploration of the illness's etiology.

Tables

Table 1. BrainSpan exon microarray sample numbers for males and females across 11 age bins and 4 brain regional clusters after data preprocessing (1: V1C-STC, 2: PFC-MSC, 3: STR-HIP-AMY, 4: MD-CBC, see Supplementary Table 1 for details). Subject numbers are shown in brackets.

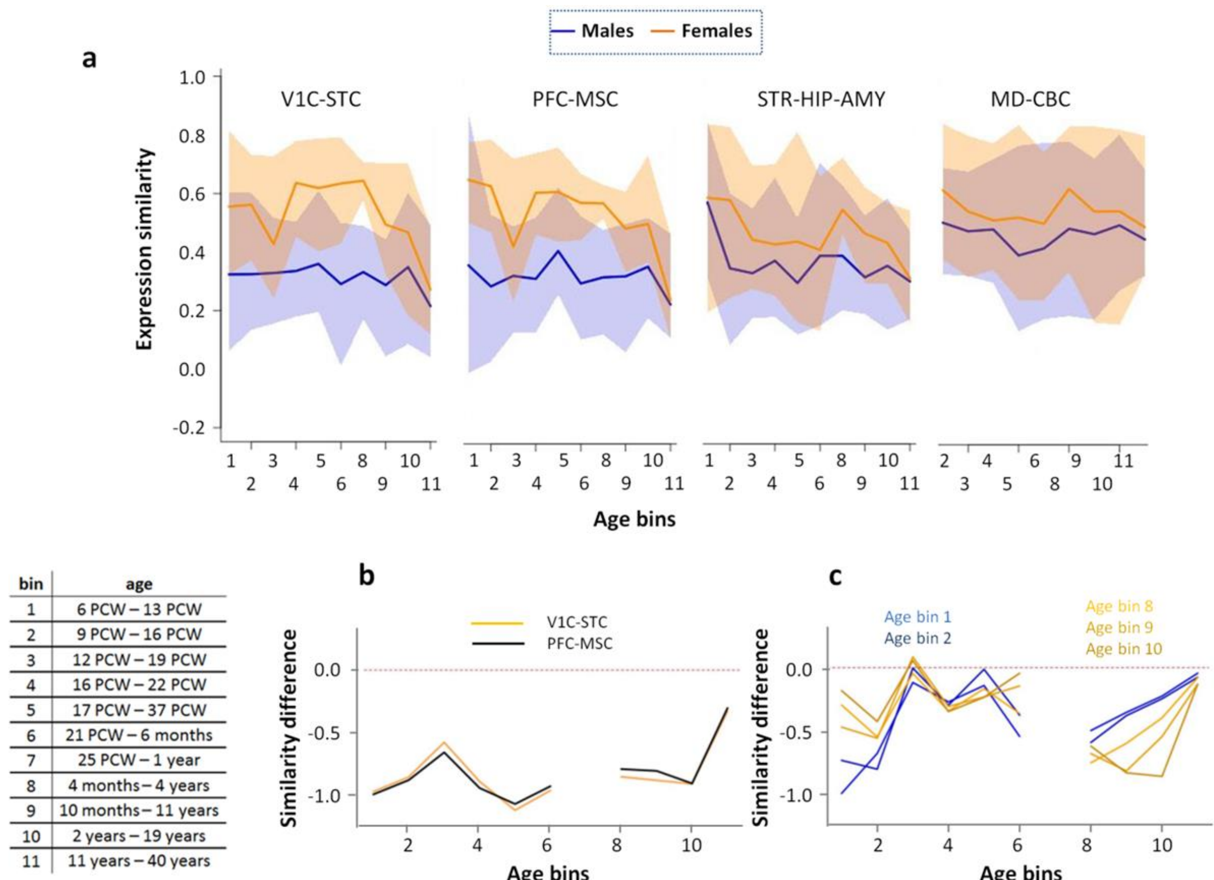
		Males				**Females**			
Age bin	Regional cluster	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
1		12 (3)	21 (4)	11 (4)	1 (1)	20 (4)	24 (4)	12 (4)	7 (4)
2		23 (5)	31 (6)	16 (6)	5 (4)	20 (4)	24 (4)	12 (4)	7 (4)
3		23 (5)	27 (5)	14 (5)	5 (4)	33 (7)	39 (7)	20 (7)	13 (7)
4		23 (5)	27 (5)	14 (5)	8 (5)	27 (6)	31 (6)	16 (6)	12 (6)
5		15 (3)	18 (3)	9 (3)	6 (3)	27 (6)	31 (6)	16 (6)	12 (6)
6		30 (6)	36 (6)	18 (6)	12 (6)	14 (3)	16 (3)	8 (3)	6 (3)
7		24 (5)	29 (5)	12 (4)	10 (5)	5 (1)	6 (1)	3 (1)	2 (1)
8		24 (5)	28 (5)	12 (4)	10 (5)	10 (2)	9 (2)	6 (2)	4 (2)
9		19 (4)	20 (4)	8 (3)	7 (4)	15 (3)	15 (3)	8 (3)	5 (3)
10		20 (4)	21 (4)	10 (4)	7 (4)	20 (4)	21 (4)	11 (4)	7 (4)
11		36 (8)	42 (8)	19 (7)	14 (8)	33 (7)	39 (7)	19 (7)	12 (7)

Figures



Subject	Sample	Age	Sex	Region
1	1	17 PCW	m	STR
1	3	17 PCW	m	AMY
2	4	19 PCW	f	STR
2	5	19 PCW	f	AMY
3	6	23 PCW	m	STR
4	7	23 PCW	f	AMY
5	9	26 PCW	m	STR
6	10	26 PCW	f	HIP
6	11	26 PCW	f	STR

Figure 1. Analysis workflow. Transcriptome-wide expression data were extracted from the BrainSpan Atlas of the Developing Human Brain for each age-bin – brain regional-cluster combination. Age-bins and regional-clusters were taken from (Willsey, Sanders et al. 2013). Using a resampling procedure, expression variability was then quantified in males and females as the mean of the pairwise correlations of transcriptome-wide expression between samples from the respective subjects. PCW, post conceptional week; V1C, primary visual cortex; ITC, inferior temporal cortex; IPC, posterior inferior parietal cortex; A1C, primary auditory cortex; STC, superior temporal cortex; M1C, primary motor cortex; S1C, primary somatosensory cortex; VFC, ventral prefrontal cortex; MFC, medial prefrontal cortex; DFC, dorsal prefrontal cortex; OFC, orbital prefrontal cortex; STR, striatum; HIP, hippocampal anlage/ hippocampus; AMY, amygdala; MD, mediodorsal nucleus of the thalamus; CBC, cerebellar cortex.



bin	age
1	6 PCW – 13 PCW
2	9 PCW – 16 PCW
3	12 PCW – 19 PCW
4	16 PCW – 22 PCW
5	17 PCW – 37 PCW
6	21 PCW – 6 months
7	25 PCW – 1 year
8	4 months – 4 years
9	10 months – 11 years
10	2 years – 19 years
11	11 years – 40 years

Figure 2. Sex differences in expression similarity in BrainSpan exon microarray data. a) Expression similarity for four brain regional-clusters: V1C-STC, PFC-MSC, STR-HIP-AMY, and MD-CBC for males (blue) and females (orange). The panels display mean estimates (solid lines) and 95% confidence intervals (shaded areas). The panels show no values for regional-cluster age-bin combinations containing data from only one donor. **b)** Expression variability for 'variability genes', identified separately for each given age-bin. In age-bin 7, data from only one donor was available for females. **c)** Expression variability profiles for variability genes derived from age-bins 9 [10 months – 11 years] and 10 [2 years – 19 years] in the PFC-MSC cluster. This panel shows variability profiles for male subjects only.

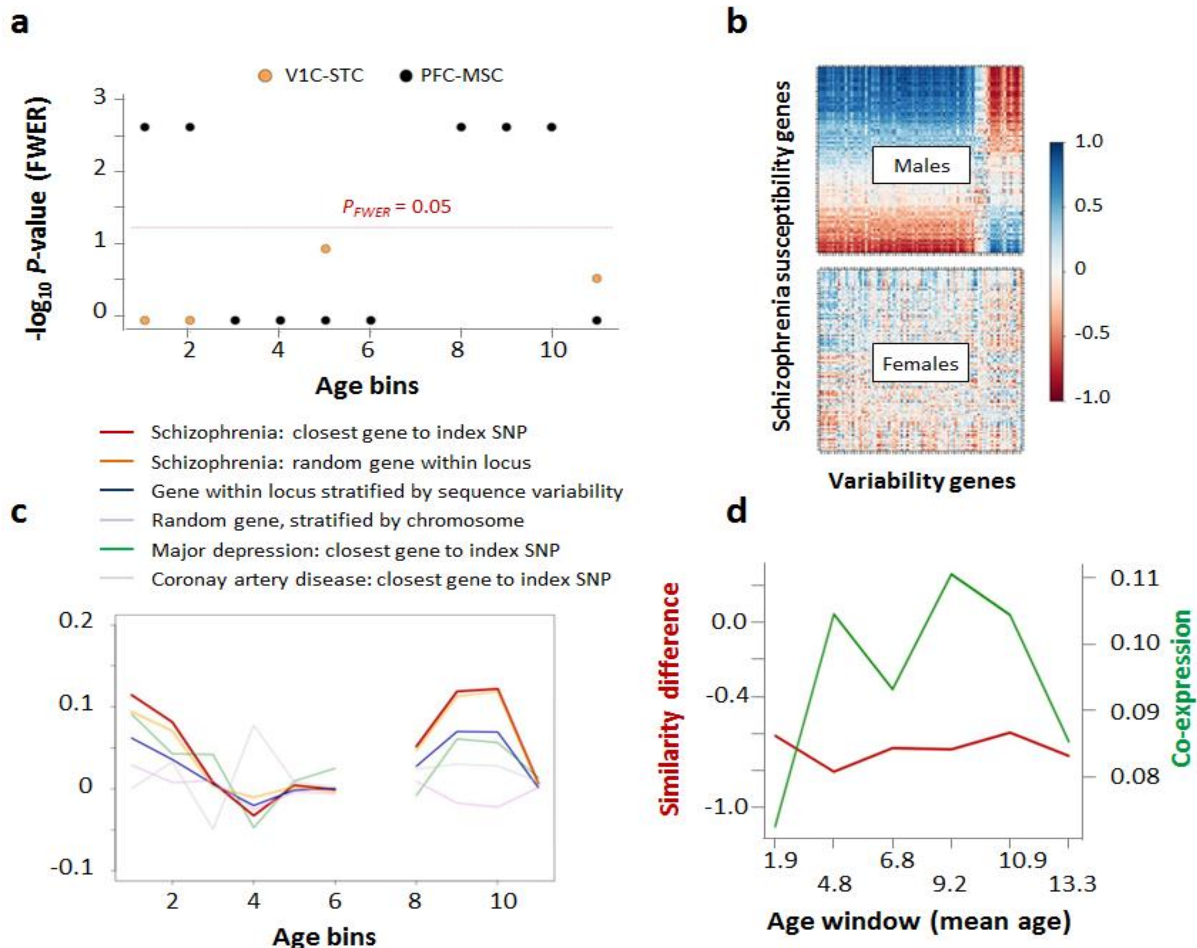


Figure 3. Co-expression between variability genes and schizophrenia susceptibility genes. a) Significance of median co-expression for variability genes determined for each age-bin in the regional-clusters V1C-STC, PFC-MSC and STR-HIP-AMY of male subjects. **b)** Co-expression in PFC-MSC cluster, age-bin 10, for males and females, respectively. Rows and columns were ordered separately based on median co-expression. **c)** Comparison of co-expression between variability genes and schizophrenia susceptibility genes chosen based on physical proximity to index SNPs (red), random selection within a given susceptibility locus (orange), randomly selected loci with comparable DNA sequence variability compared to schizophrenia loci (blue), random genes selected from the same chromosomes as schizophrenia susceptibility genes (purple), major depression susceptibility genes (green) and genes linked to a non-psychiatric phenotype (coronary artery disease, grey). **d)** Windowing of age-bins 8, 9 and 10 in the PFC-MSC cluster. The panel shows variability difference and co-expression for variability genes determined for age-bins 8 to 10. Co-expression was determined for males only.

2.2.6 Supplementary Tables

Supplementary Table 1: Brain regions and regional clusters used in the present study.

Cluster	Region	Description
1	V1C	primary visual cortex
1	ITC	inferior temporal cortex
1	IPC	posterior inferior parietal cortex
1	A1C	primary auditory cortex
1	STC	superior temporal cortex
2	M1C	primary motor cortex
2	S1C	primary somatosensory cortex
2	VFC	ventral prefrontal cortex
2	MFC	medial prefrontal cortex
2	DFC	dorsal prefrontal cortex
2	OFC	orbital prefrontal cortex
3	STR	striatum
3	HIP	hippocampal anlage (periods 1–2), hippocampus (periods 3–13)
3	AMY	amygdala
4	MD	mediodorsal nucleus of the thalamus
4	CBC	cerebellar cortex

Supplementary Table 2: Grouping of subjects into age bins as performed by Willsey et al (Willsey, Sanders et al. 2013). Periods of human brain development were taken from Kang et al (Kang, Kawasawa et al. 2011).

Age bin	Period	Ages	Description
1	1,2,3	6 PCW – 13 PCW	embryonic to early fetal
2	2,3,4	9 PCW – 16 PCW	early fetal to early mid-fetal
3	3,4,5	12 PCW – 19 PCW	early fetal to early mid-fetal
4	4,5,6	16 PCW – 22 PCW	early mid-fetal to late mid-fetal
5	5,6,7	17 PCW – 37 PCW	early mid-fetal to late fetal
6	6,7,8	21 PCW – 6 months	late mid-fetal to neonatal & early infancy
7	7,8,9	25 PCW – 1 year	Late fetal to late infancy
8	8,9,10	4 months – 4 years	neonatal & early infancy to early childhood
9	9,10,11	10 months – 11 years	late infancy to middle and late childhood
10	10,11,12	2 years – 19 years	early childhood to adolescence
11	11,12,13	11 years – 40 years	Adolescence to young adulthood

Supplementary Table 3: Schizophrenia susceptibility genes used in the present study. Susceptibility loci were taken from a study by the *Schizophrenia Working group of the Psychiatric Genomics Consortium* (Schizophrenia Working Group of the Psychiatric Genomics 2014). Rank: rank of significance of case-control difference described in (Schizophrenia Working Group of the Psychiatric Genomics 2014). Chr and position: chromosomal position of index SNP. Gene: gene in closest chromosomal position to the index SNP of a given locus. If a given locus contained multiple index SNPs (annealed locus), the gene in closest proximity to the most significant index SNP was chosen. Genes marked with # were not annotated by the R library org.Hs.eg.db. Genes marked with an asterisk were not part of the BrainSpan data.

Rank	Index SNP	Chr	Position	Gene
1	rs115329265	MHC locus position		
2	rs1702294	1	98501984	MIR137*
3	rs11191419	10	104612335	C10orf32#
4	rs2007044	12	2344960	CACNA1C
5	rs4129585	8	143312933	TSNARE1
6	rs35518360	4	103146890	SLC39A8
7	chr7_2025096_I	7	2025096	MAD1L1
8	rs4391122	5	60598543	ZSWIM6*
9	rs2851447	12	123665113	MPHOSPH9
10	chr2_200825237_I	2	200825237	C2orf47
11	rs4702	15	91426560	FURIN
12	rs75968099	3	36858583	TRANK1*
13	chr10_104957618_I	Annealed with rs11191419		
14	rs12887734	14	104046834	APOPT1*
15	rs8042374	15	78908032	CHRNA3
16	rs13240464	7	110898915	IMMP2L
17	rs10791097	11	130718630	SNX19
18	rs11693094	2	185601420	ZNF804A
19	rs1378559	X	21380266	CNKSR2
20	rs7893279	10	18745105	CACNB2
21	rs12826178	12	57622371	SHMT2
22	rs12129573	1	73768366	LRRIQ3
23	rs6704768	2	233592501	GIGYF2
24	rs55661361	11	124613957	NRGN
25	rs9636107	18	53200117	TCF4
26	chr11_46350213_D	11	46350213	DGKZ
27	rs7907645	Annealed with rs11191419		
28	chr3_180594593_I	3	180594593	FXR1
29	rs6065094	20	37453194	PPP1R16B
30	rs11682175	2	57987593	VRK2
31	rs950169	15	84706461	ADAMTSL3
32	rs72934570	18	53533189	TCF4
33	rs6434928	2	198304577	SF3B1
34	rs9607782	22	41587556	EP300
35	rs36068923	8	111485761	KCNV1
36	rs17194490	3	2547786	CNTN4

37	rs2514218	11	113392994	DRD2
38	rs75059851	11	133822569	IGSF9B
39	rs2535627	3	52845105	ITIH4
40	rs12691307	16	29939877	KCTD13
41	chr22_39987017_D	22	39987017	CACNA1I
42	rs7432375	3	136288405	STAG1
43	chr18_52749216_D	Annealed with rs9636107		
44	rs111294930	5	152177121	GRIA1
45	rs2973155	Annealed with rs111294930		
46	rs5937157	X	68377126	PJA1
47	rs4523957	17	2208899	SRR
48	rs12704290	7	86427626	GRM3
49	rs12903146	15	61854663	VPS14C[#]
50	rs11210892	1	44100084	PTPRF
51	rs2905426	19	19478022	MAU2*
52	rs140505938	1	150031490	VPS45
53	chr6_84280274_D	6	84280274	SNAP91
54	rs4648845	1	2387101	PLCH2
55	rs7405404	16	13749859	ERCC4
56	rs6466055	7	104929064	SRPK2
57	chr1_8424984_D	1	8424984	RERE
58	rs4766428	12	110723245	ATP2A2
59	rs10520163	4	170626552	CLCN3
60	rs117074560	6	96459651	FUT9
61	rs6002655	22	42603814	TCF20
62	chr2_146436222_I	No gene in proximity		
63	rs9420	11	57510294	C11orf31
64	rs11027857	11	24403620	LUZP2
65	rs1498232	No gene in proximity		
66	rs3735025	7	137074844	DGKI
67	rs11139497	9	84739941	TLE1
68	rs77149735	1	243555105	SDCCAG8
69	rs56205728	15	40567237	PAK6
70	rs2053079	19	30987423	ZNF536
71	rs16867576	5	88746331	MEF2C
72	rs4330281	3	17859366	TBC1D5
73	rs3849046	5	137851192	ETF1
74	rs2693698	14	99719219	BCL11B
75	rs2332700	14	72417326	RGS6
76	rs1501357	5	45364875	HCN1
77	rs6984242	8	60700469	CA8
78	chr1_243881945_I	Annealed with rs77149735		
79	rs79212538	Annealed with rs111294930		
80	rs3768644	2	72361505	CYP26B1
81	rs77502336	11	123394636	GRAMD1B
82	rs6704641	2	200164252	SATB2

83	rs59979824	2	193848340	PCGEM1*
84	rs1106568	4	176861301	GPM6A
85	rs10503253	8	4180844	CSMD1
86	rs10043984	Annealed with rs3849046		
87	rs11685299	2	225391296	CUL3
88	rs7819570	8	89588626	MMP16
89	rs715170	Annealed with rs72934570		
90	rs9922678	16	9946319	GRIN2A
91	rs78322266	Annealed with rs9636107		
92	rs2068012	14	30190316	PRKD1
93	rs832187	3	63833050	C3orf49
94	rs8044995	16	68189340	NFATC3
95	chr2_149429178_D	2	149429178	EPC2
96	rs8082590	17	17958402	GID4*
97	rs12148337	15	70589272	TLE3
98	rs12325245	16	58681393	CNOT1
99	rs2239063	Annealed with rs2007044		
100	rs12522290	Annealed with rs111294930		
101	rs10803138	Annealed with rs77149735		
102	rs73229090	8	27442127	CLU
103	rs324017	Annealed with rs12826178		
104	rs12845396	X	6029533	NLGN4X
105	rs55833108	Annealed with rs11191419		
106	rs9841616	Annealed with chr3_180594593_I		
107	rs76869799	Annealed with rs1702294		
108	rs1339227	6	73155701	RIMS1
109	chr7_24747494_D	7	24747494	DFNA5
110	rs4388249	5	109036066	MAN2A1
111	rs215411	4	23423603	MIR548AJ2*
112	rs11740474	5	153680747	GALNT10
113	rs1023500	Annealed with rs6002655		
114	rs12421382	11	109378071	C11orf87
115	rs211829	7	110048893	IMMP2L
116	rs679087	12	29917265	TMTC1
117	rs75575209	Annealed with rs11682175		
118	rs7801375	7	131567263	PODXL
119	rs14403	Annealed with rs77149735		
120	rs6670165	1	177280121	BRINP2*
121	rs7523273	1	207977083	CD46
122	rs7267348	20	48131036	PTGIS
123	rs4240748	12	92246786	C12orf79[#]
124	rs2909457	2	162845855	DPP4
125	rs56873913	19	50091199	PRRG2
126	rs190065944	Annealed with rs8042374		
127	rs10860964	12	103596455	C12orf42
128	chr5_140143664_I	5	140143664	PCDHA1*

Supplementary Table 4: Ontological terms associated with heterogeneity genes of age bins 8, 9 and 10 in the PFC-MSC cluster, as determined using the DAVID tool (Huang da, Sherman et al. 2009). P-values were corrected for the False Discovery Rate (FDR) according to the method of Benjamini and Hochberg.

Category	P(FDR)
calcium ion binding	2.60E-04
Synapse	3.50E-03
Epilepsy	3.80E-03
Phosphoprotein	4.00E-03
Ion channel	4.20E-03
Parkinson disease	5.10E-03
Membrane	5.90E-03
Alternative splicing	1.10E-02
Glycoprotein	2.50E-02
Ion transport	2.80E-02
Parkinsonism	3.20E-02
Cell membrane	3.40E-02
Cell adhesion	3.40E-02
Cell junction	4.10E-02
plasma membrane	4.90E-02

Supplementary Table 5: Ontological terms associated with heterogeneity genes of age bins 1 and 2 in the PFC-MSC cluster, as determined using the DAVID tool (Huang da, Sherman et al. 2009). P-values were corrected for the False Discovery Rate (FDR) according to the method of Benjamini and Hochberg.

Category	P(FDR)
Synapse	3.30E-03
Postsynaptic cell membrane	5.50E-03
dendrite	1.90E-02
cell junction	2.20E-02
Cell membrane	2.60E-02
postsynaptic membrane	3.50E-02
Membrane	4.50E-02
chemical synaptic transmission	4.60E-02

Supplementary Table 6: Ontological terms associated with heterogeneity genes of age bins 1 and 2 in the PFC-MSC cluster, as determined using the DAVID tool (Huang da, Sherman et al. 2009). P-values were corrected for the False Discovery Rate (FDR) according to the method of Benjamini and Hochberg.

Category	P(FDR)
Phosphoprotein	1.60E-07
CARM1 and Regulation of the Estrogen Receptor	7.70E-03
Cell division and chromosome partitioning	4.40E-02

Supplementary Table 7: Age and sex distribution for donors part of the age-windowing analysis. Age is shown as mean ± sd.

Age-window	age	sex (m/f)
1	1.9 ± 1.3	28/9
2	4.8 ± 3.1	20/9
3	6.8 ± 3.7	15/15
4	9.2 ± 3.5	15/15
5	10.9 ± 3.7	21/12
6	13.3 ± 3.9	16/18

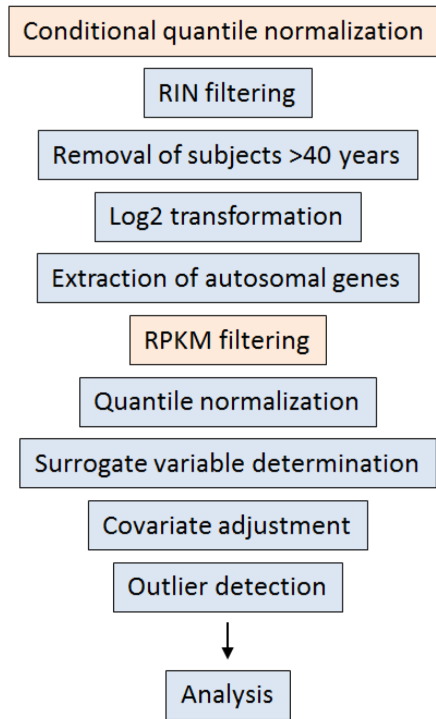
Supplementary Table 8: BrainSpan RNAseq sample numbers for males and females across 11 age bins and 4 brain regional clusters after data preprocessing (1: V1C-STC, 2: PFC-MSC, 3: STR-HIP-AMY, 4: MD-CBC, see Supplementary Table 1 for details). Subject numbers are shown in brackets.

