Genomics Yields Insights into Bipolar Disorder Architectures: Pleiotropic Genes and Polygenic Burdens

Tracey van der Veen

Division of Psychiatry
University College London

Thesis submitted for the degree of Doctor of Philosophy

29th of June 2025

Declaration

I, Tracey van der Veen confirm that the work presented in this thesis is my own carried out under the supervision of Professor Andrew McQuillin and Dr Nick Bass at the Molecular Psychiatry Laboratory (MPL), Department of Psychiatry, University College London, and Dr. Francis McMahon, Genetic Basis of Mood and Anxiety Disorders (GBMAD), Humans Genetic Branch, NIMH, USA, as part of the four-year UCL-NIMH Joint Doctoral Training Program in Neuroscience [RRDNEUS4MH01].

Tracey van der Veen
29th of June 2025

Acknowledgements

Firstly, I would like to thank my supervisor Professor Andrew McQuillin, PhD (UCL) for giving me the opportunity to pursue this research for his valuable support. Many thanks to my secondary supervisors Dr Francis McMahon, PhD, MD (NIH/NIMH) and Dr Nick Bass PhD, MD (UCL) for their critique and constructive advice of the work. I would also like to thank Dr Maria Grigoroiu-Serbanescu, PhD, Psychiatric Genetics Research Unit, Alexandru Obregia Clinical Psychiatric Hospital, Bucharest, Romania, for her advice and guidance, as well as Professor Howard Edenberg, PhD, Department of Biochemistry and Molecular Biology and the Department of Medical and Molecular Genetics, Indianapolis, Indiana, USA, for his timely advice. Lastly, I would like to thank Dr Kevin O'Connell, PhD, Centre for Precision Psychiatry, Institute of Clinical Medicine, University of Oslo, Oslo, Norway, for his direction during our collaboration. An adapted version of Tabea Schoeler's GitHub code was used when producing Figures in Chapter 4. Deze scriptie is opgedragen aan mijn man, Oscar, voor zijn niet aflatende steun en aanmoediging gedurende dit project. Hartelijke dank voor alles wat je voor me hebt gedaan.

UCL Research Paper Declaration Forms

1. For a research manuscript that has already been published (if not yet published, please skip to section 2)	
What is the title of the manuscript	
Where was the work published?	
Who published the work?	
When was the work published?	
List the manuscript's authors in the order they appear on the publication	
Was the work peer reviewed?	
Have you retained the copyright?	
Was an earlier form of the manuscript uploaded to a preprint server? (e.g. medRxiv).	
If 'No', please seek permission from the relevant publisher and check the box next to the below statement: ☐ *I acknowledge permission of the publisher named under 1d to include in this thesis portions of the publication named as included in 1c.*	
2. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)	
What is the current title of the manuscript?	Adverse Psychosocial Trajectory in Bipolar Disorder: Novel Genetic Links to ADHD and Anxiety
Has the manuscript been uploaded to a preprint server? (e.g. medRxiv; if 'Yes', please give a link or doi)	**doi:** https://doi.org/10.1101/2025.05.17.25327825
Where is the work intended to be published? (e.g. journal names)	*Psychological Medicine*
List the manuscript's authors in the intended authorship order	T. van der Veen, N. Bass, A. McQuillin
Stage of publication (e.g. in submission)	In submission
3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4)	
4. In which chapter(s) of your thesis can this material be found? Chapter 3	
5. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)	
Candidate: Tracey van der Veen *Tracey van der Veen*	Supervisor: Andrew McQuillin 
Date: 25-06-2025	*Date:* 06-06 2025