	Males				Females			
Age bin / Regional cluster	1	2	3	4	1	2	3	4
1	10 (3)	21 (4)	10 (4)	0 (0)	19 (4)	23 (4)	12 (4)	2 (2)
2	23 (5)	29 (6)	15 (6)	4 (3)	19 (4)	23 (4)	12 (4)	2 (2)
3	23 (5)	25 (5)	13 (5)	4 (3)	27 (6)	32 (6)	17 (6)	5 (4)
4	18 (4)	19 (4)	10 (4)	5 (4)	9 (3)	9 (2)	5 (2)	4 (3)
5	4 (1)	6 (1)	3 (1)	1 (1)	12 (5)	10 (3)	5 (2)	4 (3)
6	12 (3)	13 (3)	9 (3)	5 (3)	4 (2)	1 (1)	0 (0)	1 (1)
7	12 (3)	11 (3)	6 (2)	5 (3)	3 (1)	1 (1)	0 (0)	0 (0)
8	13 (4)	15 (5)	8 (3)	6 (4)	5 (2)	2 (1)	1 (1)	2 (2)
9	15 (4)	17 (5)	6 (3)	4 (4)	10 (3)	8 (2)	3 (2)	3 (3)
10	19 (5)	17 (5)	8 (5)	4 (4)	20 (5)	20 (4)	9 (4)	6 (5)
11	28 (7)	30 (6)	13 (7)	8 (6)	30 (6)	35 (6)	16 (6)	8 (6)

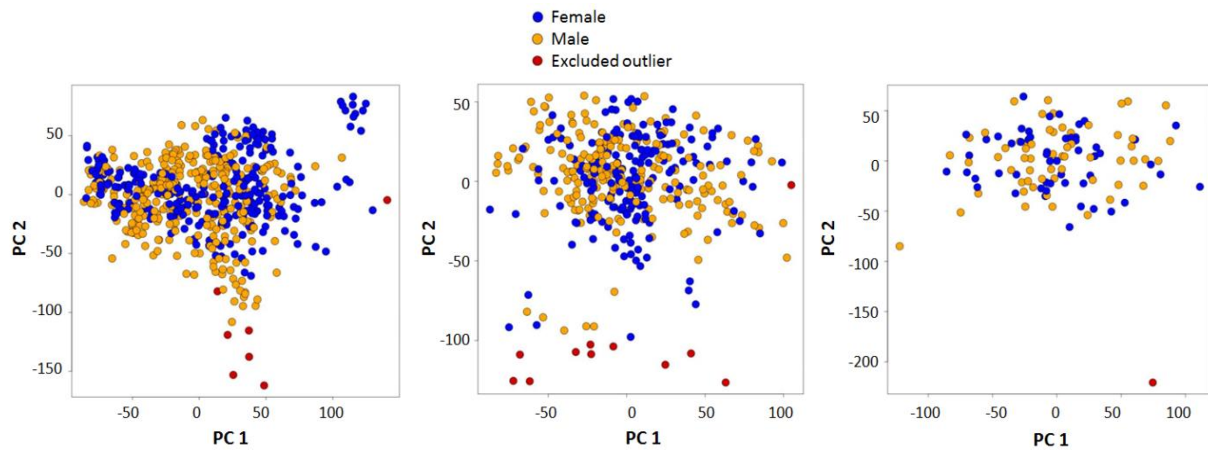
Supplementary Table 9: BrainCloud subject numbers for each age-bin after data preprocessing.

Age bin	Males	Females
1	2	2
2	8	5
3	19	19
4	17	17
5	11	14
6	6	1
7	6	1
8	10	4
9	4	4
10	30	14
11	46	14

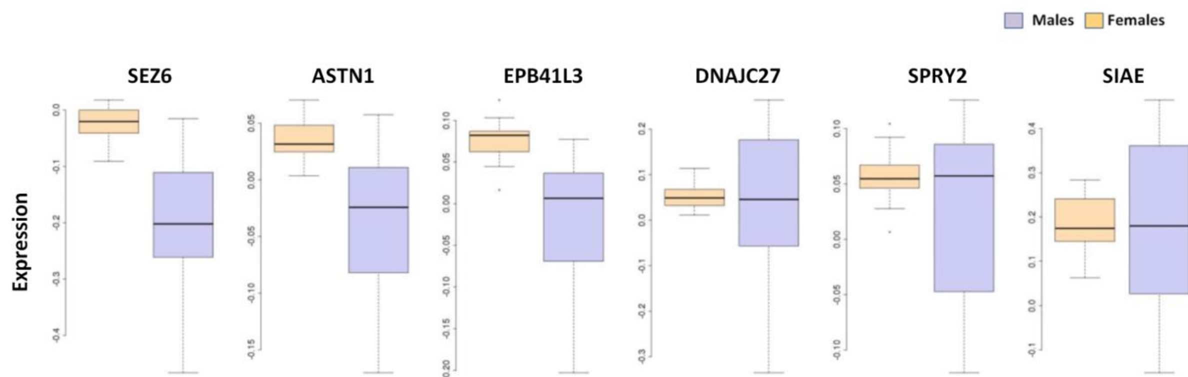
2.2.7 Supplementary Figures



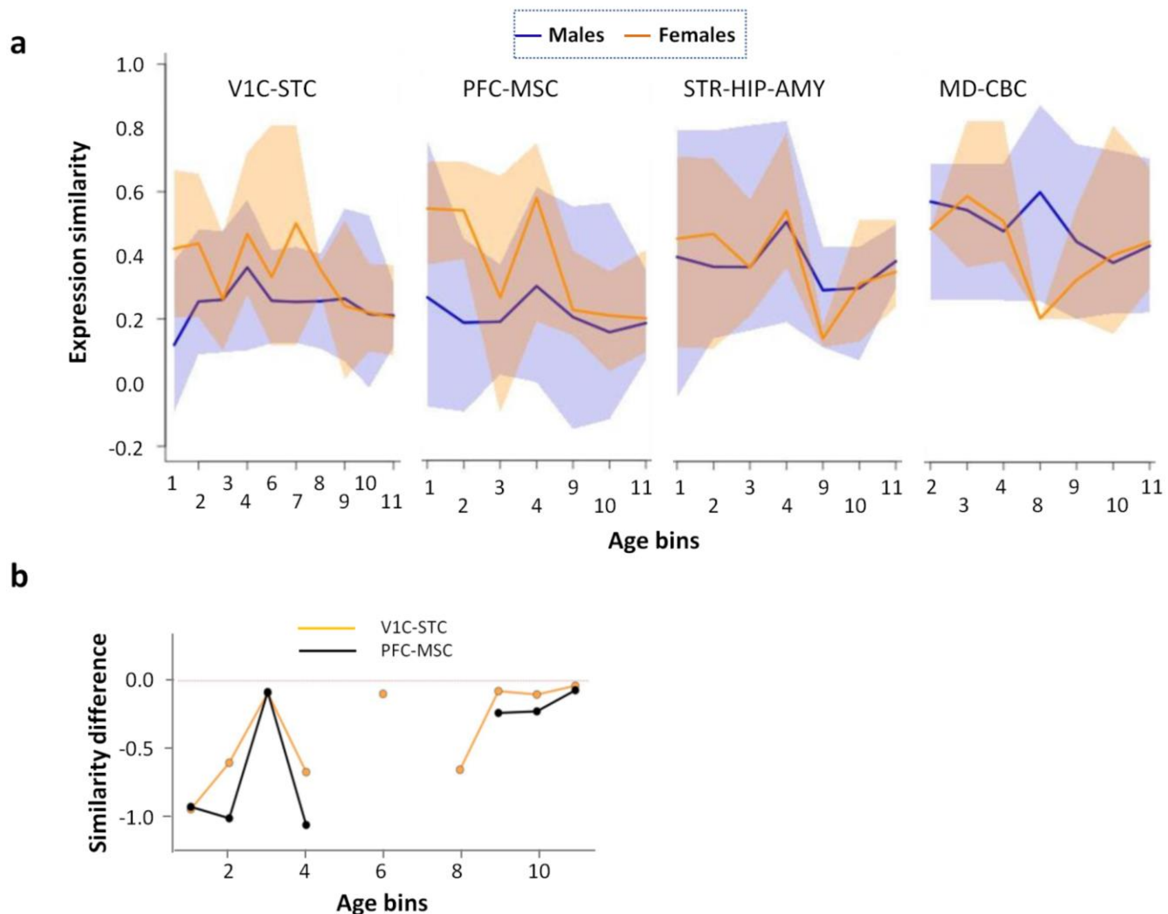
Supplementary Figure 1. Schematic overview of the preprocessing steps performed for the microarray and RNAseq data used in the present study.



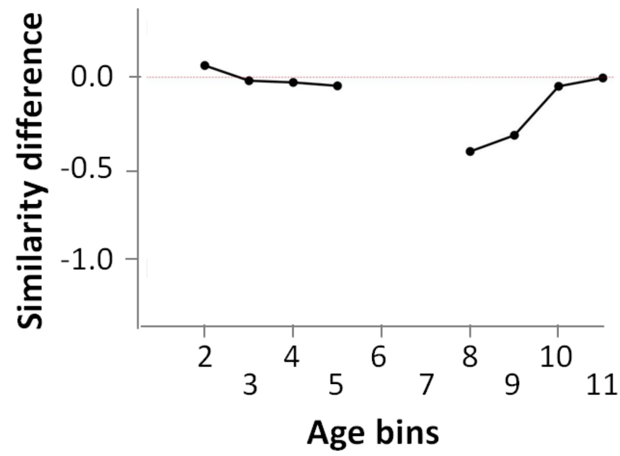
Supplementary Figure 2. PCA scores plots of BrainSpan exon microarray (left panel), BrainSpan RNAseq (middle panel) and Braincloud (right panel) data. Excluded outliers are shown in red.



Supplementary Figure 2. Examples of variability genes in the PFC-MSC cluster, age bin 9. The three panels on the left show the three genes with the most significant expression difference between males and females. The three panels on the right the genes with the least significant sex-difference in expression. Significance was determined using Wilcoxon rank-sum tests, which do, however, not account for the non-independence between the multiple samples from a given donor.



Supplementary Figure 3. Sex differences in expression similarity in BrainSpan RNAseq data. a) Expression similarity for four brain regional clusters: V1C-STC, PFC-MSC, STR-HIP-AMY, and MD-CBC for males (blue) and females (orange). The panels display mean estimates (solid lines) and 95% confidence intervals (shaded areas). **b)** Differences of expression similarity (male coexpression – female coexpression) for BrainSpan exon microarray 'variability genes', assessed separately for each given age bin in BrainSpan RNAseq data.



Supplementary Figure 4. Differences of expression similarity (male coexpression – female coexpression) for BrainSpan exon microarray 'variability genes', assessed separately for each given age bin in Braincloud data.

3 DISCUSSION

Technological advances in the omics and neuroimaging fields and fast-paced progress in the field of data science, combined with the establishment of large-scale international research consortia and the resulting increase in data availability have led to a substantial improvement of our understanding of the biological etiology of schizophrenia. Despite this progress, however, advances made have not yet been translated into the improved clinical management of this severe illness. The overarching objective of the work described in this thesis was to use contemporary data science methodologies to provide a more in-depth characterization of the (epi-)genetic landscape of schizophrenia and its relation to neural functioning relevant for the illness. An important basis for this work provided the development of a biologically-informed machine learning tool termed BioMM, which we have made publicly available in the form of a library for the statistical programming environment R. In preliminary previous work we have shown that BioMM outperforms conventional machine learning approaches when applied to high-dimensional omics data, providing an ideal foundation for the work presented in this thesis. Using this technology, we set out to identify an epigenetic signature associated with schizophrenia, characterize the relationship between this epigenetic signature and GWAS derived polygenic risk, and explore the impact of the epigenetic signature on schizophrenia-relevant brain function (Study 1). The possibility to incorporate systems-biological information into machine learning approaches and thereby improve the predictive value for complex conditions such as schizophrenia also provides the opportunity for an improved understanding of the spatiotemporal components underlying illness susceptibility as well as the potential to use this knowledge towards future, more personalized therapeutic approaches. Towards this, and based on the well-established sex-dimorphisms in onset-age and clinical course, we set out to characterize sex-specific differences in brain-region specific gene expression trajectories during human development and characterize their relationship with genetic risk for schizophrenia (Study 2). Using advanced, multimodal data science methodology, the studies described in this thesis aim to elucidate brain functionally-relevant components of the schizophrenia etiology and improve our understanding of how genetic risk translates into altered function in a spatiotemporally specific manner.

3.1 Epigenetic modulation of risk

In Study 1, we applied the BioMM procedure to identify a peripheral epigenetic risk signature (PMS). The signature identified was significantly predictive of schizophrenia case-control status and explained 10.5% to 21.8% of the variance across two independent peripheral samples, as well as 8.3% of the variance in the post-mortem DLPFC samples. These findings demonstrate the presence of a reproducible poly-epi-genetic signature associated with schizophrenia that can be detected in peripheral samples of patients with schizophrenia, and which is partly mirrored in epigenetic changes present in the DLPFC of patients. The presence of such an epigenetic signature is consistent with the strong impact of environmental risk factors on illness susceptibility and the notion that epigenetic effects may explain part of the missing heritability phenomenon (Harrison 2015). Notably, the overlap of the epigenetic differences that led to the significant cross-tissue prediction of schizophrenia is in agreement with the previously observed correlation of DNA methylation between blood and the brain (Walton, Hass et al. 2016, Edgar, Jones et al. 2017, Braun, Han et al. 2019). An important result of Study 1 was that the identified PMS was reproducibly associated with DLPFC-HC functional connectivity during two working memory tasks, suggesting a schizophrenia-related brain-

functional effect of the identified epigenetic differences. These results further support the illness-relevance of the hippocampal-dorsolateral prefrontal interaction in the pathophysiology of schizophrenia (Weinberger 1987, Bahner, Demanuele et al. 2015) and highlight its likely susceptibility to environmental risk effects.

Remarkably, the epigenetic signature did not correlate with a GWAS derived polygenic risk score, and no association was found for the interaction of PMS and polygenic risk score on DLPFC-HC connectivity. Furthermore, a significant relationship was detected between the polygenic risk score and DLPFC-HC connectivity. These findings suggest that I) the identified PMS carried illness- and brain-functionally-relevant information that was largely independent of genetic schizophrenia risk, II) that the identified epigenetic differences were not secondary to genetic susceptibility and III) that epigenetic effects did not amplify genetic schizophrenia risk in terms of its impact on DLPFC-HC connectivity. These results were further substantiated by analysis of unaffected first-degree relatives of schizophrenia, which also indicated the absence of a strong genetic effect. Notably, it has been argued that some patients may be affected by a stronger genetic load leading to psychosis irrespective of epigenetic modifications or environmental insult while others may be more affected by environmental exposure (Vitale, Matigian et al. 2017). Therefore an interesting focus of future studies is the application of multimodal data science approaches to disentangle this biological heterogeneity and identify patient subgroups where illness risk is mainly driven by genetic or environmental effects. It is important to consider that environmental risk exposure likely occurs far in advance of illness onset, with previously described risk factors including famine, prenatal stress and maternal depression, as well as toxicological exposures (Kundakovic and Jaric 2017). Therefore, the biological manifestation of environmental risk effects may have spatiotemporal specificity and compound genetic risk in brain areas and developmental periods of particular relevance for schizophrenia, such as adolescence. Disentangling such effects, including the downstream impact of epigenetic effects on gene expression will form an important part of future studies aimed a providing a more in-depth characterization of gene-environment interactions in schizophrenia.

3.2 Spatiotemporal characterization of risk

The spatiotemporal analysis in Study 2 led to the identification of a set of genes significantly co-expressed with schizophrenia genetic susceptibility genes particularly in brain regions involving the prefrontal cortex during the fetal period and adolescence, two critical periods of vulnerability for schizophrenia (Selemon and Zecevic 2015). The genes were identified based on sex-dimorphisms of their expression variance, and the risk-associated genes showed a more prominent variation in males compared to females. These findings provide deeper insight into the biological basis of sex differences in schizophrenia and may underlie the frequently observed clinical differences in onset-age, clinical course and treatment response. The results are further consistent with previous genomics and transcriptomic studies that support the role of molecular perturbations related to early brain development for schizophrenia vulnerability (Gilman, Chang et al. 2012, Xu, Ionita-Laza et al. 2012, Birnbaum, Jaffe et al. 2014, Birnbaum, Jaffe et al. 2015, Jaffe, Straub et al. 2018, Clifton, Hannon et al. 2019). It is noteworthy that DNA methylation data from post-mortem cortical tissue across the lifespan has also supported this finding (Jaffe, Gao et al. 2016) and has aided in expanding knowledge of the biological basis underlying the neurodevelopmental component of schizophrenia.

Remarkably, the identified genes in Study 2 were strongly associated with synaptic processes in line with the pathway finding in Study 1, which is in accordance with the prominent and repeatedly described role of synaptic dysfunction in schizophrenia that is supported by numerous postmortem,

brain imaging, epidemiological and clinical studies (McGlashan and Hoffman 2000, Stephan, Baldeweg et al. 2006, Yin, Chen et al. 2012, Fromer, Roussos et al. 2016, Jaffe, Straub et al. 2018). Exposure to most environmental stressors linked to schizophrenia occurs long before the occurrence of clinical symptoms, supporting an early, brain functionally-relevant manifestation of biological risk that increases illness vulnerability during later life. In the context of synaptic disturbance, two possible mechanisms were proposed by (Faludi and Mirnics 2011): less synapse production or over-pruning. (I) The number of synapses is suppressed in early life such as during the fetal or postnatal periods at the time of insult, while the function of these synapses is still maintained during the 'pre-pruning' phase. However, during the late developmental pruning that occurs during late adolescence or early adulthood, the symptoms of schizophrenia appear due to an insufficient number of functioning synapses. (II) Alternatively, over-pruning that increases susceptibility for schizophrenia during late adolescence or early adulthood may have occurred during early life possibly due to the abnormal synapse functioning. We observed that the risk-associated gene sets with sex-dimorphisms in expression variability identified in Study 2 did not overlap between the fetal and adolescent periods but converged at a functional level to synaptic processes. This suggests that the biological basis of schizophrenia risk may be better understood and explored on a systems biology basis integrating spatiotemporally-specific effects.
We show that the identified genes co-expressed with schizophrenia susceptibility genes were enriched in the prefrontal cortical region, which is a well-studied area of particular relevance for the early developmental pathology of schizophrenia (Weinberger 1987, Selemon and Zecevic 2015, Birnbaum and Weinberger 2017). Several gene expression studies focussing on the prefrontal cortex have reported that schizophrenia-linked genes involved in transcriptional regulation during fetal life are significantly co-expressed with susceptibility genes or over-represented among schizophrenia susceptibility loci (Birnbaum, Jaffe et al. 2015, Jaffe, Straub et al. 2018). Similarly, schizophrenia-associated differential CpG sites identified in the prefrontal cortex have been found to be enriched at genetic risk loci (Pidsley, Viana et al. 2014, Jaffe, Gao et al. 2016).

3.3 Synaptic and immune pathways

The application of the BioMM procedure in Study 1 has allowed the identification of pathways that contribute most to the epigenetic schizophrenia classification which comprised four pathways related to neural and synaptic functions. Pathway and gene set enrichment analyses based on GWAS data from more than 60,000 subjects have previously identified multiple immune, neuronal signaling, and synaptic pathways as being most enriched for variants linked to schizophrenia (Network, O'Dushlaine et al. 2015). Another similar large-scale GWAS summary statistics-based study also pointed to pathways linked to synaptic dysregulation reported by (Schijven, Kofink et al. 2018). Furthermore, large-scale de novo mutation and copy number variation (CNV) studies have shown the enrichment of a synapse-related gene network (Fromer, Pocklington et al. 2014). These studies are in good agreement with the synaptic pathways identified here using machine learning models based on epigenetic data.
Notably, the top pathway from Study 1 was "negative regulation of NF-kappaB transcription factor activity", which is of critical relevance to the immune system. The regulation of nuclear factor kappaB (NF-kappaB) involves a family of transcription factors that are important for inflammation, immunity, and memory, as well as the nervous system (Oeckinghaus and Ghosh 2009, Dresselhaus and Meffert 2019). This finding may further support the assumed role of the immune system for schizophrenia

(Muller and J Schwarz 2010). The identification of immune-related pathways linked to epigenetic schizophrenia risk is further consistent with the most extensive common variant genetic study by the PGC that highlighted the enrichment of identified risk loci genes expressed in tissues with vital immune functions (Schizophrenia Working Group of the Psychiatric Genomics 2014), as well as with the pathway analysis from large GWAS data that pointed to schizophrenia-relevant immune processes (Network, O'Dushlaine et al. 2015). Moreover, the role of immunologic dysfunction is also supported by a variety of omics studies (van Mierlo, Schot et al. 2019) including transcriptomics (Fillman, Cloonan et al. 2013, Gardiner, Cairns et al. 2013, Mistry, Gillis et al. 2013, Sanders, Goring et al. 2013, Bergon, Belzeaux et al. 2015, Hess, Tylee et al. 2016, Sanders, Drigalenko et al. 2017, Duan, Goring et al. 2018, Kos, Duan et al. 2018, Leirer, Iyegbe et al. 2019), and proteomics (Schwarz, Guest et al. 2012, Schwarz, van Beveren et al. 2013), as well as epigenomics (Liu, Chen et al. 2013, Aberg, McClay et al. 2014, Hannon, Dempster et al. 2016). In line with this, a recent study has investigated the bidirectional relationship between schizophrenia and 19 immune-related diseases and identified extensively shared genetic risk factors (Pouget, Consortium et al. 2019). Interestingly, NF-kappaB is also known to be involved in synaptic plasticity, memory, and navigation (Snow, Stoesz et al. 2014), which may be linked to the abnormal working memory observed in schizophrenia (Forbes, Carrick et al. 2009).
Furthermore, the analysis of randomly selected pathway sets performed in the present work further supported the robustness and specificity of the identified, top-ranked pathways, suggesting that epigenetic changes in schizophrenia converge on a specific set of synaptic and immune-related processes.

3.4 Explainable machine learning models

Machine learning applied to omics data is receiving increasing attention in psychiatric research, partially fueled by the growing amount of available omics data. An important aspect of such models is their biological interpretability, which may be challenging when models are trained on high-dimensional data. In this thesis, the critical features of our devised machine learning framework BioMM are discussed here.
First, generalizability is one of the key components of successful machine learning approaches as the trained parameters must be applicable to samples outside of the training data. BioMM has a built-in resampling procedure (cross-validation) to estimate and select model parameters at the initial model training stage. In Study 1, nested cross-validation of BioMM provided a prediction accuracy of 0.78 in AUC, which indicates a reasonably high predictive value given the clinical heterogeneity of schizophrenia. More importantly, our model was independently validated in two other datasets. BioMM performance was additionally compared with an empirical null obtained using permutation testing and the result demonstrated that no random models outperformed the real one, strengthening the robustness of our model.
Second, most machine learning applications on epigenetic data for the classification of health status generally do not provide a clear understanding of which factors drive these classifications or predictions. For example, Capper and colleagues were able to yield high prediction accuracy using the random forest algorithm based on thousands of CpGs for tumor classification (Capper, Jones et al. 2018) but the most predictive underlying features were not identified. The advantages of applying explainable machine learning approaches to genome-wide data instead of the conventional analysis of single genes are demonstrated in study 1. A central idea of the BioMM model is that the

incorporation of pathway information into model construction improves the predictive value of the combined epigenetic changes. Due to the hierarchical procedure that first builds pathway-specific models and then integrates these using a meta-learner, BioMM also provides an intrinsic ranking of pathway features. Individual pathway features can further be analyzed using univariate statistics and used for other, biologically-informed downstream analyses.- Of particular interest in the context of work described here is unsupervised learning, where pathways-level features are used to identify subgroups of patients that show similar, systems-level alterations. For example, patients may be subtyped by either synaptic or immune-related pathways based on multiple omics data.

3.5 Limitations

The work described here highlights the utility of multivariate, multimodal approaches to explore biological determinants of schizophrenia using large-scale omics data. However, there are several limitations of the presented studies that are further detailed below.

One potential limitation of our work is the sample size. In Study 1, we did not directly investigate the relationship between PMS and DLPFC-HC coupling by including patient samples because only about 30 schizophrenia cases were available. This prevented meaningful analysis of epigenetic effects and the potentially stronger interactions with genetic risk in patients. In Study 2, since we adopted stratified analysis by brain regions and age windows, the samples falling into the respective strata are limited. However, the result of the donor-wide bootstrapping based resampling strategy supported the confidence in our findings.

The second limitation is the potential presence of residual confounding despite the stringent data preprocessing and quality control steps performed in both studies. One of the most relevant potential confounders in Study 1 is antipsychotic treatment, which could not be accounted for in the analysis. As a consequence, it cannot be excluded that PMS itself was partially influenced by medication effects. However, the fact that the association between the epigenetic signature and DLPFCH-HC connectivity was identified in unmedicated healthy controls supports the findings are relevant to schizophrenia. In Study 1, we corrected for available and relevant potential confounders (20 covariates in the blood sample, and 12 in post-mortem brain sample) to construct the PMS but found residual confounding of the PMS by the smoking score, as well as some cellular component variables. These effects were a likely consequence of the fact that the random forest machine learning model integrated non-linear confounding effects that could not be removed from the data using the employed linear adjustment procedure. However, the post-hoc association results were still significant after adjusting the PMS for all confounding variables, further supporting the illness-relevance of the identified epigenetic changes.

The third limitation is the likely influence of post-mortem effects as both studies relied on post-mortem brain data (either gene expression or DNA methylation). The biological and environmental processes induced by death, and other factors such as the post-mortem interval (PMI), the post-sampling handling are known to be possible confounders (Birdsill, Walker et al. 2011, Ferreira, Muñoz-Aguirre et al. 2018, Sjöholm, Ransome et al. 2018), and information on these was not always available for adjustment in our present work. In Study 1, we performed a cross-tissue prediction of the identified epigenetic signature, showing that the PMS identified from peripheral samples could also differentiate schizophrenia patients in data from post-mortem DLPFC samples.

This suggests that the PMS captured schizophrenia-relevant epigenetic differences present in the brain and circumvented the risk for the PMS to be confounded by post-mortem artifacts.
Fourth, the choice of targeted CpGs or genes for downstream analysis may need to be improved. In Study 2, we concentrated on schizophrenia risk loci mapping to single genes. Although this kind of selection may help prevent the statistical bias, it reduced the granted amount of genetic variability and it may not be representative of the schizophrenia risk architecture. Therefore, further studies with multiple susceptibility genes per locus are needed to account for possible residual bias. In Study 1, we selected genes harboring CpGs with an extended window of -20 kb upstream and +20 kb downstream to cover possible DNA-regulatory elements such as promoters and enhancers. The choice of this proposed window size was somewhat arbitrary but, currently no strategy can capture all variations within this window. Alternative parameters may be tested to further extend or shorten the gene boundary and CpGs falling into non-coding regions can also be informative.
Lastly, our proposed biologically informed machine learning approach is computationally intensive due to the repeated bootstrapping procedure or the need for optimal parameter selection, which may become a downside when it is routinely applied to large datasets. In addition to the use of clusters with higher performance, more efficient programming languages or environments (such as GPU programming with C++) will be of help to substantially reduce the runtime.

3.6 Future work

The merit of systems biology is to cover all the facets of biological processes from molecules to cells, circuits, and organs. The integration of multiple types of omics techniques, so-called multi-omics, has become increasingly appealing in the recent years (Hasin, Seldin et al. 2017) and may play an essential role in systematically studying biological systems involved in psychiatric illness. On the one hand, the multi-omics strategy can be more reflective of the complex biological mechanism of schizophrenia as it encompasses independent signals from multiple levels. It should be noted that each omics type has its own unique conceptualization. Various advantages and disadvantages of omics technologies including genomics, transcriptomics and proteomics, as well as metabolomics are elaborately summarized by (Karahalil 2016). Identified biomarkers would represent biological changes distributed across a variety of data modalities. On the other hand, the identified biomarkers in one data type are perhaps the consequence of alterations found in another correlated data type, and therefore not be the causative factors underlying schizophrenia.

Wang and colleagues (Wang, Chen et al. 2019) recently proposed a Bayesian framework named iRIGS (integrative Risk Gene Selector), which is able to infer schizophrenia risk genes driving GWAS signals by integrating relevant information from multi-omics data (i.e. epigenomics and transcriptomic data) and a gene-gene interaction network. Such a framework could estimate a set of risk genes that are primarily expressed in brain tissue, explain a significantly higher portion of heritability, and may pinpoint novel targets for already approved drugs. The adaptation of such integrative computational approaches to schizophrenia research will allow researchers to benefit more and more from the increasing availability of multi-omics data in schizophrenia (Ayalew, Le-Niculescu et al. 2012, Wang, Shi et al. 2019).

However, thus far only few studies have explored multi-omics data acquired on the same subject cohort, and even less utilizing machine learning. Machine learning is able to perform automated learning from pooled multi-omics data for personalized risk prediction as reviewed by (Li, Wu et al. 2016, Huang, Chaudhary et al. 2017, Lin and Lane 2017, Mirza, Wang et al. 2019). For example, an unsupervised machine learning framework entitled multi-omics factor analysis, or MOFA, was developed to detect biomarkers based on multi-omics data modalities from 200 chronic lymphocytic leukaemia patients (Argelaguet, Velten et al. 2018). The omics data consisting of four different modalities: DNA mutations, RNA expression, DNA methylation and ex vivo drug response were used as input and the latent factors derived from MOFA were able to capture variation in the data that was significantly associated with diagnosis. Besides advances in multimodal factor analysis, deep learning models (Angermueller, Pärnamaa et al. 2016) have been applied on a diversity of multi-omics data albeit not in schizophrenia (Chaudhary, Poirion et al. 2018, Zhang, Lv et al. 2018, Chung, Mirza et al. 2019, Huang, Zhan et al. 2019, Sharifi-Noghabi, Zolotareva et al. 2019). Increasing predictive accuracy through deep learning usually requires large sample numbers, which is challenging particularly in the multi-omics scenario. Therefore it can be expected that deep learning approaches for analysis of multi-omics data will particularly profit from the increased availability of large sample size, facilitated by the ever increasing degree of international collaboration and data sharing.

Another promising avenue for future work on biologically informed machine learning in schizophrenia is for the advanced analysis of time-dependent data. On the one hand, gene expression and epigenetic data is more dynamic than DNA sequence information and may capture more accurately the intricate and state-dependent etiological factors of the illness. A notable example is the well-described disturbance of biological rhythm in patients with schizophrenia. A

recent study (Seney, Cahill et al. 2019) reported that patients with schizophrenia show expression of a vastly different set of diurnally rhythmic genes using RNA sequencing data in the DLPFC compared to control subjects. These genes were enriched for mitochondrial functions and showed maximum expression during the day, which subsequently decreased during the night. This finding supported that time-dependent data analysis can provide a more precise characterization of molecular changes in schizophrenia. Consequently, data capturing biological information at potentially interesting time points, such as time-of-death, time-of-birth, or time of sleeping or wake-up, or even time of day when symptoms occurs may be integrated into machine learning modeling. The inclusion of such omics-based data from a longitudinal perspective is expected to increase the power of identifying risk mechanisms and also help the patient subtype identification. Furthermore, it may aid in characterizing the transdiagnostic specificity or pleiotropic effects when utilizing developmental trajectories in conjunction with data on psychiatric or somatic comorbidities.

On the other hand, the missing heritability present in genomic data may partially be addressed by non-omics data, such as that capturing environmental risk factors or imaging data. Integrating omics and non-omics data into machine learning pipelines may therefore yield substantial gains in predictive accuracies. Notably, digital data obtained from wearable devices (Tost, Reichert et al. 2019) that capture illness-relevant, environmental components along with neuroimaging readouts may reveal dynamic variations linked to susceptibility and pinpoint protective and modifiable factors. The acquisition of large-scale longitudinal samples needed for characterization of developmental trajectories using machine learning can be aided by advances in mobile technology, such as the 5G technology. 5G technology holds promise for unprecedented speed, coverage and low latency and its adaptation may revolutionize the depth of data acquired from smartphones or other wearables. It should further facilitate the real-time, quantitative feedback of susceptibility or treatment-related outcomes to doctors, patients or at-risk subjects through cloud computing assisted prediction, moving the psychiatric field closer towards personalized medicine and digital health solutions.

Therefore, the integration of multiple data modalities, combined with adaptation of advanced data science and technological developments may not only aid in better characterizing the complex biology underlying schizophrenia, but also facilitate the long-needed clinical translation of biological insight to improve patient outcomes and, hopefully, reduce incidence.

4 SUMMARY

Extensive efforts in characterizing the biological architecture of schizophrenia have moved psychiatric research closer towards clinical application. As our understanding of psychiatric illness is slowly shifting towards a conceptualization as dimensional constructs that cut across traditional diagnostic boundaries, opportunities for personalized medicine applications that are afforded by the application of advanced data science methods on the increasingly available, large-scale and multimodal data repositories are starting to be more broadly recognized. A particularly intriguing phenomenon is the discrepancy between the high heritability of schizophrenia and the difficulty in identifying predictive genetic signatures, for which polygenic risk scores of common variants that explain approximately 18% of illness-associated variance remain the gold standard. A substantial body of research points towards two lines of investigation that may lead to a significant advance, resolve at least in part the ‘missing heritability’ phenomenon, and potentially provide the basis for more predictive, personalized clinical tools.

First, it is paramount to better understand the impact of environmental factors on illness risk and elucidate the biology underlying their impact on altered brain function in schizophrenia. This thesis aims to close a major gap in our understanding of the multivariate, epigenetic landscape associated with schizophrenia, its interaction with polygenic risk and its association with DLPFC-HC connectivity, a well-established and robust neural intermediate phenotype of schizophrenia. As a basis for this, we have developed a novel biologically-informed machine learning framework by incorporating systems-level biological domain knowledge, i.e., gene ontological pathways, entitled ‘BioMM’ using genome-wide DNA methylation data obtained from whole blood samples. An epigenetic poly-methylation score termed ‘PMS’ was estimated at the individual level using BioMM, trained and validated using a total of 2230 whole-blood samples and 244 post-mortem brain samples. The pathways contributing most to this PMS were strongly associated with synaptic, neural and immune system-related functions. The identified PMS could be successfully validated in two independent cohorts, demonstrating the robust generalizability of the identified model. Furthermore, the PMS could significantly differentiate patients with schizophrenia from healthy controls when predicted in DLPFC post-mortem brain samples, suggesting that the epigenetic landscape of schizophrenia is to a certain extent shared between the central and peripheral tissues. Importantly, the peripheral PMS was associated with an intermediate neuroimaging phenotype (i.e., DLPFC-HC functional connectivity) in two independent imaging samples under the working memory paradigm. However, we did not find sufficient evidence for a combined genetic and epigenetic effect on brain function by integrating PRS derived from GWAS data, which suggested that DLPFC-HC coupling was predominantly impacted by environmental risk components, rather than polygenic risk of common variants. The epigenetic signature was further not associated with GWAS-derived risk scores implying the observed epigenetic effect did likely not depend on the underlying genetics, and this was further substantiated by investigation of data from unaffected first-degree relatives of patients with SCZ, BD, MDD and autism. In summary, the characterization of PMS through the systems-level integration of multimodal data elucidates the multivariate impact of epigenetic effects on schizophrenia-relevant brain function and its interdependence with genetic illness risk.

Second, the limited predictive value of polygenic risk scores and the difficulty in identifying associations with heritable neural differences found in schizophrenia may be due to the possibility that the manifestation of the functional consequences of genetic risk is modulated by spatio-temporal as well as sex-specific effects. To address this, this thesis identifies sex-differences in the spatio-temporal expression trajectories during human development of genes that showed significant prefrontal co-expression with schizophrenia risk genes during the fetal phase and adolescence, consistent with a core developmental hypothesis of schizophrenia. More specifically, it was found that during these two time-periods, prefrontal expression was significantly more variable in males compared to females, a finding that could be validated in an independent data source and that was specific for schizophrenia compared to other psychiatric as well as somatic illnesses. Similar to the epigenetic differences described above, the genes underlying the risk-associated gene expression differences were significantly linked to synaptic function. Notably, individual genes with male-specific variability increases were distinct between the fetal phase and adolescence, potentially suggesting different risk associated mechanisms that converge on the shared synaptic involvement of these genes. These results provide substantial support to the hypothesis that the functional consequences of genetic risk show spatiotemporal specificity. Importantly, the temporal specificity was linked to the fetal phase and adolescence, time-periods that are thought to be of predominant importance for the brain-functional consequences of environmental risk exposure. Therefore, the presented results provide the basis for future studies exploring the polygenic risk architecture and its interaction with environmental effects in a multivariate and spatiotemporally stratified manner.

In summary, the work presented in this thesis describes multivariate, multimodal approaches to characterize the (epi-)genetic basis of schizophrenia, explores its association with a well-established neural intermediate phenotype of the illness and investigates the spatio-temporal specificity of schizophrenia-relevant gene expression effects. This work expands our knowledge of the complex biology underlying schizophrenia and provides the basis for the future development of more predictive biological algorithms that may aid in advancing personalized medicine in psychiatry.

5 REFERENCE

Abel, K. M., R. Drake and J. M. Goldstein (2010). "Sex differences in schizophrenia." Int Rev Psychiatry **22**(5): 417-428.

Aberg, K. A., Y. Liu, J. Bukszar, J. L. McClay, A. N. Khachane, O. A. Andreassen, D. Blackwood, A. Corvin, S. Djurovic, H. Gurling, R. Ophoff, C. N. Pato, M. T. Pato, B. Riley, T. Webb, K. Kendler, M. O'Donovan, N. Craddock, G. Kirov, M. Owen, D. Rujescu, D. St Clair, T. Werge, C. M. Hultman, L. E. Delisi, P. Sullivan and E. J. van den Oord (2013). "A comprehensive family-based replication study of schizophrenia genes." JAMA Psychiatry **70**(6): 573-581.

Aberg, K. A., J. L. McClay, S. Nerella, S. Clark, G. Kumar, W. Chen, A. N. Khachane, L. Xie, A. Hudson and G. Gao (2014). "Methylome-wide association study of schizophrenia: identifying blood biomarker signatures of environmental insults." JAMA psychiatry **71**(3): 255-264.

Aberg, K. A., J. L. McClay, S. Nerella, S. Clark, G. Kumar, W. Chen, A. N. Khachane, L. Xie, A. Hudson, G. Gao, A. Harada, C. M. Hultman, P. F. Sullivan, P. K. Magnusson and E. J. van den Oord (2014). "Methylome-wide association study of schizophrenia: identifying blood biomarker signatures of environmental insults." JAMA Psychiatry **71**(3): 255-264.

Aberg, K. A., J. L. McClay, S. Nerella, L. Y. Xie, S. L. Clark, A. D. Hudson, J. Bukszár, D. Adkins, S. S. Consortium and C. M. Hultman (2012). "MBD-seq as a cost-effective approach for methylome-wide association studies: demonstration in 1500 case–control samples." Epigenomics **4**(6): 605-621.

Adams, M. D., J. M. Kelley, J. D. Gocayne, M. Dubnick, M. H. Polymeropoulos, H. Xiao, C. R. Merril, A. Wu, B. Olde, R. F. Moreno and et al. (1991). "Complementary DNA sequencing: expressed sequence tags and human genome project." Science **252**(5013): 1651-1656.

Alam, M. A., H. Y. Lin, H. W. Deng, V. D. Calhoun and Y. P. Wang (2018). "A kernel machine method for detecting higher order interactions in multimodal datasets: Application to schizophrenia." J Neurosci Methods **309**: 161-174.

Aleman, A., R. S. Kahn and J.-P. Selten (2003). "Sex differences in the risk of schizophrenia: evidence from meta-analysis." Archives of general psychiatry **60**(6): 565-571.

Aleman, A., R. S. Kahn and J. P. Selten (2003). "Sex differences in the risk of schizophrenia: evidence from meta-analysis." Arch Gen Psychiatry **60**(6): 565-571.

Alemu, E. Y., J. W. Carl, Jr., H. Corrada Bravo and S. Hannenhalli (2014). "Determinants of expression variability." Nucleic Acids Res **42**(6): 3503-3514.

Allen, N. C., S. Bagade, M. B. McQueen, J. P. Ioannidis, F. K. Kavvoura, M. J. Khoury, R. E. Tanzi and L. Bertram (2008). "Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database." Nat Genet **40**(7): 827-834.

Alnæs, D., T. Kaufmann, D. van der Meer, A. Córdova-Palomera, J. Rokicki, T. Moberget, F. Bettella, I. Agartz, D. M. Barch, A. Bertolino, C. L. Brandt, S. Cervenka, S. Djurovic, N. T. Doan, S. Eisenacher, H. Fatouros-Bergman, L. Flyckt, A. Di Giorgio, B. Haatveit, E. G. Jönsson, P. Kirsch, M. J. Lund, A. Meyer-Lindenberg, G. Pergola, E. Schwarz, O. B. Smeland, T. Quarto, M. Zink, O. A. Andreassen, L. T. Westlye and f. t. K. S. P. Consortium (2019). "Brain Heterogeneity in Schizophrenia and Its Association With Polygenic Risk." JAMA Psychiatry **76**(7): 739-748.

American Psychiatric Association (2013). "American Psychiatric Association DSM-5 Task Force." Diagnostic and statistical manual of mental disorders: DSM, 5.

Andreasen, N. C. (1995). "Symptoms, signs, and diagnosis of schizophrenia." Lancet **346**(8973): 477-481.

Andreasen, N. C. (2010). "The lifetime trajectory of schizophrenia and the concept of neurodevelopment." Dialogues in clinical neuroscience **12**(3): 409.

Andreassen, O. A., S. Djurovic, W. K. Thompson, A. J. Schork, K. S. Kendler, M. C. O'Donovan, D. Rujescu, T. Werge, M. van de Bunt, A. P. Morris, M. I. McCarthy, J. C. Roddey, L. K. McEvoy, R. S. Desikan and A. M. Dale (2013). "Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors." Am J Hum Genet **92**(2): 197-209.

Angermueller, C., T. Pärnamaa, L. Parts and O. Stegle (2016). "Deep learning for computational biology." Molecular systems biology **12**(7).

Argelaguet, R., B. Velten, D. Arnol, S. Dietrich, T. Zenz, J. C. Marioni, F. Buettner, W. Huber and O. Stegle (2018). "Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets." Molecular systems biology **14**(6).

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

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin and G. Sherlock (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." Nat Genet **25**(1): 25-29.

Association, A. P. (2013). "American Psychiatric Association DSM-5 Task Force." Diagnostic and statistical manual of mental disorders: DSM, 5.

Ayalew, M., H. Le-Niculescu, D. Levey, N. Jain, B. Changala, S. Patel, E. Winiger, A. Breier, A. Shekhar and R. Amdur (2012). "Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction." Molecular psychiatry **17**(9): 887.

Bahner, F., C. Demanuele, J. Schweiger, M. F. Gerchen, V. Zamoscik, K. Ueltzhoffer, T. Hahn, P. Meyer, H. Flor, D. Durstewitz, H. Tost, P. Kirsch, M. M. Plichta and A. Meyer-Lindenberg (2015). "Hippocampal-dorsolateral prefrontal coupling as a species-conserved cognitive mechanism: a human translational imaging study." Neuropsychopharmacology **40**(7): 1674-1681.

Bandettini, P. A. (2009). "What's new in neuroimaging methods?" Ann N Y Acad Sci **1156**: 260-293.

Bassett, A. S., E. W. Chow and R. Weksberg (2000). "Chromosomal abnormalities and schizophrenia." Am J Med Genet **97**(1): 45-51.

Bauer-Mehren, A., Laura I. Furlong, and Ferran Sanz. (2009). "Pathway databases and tools for their exploitation: benefits, current limitations and challenges." Molecular systems biology **5**(1).

Bergen, S. E., C. T. O'Dushlaine, S. Ripke, P. H. Lee, D. M. Ruderfer, S. Akterin, J. L. Moran, K. D. Chambert, R. E. Handsaker, L. Backlund, U. Osby, S. McCarroll, M. Landen, E. M. Scolnick, P. K. Magnusson, P. Lichtenstein, C. M. Hultman, S. M. Purcell, P. Sklar and P. F. Sullivan (2012). "Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder." Mol Psychiatry **17**(9): 880-886.

Bergon, A., R. Belzeaux, M. Comte, F. Pelletier, M. Hervé, E. J. Gardiner, N. J. Beveridge, B. Liu, V. Carr and R. J. Scott (2015). "CX3CR1 is dysregulated in blood and brain from schizophrenia patients." Schizophrenia research **168**(1-2): 434-443.

Bertolino, A., L. Fazio, G. Caforio, G. Blasi, A. Rampino, R. Romano, A. Di Giorgio, P. Taurisano, A. Papp and J. Pinsonneault (2008). "Functional variants of the dopamine receptor D2 gene modulate prefronto-striatal phenotypes in schizophrenia." Brain **132**(2): 417-425.

Birdsill, A. C., D. G. Walker, L. Lue, L. I. Sue and T. G. Beach (2011). "Postmortem interval effect on RNA and gene expression in human brain tissue." Cell and tissue banking **12**(4): 311-318.

Birnbaum, R., A. E. Jaffe, Q. Chen, T. M. Hyde, J. E. Kleinman and D. R. Weinberger (2015). "Investigation of the prenatal expression patterns of 108 schizophrenia-associated genetic loci." Biological psychiatry **77**(11): e43-e51.

Birnbaum, R., A. E. Jaffe, T. M. Hyde, J. E. Kleinman and D. R. Weinberger (2014). "Prenatal expression patterns of genes associated with neuropsychiatric disorders." American Journal of Psychiatry **171**(7): 758-767.

Birnbaum R, W. D. (2017). "Genetic insights into the neurodevelopmental origins of schizophrenia." Nat Rev Neurosci. **18**(12): 727-740.

Birnbaum, R. and D. R. Weinberger (2017). "Genetic insights into the neurodevelopmental origins of schizophrenia." Nat Rev Neurosci **18**(12): 727-740.

Bogdan, R., B. J. Salmeron, C. E. Carey, A. Agrawal, V. D. Calhoun, H. Garavan, A. R. Hariri, A. Heinz, M. N. Hill and A. Holmes (2017). "Imaging genetics and genomics in psychiatry: a critical review of progress and potential." Biological psychiatry **82**(3): 165-175.

Børglum, A. D., D. Demontis, J. Grove, J. Pallesen, M. V. Hollegaard, C. B. Pedersen, A. Hedemand, M. Mattheisen, G. investigators, A. Uitterlinden, M. Nyegaard, T. Ørntoft, C. Wiuf, M. Didriksen, M. Nordentoft, M. M. Nöthen, M. Rietschel, R. A. Ophoff, S. Cichon, R. H. Yolken, D. M. Hougaard, P. B. Mortensen and O. Mors (2013). "Genome-wide study of association and interaction with maternal cytomegalovirus infection suggests new schizophrenia loci." Molecular Psychiatry **19**: 325.

Braun, P. R., S. Han, B. Hing, Y. Nagahama, L. N. Gaul, J. T. Heinzman, A. J. Grossbach, L. Close, B. J. Dlouhy, M. A. Howard, 3rd, H. Kawasaki, J. B. Potash and G. Shinozaki (2019). "Genome-wide DNA methylation comparison between live human brain and peripheral tissues within individuals." Transl Psychiatry **9**(1): 47.

Breiman, L. (2001). "Random forests." Machine learning **45**(1): 5-32.

Brennand, K. J., A. Simone, J. Jou, C. Gelboin-Burkhart, N. Tran, S. Sangar, Y. Li, Y. Mu, G. Chen, D. Yu, S. McCarthy, J. Sebat and F. H. Gage (2011). "Modelling schizophrenia using human induced pluripotent stem cells." Nature **473**(7346): 221-225.

Brown, A. S. (2011). "The environment and susceptibility to schizophrenia." Prog Neurobiol **93**(1): 23-58.

Buckholtz, J. W., A. Meyer-Lindenberg, R. A. Honea, R. E. Straub, L. Pezawas, M. F. Egan, R. Vakkalanka, B. Kolachana, B. A. Verchinski and S. Sust (2007). "Allelic variation in RGS4 impacts functional and structural connectivity in the human brain." Journal of Neuroscience **27**(7): 1584-1593.

Buckley, P. F. and B. J. Miller (2017). "Personalized medicine for schizophrenia." npj Schizophrenia **3**(1): 2.

Buckley, P. F., B. J. Miller, D. S. Lehrer and D. J. Castle (2009). "Psychiatric comorbidities and schizophrenia." Schizophr Bull **35**(2): 383-402.

Bzdok, D. and A. Meyer-Lindenberg (2018). "Machine Learning for Precision Psychiatry: Opportunities and Challenges." Biol Psychiatry Cogn Neurosci Neuroimaging **3**(3): 223-230.

Callicott, J. H., R. E. Straub, L. Pezawas, M. F. Egan, V. S. Mattay, A. R. Hariri, B. A. Verchinski, A. Meyer-Lindenberg, R. Balkissoon and B. Kolachana (2005). "Variation in DISC1 affects hippocampal structure and function and increases risk for schizophrenia." Proceedings of the National Academy of Sciences **102**(24): 8627-8632.

Cannon, T. D. K., J. Lönnqvist, J. Huttunen, M. Koskenvuo, M. (1998). "The genetic epidemiology of schizophrenia in a Finnish twin cohort: a population-based modeling study." Archives of general psychiatry **55**(1): 67-74.

Cao, H., A. Bertolino, H. Walter, M. Schneider, A. Schafer, P. Taurisano, G. Blasi, L. Haddad, O. Grimm, K. Otto, L. Dixson, S. Erk, S. Mohnke, A. Heinz, N. Romanczuk-Seiferth, T. W. Muhleisen, M. Mattheisen, S. H. Witt, S. Cichon, M. Noethen, M. Rietschel, H. Tost and A. Meyer-Lindenberg (2016). "Altered Functional Subnetwork During Emotional Face Processing: A Potential Intermediate Phenotype for Schizophrenia." JAMA Psychiatry **73**(6): 598-605.

Cao, H., L. Dixson, A. Meyer-Lindenberg and H. Tost (2016). "Functional connectivity measures as schizophrenia intermediate phenotypes: advances, limitations, and future directions." Current opinion in neurobiology **36**: 7-14.

Capper, D., D. T. W. Jones, M. Sill, V. Hovestadt, D. Schrimpf, D. Sturm, C. Koelsche, F. Sahm, L. Chavez, D. E. Reuss, A. Kratz, A. K. Wefers, K. Huang, K. W. Pajtler, L. Schweizer, D. Stichel, A. Olar, N. W. Engel, K. Lindenberg, P. N. Harter, A. K. Braczynski, K. H. Plate, H. Dohmen, B. K. Garvalov, R. Coras, A. Holsken, E. Hewer, M. Bewerunge-Hudler, M. Schick, R. Fischer, R. Beschorner, J. Schittenhelm, O. Staszewski, K. Wani, P. Varlet, M. Pages, P. Temming, D. Lohmann, F. Selt, H. Witt, T. Milde, O. Witt, E. Aronica, F. Giangaspero, E. Rushing, W. Scheurlen, C. Geisenberger, F. J. Rodriguez, A. Becker, M. Preusser, C. Haberler, R. Bjerkvig, J. Cryan, M. Farrell, M. Deckert, J. Hench, S. Frank, J. Serrano, K. Kannan, A. Tsirigos, W. Bruck, S. Hofer, S. Brehmer, M. Seiz-Rosenhagen, D. Hanggi, V. Hans, S. Rozsnoki, J. R. Hansford, P. Kohlhof, B. W. Kristensen, M. Lechner, B. Lopes, C. Mawrin, R. Ketter, A. Kulozik, Z. Khatib, F. Heppner, A. Koch, A. Jouvet, C. Keohane, H. Muhleisen, W. Mueller, U. Pohl, M. Prinz, A. Benner, M. Zapatka, N. G. Gottardo, P. H. Driever, C. M. Kramm, H. L. Muller, S. Rutkowski, K. von Hoff, M. C. Fruhwald, A.

Gnekow, G. Fleischhack, S. Tippelt, G. Calaminus, C. M. Monoranu, A. Perry, C. Jones, T. S. Jacques, B. Radlwimmer, M. Gessi, T. Pietsch, J. Schramm, G. Schackert, M. Westphal, G. Reifenberger, P. Wesseling, M. Weller, V. P. Collins, I. Blumcke, M. Bendszus, J. Debus, A. Huang, N. Jabado, P. A. Northcott, W. Paulus, A. Gajjar, G. W. Robinson, M. D. Taylor, Z. Jaunmuktane, M. Ryzhova, M. Platten, A. Unterberg, W. Wick, M. A. Karajannis, M. Mittelbronn, T. Acker, C. Hartmann, K. Aldape, U. Schuller, R. Buslei, P. Lichter, M. Kool, C. Herold-Mende, D. W. Ellison, M. Hasselblatt, M. Snuderl, S. Brandner, A. Korshunov, A. von Deimling and S. M. Pfister (2018). "DNA methylation-based classification of central nervous system tumours." Nature **555**(7697): 469-474.

Cecil, C. A. M., E. Walton and E. Viding (2015). "DNA Methylation, Substance Use and Addiction: a Systematic Review of Recent Animal and Human Research from a Developmental Perspective." Current Addiction Reports **2**(4): 331-346.

Charlson, F., M. van Ommeren, A. Flaxman, J. Cornett, H. Whiteford and S. Saxena (2019). "New WHO prevalence estimates of mental disorders in conflict settings: a systematic review and meta-analysis." The Lancet.

Chaudhary, K., O. B. Poirion, L. Lu and L. X. Garmire (2018). "Deep learning–based multi-omics integration robustly predicts survival in liver cancer." Clinical Cancer Research **24**(6): 1248-1259.