1. For a research manuscript that has already been published (if not yet published, please skip to section 2)	
What is the title of the manuscript	Schizophrenia polygenic risk scores, clinical variables and genetic pathways as predictors of phenotypic traits of bipolar I disorder
Where was the work published?	Journal of Affective Disorders https://doi.org/10.1016/j.jad.2024.04.066
Who published the work?	Elsevier
When was the work published?	1 July 2024
List the manuscript's authors in the order they appear on the publication	Maria Grigoroiu-Serbanescu [a 1], Tracey van der Veen [b 1], Tim Bigdeli [c], Stefan Herms [d e], Carmen C. Diaconu [f], AnaIulia Neagu [f], Nicholas Bass [b], Johan Thygesen [b g], Andreas J. Forstner [e], Markus M. Nöthen [e], Andrew McQuillin [b]
Was the work peer reviewed?	Yes
Have you retained the copyright?	Yes
Was an earlier form of the manuscript uploaded to a preprint server? (e.g. medRxiv).	No
If 'No', please seek permission from the relevant publisher and check the box next to the below statement: ☐ *I acknowledge permission of the publisher named under 1d to include in this thesis portions of the publication named as included in 1c.*	
2. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)	
What is the current title of the manuscript?	
Has the manuscript been uploaded to a preprint server? (e.g. medRxiv; if 'Yes', please give a link or doi)	
Where is the work intended to be published? (e.g. journal names)	
List the manuscript's authors in the intended authorship order	
Stage of publication (e.g. in submission)	Published
3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4)	
4. In which chapter(s) of your thesis can this material be found? Chapter 4	
5. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)	
Candidate: Tracey van der Veen *Tracey van der Veen*	Supervisor: Andrew McQuillin [redacted]
Date: 25-06-2025	*Date:* 06-06 2025

1. For a research manuscript that has already been published (if not yet published, please skip to section 2)	
What is the title of the manuscript	
Where was the work published?	
Who published the work?	
When was the work published?	
List the manuscript's authors in the order they appear on the publication	
Was the work peer reviewed?	
Have you retained the copyright?	
Was an earlier form of the manuscript uploaded to a preprint server? (e.g. medRxiv).	
If 'No', please seek permission from the relevant publisher and check the box next to the below statement: ☐ *I acknowledge permission of the publisher named under 1d to include in this thesis portions of the publication named as included in 1c.*	
2. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)	
What is the current title of the manuscript?	Immune, Developmental, and Synaptic Pathways Define Bipolar Disorder Clinical Heterogeneity
Has the manuscript been uploaded to a preprint server? (e.g. medRxiv; if 'Yes', please give a link or doi)	**doi:** https://doi.org/10.1101/2025.06.23.25330155
Where is the work intended to be published? (e.g. journal names)	*JAMA Psychiatry*
List the manuscript's authors in the intended authorship order	Bipolar Disorder Working Group of the Psychiatric Genomics Consortium, Tracey van der Veen, Markos Tesfaye…Francis J McMahon, Ole A Andreassen, Arianna Di Florio, Andrew McQuillin
Stage of publication (e.g. in submission)	In submission
3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4)	
4. In which chapter(s) of your thesis can this material be found? Chapter 5	
5. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)	
Candidate: Tracey van der Veen *Tracey van der Veen*	Supervisor: Andrew McQuillin [redacted]
Date: 25-06-2025	*Date:* 06-06 2025

1. For a research manuscript that has already been published (if not yet published, please skip to section 2)	
What is the title of the manuscript	Genomics yields biological and phenotypic insights into bipolar disorder
Where was the work published?	*Nature* https://doi.org/10.1038/s41586-024-08468-9
Who published the work?	Nature
When was the work published?	22 January 2025
List the manuscript's authors in the order they appear on the publication	Kevin S O'Connell, Maria Koromina, Tracey van der Veen…
Was the work peer reviewed?	Yes
Have you retained the copyright?	Yes
Was an earlier form of the manuscript uploaded to a preprint server? (e.g. medRxiv).	Yes
If 'No', please seek permission from the relevant publisher and check the box next to the below statement: ☒ *I acknowledge permission of the publisher named under 1d to include in this thesis portions of the publication named as included in 1c.*	
2. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)	
What is the current title of the manuscript?	
Has the manuscript been uploaded to a preprint server? (e.g. medRxiv; if 'Yes', please give a link or doi)	
Where is the work intended to be published? (e.g. journal names)	
List the manuscript's authors in the intended authorship order	
Stage of publication (e.g. in submission)	Published
3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4) The analytical team, led by K.S.O., was responsible for the main analyses presented in the paper. The bulk of the Polygenic risk score associations was conducted by T.v.d.V. (scripts written together with T.B.).	
4. In which chapter(s) of your thesis can this material be found? Chapter 6	
5. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)	
Candidate: Tracey van der Veen *Tracey van der Veen*	Supervisor Andrew McQuillin [redacted]
Date: 25-06-2025	*Date:* 06-06 2025

Thesis Abstract

Background: Genome-wide association studies (GWAS) have enhanced the understanding of the genetics of bipolar disorder (BD), yet its profound clinical and genetic heterogeneity remains a major obstacle to diagnosis and treatment. The wide range of clinical presentations, partly driven by high rates of comorbidity and individual variability, can obscure genetic discoveries and complicates the search for reliable biomarkers.