Chen, C., C. Zhang, L. Cheng, J. L. Reilly, J. R. Bishop, J. A. Sweeney, H. Y. Chen, E. S. Gershon and C. Liu (2014). "Correlation between DNA methylation and gene expression in the brains of patients with bipolar disorder and schizophrenia." Bipolar Disord **16**(8): 790-799.

Chen, J., V. D. Calhoun, G. D. Pearlson, N. Perrone-Bizzozero, J. Sui, J. A. Turner, J. R. Bustillo, S. Ehrlich, S. R. Sponheim and J. M. Cañive (2013). "Guided exploration of genomic risk for gray matter abnormalities in schizophrenia using parallel independent component analysis with reference." NeuroImage **83**: 384-396.

Chen, J., D. Lippold, J. Frank, W. Rayner, A. Meyer-Lindenberg and E. Schwarz (2018). "Gimpute: An efficient genetic data imputation pipeline." Bioinformatics.

Chen, J. and E. Schwarz (2017). "BioMM: Biologically-informed Multi-stage Machine learning for identification of epigenetic fingerprints." arXiv preprint **arXiv:1712.00336.**

Chen, J. and E. Schwarz (2017). "BioMM: Biologically-informed Multi-stage Machine learning for identification of epigenetic fingerprints." arXiv: preprint arXiv:1712.00336.

Chen, J. and E. Schwarz (2017). "The role of blood-based biomarkers in advancing personalized therapy of schizophrenia." Expert Review of Precision Medicine and Drug Development **2**(6): 363-370.

Chen, J. and E. Schwarz (2019). "BioMM: Biological-informed Multi-stage Machine learning framework for phenotype prediction using omics data." **R package version 1.1.11.**

Chen, M., J. Cho and H. Zhao (2011). "Incorporating biological pathways via a Markov random field model in genome-wide association studies." PLoS Genet **7**(4): e1001353.

Chen, Q., G. Ursini, A. L. Romer, A. R. Knodt, K. Mezeivtch, E. Xiao, G. Pergola, G. Blasi, R. E. Straub and J. H. Callicott (2018). "Schizophrenia polygenic risk score predicts mnemonic hippocampal activity." Brain **141**(4): 1218-1228.

Chung, N. C., B. Mirza, H. Choi, J. Wang, D. Wang, P. Ping and W. Wang (2019). "Unsupervised classification of multi-omics data during cardiac remodeling using deep learning." Methods.

Clifton, N. E., E. Hannon, J. C. Harwood, A. Di Florio, K. L. Thomas, P. A. Holmans, J. T. R. Walters, M. C. O'Donovan, M. J. Owen, A. J. Pocklington and J. Hall (2019). "Dynamic expression of genes associated with schizophrenia and bipolar disorder across development." Transl Psychiatry **9**(1): 74.

Clifton, N. E., E. Hannon, J. C. Harwood, A. Di Florio, K. L. Thomas, P. A. Holmans, J. T. R. Walters, M. C. O'Donovan, M. J. Owen, A. J. Pocklington and J. Hall (2019). "Dynamic expression of genes associated with schizophrenia and bipolar disorder across development." Translational Psychiatry **9**(1): 74.

Clifton, N. E., E. s. Hannon, J. C. Harwood, A. Di Florio, K. L. Thomas, P. A. Holmans, J. T. Walters, M. C. O'Donovan, M. J. Owen and A. J. Pocklington (2019). "Dynamic expression of genes associated with schizophrenia and bipolar disorder across development." Translational psychiatry **9**(1): 74.

Colantuoni, C., B. K. Lipska, T. Ye, T. M. Hyde, R. Tao, J. T. Leek, E. A. Colantuoni, A. G. Elkahloun, M. M. Herman and D. R. Weinberger (2011). "Temporal dynamics and genetic control of transcription in the human prefrontal cortex." Nature **478**(7370): 519.

Colantuoni, C., B. K. Lipska, T. Ye, T. M. Hyde, R. Tao, J. T. Leek, E. A. Colantuoni, A. G. Elkahloun, M. M. Herman, D. R. Weinberger and J. E. Kleinman (2011). "Temporal dynamics and genetic control of transcription in the human prefrontal cortex." Nature **478**(7370): 519-523.

Consortium, C.-D. G. o. t. P. G. (2013). "Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs." Nature genetics **45**(9): 984.

Cortes, C. and V. Vapnik (1995). "Support-vector networks." Machine learning **20**(3): 273-297.

Cross-Disorder Group of the Psychiatric Genomics, C. (2013). "Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis." Lancet **381**(9875): 1371-1379.

Cun, Y. and H. Fröhlich (2012). "Prognostic gene signatures for patient stratification in breast cancer-accuracy, stability and interpretability of gene selection approaches using prior knowledge on protein-protein interactions." BMC bioinformatics **13**(1): 69.

Dainis, A. M. and E. A. Ashley (2018). "Cardiovascular Precision Medicine in the Genomics Era." JACC Basic Transl Sci **3**(2): 313-326.

Dalvie, S., N. Koen, N. McGregor, K. O'Connell, L. Warnich, R. Ramesar, C. M. Nievergelt and D. J. Stein (2016). "Toward a Global Roadmap for Precision Medicine in Psychiatry: Challenges and Opportunities." OMICS **20**(10): 557-564.

Datta, S. R., A. McQuillin, M. Rizig, E. Blaveri, S. Thirumalai, G. Kalsi, J. Lawrence, N. J. Bass, V. Puri, K. Choudhury, J. Pimm, C. Crombie, G. Fraser, N. Walker, D. Curtis, M. Zvelebil, A. Pereira, R. Kandaswamy, D. St Clair and H. M. Gurling (2010). "A threonine to isoleucine missense mutation in the pericentriolar material 1 gene is strongly associated with schizophrenia." Mol Psychiatry **15**(6): 615-628.

de Jong, S., M. P. Boks, T. F. Fuller, E. Strengman, E. Janson, C. G. de Kovel, A. P. Ori, N. Vi, F. Mulder, J. D. Blom, B. Glenthoj, C. D. Schubart, W. Cahn, R. S. Kahn, S. Horvath and R. A. Ophoff (2012). "A gene co-expression network in whole blood of schizophrenia patients is independent of antipsychotic-use and enriched for brain-expressed genes." PLoS One **7**(6): e39498.

De La Vega, F. M. and C. D. Bustamante (2018). "Polygenic risk scores: a biased prediction?" Genome Med **10**(1): 100.

de Leon, J. and F. J. Diaz (2005). "A meta-analysis of worldwide studies demonstrates an association between schizophrenia and tobacco smoking behaviors." Schizophr Res **76**(2-3): 135-157.

De Peri, L., A. Crescini, G. Deste, P. Fusar-Poli, E. Sacchetti and A. Vita (2012). "Brain structural abnormalities at the onset of schizophrenia and bipolar disorder: a meta-analysis of controlled magnetic resonance imaging studies." Current pharmaceutical design **18**(4): 486-494.

Dean, K. and R. M. Murray (2005). "Environmental risk factors for psychosis." Dialogues Clin Neurosci **7**(1): 69-80.

Delaneau, O., J. F. Zagury and J. Marchini (2013). "Improved whole-chromosome phasing for disease and population genetic studies." Nat Methods **10**(1): 5-6.

DeLisi, L. E. and W. W. Fleischhacker (2016). "How precise is precision medicine for schizophrenia?" Current opinion in psychiatry **29**(3): 187-189.

Dempster, E. L., R. Pidsley, L. C. Schalkwyk, S. Owens, A. Georgiades, F. Kane, S. Kalidindi, M. Picchioni, E. Kravariti, T. Toulopoulou, R. M. Murray and J. Mill (2011). "Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder." Human Molecular Genetics **20**(24): 4786-4796.

Dima, D. and G. Breen (2015). "Polygenic risk scores in imaging genetics: usefulness and applications." Journal of Psychopharmacology **29**(8): 867-871.

Dixson, L., H. Tost and A. Meyer-Lindenberg (2018). "Imaging Genetics." Psychiatric Genetics: A Primer for Clinical and Basic Scientists: 9.

Dresselhaus, E. C. and M. K. Meffert (2019). "Cellular Specificity of NF-κB Function in the Nervous System." Frontiers in immunology **10**: 1043-1043.

Duan, J., H. H. H. Goring, A. R. Sanders, W. Moy, J. Freda, E. I. Drigalenko, M. Kos, D. He, P. V. Gejman and Mgs (2018). "Transcriptomic signatures of schizophrenia revealed by dopamine perturbation in an ex vivo model." Transl Psychiatry **8**(1): 158.

Duan, J., A. R. Sanders, W. Moy, E. I. Drigalenko, E. C. Brown, J. Freda, C. Leites, H. H. Goring, Mgs and P. V. Gejman (2015). "Transcriptome outlier analysis implicates schizophrenia susceptibility genes and enriches putatively functional rare genetic variants." Hum Mol Genet **24**(16): 4674-4685.

Duncan, L. E. and M. C. Keller (2011). "A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry." American Journal of Psychiatry **168**(10): 1041-1049.

Durinck, S., P. T. Spellman, E. Birney and W. Huber (2009). "Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt." Nat Protoc **4**(8): 1184-1191.

Edgar, R., M. Domrachev and A. E. Lash (2002). "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository." Nucleic Acids Res **30**(1): 207-210.

Edgar, R. D., M. J. Jones, M. J. Meaney, G. Turecki and M. S. Kobor (2017). "BECon: a tool for interpreting DNA methylation findings from blood in the context of brain." Transl Psychiatry **7**(8): e1187.

Egan, M. F., T. E. Goldberg, B. S. Kolachana, J. H. Callicott, C. M. Mazzanti, R. E. Straub, D. Goldman and D. R. Weinberger (2001). "Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia." Proceedings of the National Academy of Sciences **98**(12): 6917-6922.

Elliott, H. R., T. Tillin, W. L. McArdle, K. Ho, A. Duggirala, T. M. Frayling, G. Davey Smith, A. D. Hughes, N. Chaturvedi and C. L. Relton (2014). "Differences in smoking associated DNA methylation patterns in South Asians and Europeans." Clin Epigenetics **6**(1): 4.

Erickson, K. I., M. W. Voss, R. S. Prakash, C. Basak, A. Szabo, L. Chaddock, J. S. Kim, S. Heo, H. Alves, S. M. White, T. R. Wojcicki, E. Mailey, V. J. Vieira, S. A. Martin, B. D. Pence, J. A. Woods, E. McAuley and A. F. Kramer (2011). "Exercise training increases size of hippocampus and improves memory." Proceedings of the National Academy of Sciences **108**(7): 3017-3022.

Erk, S., S. Mohnke, S. Ripke, T. A. Lett, I. M. Veer, C. Wackerhagen, O. Grimm, N. Romanczuk-Seiferth, F. Degenhardt, H. Tost, M. Mattheisen, T. W. Mühleisen, K. Charlet, N. Skarabis, F. Kiefer, S. Cichon, S. H. Witt, M. M. Nöthen, M. Rietschel, A. Heinz, A. Meyer-Lindenberg and H. Walter (2017). "Functional neuroimaging effects of recently discovered genetic risk loci for schizophrenia and polygenic risk profile in five RDoC subdomains." Transl Psychiatry **7**(1): e997.

Esslinger, C., H. Walter, P. Kirsch, S. Erk, K. Schnell, C. Arnold, L. Haddad, D. Mier, C. Opitz von Boberfeld, K. Raab, S. H. Witt, M. Rietschel, S. Cichon and A. Meyer-Lindenberg (2009). "Neural mechanisms of a genome-wide supported psychosis variant." Science **324**(5927): 605.

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

European Network of National Networks studying Gene-Environment Interactions in Schizophrenia, J. van Os, B. P. Rutten, I. Myin-Germeys, P. Delespaul, W. Viechtbauer, C. van Zelst, R. Bruggeman, U. Reininghaus, C. Morgan, R. M. Murray, M. Di Forti, P. McGuire, L. R. Valmaggia, M. J. Kempton, C. Gayer-Anderson, K. Hubbard, S. Beards, S. A. Stilo, A. Onyejiaka, F. Bourque, G. Modinos, S. Tognin, M. Calem, M. C. O'Donovan, M. J. Owen, P. Holmans, N. Williams, N. Craddock, A. Richards, I. Humphreys, A. Meyer-Lindenberg, F. M. Leweke, H. Tost, C. Akdeniz, C. Rohleder, J. M. Bumb, E. Schwarz, K. Alptekin, A. Ucok, M. C. Saka, E. C. Atbasoglu, S. Guloksuz, G. Gumus-Akay, B. Cihan, H. Karadag, H. Soygur, E. S. Cankurtaran, S. Ulusoy, B. Akdede, T. Binbay, A. Ayer, H. Noyan, G. Karadayi, E. Akturan, H. Ulas, C. Arango, M. Parellada, M. Bernardo, J. Sanjuan, J. Bobes, M. Arrojo, J. L. Santos, P. Cuadrado, J. J. Rodriguez Solano, A. Carracedo, E. Garcia Bernardo, L. Roldan, G. Lopez, B. Cabrera, S. Cruz, E. M. Diaz Mesa, M. Pouso, E. Jimenez, T. Sanchez, M. Rapado, E. Gonzalez, C. Martinez, E. Sanchez, M. S. Olmeda, L. de Haan, E. Velthorst, M. van der Gaag, J. P. Selten, D. van Dam, E. van der Ven, F. van der Meer, E. Messchaert, T. Kraan, N. Burger, M. Leboyer, A. Szoke, F. Schurhoff, P. M. Llorca, S. Jamain, A. Tortelli, F. Frijda, J. Vilain, A. M. Galliot, G. Baudin, A. Ferchiou, J. R. Richard, E. Bulzacka, T.

Charpeaud, A. M. Tronche, M. De Hert, R. van Winkel, J. Decoster, C. Derom, E. Thiery, N. C. Stefanis, G. Sachs, H. Aschauer, I. Lasser, B. Winklbaur, M. Schlogelhofer, A. Riecher-Rossler, S. Borgwardt, A. Walter, F. Harrisberger, R. Smieskova, C. Rapp, S. Ittig, F. Soguel-dit-Piquard, E. Studerus, J. Klosterkotter, S. Ruhrmann, J. Paruch, D. Julkowski, D. Hilboll, P. C. Sham, S. S. Cherny, E. Y. Chen, D. D. Campbell, M. Li, C. M. Romeo-Casabona, A. Emaldi Cirion, A. Urruela Mora, P. Jones, J. Kirkbride, M. Cannon, D. Rujescu, I. Tarricone, D. Berardi, E. Bonora, M. Seri, T. Marcacci, L. Chiri, F. Chierzi, V. Storbini, M. Braca, M. G. Minenna, I. Donegani, A. Fioritti, D. La Barbera, C. E. La Cascia, A. Mule, L. Sideli, R. Sartorio, L. Ferraro, G. Tripoli, F. Seminerio, A. M. Marinaro, P. McGorry, B. Nelson, G. P. Amminger, C. Pantelis, P. R. Menezes, C. M. Del-Ben, S. H. Gallo Tenan, R. Shuhama, M. Ruggeri, S. Tosato, A. Lasalvia, C. Bonetto, E. Ira, M. Nordentoft, M. O. Krebs, N. Barrantes-Vidal, P. Cristobal, T. R. Kwapil, E. Brietzke, R. A. Bressan, A. Gadelha, N. P. Maric, S. Andric, M. Mihaljevic and T. Mirjanic (2014). "Identifying gene-environment interactions in schizophrenia: contemporary challenges for integrated, large-scale investigations." Schizophr Bull **40**(4): 729-736.

Evans, G. W. and M. A. Schamberg (2009). "Childhood poverty, chronic stress, and adult working memory." Proc Natl Acad Sci U S A **106**(16): 6545-6549.

Fagiolini, M., C. L. Jensen and F. A. Champagne (2009). "Epigenetic influences on brain development and plasticity." Curr Opin Neurobiol **19**(2): 207-212.

Faludi, G. and K. Mirnics (2011). "Synaptic changes in the brain of subjects with schizophrenia." International Journal of Developmental Neuroscience **29**(3): 305-309.

Fatemi, S. H. and T. D. Folsom (2009). "The neurodevelopmental hypothesis of schizophrenia, revisited." Schizophr Bull **35**(3): 528-548.

Ferreira, P. G., M. Muñoz-Aguirre, F. Reverter, C. P. S. Godinho, A. Sousa, A. Amadoz, R. Sodaei, M. R. Hidalgo, D. Pervouchine and J. Carbonell-Caballero (2018). "The effects of death and post-mortem cold ischemia on human tissue transcriptomes." Nature communications **9**(1): 490.

Fikri, R. M. N., A. T. Norlelawati, A. R. N. El-Huda, M. N. Hanisah, A. Kartini, K. Norsidah and A. N. Zamzila (2017). "Reelin (RELN) DNA methylation in the peripheral blood of schizophrenia." Journal of psychiatric research **88**: 28-37.

Fillman, S. G., N. Cloonan, V. S. Catts, L. C. Miller, J. Wong, T. McCrossin, M. Cairns and C. S. Weickert (2013). "Increased inflammatory markers identified in the dorsolateral prefrontal cortex of individuals with schizophrenia." Mol Psychiatry **18**(2): 206-214.

Floresco, S. B., M. A. Geyer, L. H. Gold and A. A. Grace (2005). "Developing predictive animal models and establishing a preclinical trials network for assessing treatment effects on cognition in schizophrenia." Schizophrenia Bulletin **31**(4): 888-894.

Forbes, N., L. Carrick, A. McIntosh and S. Lawrie (2009). "Working memory in schizophrenia: a meta-analysis." Psychological medicine **39**(6): 889-905.

Fromer, M., A. J. Pocklington, D. H. Kavanagh, H. J. Williams, S. Dwyer, P. Gormley, L. Georgieva, E. Rees, P. Palta and D. M. Ruderfer (2014). "De novo mutations in schizophrenia implicate synaptic networks." Nature **506**(7487): 179.

Fromer, M., P. Roussos, S. K. Sieberts, J. S. Johnson, D. H. Kavanagh, T. M. Perumal, D. M. Ruderfer, E. C. Oh, A. Topol, H. R. Shah, L. L. Klei, R. Kramer, D. Pinto, Z. H. Gumus, A. E. Cicek, K. K. Dang, A. Browne, C. Lu, L. Xie, B. Readhead, E. A. Stahl, J. Xiao, M. Parvizi, T. Hamamsy, J. F. Fullard, Y. C. Wang, M. C. Mahajan, J. M. Derry, J. T. Dudley, S. E. Hemby, B. A. Logsdon, K. Talbot, T. Raj, D. A. Bennett, P. L. De Jager, J. Zhu, B. Zhang, P. F. Sullivan, A. Chess, S. M. Purcell, L. A. Shinobu, L. M. Mangravite, H. Toyoshiba, R. E. Gur, C. G. Hahn, D. A. Lewis, V. Haroutunian, M. A. Peters, B. K. Lipska, J. D. Buxbaum, E. E. Schadt, K. Hirai, K. Roeder, K. J. Brennand, N. Katsanis, E. Domenici, B. Devlin and P. Sklar (2016). "Gene expression elucidates functional impact of polygenic risk for schizophrenia." Nat Neurosci **19**(11): 1442-1453.

Fung, S. J., M. J. Webster, S. Sivagnanasundaram, C. Duncan, M. Elashoff and C. S. Weickert (2010). "Expression of interneuron markers in the dorsolateral prefrontal cortex of the developing human and in schizophrenia." American Journal of Psychiatry **167**(12): 1479-1488.

Gamazon, E. R., H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, A. E. Eyler, J. C. Denny, G. T. Consortium, D. L. Nicolae, N. J. Cox and H. K. Im (2015). "A gene-based

association method for mapping traits using reference transcriptome data." Nat Genet **47**(9): 1091-1098.

Gandal, M. J., V. Leppa, H. Won, N. N. Parikshak and D. H. Geschwind (2016). "The road to precision psychiatry: translating genetics into disease mechanisms." Nat Neurosci **19**(11): 1397-1407.

Gandal, M. J., P. Zhang, E. Hadjimichael, R. L. Walker, C. Chen, S. Liu, H. Won, H. Van Bakel, M. Varghese and Y. Wang (2018). "Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder." Science **362**(6420): eaat8127.

Garcia-Campos, M. A., J. Espinal-Enriquez and E. Hernandez-Lemus (2015). "Pathway Analysis: State of the Art." Front Physiol **6**: 383.

Gardiner, E. J., M. J. Cairns, B. Liu, N. J. Beveridge, V. Carr, B. Kelly, R. J. Scott and P. A. Tooney (2013). "Gene expression analysis reveals schizophrenia-associated dysregulation of immune pathways in peripheral blood mononuclear cells." Journal of psychiatric research **47**(4): 425-437.

Geiger, L. S., C. Moessnang, A. Schafer, Z. Zang, M. Zangl, H. Cao, T. R. van Raalten, A. Meyer-Lindenberg and H. Tost (2018). "Novelty modulates human striatal activation and prefrontal-striatal effective connectivity during working memory encoding." Brain Struct Funct **223**(7): 3121-3132.

Gejman, P. V., A. R. Sanders and K. S. Kendler (2011). "Genetics of schizophrenia: new findings and challenges." Annu Rev Genomics Hum Genet **12**: 121-144.

Gene Ontology, C., J. A. Blake, M. Dolan, H. Drabkin, D. P. Hill, N. Li, D. Sitnikov, S. Bridges, S. Burgess, T. Buza, F. McCarthy, D. Peddinti, L. Pillai, S. Carbon, H. Dietze, A. Ireland, S. E. Lewis, C. J. Mungall, P. Gaudet, R. L. Chrisholm, P. Fey, W. A. Kibbe, S. Basu, D. A. Siegele, B. K. McIntosh, D. P. Renfro, A. E. Zweifel, J. C. Hu, N. H. Brown, S. Tweedie, Y. Alam-Faruque, R. Apweiler, A. Auchinchloss, K. Axelsen, B. Bely, M. Blatter, C. Bonilla, L. Bouguerleret, E. Boutet, L. Breuza, A. Bridge, W. M. Chan, G. Chavali, E. Coudert, E. Dimmer, A. Estreicher, L. Famiglietti, M. Feuermann, A. Gos, N. Gruaz-Gumowski, R. Hieta, C. Hinz, C. Hulo, R. Huntley, J. James, F. Jungo, G. Keller, K. Laiho, D. Legge, P. Lemercier, D. Lieberherr, M. Magrane, M. J. Martin, P. Masson, P. Mutowo-Muellenet, C. O'Donovan, I. Pedruzzi, K. Pichler, D. Poggioli, P. Porras Millan, S. Poux, C. Rivoire, B. Roechert, T. Sawford, M. Schneider, A. Stutz, S. Sundaram, M. Tognolli, I. Xenarios, R. Foulgar, J. Lomax, P. Roncaglia, V. K. Khodiyar, R. C. Lovering, P. J. Talmud, M. Chibucos, M. G. Giglio, H. Chang, S. Hunter, C. McAnulla, A. Mitchell, A. Sangrador, R. Stephan, M. A. Harris, S. G. Oliver, K. Rutherford, V. Wood, J. Bahler, A. Lock, P. J. Kersey, D. M. McDowall, D. M. Staines, M. Dwinell, M. Shimoyama, S. Laulederkind, T. Hayman, S. Wang, V. Petri, T. Lowry, P. D'Eustachio, L. Matthews, R. Balakrishnan, G. Binkley, J. M. Cherry, M. C. Costanzo, S. S. Dwight, S. R. Engel, D. G. Fisk, B. C. Hitz, E. L. Hong, K. Karra, S. R. Miyasato, R. S. Nash, J. Park, M. S. Skrzypek, S. Weng, E. D. Wong, T. Z. Berardini, E. Huala, H. Mi, P. D. Thomas, J. Chan, R. Kishore, P. Sternberg, K. Van Auken, D. Howe and M. Westerfield (2013). "Gene Ontology annotations and resources." Nucleic Acids Res **41**(Database issue): D530-535.

Gibney, E. R. and C. M. Nolan (2010). "Epigenetics and gene expression." Heredity **105**(1): 4-13.

Gilman, S. R., J. Chang, B. Xu, T. S. Bawa, J. A. Gogos, M. Karayiorgou and D. Vitkup (2012). "Diverse types of genetic variation converge on functional gene networks involved in schizophrenia." Nature neuroscience **15**(12): 1723.

Gilman, S. R., J. Chang, B. Xu, T. S. Bawa, J. A. Gogos, M. Karayiorgou and D. Vitkup (2012). "Diverse types of genetic variation converge on functional gene networks involved in schizophrenia." Nat Neurosci **15**(12): 1723-1728.

Glatt, S. J., Chandler, S. D., Bousman, C. A., Chana, G., Lucero, G. R., Tatro, E., ... & Tsuang, M. T. (2009). "Alternatively Spliced Genes as Biomarkers for Schizophrenia, Bipolar Disorder and Psychosis: A Blood-Based Spliceome-Profiling Exploratory Study (Supplementry Table)." Current Pharmacogenomics and Personalized Medicine (Formerly Current Pharmacogenomics) **7**(3): 164-188.

Glausier, J. R. and D. A. Lewis (2013). "Dendritic spine pathology in schizophrenia." Neuroscience **251**: 90-107.

Goldstein, J. M., S. Cherkerzian, M. T. Tsuang and T. L. Petryshen (2013). "Sex differences in the genetic risk for schizophrenia: history of the evidence for sex-specific and sex-dependent effects." Am J Med Genet B Neuropsychiatr Genet **162B**(7): 698-710.

Good, P. (2013). Permutation tests: a practical guide to resampling methods for testing hypotheses, Springer Science & Business Media.

Gottesman, II and T. D. Gould (2003). "The endophenotype concept in psychiatry: etymology and strategic intentions." Am J Psychiatry **160**(4): 636-645.

Gottesman, II and J. Shields (1973). "Genetic theorizing and schizophrenia." Br J Psychiatry **122**(566): 15-30.

Goyal, M. S., M. Hawrylycz, J. A. Miller, A. Z. Snyder and M. E. Raichle (2014). "Aerobic glycolysis in the human brain is associated with development and neotenous gene expression." Cell Metab **19**(1): 49-57.

Grayson, D. R., X. Jia, Y. Chen, R. P. Sharma, C. P. Mitchell, A. Guidotti and E. Costa (2005). "Reelin promoter hypermethylation in schizophrenia." Proceedings of the National Academy of Sciences **102**(26): 9341-9346.

Greenwood, T. A., D. L. Braff, G. A. Light, K. S. Cadenhead, M. E. Calkins, D. J. Dobie, R. Freedman, M. F. Green, R. E. Gur, R. C. Gur, J. Mintz, K. H. Nuechterlein, A. Olincy, A. D. Radant, L. J. Seidman, L. J. Siever, J. M. Silverman, W. S. Stone, N. R. Swerdlow, D. W. Tsuang, M. T. Tsuang, B. I. Turetsky and N. J. Schork (2007). "Initial heritability analyses of endophenotypic measures for schizophrenia: the consortium on the genetics of schizophrenia." Arch Gen Psychiatry **64**(11): 1242-1250.

Gulsuner, S., T. Walsh, A. C. Watts, M. K. Lee, A. M. Thornton, S. Casadei, C. Rippey, H. Shahin, V. L. Nimgaonkar, R. C. Go, R. M. Savage, N. R. Swerdlow, R. E. Gur, D. L. Braff, M. C. King and J. M. McClellan (2013). "Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network." Cell **154**(3): 518-529.

Gusev, A., N. Mancuso, H. Won, M. Kousi, H. K. Finucane, Y. Reshef, L. Song, A. Safi, S. McCarroll, B. M. Neale, R. A. Ophoff, M. C. O'Donovan, G. E. Crawford, D. H. Geschwind, N. Katsanis, P. F. Sullivan, B. Pasaniuc and A. L. Price (2018). "Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights." Nat Genet **50**(4): 538-548.

Hafner, H. (1998). "Onset and course of the first schizophrenic episode." The Kaohsiung journal of medical sciences **14**(7): 413-431.

Häfner, H. (2019). "From Onset and Prodromal Stage to a Life-Long Course of Schizophrenia and Its Symptom Dimensions: How Sex, Age, and Other Risk Factors Influence Incidence and Course of Illness." Psychiatry journal **2019**.

Häfner, H. and W. an der Heiden (1997). "Epidemiology of schizophrenia." The Canadian Journal of Psychiatry **42**(2): 139-151.

Hall, J., H. C. Whalley, D. E. Job, B. J. Baig, A. M. McIntosh, K. L. Evans, P. A. Thomson, D. J. Porteous, D. G. Cunningham-Owens and E. C. Johnstone (2006). "A neuregulin 1 variant associated with abnormal cortical function and psychotic symptoms." Nature neuroscience **9**(12): 1477.

Hannon, E., E. Dempster, J. Viana, J. Burrage, A. R. Smith, R. Macdonald, D. St Clair, C. Mustard, G. Breen, S. Therman, J. Kaprio, T. Toulopoulou, H. E. Hulshoff Pol, M. M. Bohlken, R. S. Kahn, I. Nenadic, C. M. Hultman, R. M. Murray, D. A. Collier, N. Bass, H. Gurling, A. McQuillin, L. Schalkwyk and J. Mill (2016). "An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation." Genome Biol **17**(1): 176.

Hannon, E., K. Lunnon, L. Schalkwyk and J. Mill (2015). "Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes." Epigenetics **10**(11): 1024-1032.

Hannon, E., H. Spiers, J. Viana, R. Pidsley, J. Burrage, T. M. Murphy, C. Troakes, G. Turecki, M. C. O'Donovan, L. C. Schalkwyk, N. J. Bray and J. Mill (2015). "Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci." Nature Neuroscience **19**: 48.

Hansen, K. D., R. A. Irizarry and Z. Wu (2012). "Removing technical variability in RNA-seq data using conditional quantile normalization." Biostatistics **13**(2): 204-216.

Harrison, P. J. (2015). "Recent genetic findings in schizophrenia and their therapeutic relevance." J Psychopharmacol **29**(2): 85-96.

Harrison, P. J. and D. R. Weinberger (2005). "Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence." Mol Psychiatry **10**(1): 40-68; image 45.

Hasegawa, Y., D. Taylor, D. A. Ovchinnikov, E. J. Wolvetang, L. de Torrente and J. C. Mar (2015). "Variability of Gene Expression Identifies Transcriptional Regulators of Early Human Embryonic Development." PLoS Genet **11**(8): e1005428.

Hasin, Y., M. Seldin and A. Lusis (2017). "Multi-omics approaches to disease." Genome Biol **18**(1): 83.

Hawrylycz, M. J., E. S. Lein, A. L. Guillozet-Bongaarts, E. H. Shen, L. Ng, J. A. Miller, L. N. Van De Lagemaat, K. A. Smith, A. Ebbert and Z. L. Riley (2012). "An anatomically comprehensive atlas of the adult human brain transcriptome." Nature **489**(7416): 391.

Hess, J. L., D. S. Tylee, R. Barve, S. de Jong, R. A. Ophoff, N. Kumarasinghe, P. Tooney, U. Schall, E. Gardiner, N. J. Beveridge, R. J. Scott, S. Yasawardene, A. Perera, J. Mendis, V. Carr, B. Kelly, M. Cairns, U. Neurobehavioural Genetics, M. T. Tsuang and S. J. Glatt (2016). "Transcriptome-wide mega-analyses reveal joint dysregulation of immunologic genes and transcription regulators in brain and blood in schizophrenia." Schizophr Res **176**(2-3): 114-124.

Ho, B.-C., P. Milev, D. S. O'Leary, A. Librant, N. C. Andreasen and T. H. Wassink (2006). "Cognitive and magnetic resonance imaging brain morphometric correlates of brain-derived neurotrophic factor Val66Met gene polymorphism in patients with schizophrenia and healthy volunteers." Archives of general psychiatry **63**(7): 731-740.

Ho, J. W., M. Stefani, C. G. dos Remedios and M. A. Charleston (2008). "Differential variability analysis of gene expression and its application to human diseases." Bioinformatics **24**(13): i390-398.

Hoffman, G. E., B. J. Hartley, E. Flaherty, I. Ladran, P. Gochman, D. M. Ruderfer, E. A. Stahl, J. Rapoport, P. Sklar and K. J. Brennand (2017). "Transcriptional signatures of schizophrenia in hiPSC-derived NPCs and neurons are concordant with post-mortem adult brains." Nat Commun **8**(1): 2225.

Horvath, S. (2013). "DNA methylation age of human tissues and cell types." Genome Biol **14**(10): R115.

Horváth, S. and K. Mirnics (2015). "Schizophrenia as a disorder of molecular pathways." Biol Psychiatry **77**(1): 22-28.

Horwitz, T., K. Lam, Y. Chen, Y. Xia and C. Liu (2019). "A decade in psychiatric GWAS research." Molecular Psychiatry **24**(3): 378-389.

Howie, B., C. Fuchsberger, M. Stephens, J. Marchini and G. R. Abecasis (2012). "Fast and accurate genotype imputation in genome-wide association studies through pre-phasing." Nat Genet **44**(8): 955-959.

Howie, B. N., P. Donnelly and J. Marchini (2009). "A flexible and accurate genotype imputation method for the next generation of genome-wide association studies." PLoS Genet **5**(6): e1000529.

Huang da, W., B. T. Sherman and R. A. Lempicki (2009). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." Nat Protoc **4**(1): 44-57.

Huang, S., K. Chaudhary and L. X. Garmire (2017). "More is better: recent progress in multi-omics data integration methods." Frontiers in genetics **8**: 84.

Huang, Z., X. Zhan, S. Xiang, T. S. Johnson, B. Helm, C. Y. Yu, J. Zhang, P. Salama, M. Rizkalla and Z. Han (2019). "SALMON: Survival Analysis Learning With Multi-Omics Neural Networks on Breast Cancer." Frontiers in genetics **10**: 166.

Huckins, L. M., A. Dobbyn, D. M. Ruderfer, G. Hoffman, W. Wang, A. F. Pardiñas, V. M. Rajagopal, T. D. Als, H. T. Nguyen, K. Girdhar, J. Boocock, P. Roussos, M. Fromer, R. Kramer, E. Domenici, E. R. Gamazon, S. Purcell, J. S. Johnson, H. R. Shah, L. L. Klein, K. K. Dang, B. A. Logsdon, M. C. Mahajan, L. M. Mangravite, H. Toyoshiba, R. E. Gur, C.-G. Hahn, E. Schadt, D. A. Lewis, V. Haroutunian, M. A. Peters, B. K. Lipska, J. D. Buxbaum, K. Hirai, T. M. Perumal, L. Essioux, S. Ripke, B. M. Neale, A. Corvin, J. T. R. Walters, K.-H. Farh, P. A. Holmans, P. Lee, B. Bulik-Sullivan,

D. A. Collier, H. Huang, T. H. Pers, I. Agartz, E. Agerbo, M. Albus, M. Alexander, F. Amin, S. A. Bacanu, M. Begemann, R. A. Belliveau, J. Bene, S. E. Bergen, E. Bevilacqua, T. B. Bigdeli, D. W. Black, R. Bruggeman, N. G. Buccola, R. L. Buckner, W. Byerley, W. Cahn, G. Cai, D. Campion, R. M. Cantor, V. J. Carr, N. Carrera, S. V. Catts, K. D. Chambert, R. C. K. Chan, R. Y. L. Chen, E. Y. H. Chen, W. Cheng, E. F. C. Cheung, S. A. Chong, C. R. Cloninger, D. Cohen, N. Cohen, P. Cormican, N. Craddock, J. J. Crowley, D. Curtis, M. Davidson, K. L. Davis, F. Degenhardt, J. Del Favero, D. Demontis, D. Dikeos, T. Dinan, S. Djurovic, G. Donohoe, E. Drapeau, J. Duan, F. Dudbridge, N. Durmishi, P. Eichhammer, J. Eriksson, V. Escott-Price, L. Essioux, A. H. Fanous, M. S. Farrell, J. Frank, L. Franke, R. Freedman, N. B. Freimer, M. Friedl, J. I. Friedman, M. Fromer, G. Genovese, L. Georgieva, I. Giegling, P. Giusti-Rodríguez, S. Godard, J. I. Goldstein, V. Golimbet, S. Gopal, J. Gratten, L. de Haan, C. Hammer, M. L. Hamshere, M. Hansen, T. Hansen, V. Haroutunian, A. M. Hartmann, F. A. Henskens, S. Herms, J. N. Hirschhorn, P. Hoffmann, A. Hofman, M. V. Hollegaard, D. M. Hougaard, M. Ikeda, I. Joa, A. Julia, R. S. Kahn, L. Kalaydjieva, S. Karachanak-Yankova, J. Karjalainen, D. Kavanagh, M. C. Keller, J. L. Kennedy, A. Khrunin, Y. Kim, J. Klovins, J. A. Knowles, B. Konte, V. Kucinskas, Z. A. Kucinskiene, H. Kuzelova-Ptackova, A. K. Kahler, C. Laurent, J. L. C. Keong, S. H. Lee, S. E. Legge, B. Lerer, M. Li, T. Li, K.-Y. Liang, J. Lieberman, S. Limborska, C. M. Loughland, J. Lubinski, J. Lonnqvist, M. Macek, P. K. E. Magnusson, B. S. Maher, W. Maier, J. Mallet, S. Marsal, M. Mattheisen, M. Mattingsdal, R. W. McCarley, C. McDonald, A. M. McIntosh, S. Meier, C. J. Meijer, B. Melegh, I. Melle, R. I. Mesholam-Gately, A. Metspalu, P. T. Michie, L. Milani, V. Milanova, Y. Mokrab, D. W. Morris, O. Mors, K. C. Murphy, R. M. Murray, I. Myin-Germeys, B. Muller-Myhsok, M. Nelis, I. Nenadic, D. A. Nertney, G. Nestadt, K. K. Nicodemus, L. Nikitina-Zake, L. Nisenbaum, A. Nordin, E. O'Callaghan, C. O'Dushlaine, F. A. O'Neill, S.-Y. Oh, A. Olincy, L. Olsen, J. Van Os, C. Pantelis, G. N. Papadimitriou, S. Papiol, E. Parkhomenko, M. T. Pato, T. Paunio, M. Pejovic-Milovancevic, D. O. Perkins, O. Pietiläinen, J. Pimm, A. J. Pocklington, J. Powell, A. Price, A. E. Pulver, S. M. Purcell, D. Quested, H. B. Rasmussen, A. Reichenberg, M. A. Reimers, A. L. Richards, J. L. Roffman, P. Roussos, D. M. Ruderfer, V. Salomaa, A. R. Sanders, U. Schall, C. R. Schubert, T. G. Schulze, S. G. Schwab, E. M. Scolnick, R. J. Scott, L. J. Seidman, J. Shi, E. Sigurdsson, T. Silagadze, J. M. Silverman, K. Sim, P. Slominsky, J. W. Smoller, H.-C. So, C. C. A. Spencer, E. A. Stahl, H. Stefansson, S. Steinberg, E. Stogmann, R. E. Straub, E. Strengman, J. Strohmaier, T. S. Stroup, M. Subramaniam, J. Suvisaari, D. M. Svrakic, J. P. Szatkiewicz, E. Soderman, S. Thirumalai, D. Toncheva, S. Tosato, J. Veijola, J. Waddington, D. Walsh, D. Wang, Q. Wang, B. T. Webb, M. Weiser, D. B. Wildenauer, N. M. Williams, S. Williams, S. H. Witt, A. R. Wolen, E. H. M. Wong, B. K. Wormley, H. S. Xi, C. C. Zai, X. Zheng, F. Zimprich, N. R. Wray, K. Stefansson, P. M. Visscher, R. Adolfsson, O. A. Andreassen, D. H. R. Blackwood, E. Bramon, J. D. Buxbaum, A. D. Børglum, S. Cichon, A. Darvasi, E. Domenici, H. Ehrenreich, T. Esko, P. V. Gejman, C. CommonMind and C. The Schizophrenia Working Group of the Psychiatric Genomics (2019). "Gene expression imputation across multiple brain regions provides insights into schizophrenia risk." Nature Genetics **51**(4): 659-674.

Insel, T., B. Cuthbert, M. Garvey, R. Heinssen, D. S. Pine, K. Quinn, C. Sanislow and P. Wang (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders, Am Psychiatric Assoc.

International Schizophrenia, C., S. M. Purcell, N. R. Wray, J. L. Stone, P. M. Visscher, M. C. O'Donovan, P. F. Sullivan and P. Sklar (2009). "Common polygenic variation contributes to risk of schizophrenia and bipolar disorder." Nature **460**(7256): 748-752.

Iyegbe, C., D. Campbell, A. Butler, O. Ajnakina and P. Sham (2014). "The emerging molecular architecture of schizophrenia, polygenic risk scores and the clinical implications for GxE research." Social psychiatry and psychiatric epidemiology **49**(2): 169-182.

Jaenisch, R. and A. Bird (2003). "Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals." Nature genetics **33**(3): 245-254.

Jaffe, A. E., Y. Gao, A. Deep-Soboslay, R. Tao, T. M. Hyde, D. R. Weinberger and J. E. Kleinman (2016). "Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex." Nat Neurosci **19**(1): 40-47.

Jaffe, A. E., J. Shin, L. Collado-Torres, J. T. Leek, R. Tao, C. Li, Y. Gao, Y. Jia, B. J. Maher, T. M. Hyde, J. E. Kleinman and D. R. Weinberger (2015). "Developmental regulation of human cortex transcription and its clinical relevance at single base resolution." Nat Neurosci **18**(1): 154-161.

Jaffe, A. E., R. E. Straub, J. H. Shin, R. Tao, Y. Gao, L. Collado-Torres, T. Kam-Thong, H. S. Xi, J. Quan and Q. Chen (2018). "Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis." Nat Neurosci **21**(8): 1117-1125.

Jaffe, A. E., R. E. Straub, J. H. Shin, R. Tao, Y. Gao, L. Collado-Torres, T. Kam-Thong, H. S. Xi, J. Quan, Q. Chen, C. Colantuoni, W. S. Ulrich, B. J. Maher, A. Deep-Soboslay, C. BrainSeq, A. J. Cross, N. J. Brandon, J. T. Leek, T. M. Hyde, J. E. Kleinman and D. R. Weinberger (2018). "Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis." Nature neuroscience **21**(8): 1117-1125.

Jaffe, A. E., R. E. Straub, J. H. Shin, R. Tao, Y. Gao, L. Collado-Torres, T. Kam-Thong, H. S. Xi, J. Quan, Q. Chen, C. Colantuoni, W. S. Ulrich, B. J. Maher, A. Deep-Soboslay, C. BrainSeq, A. J. Cross, N. J. Brandon, J. T. Leek, T. M. Hyde, J. E. Kleinman and D. R. Weinberger (2018). "Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis." Nat Neurosci **21**(8): 1117-1125.

Janca, A., T. B. Ustun, T. S. Early and N. Sartorius (1993). "The ICD-10 symptom checklist: a companion to the ICD-10 classification of mental and behavioural disorders." Soc Psychiatry Psychiatr Epidemiol **28**(5): 239-242.

Jia, G., Y. Fu and C. He (2013). "Reversible RNA adenosine methylation in biological regulation." Trends Genet **29**(2): 108-115.

Jiang, F., Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen and Y. Wang (2017). "Artificial intelligence in healthcare: past, present and future." Stroke Vasc Neurol **2**(4): 230-243.

Jin, L., X. Y. Zuo, W. Y. Su, X. L. Zhao, M. Q. Yuan, L. Z. Han, X. Zhao, Y. D. Chen and S. Q. Rao (2014). "Pathway-based analysis tools for complex diseases: a review." Genomics Proteomics Bioinformatics **12**(5): 210-220.

Kakimi, K., T. Karasaki, H. Matsushita and T. Sugie (2017). "Advances in personalized cancer immunotherapy." Breast Cancer **24**(1): 16-24.

Kang, H. J., Y. I. Kawasawa, F. Cheng, Y. Zhu, X. Xu, M. Li, A. M. Sousa, M. Pletikos, K. A. Meyer, G. Sedmak, T. Guennel, Y. Shin, M. B. Johnson, Z. Krsnik, S. Mayer, S. Fertuzinhos, S. Umlauf, S. N. Lisgo, A. Vortmeyer, D. R. Weinberger, S. Mane, T. M. Hyde, A. Huttner, M. Reimers, J. E. Kleinman and N. Sestan (2011). "Spatio-temporal transcriptome of the human brain." Nature **478**(7370): 483-489.

Kang, T., W. Ding, L. Zhang, D. Ziemek and K. Zarringhalam (2017). "A biological network-based regularized artificial neural network model for robust phenotype prediction from gene expression data." BMC Bioinformatics **18**(1): 565.

Karahalil, B. (2016). "Overview of systems biology and omics technologies." Current medicinal chemistry **23**(37): 4221-4230.

Kaufmann, T., D. Alnaes, N. T. Doan, C. L. Brandt, O. A. Andreassen and L. T. Westlye (2017). "Delayed stabilization and individualization in connectome development are related to psychiatric disorders." Nat Neurosci **20**(4): 513-515.

Kauppi, K., L. T. Westlye, M. Tesli, F. Bettella, C. L. Brandt, M. Mattingsdal, T. Ueland, T. Espeseth, I. Agartz and I. Melle (2014). "Polygenic risk for schizophrenia associated with working memory-related prefrontal brain activation in patients with schizophrenia and healthy controls." Schizophrenia bulletin **41**(3): 736-743.

Kelly, T. K., D. D. De Carvalho and P. A. Jones (2010). "Epigenetic modifications as therapeutic targets." Nat Biotechnol **28**(10): 1069-1078.

Kendler, K. S. and S. R. Diehl (1993). "The genetics of schizophrenia: a current, genetic-epidemiologic perspective." Schizophr Bull **19**(2): 261-285.

Khatri, P., M. Sirota and A. J. Butte (2012). "Ten years of pathway analysis: current approaches and outstanding challenges." PLoS Comput Biol **8**(2): e1002375.

Kirkpatrick, B., R. W. Buchanan, D. E. Ross and W. T. Carpenter (2001). "A separate disease within the syndrome of schizophrenia." Archives of general psychiatry **58**(2): 165-171.

Kirov, G., E. Rees, J. T. Walters, V. Escott-Price, L. Georgieva, A. L. Richards, K. D. Chambert, G. Davies, S. E. Legge, J. L. Moran, S. A. McCarroll, M. C. O'Donovan and M. J. Owen (2014). "The penetrance of copy number variations for schizophrenia and developmental delay." Biol Psychiatry **75**(5): 378-385.

Klengel, T., J. Pape, E. B. Binder and D. Mehta (2014). "The role of DNA methylation in stress-related psychiatric disorders." Neuropharmacology **80**: 115-132.

Kordi-Tamandani, D. M., R. Sahranavard and A. Torkamanzehi (2012). "DNA methylation and expression profiles of the brain-derived neurotrophic factor (BDNF) and dopamine transporter (DAT1) genes in patients with schizophrenia." Molecular biology reports **39**(12): 10889-10893.

Kos, M. Z., J. Duan, A. R. Sanders, L. Blondell, E. I. Drigalenko, M. A. Carless, P. V. Gejman, H. H. H. Göring, P. V. Gejman, A. R. Sanders, J. Duan, D. F. Levinson, J. Shi, N. G. Buccola, B. J. Mowry, R. Freedman, A. Olincy, F. Amin, D. W. Black, J. M. Silverman, W. F. Byerley, C. R. Cloninger, D. M. Svrakic and Mgs (2018). "Dopamine perturbation of gene co-expression networks reveals differential response in schizophrenia for translational machinery." Translational Psychiatry **8**(1): 278.

Kotrla, M., K. J. and M. Daniel R. Weinberger (1995). "Brain imaging in schizophrenia." Annual review of medicine **46**(1): 113-122.

Krzyszczyk, P., A. Acevedo, E. J. Davidoff, L. M. Timmins, I. Marrero-Berrios, M. Patel, C. White, C. Lowe, J. J. Sherba, C. Hartmanshenn, K. M. O'Neill, M. L. Balter, Z. R. Fritz, I. P. Androulakis, R. S. Schloss and M. L. Yarmush (2018). "The growing role of precision and personalized medicine for cancer treatment." Technology (Singap World Sci) **6**(3-4): 79-100.

Kundakovic, M. and I. Jaric (2017). "The epigenetic link between prenatal adverse environments and neurodevelopmental disorders." Genes **8**(3): 104.

Laurent, L., E. Wong, G. Li, T. Huynh, A. Tsirigos, C. T. Ong, H. M. Low, K. W. Kin Sung, I. Rigoutsos, J. Loring and C. L. Wei (2010). "Dynamic changes in the human methylome during differentiation." Genome Res **20**(3): 320-331.

Lee, H., D. Dvorak, H. Y. Kao, A. M. Duffy, H. E. Scharfman and A. A. Fenton (2012). "Early cognitive experience prevents adult deficits in a neurodevelopmental schizophrenia model." Neuron **75**(4): 714-724.

Lee, S. H., S. Ripke, B. M. Neale, S. V. Faraone, S. M. Purcell, R. H. Perlis, B. J. Mowry, A. Thapar, M. E. Goddard and J. S. Witte (2013). "Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs." Nature genetics **45**(9): 984.

Leek, J. T. (2011). "Asymptotic conditional singular value decomposition for high-dimensional genomic data." Biometrics **67**(2): 344-352.

Leek, J. T. and J. D. Storey (2007). "Capturing heterogeneity in gene expression studies by surrogate variable analysis." PLoS Genet **3**(9): 1724-1735.

Leirer, D. J., C. O. Iyegbe, M. Di Forti, H. Patel, E. Carra, S. Fraietta, M. Colizzi, V. Mondelli, D. Quattrone and J. Lally (2019). "Differential gene expression analysis in blood of first episode psychosis patients." Schizophrenia research.

Lengauer, T., N. Pfeifer and R. Kaiser (2014). "Personalized HIV therapy to control drug resistance." Drug Discov Today Technol **11**: 57-64.

Lenzenweger, M. F. (2013). "Endophenotype, intermediate phenotype, biomarker: definitions, concept comparisons, clarifications." Depress Anxiety **30**(3): 185-189.

Leopold, J. A. and J. Loscalzo (2018). "Emerging Role of Precision Medicine in Cardiovascular Disease." Circ Res **122**(9): 1302-1315.

Leung, A. and P. Chue (2000). "Sex differences in schizophrenia, a review of the literature." Acta Psychiatr Scand Suppl **401**: 3-38.

Lewis, C. R., A. Henderson-Smith, R. S. Breitenstein, H. A. Sowards, I. S. Piras, M. J. Huentelman, L. D. Doane and K. Lemery-Chalfant (2019). "Dopaminergic gene methylation is associated with cognitive performance in a childhood monozygotic twin study." Epigenetics **14**(3): 310-323.

Lewis, D. A. and P. Levitt (2002). "Schizophrenia as a disorder of neurodevelopment." Annu Rev Neurosci **25**: 409-432.

Li, J., T. Cai, Y. Jiang, H. Chen, X. He, C. Chen, X. Li, Q. Shao, X. Ran and Z. Li (2016). "Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database." Molecular psychiatry **21**(2): 290.

Li, J., Y. Liu, T. Kim, R. Min and Z. Zhang (2010). "Gene expression variability within and between human populations and implications toward disease susceptibility." PLoS Comput Biol **6**(8).

Li, Y., F.-X. Wu and A. Ngom (2016). "A review on machine learning principles for multi-view biological data integration." Briefings in Bioinformatics **19**(2): 325-340.

Lim, U. and M. A. Song (2012). "Dietary and lifestyle factors of DNA methylation." Methods Mol Biol **863**: 359-376.

Lin, D., J. Chen, N. Perrone-Bizzozero, J. R. Bustillo, Y. Du, V. D. Calhoun and J. Liu (2018). "Characterization of cross-tissue genetic-epigenetic effects and their patterns in schizophrenia." Genome Medicine **10**(1): 13.

Lin, E. and H. Y. Lane (2017). "Machine learning and systems genomics approaches for multi-omics data." Biomark Res **5**: 2.

Lipska, B. K. and D. R. Weinberger (2000). "To model a psychiatric disorder in animals: schizophrenia as a reality test." Neuropsychopharmacology **23**(3): 223-239.

Lipska, B. K. and D. R. Weinberger (2000). "To model a psychiatric disorder in animals: schizophrenia as a reality test." Neuropsychopharmacology **23**(3): 223.

Lister, R., M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren and J. R. Ecker (2009). "Human DNA methylomes at base resolution show widespread epigenomic differences." Nature **462**(7271): 315-322.