Aims: This thesis aims to deconstruct the clinical heterogeneity of BD by identifying novel dimensional frameworks by dissecting the genetic architecture of specific clinical subphenotypes. The central goal is to identify distinct genetic mechanisms and biological pathways that can inform biomarker discovery, advance precision medicine, and identify potentially new functional genomic targets.

Methods: This research employed a multi-stage approach, beginning with the development of a dimensional model of BD psychopathology that integrated premorbid factors (Chapter 3). Subsequent analyses utilized large-scale genetic data to assess transdiagnostic risk from schizophrenia (Chapter 4), delineate the genetic architecture of 11 distinct clinical subphenotypes using Multi-Trait Analysis of GWAS (MTAG) (Chapter 5), and evaluate the impact of ascertainment and ancestry on polygenic prediction (Chapter 6).

Results: A novel 'Adverse Chronic Trajectory' (ACT) dimension was identified, potentially linking premorbid neurodevelopmental deficits to chronic BD outcomes; this dimension was genetically associated with polygenic risk for ADHD and anxiety, not core BD. Multi-trait analyses of eleven subphenotypes revealed four underlying genetic dimensions, including a 'Severe Illness' dimension defined by a unique neuro-immune signature (a protective association with the human Major Histocompatibility Complex (MHC) Human Leukocyte Antigen, Class II, DM Alpha gene *(HLA-DMA*) and specific risk loci Sodium Voltage-Gated Channel Alpha Subunit 2 (*SCN2A*), and a 'Comorbidity' dimension linked to neurodevelopmental genes such as Deleted in Colorectal Carcinoma (*DCC*). Further analyses demonstrated that the predictive power of polygenic scores is substantially influenced by both patient ascertainment strategies and genetic ancestry.

Conclusions: This thesis advances the understanding of BD's genetic architecture by providing a biological framework that helps explain its clinical diversity. The identification of distinct genetic dimensions and subphenotype-specific pathways begins to address the "hidden heritability" challenge by revealing previously obscured genetic mechanisms. These findings offer novel, biologically grounded hypotheses for future research and lay the groundwork for developing stratified, personalized treatment strategies in the pursuit of precision psychiatry.

Impact Statement

The profound clinical and genetic heterogeneity of bipolar disorder (BD) presents a formidable challenge for research and treatment, a problem now being addressed by large-scale genomic analyses that provide new biological insights. As a leading cause of disability worldwide, individuals with BD experience a suicide risk many times higher than that of the general population, and a reduction in life expectancy [7]. The disorder's complex aetiology, involving substantial genetic contributions (estimated at 85-89% of heritability) [1], and environmental factors, presents challenges. The clinical course of the illness underscores these difficulties, as many patients in long-term outpatient care experience high rates of relapse and struggle to achieve full functional recovery. Treatment is also complicated by high rates of comorbidity; most individuals reported one or more other psychiatric or medical conditions in a comprehensive survey of BD. Lifetime, (and 12-month) prevalence estimates are 1.0% (.6%) for bipolar disorder I (BD1), 1.1% (.8%) for bipolar disorder II (BD2), and 2.4% (1.4%) for subthreshold symptoms [53]. Consequently, many patients do not achieve an adequate response to first- or second-line medications. While research has identified numerous genetic variants associated with BD, the disorder's profound clinical heterogeneity makes pinpointing causal genes and developing targeted treatments exceptionally difficult.

This thesis directly confronts this challenge by investigating the clinical diversity of BD and its genetic underpinnings. By seeking to elucidate the aetiology of the disorder, the overarching aim is to lay the scientific groundwork for future advancements in prevention strategies, diagnostic precision, and treatment options, ultimately to enhance the quality of life (QoL) for those affected.

The key contributions of this research are summarised below. Methodologically, this thesis demonstrates that a subphenotypic approach can advance genomic discovery in BD. By leveraging a multi-trait analysis and deconstructing the disorder's heterogeneity, this work yielded 53 novel risk loci and incrementally improved polygenic risk prediction, providing a direct, evidence-based strategy for addressing the 'missing heritability' in BD.