Liu, B., X. Zhang, Y. Cui, W. Qin, Y. Tao, J. Li, C. Yu and T. Jiang (2017). "Polygenic risk for schizophrenia influences cortical gyrification in 2 independent general populations." Schizophrenia bulletin **43**(3): 673-680.

Liu, C., C. A. Bousman, C. Pantelis, E. Skafidas, D. Zhang, W. Yue and I. P. Everall (2017). "Pathway-wide association study identifies five shared pathways associated with schizophrenia in three ancestral distinct populations." Transl Psychiatry **7**(2): e1037.

Liu, J. and V. Calhoun (2014). "A review of multivariate analyses in imaging genetics." Frontiers in Neuroinformatics **8**(29).

Liu, J., J. Chen, S. Ehrlich, E. Walton, T. White, N. Perrone-Bizzozero, J. Bustillo, J. A. Turner and V. D. Calhoun (2013). "Methylation Patterns in Whole Blood Correlate With Symptoms in Schizophrenia Patients." Schizophrenia Bulletin **40**(4): 769-776.

Liu, J., K. Hutchison, N. Perrone-Bizzozero, M. Morgan, J. Sui and V. Calhoun (2010). "Identification of genetic and epigenetic marks involved in population structure." PLoS One **5**(10): e13209.

Lowe, R., N. Shirley, M. Bleackley, S. Dolan and T. Shafee (2017). "Transcriptomics technologies." PLoS Comput Biol **13**(5): e1005457.

Mailman, M. D., M. Feolo, Y. Jin, M. Kimura, K. Tryka, R. Bagoutdinov, L. Hao, A. Kiang, J. Paschall and L. Phan (2007). "The NCBI dbGaP database of genotypes and phenotypes." Nature genetics **39**(10): 1181.

Malhotra, D. and J. Sebat (2012). "CNVs: harbingers of a rare variant revolution in psychiatric genetics." Cell **148**(6): 1223-1241.

Malki, K., M. G. Tosto, H. Mouriño-Talín, S. Rodríguez-Lorenzo, O. Pain, I. Jumhaboy, T. Liu, P. Parpas, S. Newman, A. Malykh, L. Carboni, R. Uher, P. McGuffin, L. C. Schalkwyk, K. Bryson and M. Herbster (2017). "Highly polygenic architecture of antidepressant treatment response: Comparative analysis of SSRI and NRI treatment in an animal model of depression." American Journal of Medical Genetics Part B: Neuropsychiatric Genetics **174**(3): 235-250.

Mannuzza, S., A. J. Fyer, D. F. Klein and J. Endicott (1986). "Schedule for Affective Disorders and Schizophrenia--Lifetime Version modified for the study of anxiety disorders (SADS-LA): rationale and conceptual development." J Psychiatr Res **20**(4): 317-325.

Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, J. H. Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G.

Gibson, J. L. Haines, T. F. Mackay, S. A. McCarroll and P. M. Visscher (2009). "Finding the missing heritability of complex diseases." Nature **461**(7265): 747-753.

Marco, E. M., S. Macri and G. Laviola (2011). "Critical age windows for neurodevelopmental psychiatric disorders: evidence from animal models." Neurotox Res **19**(2): 286-307.

Maric, N., L. Krabbendam, W. Vollebergh, R. de Graaf and J. van Os (2003). "Sex differences in symptoms of psychosis in a non-selected, general population sample." Schizophr Res **63**(1-2): 89-95.

Markham, J. A. (2012). "Sex steroids and schizophrenia." Rev Endocr Metab Disord **13**(3): 187-207.

Martin, S. A., C. E. Dehler and E. Krol (2016). "Transcriptomic responses in the fish intestine." Dev Comp Immunol **64**: 103-117.

Maschietto, M., A. C. Tahira, R. Puga, L. Lima, D. Mariani, B. da Silveira Paulsen, P. Belmonte-de-Abreu, H. Vieira, A. C. V. Krepischi, D. M. Carraro, J. A. Palha, S. Rehen and H. Brentani (2015). "Co-expression network of neural-differentiation genes shows specific pattern in schizophrenia." BMC Medical Genomics **8**(1): 23.

McDonald, C., E. T. Bullmore, P. C. Sham, X. Chitnis, H. Wickham, E. Bramon and R. M. Murray (2004). "Association of genetic risks for schizophrenia and bipolar disorderwith specific and generic brain structural endophenotypes." Archives of general psychiatry **61**(10): 974-984.

McGlashan, T. H. and R. E. Hoffman (2000). "Schizophrenia as a disorder of developmentally reduced synaptic connectivity." Archives of general psychiatry **57**(7): 637-648.

McGlashan, T. H. and R. E. Hoffman (2000). "Schizophrenia as a disorder of developmentally reduced synaptic connectivity." Arch Gen Psychiatry **57**(7): 637-648.

McGrath, J., S. Saha, D. Chant and J. Welham (2008). "Schizophrenia: a concise overview of incidence, prevalence, and mortality." Epidemiol Rev **30**: 67-76.

McGuffin, P., A. E. Farmer, Gottesman, II, R. M. Murray and A. M. Reveley (1984). "Twin concordance for operationally defined schizophrenia. Confirmation of familiality and heritability." Arch Gen Psychiatry **41**(6): 541-545.

Meadows, J. P., M. C. Guzman-Karlsson, S. Phillips, C. Holleman, J. L. Posey, J. J. Day, J. J. Hablitz and J. D. Sweatt (2015). "DNA methylation regulates neuronal glutamatergic synaptic scaling." Sci Signal **8**(382): ra61.

Messias, E. L., C. Y. Chen and W. W. Eaton (2007). "Epidemiology of schizophrenia: review of findings and myths." Psychiatr Clin North Am **30**(3): 323-338.

Meyer-Lindenberg, A. (2010). "From maps to mechanisms through neuroimaging of schizophrenia." Nature **468**(7321): 194.

Meyer-Lindenberg, A. (2010). "From maps to mechanisms through neuroimaging of schizophrenia." Nature **468**(7321): 194-202.

Meyer-Lindenberg, A. (2010). "Imaging genetics of schizophrenia." Dialogues in clinical neuroscience **12**(4): 449.

Meyer-Lindenberg, A., R. E. Straub, B. K. Lipska, B. A. Verchinski, T. Goldberg, J. H. Callicott, M. F. Egan, S. S. Huffaker, V. S. Mattay and B. Kolachana (2007). "Genetic evidence implicating DARPP-32 in human frontostriatal structure, function, and cognition." The Journal of clinical investigation **117**(3): 672-682.

Meyer-Lindenberg, A. and D. R. Weinberger (2006). "Intermediate phenotypes and genetic mechanisms of psychiatric disorders." Nature Reviews Neuroscience **7**(10): 818.

Meyer-Lindenberg, A. S., R. K. Olsen, P. D. Kohn, T. Brown, M. F. Egan, D. R. Weinberger and K. F. Berman (2005). "Regionally specific disturbance of dorsolateral prefrontal-hippocampal functional connectivity in schizophrenia." Arch Gen Psychiatry **62**(4): 379-386.

Mill, J., T. Tang, Z. Kaminsky, T. Khare, S. Yazdanpanah, L. Bouchard, P. Jia, A. Assadzadeh, J. Flanagan, A. Schumacher, S. C. Wang and A. Petronis (2008). "Epigenomic profiling reveals DNA-methylation changes associated with major psychosis." Am J Hum Genet **82**(3): 696-711.

Millan, M. J., A. Andrieux, G. Bartzokis, K. Cadenhead, P. Dazzan, P. Fusar-Poli, J. Gallinat, J. Giedd, D. R. Grayson and M. Heinrichs (2016). "Altering the course of schizophrenia: progress and perspectives." Nature Reviews Drug Discovery **15**(7): 485.

Millan, M. J., G. M. Goodwin, A. Meyer-Lindenberg and S. O. Ögren (2015). "Learning from the past and looking to the future: emerging perspectives for improving the treatment of psychiatric disorders." European Neuropsychopharmacology **25**(5): 599-656.

Miller, J. A., S.-L. Ding, S. M. Sunkin, K. A. Smith, L. Ng, A. Szafer, A. Ebbert, Z. L. Riley, J. J. Royall and K. Aiona (2014). "Transcriptional landscape of the prenatal human brain." Nature **508**(7495): 199.

Miller, J. A., M. A. Scult, E. D. Conley, Q. Chen, D. R. Weinberger and A. R. Hariri (2017). "Effects of schizophrenia polygenic risk scores on brain activity and performance during working memory subprocesses in healthy young adults." Schizophrenia bulletin **44**(4): 844-853.

Miller, J. A., M. A. Scult, E. D. Conley, Q. Chen, D. R. Weinberger and A. R. Hariri (2018). "Effects of Schizophrenia Polygenic Risk Scores on Brain Activity and Performance During Working Memory Subprocesses in Healthy Young Adults." Schizophr Bull **44**(4): 844-853.

Mirza, B., W. Wang, J. Wang, H. Choi, N. C. Chung and P. Ping (2019). "Machine learning and integrative analysis of biomedical big data." Genes **10**(2): 87.

Mistry, M., J. Gillis and P. Pavlidis (2013). "Genome-wide expression profiling of schizophrenia using a large combined cohort." Mol Psychiatry **18**(2): 215-225.

Modinos, G., C. Iyegbe, D. Prata, M. Rivera, M. J. Kempton, L. R. Valmaggia, P. C. Sham, J. van Os and P. McGuire (2013). "Molecular genetic gene–environment studies using candidate genes in schizophrenia: a systematic review." Schizophrenia research **150**(2-3): 356-365.

Moghadam, B. T., M. Etemadikhah, G. Rajkowska, C. Stockmeier, M. Grabherr, J. Komorowski, L. Feuk and E. L. Carlström (2019). "Analyzing DNA methylation patterns in subjects diagnosed with schizophrenia using machine learning methods." Journal of psychiatric research **114**: 41-47.

Montano, C., M. A. Taub, A. Jaffe, E. Briem, J. I. Feinberg, R. Trygvadottir, A. Idrizi, A. Runarsson, B. Berndsen, R. C. Gur, T. M. Moore, R. T. Perry, D. Fugman, S. Sabunciyan, R. H. Yolken, T. M. Hyde, J. E. Kleinman, J. L. Sobell, C. N. Pato, M. T. Pato, R. C. Go, V. Nimgaonkar, D. R. Weinberger, D. Braff, R. E. Gur, M. D. Fallin and A. P. Feinberg (2016). "Association of DNA Methylation Differences With Schizophrenia in an Epigenome-Wide Association Study." JAMA Psychiatry **73**(5): 506-514.

Morgan VA, C. D., Jablensky AV. (2008). "Do women express and experience psychosis differently from men? Epidemiological evidence from the Australian National Study of Low Prevalence (Psychotic) Disorders." Aust N Z J Psychiatry **42**(1): 74-82.

Morris, S. E. and B. N. Cuthbert (2012). "Research Domain Criteria: cognitive systems, neural circuits, and dimensions of behavior." Dialogues in clinical neuroscience **14**(1): 29.

Mu, Y., S. Kodidela, Y. Wang, S. Kumar and T. J. Cory (2018). "The dawn of precision medicine in HIV: state of the art of pharmacotherapy." Expert Opin Pharmacother **19**(14): 1581-1595.

Mufford, M. S., D. J. Stein, S. Dalvie, N. A. Groenewold, P. M. Thompson and N. Jahanshad (2017). "Neuroimaging genomics in psychiatry—a translational approach." Genome medicine **9**(1): 1-12.

Muller, N. and M. J Schwarz (2010). "The role of immune system in schizophrenia." Current immunology reviews **6**(3): 213-220.

Murphy, B. C., R. L. O'Reilly and S. M. Singh (2005). "Site-specific cytosine methylation in S-COMT promoter in 31 brain regions with implications for studies involving schizophrenia." American Journal of Medical Genetics Part B: Neuropsychiatric Genetics **133**(1): 37-42.

Neilson, E., C. Bois, J. Gibson, B. Duff, A. Watson, N. Roberts, N. J. Brandon, J. Dunlop, J. Hall and A. M. McIntosh (2017). "Effects of environmental risks and polygenic loading for schizophrenia on cortical thickness." Schizophrenia research **184**: 128-136.

Nelder, J. A. and R. W. Wedderburn (1972). "Generalized Linear Models." Journal of the Royal Statistical Society: Series A (General) **135**(3): 370-384.

Network and C. Pathway Analysis Subgroup of Psychiatric Genomics (2015). "Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways." Nat Neurosci **18**(2): 199-209.

Network, T., C. O'Dushlaine, L. Rossin, P. H. Lee, L. Duncan, N. N. Parikshak, S. Newhouse, S. Ripke, B. M. Neale and S. M. Purcell (2015). "Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways." Nature neuroscience **18**(2): 199.

Nickl-Jockschat, T., H. Janouschek, S. B. Eickhoff and C. R. Eickhoff (2015). "Lack of meta-analytic evidence for an impact of COMT Val158Met genotype on brain activation during working memory tasks." Biological psychiatry **78**(11): e43-e46.

Nishioka, M., M. Bundo, K. Kasai and K. Iwamoto (2012). "DNA methylation in schizophrenia: progress and challenges of epigenetic studies." Genome Medicine **4**(12): 96.

Nishioka, M., M. Bundo, S. Koike, R. Takizawa, C. Kakiuchi, T. Araki, K. Kasai and K. Iwamoto (2013). "Comprehensive DNA methylation analysis of peripheral blood cells derived from patients with first-episode schizophrenia." Journal of human genetics **58**(2): 91.

Nishioka, M. B., M. Kasai, K. Iwamoto, K. (2012). "DNA methylation in schizophrenia: progress and challenges of epigenetic studies. ." Genome medicine **4**(12): 96.

O'Donovan, M. C. (2015). "What have we learned from the Psychiatric Genomics Consortium." World Psychiatry **14**(3): 291-293.

O'donovan, M. C., N. Craddock, N. Norton, H. Williams, T. Peirce, V. Moskvina, I. Nikolov, M. Hamshere, L. Carroll and L. Georgieva (2008). "Identification of loci associated with schizophrenia by genome-wide association and follow-up." Nature genetics **40**(9): 1053.

OECD and E. Union (2018). Health at a Glance: Europe 2018.

Oeckinghaus, A. and S. Ghosh (2009). "The NF-kappaB family of transcription factors and its regulation." Cold Spring Harbor perspectives in biology **1**(4): a000034-a000034.

Oldham, M. C., G. Konopka, K. Iwamoto, P. Langfelder, T. Kato, S. Horvath and D. H. Geschwind (2008). "Functional organization of the transcriptome in human brain." Nature neuroscience **11**(11): 1271.

Organization, W. H. (1992). "The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines." Geneva: World Health Organization.

Organization, W. H. (2002). "Genomics and world health: Report of the Advisory Committee on Health Research."

Orru, G., W. Pettersson-Yeo, A. F. Marquand, G. Sartori and A. Mechelli (2012). "Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review." Neurosci Biobehav Rev **36**(4): 1140-1152.

Osimo, E. F., K. Beck, T. Reis Marques and O. D. Howes (2019). "Synaptic loss in schizophrenia: a meta-analysis and systematic review of synaptic protein and mRNA measures." Mol Psychiatry **24**(4): 549-561.

Ottman, R. (1996). "Gene-environment interaction: definitions and study designs." Prev Med **25**(6): 764-770.

Owen, M. J. (2014). "New approaches to psychiatric diagnostic classification." Neuron **84**(3): 564-571.

Owen, M. J. and M. C. O'Donovan (2017). "Schizophrenia and the neurodevelopmental continuum: evidence from genomics." World Psychiatry **16**(3): 227-235.

Pacifico, R. and R. L. Davis (2017). "Transcriptome sequencing implicates dorsal striatum-specific gene network, immune response and energy metabolism pathways in bipolar disorder." Mol Psychiatry **22**(3): 441-449.

Pan, Q., T. Hu, J. D. Malley, A. S. Andrew, M. R. Karagas and J. H. Moore (2014). "A system-level pathway-phenotype association analysis using synthetic feature random forest." Genet Epidemiol **38**(3): 209-219.

Pardiñas, A. F., P. Holmans, A. J. Pocklington, V. Escott-Price, S. Ripke, N. Carrera, S. E. Legge, S. Bishop, D. Cameron and M. L. Hamshere (2018). "Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection." Nature genetics **50**(3): 381.

Paulus, F. M., J. Bedenbender, S. Krach, M. Pyka, A. Krug, J. Sommer, M. Mette, M. M. Nothen, S. H. Witt, M. Rietschel, T. Kircher and A. Jansen (2014). "Association of rs1006737 in CACNA1C with alterations in prefrontal activation and fronto-hippocampal connectivity." Hum Brain Mapp **35**(4): 1190-1200.

Peng, J., X. Wang and X. Shang (2019). "Combining gene ontology with deep neural networks to enhance the clustering of single cell RNA-Seq data." BMC Bioinformatics **20**(8): 284.

Pergola, G., P. Di Carlo, A. E. Jaffe, M. Papalino, Q. Chen, T. M. Hyde, J. E. Kleinman, J. H. Shin, A. Rampino and G. Blasi (2019). "Prefrontal Coexpression of Schizophrenia Risk Genes Is Associated With Treatment Response in Patients." Biological psychiatry.

Perlich, C. and G. Swirszcz (2011). "On cross-validation and stacking: Building seemingly predictive models on random data." ACM SIGKDD Explorations Newsletter **12**(2): 11-15.

Pfefferbaum, A. and R. B. Zipursky (1991). "Neuroimaging studies of schizophrenia." Schizophr Res **4**(2): 193-208.

Phillips, K. G., U. Bartsch, A. P. McCarthy, D. M. Edgar, M. D. Tricklebank, K. A. Wafford and M. W. Jones (2012). "Decoupling of sleep-dependent cortical and hippocampal interactions in a neurodevelopmental model of schizophrenia." Neuron **76**(3): 526-533.

Pidsley, R., J. Viana, E. Hannon, H. Spiers, C. Troakes, S. Al-Saraj, N. Mechawar, G. Turecki, L. C. Schalkwyk, N. J. Bray and J. Mill (2014). "Methylomic profiling of human brain tissue supports a neurodevelopmental origin for schizophrenia." Genome Biol **15**(10): 483.

Pinals, D. A., A. K. Malhotra, C. D. Missar, D. Pickar and A. Breier (1996). "Lack of gender differences in neuroleptic response in patients with schizophrenia." Schizophr Res **22**(3): 215-222.

Pouget, J. G., S. W. G. o. t. P. G. Consortium, B. Han, Y. Wu, E. Mignot, H. M. Ollila, J. Barker, S. Spain, N. Dand, R. Trembath, J. Martin, M. D. Mayes, L. Bossini-Castillo, E. López-Isac, Y. Jin, S. A. Santorico, R. A. Spritz, H. Hakonarson, C. Polychronakos, S. Raychaudhuri and J. Knight (2019). "Cross-disorder analysis of schizophrenia and 19 immune-mediated diseases identifies shared genetic risk." Human Molecular Genetics.

Pries, L.-K., S. Gülöksüz and G. Kenis (2017). DNA methylation in schizophrenia. Neuroepigenomics in Aging and Disease, Springer**:** 211-236.

Purcell, S. M., J. L. Moran, M. Fromer, D. Ruderfer, N. Solovieff, P. Roussos, C. O'Dushlaine, K. Chambert, S. E. Bergen, A. Kahler, L. Duncan, E. Stahl, G. Genovese, E. Fernandez, M. O. Collins, N. H. Komiyama, J. S. Choudhary, P. K. Magnusson, E. Banks, K. Shakir, K. Garimella, T. Fennell, M. DePristo, S. G. Grant, S. J. Haggarty, S. Gabriel, E. M. Scolnick, E. S. Lander, C. M. Hultman, P. F. Sullivan, S. A. McCarroll and P. Sklar (2014). "A polygenic burden of rare disruptive mutations in schizophrenia." Nature **506**(7487): 185-190.

Purcell, S. M., N. R. Wray, J. L. Stone, P. M. Visscher, M. C. O'Donovan, P. F. Sullivan and P. Sklar (2009). "Common polygenic variation contributes to risk of schizophrenia and bipolar disorder." Nature **460**(7256): 748-752.

Qin, W., C. Liu, M. Sodhi and H. Lu (2016). "Meta-analysis of sex differences in gene expression in schizophrenia." BMC Syst Biol **10 Suppl 1**(Suppl 1): 9.

Ramaker, R. C., K. M. Bowling, B. N. Lasseigne, M. H. Hagenauer, A. A. Hardigan, N. S. Davis, J. Gertz, P. M. Cartagena, D. M. Walsh, M. P. Vawter, E. G. Jones, A. F. Schatzberg, J. D. Barchas, S. J. Watson, B. G. Bunney, H. Akil, W. E. Bunney, J. Z. Li, S. J. Cooper and R. M. Myers (2017). "Post-mortem molecular profiling of three psychiatric disorders." Genome Med **9**(1): 72.

Ramchandani, S., S. K. Bhattacharya, N. Cervoni and M. Szyf (1999). "DNA methylation is a reversible biological signal." Proc Natl Acad Sci U S A **96**(11): 6107-6112.

Ramsey, J. M., E. Schwarz, P. C. Guest, N. J. van Beveren, F. M. Leweke, M. Rothermundt, B. Bogerts, J. Steiner and S. Bahn (2013). "Distinct molecular phenotypes in male and female schizophrenia patients." PLoS One **8**(11): e78729.

Ran, D. and Z. J. Daye (2017). "Gene expression variability and the analysis of large-scale RNA-seq studies with the MDSeq." Nucleic Acids Res **45**(13): e127.

Rapoport, J. L., J. N. Giedd and N. Gogtay (2012). "Neurodevelopmental model of schizophrenia: update 2012." Mol Psychiatry **17**(12): 1228-1238.

Raser, J. M. and E. K. O'Shea (2004). "Control of stochasticity in eukaryotic gene expression." Science **304**(5678): 1811-1814.

Raser, J. M. and E. K. O'Shea (2005). "Noise in gene expression: origins, consequences, and control." Science **309**(5743): 2010-2013.

Rasetti, R., F. Sambataro, Q. Chen, J. H. Callicott, V. S. Mattay and D. R. Weinberger (2011). "Altered cortical network dynamics: a potential intermediate phenotype for schizophrenia and association with ZNF804A." Arch Gen Psychiatry **68**(12): 1207-1217.

Rasetti, R. and D. R. Weinberger (2011). "Intermediate phenotypes in psychiatric disorders." Current opinion in genetics & development **21**(3): 340-348.

Reichstein, M., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler and N. Carvalhais (2019). "Deep learning and process understanding for data-driven Earth system science." Nature **566**(7743): 195-204.

Ribeiro, M. T., S. Singh and C. Guestrin (2016). Why should i trust you?: Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, ACM.

Richfield, O., M. A. Alam, V. Calhoun and Y.-P. Wang (2016). Learning schizophrenia imaging genetics data via multiple kernel canonical correlation analysis. 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE.

Ripke, S., C. O'Dushlaine, K. Chambert, J. L. Moran, A. K. Kähler, S. Akterin, S. E. Bergen, A. L. Collins, J. J. Crowley and M. Fromer (2013). "Genome-wide association analysis identifies 13 new risk loci for schizophrenia." Nature genetics **45**(10): 1150.

Ripke, S., C. O'Dushlaine, K. Chambert, J. L. Moran, A. K. Kahler, S. Akterin, S. E. Bergen, A. L. Collins, J. J. Crowley, M. Fromer, Y. Kim, S. H. Lee, P. K. Magnusson, N. Sanchez, E. A. Stahl, S. Williams, N. R. Wray, K. Xia, F. Bettella, A. D. Borglum, B. K. Bulik-Sullivan, P. Cormican, N. Craddock, C. de Leeuw, N. Durmishi, M. Gill, V. Golimbet, M. L. Hamshere, P. Holmans, D. M. Hougaard, K. S. Kendler, K. Lin, D. W. Morris, O. Mors, P. B. Mortensen, B. M. Neale, F. A. O'Neill, M. J. Owen, M. P. Milovancevic, D. Posthuma, J. Powell, A. L. Richards, B. P. Riley, D. Ruderfer, D. Rujescu, E. Sigurdsson, T. Silagadze, A. B. Smit, H. Stefansson, S. Steinberg, J. Suvisaari, S. Tosato, M. Verhage, J. T. Walters, C. Multicenter Genetic Studies of Schizophrenia, D. F. Levinson, P. V. Gejman, K. S. Kendler, C. Laurent, B. J. Mowry, M. C. O'Donovan, M. J. Owen, A. E. Pulver, B. P. Riley, S. G. Schwab, D. B. Wildenauer, F. Dudbridge, P. Holmans, J. Shi, M. Albus, M. Alexander, D. Campion, D. Cohen, D. Dikeos, J. Duan, P. Eichhammer, S. Godard, M. Hansen, F. B. Lerer, K. Y. Liang, W. Maier, J. Mallet, D. A. Nertney, G. Nestadt, N. Norton, F. A. O'Neill, G. N. Papadimitriou, R. Ribble, A. R. Sanders, J. M. Silverman, D. Walsh, N. M. Williams, B. Wormley, C. Psychosis Endophenotypes International, M. J. Arranz, S. Bakker, S. Bender, E. Bramon, D. Collier, B. Crespo-Facorro, J. Hall, C. Iyegbe, A. Jablensky, R. S. Kahn, L. Kalaydjieva, S. Lawrie, C. M. Lewis, K. Lin, D. H. Linszen, I. Mata, A. McIntosh, R. M. Murray, R. A. Ophoff, J. Powell, D. Rujescu, J. Van Os, M. Walshe, M. Weisbrod, D. Wiersma, C. Wellcome Trust Case Control, P. Donnelly, I. Barroso, J. M. Blackwell, E. Bramon, M. A. Brown, J. P. Casas, A. P. Corvin, P. Deloukas, A. Duncanson, J. Jankowski, H. S. Markus, C. G. Mathew, C. N. Palmer, R. Plomin, A. Rautanen, S. J. Sawcer, R. C. Trembath, A. C. Viswanathan, N. W. Wood, C. C. Spencer, G. Band, C. Bellenguez, C. Freeman, G. Hellenthal, E. Giannoulatou, M. Pirinen, R. D. Pearson, A. Strange, Z. Su, D. Vukcevic, P. Donnelly, C. Langford, S. E. Hunt, S. Edkins, R. Gwilliam, H. Blackburn, S. J. Bumpstead, S. Dronov, M. Gillman, E. Gray, N. Hammond, A. Jayakumar, O. T. McCann, J. Liddle, S. C. Potter, R. Ravindrarajah, M. Ricketts, A. Tashakkori-Ghanbaria, M. J. Waller, P. Weston, S. Widaa, P. Whittaker, I. Barroso, P. Deloukas, C. G. Mathew, J. M. Blackwell, M. A. Brown, A. P. Corvin, M. I. McCarthy, C. C. Spencer, E. Bramon, A. P. Corvin, M. C. O'Donovan, K. Stefansson, E. Scolnick, S. Purcell, S. A. McCarroll, P. Sklar, C. M. Hultman and P. F. Sullivan (2013). "Genome-wide association analysis identifies 13 new risk loci for schizophrenia." Nat Genet **45**(10): 1150-1159.

Risch, N. and K. Merikangas (1996). "The future of genetic studies of complex human diseases." Science **273**(5281): 1516-1517.

Roscher, R., B. Bohn, M. F. Duarte and J. Garcke (2019). "Explainable Machine Learning for Scientific Insights and Discoveries." arXiv preprint arXiv:1905.08883.

Roussos, P., B. Guennewig, D. C. Kaczorowski, G. Barry and K. J. Brennand (2016). "Activity-Dependent Changes in Gene Expression in Schizophrenia Human-Induced Pluripotent Stem Cell Neurons." JAMA Psychiatry **73**(11): 1180-1188.

Roussos, P., P. Katsel, K. L. Davis, S. G. Giakoumaki, T. Lencz, A. K. Malhotra, L. J. Siever, P. Bitsios and V. Haroutunian (2013). "Convergent findings for abnormalities of the NF-kappaB signaling pathway in schizophrenia." Neuropsychopharmacology **38**(3): 533-539.

Roussos, P., P. Katsel, K. L. Davis, L. J. Siever and V. Haroutunian (2012). "A system-level transcriptomic analysis of schizophrenia using postmortem brain tissue samples." Arch Gen Psychiatry **69**(12): 1205-1213.

Saeys, Y., I. Inza and P. Larranaga (2007). "A review of feature selection techniques in bioinformatics." Bioinformatics **23**(19): 2507-2517.

Saha, S., D. Chant, J. Welham and J. McGrath (2005). "A systematic review of the prevalence of schizophrenia." PLoS Med **2**(5): e141.

Sanders, A. R., E. I. Drigalenko, J. Duan, W. Moy, J. Freda, H. H. H. Goring and P. V. Gejman (2017). "Transcriptome sequencing study implicates immune-related genes differentially expressed in schizophrenia: new data and a meta-analysis." Transl Psychiatry **7**(4): e1093.

Sanders, A. R., H. H. Goring, J. Duan, E. I. Drigalenko, W. Moy, J. Freda, D. He, J. Shi, Mgs and P. V. Gejman (2013). "Transcriptome study of differential expression in schizophrenia." Hum Mol Genet **22**(24): 5001-5014.

Saunders, R. C., B. S. Kolachana, J. Bachevalier and D. R. Weinberger (1998). "Neonatal lesions of the medial temporal lobe disrupt prefrontal cortical regulation of striatal dopamine." Nature **393**(6681): 169.

Schijven, D., D. Kofink, V. Tragante, M. Verkerke, S. L. Pulit, R. S. Kahn, J. H. Veldink, C. H. Vinkers, M. P. Boks and J. J. Luykx (2018). "Comprehensive pathway analyses of schizophrenia risk loci point to dysfunctional postsynaptic signaling." Schizophrenia research **199**: 195-202.

Schizophrenia Working Group of the Psychiatric Genomics, C. (2014). "Biological insights from 108 schizophrenia-associated genetic loci." Nature **511**(7510): 421-427.

Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). "Biological insights from 108 schizophrenia-associated genetic loci." Nature **511**(7510): 421-427.

Schneider, M., H. Walter, C. Moessnang, A. Schafer, S. Erk, S. Mohnke, L. Romund, M. Garbusow, L. Dixson, A. Heinz, N. Romanczuk-Seiferth, A. Meyer-Lindenberg and H. Tost (2017). "Altered DLPFC-Hippocampus Connectivity During Working Memory: Independent Replication and Disorder Specificity of a Putative Genetic Risk Phenotype for Schizophrenia." Schizophr Bull **43**(5): 1114-1122.

Schork, N. J. (2015). "Personalized medicine: Time for one-person trials." Nature **520**(7549): 609-611.

Schunkert, H., I. R. Konig, S. Kathiresan, M. P. Reilly, T. L. Assimes, H. Holm, M. Preuss, A. F. Stewart, M. Barbalic, C. Gieger, D. Absher, Z. Aherrahrou, H. Allayee, D. Altshuler, S. S. Anand, K. Andersen, J. L. Anderson, D. Ardissino, S. G. Ball, A. J. Balmforth, T. A. Barnes, D. M. Becker, L. C. Becker, K. Berger, J. C. Bis, S. M. Boekholdt, E. Boerwinkle, P. S. Braund, M. J. Brown, M. S. Burnett, I. Buysschaert, Cardiogenics, J. F. Carlquist, L. Chen, S. Cichon, V. Codd, R. W. Davies, G. Dedoussis, A. Dehghan, S. Demissie, J. M. Devaney, P. Diemert, R. Do, A. Doering, S. Eifert, N. E. Mokhtari, S. G. Ellis, R. Elosua, J. C. Engert, S. E. Epstein, U. de Faire, M. Fischer, A. R. Folsom, J. Freyer, B. Gigante, D. Girelli, S. Gretarsdottir, V. Gudnason, J. R. Gulcher, E. Halperin, N. Hammond, S. L. Hazen, A. Hofman, B. D. Horne, T. Illig, C. Iribarren, G. T. Jones, J. W. Jukema, M. A. Kaiser, L. M. Kaplan, J. J. Kastelein, K. T. Khaw, J. W. Knowles, G. Kolovou, A. Kong, R. Laaksonen, D. Lambrechts, K. Leander, G. Lettre, M. Li, W. Lieb, C. Loley, A. J. Lotery, P. M. Mannucci, S. Maouche, N. Martinelli, P. P. McKeown, C. Meisinger, T. Meitinger, O. Melander, P. A. Merlini, V. Mooser, T. Morgan, T. W. Muhleisen, J. B. Muhlestein, T. Munzel, K. Musunuru, J. Nahrstaedt, C. P. Nelson, M. M. Nothen, O. Olivieri, R. S. Patel, C. C. Patterson, A. Peters, F. Peyvandi, L. Qu, A. A. Quyyumi, D. J. Rader, L. S. Rallidis, C. Rice, F. R. Rosendaal, D. Rubin, V. Salomaa, M. L. Sampietro, M. S. Sandhu, E. Schadt, A. Schafer, A. Schillert, S. Schreiber, J. Schrezenmeir, S. M. Schwartz, D. S. Siscovick, M. Sivananthan, S. Sivapalaratnam, A. Smith, T. B. Smith, J. D. Snoep, N. Soranzo, J. A. Spertus, K. Stark, K. Stirrups, M. Stoll, W. H. Tang, S. Tennstedt, G. Thorgeirsson, G. Thorleifsson, M. Tomaszewski, A. G. Uitterlinden, A. M. van Rij, B. F. Voight, N. J. Wareham, G. A. Wells, H. E. Wichmann, P. S. Wild, C. Willenborg, J. C. Witteman, B. J. Wright, S. Ye, T. Zeller, A. Ziegler, F. Cambien, A. H. Goodall, L. A. Cupples, T. Quertermous, W. Marz, C. Hengstenberg, S. Blankenberg, W. H. Ouwehand, A. S. Hall, P. Deloukas, J. R. Thompson, K. Stefansson, R. Roberts, U. Thorsteinsdottir, C. J. O'Donnell, R. McPherson, J.

Erdmann, C. A. Consortium and N. J. Samani (2011). "Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease." Nat Genet **43**(4): 333-338.

Schwarz, E., P. C. Guest, H. Rahmoune, L. W. Harris, L. Wang, F. M. Leweke, M. Rothermundt, B. Bogerts, D. Koethe, L. Kranaster, P. Ohrmann, T. Suslow, G. McAllister, M. Spain, A. Barnes, N. J. M. van Beveren, S. Baron-Cohen, J. Steiner, F. E. Torrey, R. H. Yolken and S. Bahn (2012). "Identification of a biological signature for schizophrenia in serum." Molecular Psychiatry **17**(5): 494-502.

Schwarz, E., P. C. Guest, H. Rahmoune, D. Martins-de-Souza, D. W. Niebuhr, N. S. Weber, D. N. Cowan, R. H. Yolken, M. Spain and A. Barnes (2012). "Identification of a blood-based biological signature in subjects with psychiatric disorders prior to clinical manifestation." The World Journal of Biological Psychiatry **13**(8): 627-632.

Schwarz, E., R. Izmailov, P. Lio and A. Meyer-Lindenberg (2016). "Protein Interaction Networks Link Schizophrenia Risk Loci to Synaptic Function." Schizophr Bull **42**(6): 1334-1342.

Schwarz, E., H. Tost and A. Meyer-Lindenberg (2016). "Working memory genetics in schizophrenia and related disorders: An RDoC perspective." Am J Med Genet B Neuropsychiatr Genet **171B**(1): 121-131.

Schwarz, E., N. J. van Beveren, J. Ramsey, F. M. Leweke, M. Rothermundt, B. Bogerts, J. Steiner, P. C. Guest and S. Bahn (2013). "Identification of subgroups of schizophrenia patients with changes in either immune or growth factor and hormonal pathways." Schizophrenia bulletin **40**(4): 787-795.

Schwarz, K., C. Moessnang, J. I. Schweiger, S. Baumeister, M. M. Plichta, D. Brandeis, T. Banaschewski, C. Wackerhagen, S. Erk, H. Walter, H. Tost and A. Meyer-Lindenberg (2019). "Transdiagnostic Prediction of Affective, Cognitive, and Social Function Through Brain Reward Anticipation in Schizophrenia, Bipolar Disorder, Major Depression, and Autism Spectrum Diagnoses." Schizophrenia Bulletin.

Selemon, L. D. and N. Zecevic (2015). "Schizophrenia: a tale of two critical periods for prefrontal cortical development." Translational Psychiatry **5**(8): e623-e623.

Semick, S. A., L. Collado-Torres, C. A. Markunas, J. H. Shin, A. Deep-Soboslay, R. Tao, M. A. Huestis, L. J. Bierut, B. S. Maher, E. O. Johnson, T. M. Hyde, D. R. Weinberger, D. B. Hancock, J. E. Kleinman and A. E. Jaffe (2018). "Developmental effects of maternal smoking during pregnancy on the human frontal cortex transcriptome." Mol Psychiatry.

Seney, M. L., K. Cahill, J. F. Enwright, R. W. Logan, Z. Huo, W. Zong, G. Tseng and C. A. McClung (2019). "Diurnal rhythms in gene expression in the prefrontal cortex in schizophrenia." Nature Communications **10**(1): 3355.

Sharifi-Noghabi, H., O. Zolotareva, C. C. Collins and M. Ester (2019). "MOLI: Multi-Omics Late Integration with deep neural networks for drug response prediction." bioRxiv: 531327.

Sharp, A. J., H. C. Mefford, K. Li, C. Baker, C. Skinner, R. E. Stevenson, R. J. Schroer, F. Novara, M. De Gregori, R. Ciccone, A. Broomer, I. Casuga, Y. Wang, C. Xiao, C. Barbacioru, G. Gimelli, B. D. Bernardina, C. Torniero, R. Giorda, R. Regan, V. Murday, S. Mansour, M. Fichera, L. Castiglia, P. Failla, M. Ventura, Z. Jiang, G. M. Cooper, S. J. Knight, C. Romano, O. Zuffardi, C. Chen, C. E. Schwartz and E. E. Eichler (2008). "A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures." Nat Genet **40**(3): 322-328.

Shi, J., D. F. Levinson, J. Duan, A. R. Sanders, Y. Zheng, I. Pe'Er, F. Dudbridge, P. A. Holmans, A. S. Whittemore and B. J. Mowry (2009). "Common variants on chromosome 6p22. 1 are associated with schizophrenia." Nature **460**(7256): 753.

Shi, L., Z. Zhang and B. Su (2016). "Sex Biased Gene Expression Profiling of Human Brains at Major Developmental Stages." Sci Rep **6**: 21181.

Sjöholm, L. K., Y. Ransome, T. J. Ekström and O. Karlsson (2018). "Evaluation of post-mortem effects on global brain DNA methylation and hydroxymethylation." Basic & clinical pharmacology & toxicology **122**(2): 208-213.

Smith, A. K., V. Kilaru, T. Klengel, K. B. Mercer, B. Bradley, K. N. Conneely, K. J. Ressler and E. B. Binder (2015). "DNA extracted from saliva for methylation studies of psychiatric traits: evidence tissue specificity and relatedness to brain." Am J Med Genet B Neuropsychiatr Genet **168B**(1): 36-44.

Snow, W. M., B. M. Stoesz, D. M. Kelly and B. C. Albensi (2014). "Roles for NF-κB and gene targets of NF-κB in synaptic plasticity, memory, and navigation." Molecular neurobiology **49**(2): 757-770.

Stahl, E. A., G. Breen, A. J. Forstner, A. McQuillin, S. Ripke, V. Trubetskoy, M. Mattheisen, Y. Wang, J. R. I. Coleman, H. A. Gaspar, C. A. de Leeuw, S. Steinberg, J. M. W. Pavlides, M. Trzaskowski, E. M. Byrne, T. H. Pers, P. A. Holmans, A. L. Richards, L. Abbott, E. Agerbo, H. Akil, D. Albani, N. Alliey-Rodriguez, T. D. Als, A. Anjorin, V. Antilla, S. Awasthi, J. A. Badner, M. Baekvad-Hansen, J. D. Barchas, N. Bass, M. Bauer, R. Belliveau, S. E. Bergen, C. B. Pedersen, E. Boen, M. P. Boks, J. Boocock, M. Budde, W. Bunney, M. Burmeister, J. Bybjerg-Grauholm, W. Byerley, M. Casas, F. Cerrato, P. Cervantes, K. Chambert, A. W. Charney, D. Chen, C. Churchhouse, T. K. Clarke, W. Coryell, D. W. Craig, C. Cruceanu, D. Curtis, P. M. Czerski, A. M. Dale, S. de Jong, F. Degenhardt, J. Del-Favero, J. R. DePaulo, S. Djurovic, A. L. Dobbyn, A. Dumont, T. Elvsashagen, V. Escott-Price, C. C. Fan, S. B. Fischer, M. Flickinger, T. M. Foroud, L. Forty, J. Frank, C. Fraser, N. B. Freimer, L. Frisen, K. Gade, D. Gage, J. Garnham, C. Giambartolomei, M. G. Pedersen, J. Goldstein, S. D. Gordon, K. Gordon-Smith, E. K. Green, M. J. Green, T. A. Greenwood, J. Grove, W. Guan, J. Guzman-Parra, M. L. Hamshere, M. Hautzinger, U. Heilbronner, S. Herms, M. Hipolito, P. Hoffmann, D. Holland, L. Huckins, S. Jamain, J. S. Johnson, A. Jureus, R. Kandaswamy, R. Karlsson, J. L. Kennedy, S. Kittel-Schneider, J. A. Knowles, M. Kogevinas, A. C. Koller, R. Kupka, C. Lavebratt, J. Lawrence, W. B. Lawson, M. Leber, P. H. Lee, S. E. Levy, J. Z. Li, C. Liu, S. Lucae, A. Maaser, D. J. MacIntyre, P. B. Mahon, W. Maier, L. Martinsson, S. McCarroll, P. McGuffin, M. G. McInnis, J. D. McKay, H. Medeiros, S. E. Medland, F. Meng, L. Milani, G. W. Montgomery, D. W. Morris, T. W. Muhleisen, N. Mullins, H. Nguyen, C. M. Nievergelt, A. N. Adolfsson, E. A. Nwulia, C. O'Donovan, L. M. O. Loohuis, A. P. S. Ori, L. Oruc, U. Osby, R. H. Perlis, A. Perry, A. Pfennig, J. B. Potash, S. M. Purcell, E. J. Regeer, A. Reif, C. S. Reinbold, J. P. Rice, F. Rivas, M. Rivera, P. Roussos, D. M. Ruderfer, E. Ryu, C. Sanchez-Mora, A. F. Schatzberg, W. A. Scheftner, N. J. Schork, C. Shannon Weickert, T. Shehktman, P. D. Shilling, E. Sigurdsson, C. Slaney, O. B. Smeland, J. L. Sobell, C. Soholm Hansen, A. T. Spijker, D. St Clair, M. Steffens, J. S. Strauss, F. Streit, J. Strohmaier, S. Szelinger, R. C. Thompson, T. E. Thorgeirsson, J. Treutlein, H. Vedder, W. Wang, S. J. Watson, T. W. Weickert, S. H. Witt, S. Xi, W. Xu, A. H. Young, P. Zandi, P. Zhang, S. Zollner, Q. C. e, B. Consortium, R. Adolfsson, I. Agartz, M. Alda, L. Backlund, B. T. Baune, F. Bellivier, W. H. Berrettini, J. M. Biernacka, D. H. R. Blackwood, M. Boehnke, A. D. Borglum, A. Corvin, N. Craddock, M. J. Daly, U. Dannlowski, T. Esko, B. Etain, M. Frye, J. M. Fullerton, E. S. Gershon, M. Gill, F. Goes, M. Grigoroiu-Serbanescu, J. Hauser, D. M. Hougaard, C. M. Hultman, I. Jones, L. A. Jones, R. S. Kahn, G. Kirov, M. Landen, M. Leboyer, C. M. Lewis, Q. S. Li, J. Lissowska, N. G. Martin, F. Mayoral, S. L. McElroy, A. M. McIntosh, F. J. McMahon, I. Melle, A. Metspalu, P. B. Mitchell, G. Morken, O. Mors, P. B. Mortensen, B. Muller-Myhsok, R. M. Myers, B. M. Neale, V. Nimgaonkar, M. Nordentoft, M. M. Nothen, M. C. O'Donovan, K. J. Oedegaard, M. J. Owen, S. A. Paciga, C. Pato, M. T. Pato, D. Posthuma, J. A. Ramos-Quiroga, M. Ribases, M. Rietschel, G. A. Rouleau, M. Schalling, P. R. Schofield, T. G. Schulze, A. Serretti, J. W. Smoller, H. Stefansson, K. Stefansson, E. Stordal, P. F. Sullivan, G. Turecki, A. E. Vaaler, E. Vieta, J. B. Vincent, T. Werge, J. I. Nurnberger, N. R. Wray, A. Di Florio, H. J. Edenberg, S. Cichon, R. A. Ophoff, L. J. Scott, O. A. Andreassen, J. Kelsoe, P. Sklar and C. Bipolar Disorder Working Group of the Psychiatric Genomics (2019). "Genome-wide association study identifies 30 loci associated with bipolar disorder." Nat Genet **51**(5): 793-803.

Stephan, K. E., T. Baldeweg and K. J. Friston (2006). "Synaptic plasticity and dysconnection in schizophrenia." Biological psychiatry **59**(10): 929-939.

Stevens, E. L., G. Heckenberg, E. D. Roberson, J. D. Baugher, T. J. Downey and J. Pevsner (2011). "Inference of relationships in population data using identity-by-descent and identity-by-state." PLoS Genet **7**(9): e1002287.

Stilo, S. A., M. Di Forti and R. M. Murray (2011). "Environmental risk factors for schizophrenia: implications for prevention." Neuropsychiatry **1**(5): 457-466.

Stone, J. L., M. C. O'Donovan, H. Gurling, G. K. Kirov, D. H. R. Blackwood, A. Corvin, N. J. Craddock, M. Gill, C. M. Hultman, P. Lichtenstein, A. McQuillin, C. N. Pato, D. M. Ruderfer, M. J. Owen, D. St Clair, P. F. Sullivan, P. Sklar, S. M. Purcell, J. L. Stone, D. M. Ruderfer, J. Korn, G. K. Kirov, S.

Macgregor, A. McQuillin, D. W. Morris, C. T. O'Dushlaine, M. J. Daly, P. M. Visscher, P. A. Holmans, M. C. O'Donovan, P. F. Sullivan, P. Sklar, S. M. Purcell, H. Gurling, A. Corvin, D. H. R. Blackwood, N. J. Craddock, M. Gill, C. M. Hultman, G. K. Kirov, P. Lichtenstein, A. McQuillin, M. C. O'Donovan, M. J. Owen, C. N. Pato, S. M. Purcell, E. M. Scolnick, D. St Clair, J. L. Stone, P. F. Sullivan, P. Sklar, M. C. O'Donovan, G. K. Kirov, N. J. Craddock, P. A. Holmans, N. M. Williams, L. Georgieva, I. Nikolov, N. Norton, H. Williams, D. Toncheva, V. Milanova, M. J. Owen, C. M. Hultman, P. Lichtenstein, E. F. Thelander, P. Sullivan, D. W. Morris, C. T. O'Dushlaine, E. Kenny, J. L. Waddington, M. Gill, A. Corvin, A. McQuillin, K. Choudhury, S. Datta, J. Pimm, S. Thirumalai, V. Puri, R. Krasucki, J. Lawrence, D. Quested, N. Bass, D. Curtis, H. Gurling, C. Crombie, G. Fraser, S. Leh Kwan, N. Walker, D. St Clair, D. H. R. Blackwood, W. J. Muir, K. A. McGhee, B. Pickard, P. Malloy, A. W. Maclean, M. Van Beck, P. M. Visscher, S. Macgregor, M. T. Pato, H. Medeiros, F. Middleton, C. Carvalho, C. Morley, A. Fanous, D. Conti, J. A. Knowles, C. Paz Ferreira, A. Macedo, M. Helena Azevedo, C. N. Pato, J. L. Stone, D. M. Ruderfer, J. Korn, S. A. McCarroll, M. Daly, S. M. Purcell, P. Sklar, S. M. Purcell, J. L. Stone, K. Chambert, D. M. Ruderfer, J. Korn, S. A. McCarroll, C. Gates, S. B. Gabriel, S. Mahon, K. Ardlie, M. J. Daly, E. M. Scolnick, P. Sklar, C. The International Schizophrenia, p. Manuscript, a. Data, c. Management, U. Cardiff, H. Karolinska Institutet/University of North Carolina at Chapel, D. Trinity College, L. University College, A. University of, E. University of, R. Queensland Institute of Medical, C. University of Southern, H. Massachusetts General, R. Stanley Center for Psychiatric, M. I. T. Broad Institute of and Harvard (2008). "Rare chromosomal deletions and duplications increase risk of schizophrenia." Nature **455**(7210): 237-241.

Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander and J. P. Mesirov (2005). "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles." Proc Natl Acad Sci U S A **102**(43): 15545-15550.

Sullivan, P. F., M. J. Daly and M. O'Donovan (2012). "Genetic architectures of psychiatric disorders: the emerging picture and its implications." Nat Rev Genet **13**(8): 537-551.

Sullivan, P. F., K. S. Kendler and M. C. Neale (2003). "Schizophrenia as a Complex Trait: Evidence From a Meta-analysis of Twin Studies." JAMA Psychiatry **60**(12): 1187-1192.

Sullivan, P. F., K. S. Kendler and M. C. Neale (2003). "Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies." Arch Gen Psychiatry **60**(12): 1187-1192.

Teroganova, N., L. Girshkin, C. M. Suter and M. J. Green (2016). "DNA methylation in peripheral tissue of schizophrenia and bipolar disorder: a systematic review." BMC Genet **17**: 27.

The Gene Ontology Consortium (2018). "The Gene Ontology Resource: 20 years and still GOing strong." Nucleic Acids Research **47**(D1): D330-D338.

Thompson, P. M., O. A. Andreassen, A. Arias-Vasquez, C. E. Bearden, P. S. Boedhoe, R. M. Brouwer, R. L. Buckner, J. K. Buitelaar, K. B. Bulayeva, D. M. Cannon, R. A. Cohen, P. J. Conrod, A. M. Dale, I. J. Deary, E. L. Dennis, M. A. de Reus, S. Desrivieres, D. Dima, G. Donohoe, S. E. Fisher, J. P. Fouche, C. Francks, S. Frangou, B. Franke, H. Ganjgahi, H. Garavan, D. C. Glahn, H. J. Grabe, T. Guadalupe, B. A. Gutman, R. Hashimoto, D. P. Hibar, D. Holland, M. Hoogman, H. E. H. Pol, N. Hosten, N. Jahanshad, S. Kelly, P. Kochunov, W. S. Kremen, P. H. Lee, S. Mackey, N. G. Martin, B. Mazoyer, C. McDonald, S. E. Medland, R. A. Morey, T. E. Nichols, T. Paus, Z. Pausova, L. Schmaal, G. Schumann, L. Shen, S. M. Sisodiya, D. J. A. Smit, J. W. Smoller, D. J. Stein, J. L. Stein, R. Toro, J. A. Turner, M. P. van den Heuvel, O. L. van den Heuvel, T. G. M. van Erp, D. van Rooij, D. J. Veltman, H. Walter, Y. Wang, J. M. Wardlaw, C. D. Whelan, M. J. Wright, J. Ye and E. Consortium (2017). "ENIGMA and the individual: Predicting factors that affect the brain in 35 countries worldwide." Neuroimage **145**(Pt B): 389-408.

Thompson, P. M., J. L. Stein, S. E. Medland, D. P. Hibar, A. A. Vasquez, M. E. Renteria, R. Toro, N. Jahanshad, G. Schumann and B. Franke (2014). "The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data." Brain imaging and behavior **8**(2): 153-182.

Tiihonen, J., M. Koskuvi, M. Storvik, I. Hyötyläinen, Y. Gao, K. A. Puttonen, R. Giniatullina, E. Poguzhelskaya, I. Ojansuu, O. Vaurio, T. D. Cannon, J. Lönnqvist, S. Therman, J. Suvisaari, J.

Kaprio, L. Cheng, A. F. Hill, M. Lähteenvuo, J. Tohka, R. Giniatullin, Š. Lehtonen and J. Koistinaho (2019). "Sex-specific transcriptional and proteomic signatures in schizophrenia." Nature Communications **10**(1): 3933.