This granular approach allowed for the identification of several distinct clinical-genetic profiles. The results provide strong evidence for an 'Adverse Chronic Trajectory' (ACT) (Chapter 3), a dimension linking premorbid factors to a chronic course, which was uniquely predicted by polygenic risk for attention-deficit/hyperactivity disorder (ADHD) and anxiety rather than core BD. This suggests a distinct biological basis for this challenging trajectory. Furthermore, this research advances risk stratification by showing that schizophrenia (SCZ) polygenic risk scores can predict severe outcomes including psychosis and earlier onset in BD. Individual-level pathway analysis of these findings implicates specific biological mechanisms, such as mitochondrial dysfunction, as potential markers of severe illness (Chapter 4).

Ultimately, this thesis proposes a new framework for understanding BD, deconstructing it into genetically-informed dimensions distinguished by unique biological signatures, such as a neuro-immune profile for severe illness and specific neurodevelopmental pathways for comorbid forms (Chapter 5). By moving beyond a monolithic view of the disorder (Chapter 6), this work lays a crucial foundation for more comprehensive etiological models. The implications for designing targeted, genetically-informed clinical trials and enhancing public health awareness of BD's complexity underscore the broad relevance of this research and warrant its dissemination to the wider scientific community.

Note: For references see Section 8.1.

This thesis presents my own account of investigations, the entirety of which were undertaken during the period of research supervision. This demonstrates my ability to design and implement several independent research projects, outlined in Chapters 3 to 6.

List Of Publications Resulting From PhD

Genomics yields biological and phenotypic insights into bipolar disorder

Nature
2025-01-22 | Journal article
DOI: 10.1038/s41586-024-08468-9
Part of ISSN: 0028-0836
CONTRIBUTORS: Kevin S O'Connell; Maria Koromina; Tracey van der Veen; Toni Boltz; Friederike S. David; Jessica Mei Kay Yang; Keng-Han Lin; Xin Wang; Jonathan R. I. Coleman; Brittany L. Mitchell *et al*.

Schizophrenia polygenic risk scores, clinical variables and genetic pathways as predictors of phenotypic traits of bipolar I disorder

Journal of affective disorders
2024-04-18 | Journal article
DOI: 10.1016/j.jad.2024.04.066
PMID: 38640977
CONTRIBUTORS: Grigoroiu-Serbanescu M; van der Veen T; Bigdeli T; Herms S; Diaconu CC; Ana Neagu; Bass N; Thygesen J; Forstner AJ; Nöthen MM *et al*.

Predicting ADHD in alcohol dependence using polygenic risk scores for ADHD

American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics
2023-11-09 | Journal article
DOI: 10.1002/ajmg.b.32967
PMID: 37946686
CONTRIBUTORS: Kejal Harish Sachin Patel; G. Bragi Walters; Stefánsson H; Stefánsson K; Degenhardt F; Nothen M; Van Der Veen T; Ditte Demontis; Borglum A; Kristiansen M *et al.*

Multiple psychiatric polygenic risk scores predict associations between childhood adversity and bipolar disorder

Journal of affective disorders
2023-08-27 | Journal article
DOI: 10.1016/j.jad.2023.08.116
PMID: 37643680
CONTRIBUTORS: Kai Yao; van der Veen T; Thygesen J; Bass N; McQuillin A

GWAS IDENTIFIES NOVEL LOCI ASSOCIATED WITH SPECIFIERS OF BIPOLAR DISORDER

European Neuropsychopharmacology
2024-10-01 | Abstract.
DOI:10.1016/j.euroneuro.2024.08.189

CONTRIBUTORS: Shane Crinion, Tracey van der Veen, Markos Tesfaye, Arianna Di Florio, Andrew McQuillin, Niahm Mullins, Psychiatric Genomics Consortium Bipolar Disorder Working Group

Publications submitted or in preprint:

Immune, Developmental, and Synaptic Pathways Define Bipolar Disorder Clinical Heterogeneity

2025-06-26 | Preprint. Submitted to *JAMA Psychiatry.*
DOI: https://doi.org/10.1101/2025.06.23.25330155
CONTRIBUTORS: Tracey van der Veen, Markos Tesfaye, Arianna Di Florio, Andrew McQuillin, Psychiatric Genomics Consortium Bipolar Disorder Working Group

Implication of the *ADCY1* Gene in Lithium Response in Bipolar Disorder by Genome-wide Association Meta-analysis