Titus, A. J., R. M. Gallimore, L. A. Salas and B. C. Christensen (2017). "Cell-type deconvolution from DNA methylation: a review of recent applications." Hum Mol Genet **26**(R2): R216-R224.

Torkamani, A., B. Dean, N. J. Schork and E. A. Thomas (2010). "Coexpression network analysis of neural tissue reveals perturbations in developmental processes in schizophrenia." Genome Res **20**(4): 403-412.

Torkamani, A., N. E. Wineinger and E. J. Topol (2018). "The personal and clinical utility of polygenic risk scores." Nature Reviews Genetics **19**(9): 581-590.

Tost, H., M. Reichert, U. Braun, I. Reinhard, R. Peters, S. Lautenbach, A. Hoell, E. Schwarz, U. Ebner-Priemer and A. Zipf (2019). "Neural correlates of individual differences in affective benefit of real-life urban green space exposure." Nature neuroscience **22**(9): 1389-1393.

Trabzuni, D., A. Ramasamy, S. Imran, R. Walker, C. Smith, M. E. Weale, J. Hardy, M. Ryten and C. North American Brain Expression (2013). "Widespread sex differences in gene expression and splicing in the adult human brain." Nat Commun **4**: 2771.

Tsai, G. and J. T. Coyle (2002). "Glutamatergic mechanisms in schizophrenia." Annu Rev Pharmacol Toxicol **42**: 165-179.

Tsuang, M. (2000). "Schizophrenia: genes and environment." Biol Psychiatry **47**(3): 210-220.

Ursini, G., V. Bollati, L. Fazio, A. Porcelli, L. Iacovelli, A. Catalani, L. Sinibaldi, B. Gelao, R. Romano, A. Rampino, P. Taurisano, M. Mancini, A. Di Giorgio, T. Popolizio, A. Baccarelli, A. De Blasi, G. Blasi and A. Bertolino (2011). "Stress-related methylation of the catechol-O-methyltransferase Val 158 allele predicts human prefrontal cognition and activity." J Neurosci **31**(18): 6692-6698.

Ursini, G., T. Cavalleri, L. Fazio, T. Angrisano, L. Iacovelli, A. Porcelli, G. Maddalena, G. Punzi, M. Mancini, B. Gelao, R. Romano, R. Masellis, F. Calabrese, A. Rampino, P. Taurisano, A. Di Giorgio, S. Keller, L. Tarantini, L. Sinibaldi, T. Quarto, T. Popolizio, G. Caforio, G. Blasi, M. A. Riva, A. De Blasi, L. Chiariotti, V. Bollati and A. Bertolino (2016). "BDNF rs6265 methylation and genotype interact on risk for schizophrenia." Epigenetics **11**(1): 11-23.

van Dam, S., U. Vosa, A. van der Graaf, L. Franke and J. P. de Magalhaes (2018). "Gene co-expression analysis for functional classification and gene-disease predictions." Brief Bioinform **19**(4): 575-592.

Van Der Maaten, L., E. Postma and J. Van den Herik (2009). "Dimensionality Reduction: A Comparative Review." J Mach Learn Res **10**(66-71): 13.

Van Erp, T. G., E. Walton, D. P. Hibar, L. Schmaal, W. Jiang, D. C. Glahn, G. D. Pearlson, N. Yao, M. Fukunaga and R. Hashimoto (2018). "Cortical brain abnormalities in 4474 individuals with schizophrenia and 5098 control subjects via the Enhancing Neuro Imaging Genetics Through Meta Analysis (ENIGMA) Consortium." Biological psychiatry **84**(9): 644-654.

van Erp, T. G. M., D. P. Hibar, J. M. Rasmussen, D. C. Glahn, G. D. Pearlson, O. A. Andreassen, I. Agartz, L. T. Westlye, U. K. Haukvik, A. M. Dale, I. Melle, C. B. Hartberg, O. Gruber, B. Kraemer, D. Zilles, G. Donohoe, S. Kelly, C. McDonald, D. W. Morris, D. M. Cannon, A. Corvin, M. W. J. Machielsen, L. Koenders, L. de Haan, D. J. Veltman, T. D. Satterthwaite, D. H. Wolf, R. C. Gur, R. E. Gur, S. G. Potkin, D. H. Mathalon, B. A. Mueller, A. Preda, F. Macciardi, S. Ehrlich, E. Walton, J. Hass, V. D. Calhoun, H. J. Bockholt, S. R. Sponheim, J. M. Shoemaker, N. E. M. van Haren, H. E. H. Pol, R. A. Ophoff, R. S. Kahn, R. Roiz-Santiañez, B. Crespo-Facorro, L. Wang, K. I. Alpert, E. G. Jönsson, R. Dimitrova, C. Bois, H. C. Whalley, A. M. McIntosh, S. M. Lawrie, R. Hashimoto, P. M. Thompson, J. A. Turner and E. S. W. G. for the (2016). "Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium." Molecular Psychiatry **21**(4): 547-553.

van Mierlo, H. C., A. Schot, M. P. Boks and L. D. de Witte (2019). "The association between schizophrenia and the immune system: Review of the evidence from unbiased 'omic-studies'." Schizophrenia research.

van Os, J., G. Kenis and B. P. Rutten (2010). "The environment and schizophrenia." Nature **468**(7321): 203-212.

van Os, J., B. P. Rutten and R. Poulton (2008). "Gene-environment interactions in schizophrenia: review of epidemiological findings and future directions." Schizophr Bull **34**(6): 1066-1082.

Vargas, T., P. H. Lam, M. Azis, K. J. Osborne, A. Lieberman and V. A. Mittal (2018). "Childhood Trauma and Neurocognition in Adults With Psychotic Disorders: A Systematic Review and Meta-analysis." Schizophr Bull.

Visscher, P. M., N. R. Wray, Q. Zhang, P. Sklar, M. I. McCarthy, M. A. Brown and J. Yang (2017). "10 years of GWAS discovery: biology, function, and translation." The American Journal of Human Genetics **101**(1): 5-22.

Vitale, A. M., N. A. Matigian, A. S. Cristino, K. Nones, S. Ravishankar, B. Bellette, Y. Fan, S. A. Wood, E. Wolvetang and A. Mackay-Sim (2017). "DNA methylation in schizophrenia in different patient-derived cell types." npj Schizophrenia **3**(1): 6.

Vivian-Griffiths, T., E. Baker, K. M. Schmidt, M. Bracher-Smith, J. Walters, A. Artemiou, P. Holmans, M. C. O'Donovan, M. J. Owen, A. Pocklington and V. Escott-Price (2019). "Predictive modeling of schizophrenia from genomic data: Comparison of polygenic risk score with kernel support vector machines approach." Am J Med Genet B Neuropsychiatr Genet **180**(1): 80-85.

Voisin, S., Eynon, N., Yan, X., & Bishop, D. J. (2015). "Exercise training and DNA methylation in humans." Acta Physiologica **213**(1): 39-59.

von Rueden, L., S. Mayer, J. Garcke, C. Bauckhage and J. Schuecker (2019). "Informed Machine Learning-Towards a Taxonomy of Explicit Integration of Knowledge into Machine Learning." arXiv preprint arXiv:1903.12394.

Walton, E., J. Hass, J. Liu, J. L. Roffman, F. Bernardoni, V. Roessner, M. Kirsch, G. Schackert, V. Calhoun and S. Ehrlich (2016). "Correspondence of DNA Methylation Between Blood and Brain Tissue and Its Application to Schizophrenia Research." Schizophr Bull **42**(2): 406-414.

Walton, E., J. Liu, J. Hass, T. White, M. Scholz, V. Roessner, R. Gollub, V. D. Calhoun and S. Ehrlich (2014). "MB-COMT promoter DNA methylation is associated with working-memory processing in schizophrenia patients and healthy controls." Epigenetics **9**(8): 1101-1107.

Walton, E., J. Turner, R. L. Gollub, D. S. Manoach, A. Yendiki, B.-C. Ho, S. R. Sponheim, V. D. Calhoun and S. Ehrlich (2012). "Cumulative genetic risk and prefrontal activity in patients with schizophrenia." Schizophrenia bulletin **39**(3): 703-711.

Wang, Q., R. Chen, F. Cheng, Q. Wei, Y. Ji, H. Yang, X. Zhong, R. Tao, Z. Wen, J. S. Sutcliffe, C. Liu, E. H. Cook, N. J. Cox and B. Li (2019). "A Bayesian framework that integrates multi-omics data and gene networks predicts risk genes from schizophrenia GWAS data." Nat Neurosci **22**(5): 691-699.

Wang, S., X. Shi, M. Wu and S. Ma (2019). "Horizontal and vertical integrative analysis methods for mental disorders omics data." Sci Rep **9**(1): 13430.

Wang, Z., M. Gerstein and M. Snyder (2009). "RNA-Seq: a revolutionary tool for transcriptomics." Nat Rev Genet **10**(1): 57-63.

Weinberger, D. R. (1987). "Implications of Normal Brain Development for the Pathogenesis of Schizophrenia." JAMA Psychiatry **44**(7): 660-669.

Weinberger, D. R. (1987). "Implications of normal brain development for the pathogenesis of schizophrenia." Arch Gen Psychiatry **44**(7): 660-669.

Wen, Z., H. N. Nguyen, Z. Guo, M. A. Lalli, X. Wang, Y. Su, N. S. Kim, K. J. Yoon, J. Shin, C. Zhang, G. Makri, D. Nauen, H. Yu, E. Guzman, C. H. Chiang, N. Yoritomo, K. Kaibuchi, J. Zou, K. M. Christian, L. Cheng, C. A. Ross, R. L. Margolis, G. Chen, K. S. Kosik, H. Song and G. L. Ming (2014). "Synaptic dysregulation in a human iPS cell model of mental disorders." Nature **515**(7527): 414-418.

Werling, D. M., N. N. Parikshak and D. H. Geschwind (2016). "Gene expression in human brain implicates sexually dimorphic pathways in autism spectrum disorders." Nat Commun **7**: 10717.

Whalley, H. C., L. Hall, L. Romaniuk, A. Macdonald, S. M. Lawrie, J. E. Sussmann and A. M. McIntosh (2015). "Impact of cross-disorder polygenic risk on frontal brain activation with specific effect of schizophrenia risk." Schizophrenia research **161**(2-3): 484-489.

Whalley, H. C., E. Simonotto, I. Marshall, D. G. Owens, N. H. Goddard, E. C. Johnstone and S. M. Lawrie (2005). "Functional disconnectivity in subjects at high genetic risk of schizophrenia." Brain **128**(Pt 9): 2097-2108.

Willsey, A. J., S. J. Sanders, M. Li, S. Dong, A. T. Tebbenkamp, R. A. Muhle, S. K. Reilly, L. Lin, S. Fertuzinhos, J. A. Miller, M. T. Murtha, C. Bichsel, W. Niu, J. Cotney, A. G. Ercan-Sencicek, J. Gockley, A. R. Gupta, W. Han, X. He, E. J. Hoffman, L. Klei, J. Lei, W. Liu, L. Liu, C. Lu, X. Xu, Y. Zhu, S. M. Mane, E. S. Lein, L. Wei, J. P. Noonan, K. Roeder, B. Devlin, N. Sestan and M. W. State (2013). "Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism." Cell **155**(5): 997-1007.

Wockner, L. F., E. P. Noble, B. R. Lawford, R. M. Young, C. P. Morris, V. L. J. Whitehall and J. Voisey (2014). "Genome-wide DNA methylation analysis of human brain tissue from schizophrenia patients." Translational Psychiatry **4**(1): e339-e339.

Wray, N. R., S. Ripke, M. Mattheisen, M. Trzaskowski, E. M. Byrne, A. Abdellaoui, M. J. Adams, E. Agerbo, T. M. Air, T. M. F. Andlauer, S. A. Bacanu, M. Baekvad-Hansen, A. F. T. Beekman, T. B. Bigdeli, E. B. Binder, D. R. H. Blackwood, J. Bryois, H. N. Buttenschon, J. Bybjerg-Grauholm, N. Cai, E. Castelao, J. H. Christensen, T. K. Clarke, J. I. R. Coleman, L. Colodro-Conde, B. Couvy-Duchesne, N. Craddock, G. E. Crawford, C. A. Crowley, H. S. Dashti, G. Davies, I. J. Deary, F. Degenhardt, E. M. Derks, N. Direk, C. V. Dolan, E. C. Dunn, T. C. Eley, N. Eriksson, V. Escott-Price, F. H. F. Kiadeh, H. K. Finucane, A. J. Forstner, J. Frank, H. A. Gaspar, M. Gill, P. Giusti-Rodriguez, F. S. Goes, S. D. Gordon, J. Grove, L. S. Hall, E. Hannon, C. S. Hansen, T. F. Hansen, S. Herms, I. B. Hickie, P. Hoffmann, G. Homuth, C. Horn, J. J. Hottenga, D. M. Hougaard, M. Hu, C. L. Hyde, M. Ising, R. Jansen, F. Jin, E. Jorgenson, J. A. Knowles, I. S. Kohane, J. Kraft, W. W. Kretzschmar, J. Krogh, Z. Kutalik, J. M. Lane, Y. Li, Y. Li, P. A. Lind, X. Liu, L. Lu, D. J. MacIntyre, D. F. MacKinnon, R. M. Maier, W. Maier, J. Marchini, H. Mbarek, P. McGrath, P. McGuffin, S. E. Medland, D. Mehta, C. M. Middeldorp, E. Mihailov, Y. Milaneschi, L. Milani, J. Mill, F. M. Mondimore, G. W. Montgomery, S. Mostafavi, N. Mullins, M. Nauck, B. Ng, M. G. Nivard, D. R. Nyholt, P. F. O'Reilly, H. Oskarsson, M. J. Owen, J. N. Painter, C. B. Pedersen, M. G. Pedersen, R. E. Peterson, E. Pettersson, W. J. Peyrot, G. Pistis, D. Posthuma, S. M. Purcell, J. A. Quiroz, P. Qvist, J. P. Rice, B. P. Riley, M. Rivera, S. Saeed Mirza, R. Saxena, R. Schoevers, E. C. Schulte, L. Shen, J. Shi, S. I. Shyn, E. Sigurdsson, G. B. C. Sinnamon, J. H. Smit, D. J. Smith, H. Stefansson, S. Steinberg, C. A. Stockmeier, F. Streit, J. Strohmaier, K. E. Tansey, H. Teismann, A. Teumer, W. Thompson, P. A. Thomson, T. E. Thorgeirsson, C. Tian, M. Traylor, J. Treutlein, V. Trubetskoy, A. G. Uitterlinden, D. Umbricht, S. Van der Auwera, A. M. van Hemert, A. Viktorin, P. M. Visscher, Y. Wang, B. T. Webb, S. M. Weinsheimer, J. Wellmann, G. Willemsen, S. H. Witt, Y. Wu, H. S. Xi, J. Yang, F. Zhang, eQtlgen, andMe, V. Arolt, B. T. Baune, K. Berger, D. I. Boomsma, S. Cichon, U. Dannlowski, E. C. J. de Geus, J. R. DePaulo, E. Domenici, K. Domschke, T. Esko, H. J. Grabe, S. P. Hamilton, C. Hayward, A. C. Heath, D. A. Hinds, K. S. Kendler, S. Kloiber, G. Lewis, Q. S. Li, S. Lucae, P. F. A. Madden, P. K. Magnusson, N. G. Martin, A. M. McIntosh, A. Metspalu, O. Mors, P. B. Mortensen, B. Muller-Myhsok, M. Nordentoft, M. M. Nothen, M. C. O'Donovan, S. A. Paciga, N. L. Pedersen, B. Penninx, R. H. Perlis, D. J. Porteous, J. B. Potash, M. Preisig, M. Rietschel, C. Schaefer, T. G. Schulze, J. W. Smoller, K. Stefansson, H. Tiemeier, R. Uher, H. Volzke, M. M. Weissman, T. Werge, A. R. Winslow, C. M. Lewis, D. F. Levinson, G. Breen, A. D. Borglum, P. F. Sullivan and C. Major Depressive Disorder Working Group of the Psychiatric Genomics (2018). "Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression." Nat Genet **50**(5): 668-681.

Wu, C., R. M. Bendriem, S. P. Garamszegi, L. Song and C. T. Lee (2017). "RNA sequencing in post-mortem human brains of neuropsychiatric disorders." Psychiatry Clin Neurosci **71**(10): 663-672.

Wu, H. and Y. Zhang (2014). "Reversing DNA methylation: mechanisms, genomics, and biological functions." Cell **156**(1-2): 45-68.

Xu, B., I. Ionita-Laza, J. L. Roos, B. Boone, S. Woodrick, Y. Sun, S. Levy, J. A. Gogos and M. Karayiorgou (2012). "De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia." Nature genetics **44**(12): 1365.

Xu, B., I. Ionita-Laza, J. L. Roos, B. Boone, S. Woodrick, Y. Sun, S. Levy, J. A. Gogos and M. Karayiorgou (2012). "De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia." Nat Genet **44**(12): 1365-1369.

Xu, R. and Q. Wang (2015). "PhenoPredict: a disease phenome-wide drug repositioning approach towards schizophrenia drug discovery." Journal of biomedical informatics **56**: 348-355.

Yin, D. M., Y. J. Chen, A. Sathyamurthy, W. C. Xiong and L. Mei (2012). "Synaptic dysfunction in schizophrenia." Adv Exp Med Biol **970**: 493-516.

Yu, K.-H., A. L. Beam and I. S. Kohane (2018). "Artificial intelligence in healthcare." Nature biomedical engineering **2**(10): 719-731.

Zakhari, S. (2013). "Alcohol metabolism and epigenetics changes." Alcohol Res **35**(1): 6-16.

Zang, Z., L. S. Geiger, U. Braun, H. Cao, M. Zangl, A. Schäfer, C. Moessnang, M. Ruf, J. Reis, J. I. Schweiger, L. Dixson, A. Moscicki, E. Schwarz, A. Meyer-Lindenberg and H. Tost (2018). "Resting-state brain network features associated with short-term skill learning ability in humans and the influence of N-methyl-d-aspartate receptor antagonism." Network Neuroscience **2**(4): 464-480.

Zarringhalam, K., D. Degras, C. Brockel and D. Ziemek (2018). "Robust phenotype prediction from gene expression data using differential shrinkage of co-regulated genes." Sci Rep **8**(1): 1237.

Zeilinger, S., B. Kuhnel, N. Klopp, H. Baurecht, A. Kleinschmidt, C. Gieger, S. Weidinger, E. Lattka, J. Adamski, A. Peters, K. Strauch, M. Waldenberger and T. Illig (2013). "Tobacco smoking leads to extensive genome-wide changes in DNA methylation." PLoS One **8**(5): e63812.

Zhang, C., Y. Mao and L. Song (2018). "Precise treatments for schizophrenia: where is the way forward?" General psychiatry **31**(1).

Zhang, H. and J. Gelernter (2017). "Review: DNA methylation and alcohol use disorders: Progress and challenges." Am J Addict **26**(5): 502-515.

Zhang, L., C. Lv, Y. Jin, G. Cheng, Y. Fu, D. Yuan, Y. Tao, Y. Guo, X. Ni and T. Shi (2018). "Deep Learning-Based Multi-Omics Data Integration Reveals Two Prognostic Subtypes in High-Risk Neuroblastoma." Front Genet **9**: 477.

Zhang, W., S. Emrich and E. Zeng (2010). "A two-stage machine learning approach for pathway analysis." IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE: 274-279.

Zhang, X., J. Yazaki, A. Sundaresan, S. Cokus, S. W. Chan, H. Chen, I. R. Henderson, P. Shinn, M. Pellegrini, S. E. Jacobsen and J. R. Ecker (2006). "Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis." Cell **126**(6): 1189-1201.

Ziller, M. J., H. Gu, F. Muller, J. Donaghey, L. T. Tsai, O. Kohlbacher, P. L. De Jager, E. D. Rosen, D. A. Bennett, B. E. Bernstein, A. Gnirke and A. Meissner (2013). "Charting a dynamic DNA methylation landscape of the human genome." Nature **500**(7463): 477-481.

6 PUBLICATION

Manuscripts under review

Chen, J., Wang, H. and Schwarz, E. (2019). BioMM: Biologically-informed multi-stage machine learning for phenotype prediction using epigenetic data. BMC bioinformatics

Braun, U., Harneit, A., Pergola, G., Menara, T., Schaefer, A., Betzel, R.F., Zang, Z., Schweiger, J.I., Schwarz, K., **Chen, J.** and Blasi, G., 2019. Brain state stability during working memory is explained by network control theory, modulated by dopamine D1/D2 receptor function, and diminished in schizophrenia. Nature Communication.

Zhang, X., Braun, U., Harneit, A., Zang, Z., Geiger, L., Betzel, R., **Chen, J**., Schweiger, J., Schwarz, K., Reinwald, J.R. and Fritze, S., 2019. Generative network models identify biological mechanisms of altered structural brain connectivity in schizophrenia. Schizophrenia Bulletin

Published manuscripts

Chen, J., Cao, H., Kaufmann, T., Westlye, L. T., Tost, H., Meyer-Lindenberg, A., & Schwarz, E. (2020). Identification of Reproducible BCL11A Alterations in Schizophrenia Through Individual-Level Prediction of Coexpression. Schizophrenia Bulletin.

Chen, J., Zang, Z., Braun, U., Schwarz, K., Harneit, A., Kremer, T., ... & Schwarz, E. (2020). Association of a Reproducible Epigenetic Risk Profile for Schizophrenia With Brain Methylation and Function. JAMA psychiatry.

Foo, J. C., Trautmann, N., Sticht, C., Treutlein, J., Frank, J., Streit, F., ... & Chen, J. (2019). Longitudinal transcriptome-wide gene expression analysis of sleep deprivation treatment shows involvement of circadian genes and immune pathways. Translational psychiatry, 9(1), 1-10.

Peterson, R.E., Kuchenbaecker, K., Walters, R.K., Chen, C.Y., Popejoy, A.B., Periyasamy, S., Lam, M., Iyegbe, C., Strawbridge, R.J., Brick, L. **Chen, J.** ... and Carey, C.E., 2019. Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. Cell.

Chen, J., Cao, H., Meyer-Lindenberg, A. and Schwarz, E., 2018. Male increase in brain gene expression variability is linked to genetic risk for schizophrenia. *Translational psychiatry*, *8*(1), p.140.

Chen, J., Lippold, D., Frank, J., Rayner, W., Meyer-Lindenberg, A. and Schwarz, E., 2018. Gimpute: an efficient genetic data imputation pipeline. *Bioinformatics*, *35*(8), pp.1433-1435.

Chen, J. and Schwarz, E., 2017. BioMM: Biologically-informed Multi-stage Machine learning for identification of epigenetic fingerprints. *arXiv preprint arXiv:1712.00336*. (Submission to NIPS ML4H 2017)

Chen, J. and Schwarz, E., 2017. Opportunities and challenges of multiplex assays: a machine learning perspective. In *Multiplex Biomarker Techniques* (pp. 115-122). Humana Press, New York, NY.

Cao, H., **Chen, J.**, Meyer-Lindenberg, A. and Schwarz, E., 2017. A polygenic score for schizophrenia predicts glycemic control. *Translational psychiatry*, *7*(12), pp.1-9.

Chen, J. and Schwarz, E., 2017. The role of blood-based biomarkers in advancing personalized therapy of schizophrenia. *Expert Review of Precision Medicine and Drug Development*, *2*(6), pp.363-370.

Chen, J., Guest, P.C. and Schwarz, E., 2017. The utility of multiplex assays for identification of proteomic signatures in psychiatry. In *Proteomic Methods in Neuropsychiatric Research* (pp. 131-138). Springer, Cham.

Chen, J., Lutsik, P., Akulenko, R., Walter, J. and Helms, V., 2014. AKSmooth: Enhancing low-coverage bisulfite sequencing data via kernel-based smoothing. *Journal of bioinformatics and computational biology*, *12*(06), p.1442005.

7 CURRICULUM VITAE

Personal Information

Name und Vorname: Junfang Chen

Geburtsdatum: January 24th 1986

Geburtsort: Putian, China

Familienstand: Single

Vater: Jinxu Chen

Mutter: Jinying Chen

Acedemic Education

02/2015 – present	Graduate student, Research group of translational bioinformatics in psychiatry, Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University Supervisor: Dr. Emanuel Schwarz
10/2010 – 08/2014	MSc. in Bioinformatics, Saarland University, Germany, Grade : 1.9, Good. Master thesis title: Whole human colon methylome profiling using low-coverage bisulfite sequencing data.
10/2005 – 07/2009	BSc. in Bioinformatics, Fujian Agriculture and Forest University, China, Grade: 82/100, Good.

8 ACKNOWLEDGEMENT

First and foremost, I would like to express my sincere gratitude to my fantastic mentor, Emanuel Schwarz. I am extremely honored to your student and work on this cutting-edge project under your supervision over the past five years. I have learned so much from your scientific ability and knowledge; I'm impressed by your speedy feedback, guidance and thorough advices to my manuscripts and thesis. What I highly appreciate is that you have provided sufficient freedom during my research.

I also dedicate my gratitude to Andreas Meyer-Lindenberg for your supervision and scientific support. In particular, your instructive feedback and comments on the kickoff meeting and the manuscripts.

I want to acknowledge Heike Tost for providing very nice and highly interactive scientific environment, and a diversity of genetic and imaging data. Without your support, my projects would not have been possible.

I am especially grateful to Marcella Rietschel for giving insightful suggestions and providing me opportunities to work on 'big GWAS data' and the possibility of intriguing collaborations.

I want to thank all members from the RG systems Neuroscience and friends for their support and cooperation. In particular, I would thank Urs for sharing valuable suggestions and Zhenxiang for the imaging data analysis. I own special thanks to Fabian and Jerome for your precious advices and critical revision. I want to thank Luanna for your revision and proofreading. I would like to thank Edda, Mimi, Anais, Janina, Caroline, Thomas, Xiaolong, Ren, Lena, Josef, Zhimin, Hongcai, Dietmar, Haohan for your discussion and help during my PhD.

Special thanks is dedicated to my wonderful friends who I met in Saarland and old friends in China for you keeping my spirits up and insightful discussion.

Last but most, I owe deep thanks to my dearest family/relatives for your persistent and priceless support.