2024-03-12 | Preprint.
DOI: 10.21203/rs.3.rs-4000581/v1
CONTRIBUTORS: Andrew McQuillin; Kai Yao; Ayeda Nadeem; Tracey Van Der Veen; Johan H Thygesen; Lina Jonsson; Mikael Landén; Jie Song; Nick Bass

Adverse Psychosocial Trajectory in Bipolar Disorder: Novel Genetic Links to ADHD and anxiety

2025-05-18 | Preprint. Submitted to *Psychological Medicine.*
DOI: https://doi.org/10.1101/2025.05.17.25327825
CONTRIBUTORS: Tracey van der Veen; Nick Bass; Andrew McQuillin

Table of Contents

List of Figures

List of Tables

List of Abbreviations

1H-MRS: Proton Magnetic Resonance Spectroscopy **2-AG**: 2-Arachidonoylglycerol **5-HIAA**: 5-Hydroxyindoleacetic Acid **5-HT3AR**: 5-Hydroxytryptamine 3A Receptor **ACC**: Anterior Cingulate Cortex **ACTH**: Adrenocorticotropic Hormone ***ADCY2***: Adenylate Cyclase 2 **ADHD**: Attention-Deficit/Hyperactivity Disorder **AFR**: African (ancestry) **AI**: Anterior Insula ***AKAP11***: A-Kinase Anchoring Protein 11 **AlcSUD**: Alcohol or Substance Use Disorder ***ANK3***: Ankyrin-G **ANX**: Anxiety Disorders **AO**: Age of Onset **AOO**: Age of Onset (used interchangeably with AO) **APA**: American Psychiatric Association **ACT**: Adverse Chronic Trajectory **ASD**: Autism Spectrum Disorder **ASPD**: Antisocial Personality Disorder **ATP**: Adenosine Triphosphate **AUC**: Area Under the ROC curve **AUD**: Alcohol Use Dependency **BD**: Bipolar Disorder **BD1**: Bipolar Disorder I **BD2**: Bipolar Disorder II ***BDNF***: Brain-Derived Neurotrophic Factor **BD-NOS**: Bipolar Disorder Not Otherwise Specified **BDWG**: Bipolar Disorder Working Group (PGC) **BipEx**: Bipolar Exome consortium **BPD**: Borderline Personality Disorder **CAD**: Cardiovascular Disease (or Coronary Artery Disease, context-dependent) **CADD**: Combined Annotation-Dependent Depletion **cAMP**: Cyclic Adenosine Monophosphate **CASTom-iGEx**: Context-Aware Stratification based on Tissue-specific imputed Gene Expression **CBT**: Cognitive Behavioural Therapy **CCBs**: Calcium Channel Blockers **CFA**: Confirmatory Factor Analysis **CFI**: Comparative Fit Index **CGI**: CpG Island **CI**: Confidence Interval **CNS**: Central Nervous System **CNV**: Copy Number Variant **CpG2**: Candidate Plasticity Gene 2 ***CREB***: cAMP Response Element-Binding Protein **CS**: Continuous Shrinkage (PRS method) **CSF**: Cerebrospinal fluid **CTG-VL**: Complex Trait Genetics Virtual Lab platform **CVD**: Cardiovascular Diseases **dACC**: Dorsal Anterior Cingulate Cortex **DAG**: Diacylglycerol ***DAGLA***: Diacylglycerol Lipase Alpha ***DAOA***: D-amino Acid Oxidase Activator ***DAT***: Dopamine Transporter **DBT**: Dialectical Behaviour Therapy ***DCC***: Deleted in Colorectal Carcinoma (gene) ***DCLK3***: Doublecortin Like Kinase 3 **DCNN**: Deep Convolutional Neural Network **DE**: Differentially Expressed **DEEN**: Deep Ensemble Encoder Networks ***DGKH***: Diacylglycerol Kinase Eta **DIGS**: Diagnostic Interview for Genetic Studies ***DISC1***: Disrupted in Schizophrenia 1 **DLPFC**: Dorsolateral Prefrontal Cortex **DNA**: Deoxyribonucleic Acid **DNMs**: De Novo Mutations ***DRD4***: Dopamine Receptor D4 **DSM**: Diagnostic and Statistical Manual of Mental Disorders **DSM-IV**: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition **DSM-5**: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition **DUP**: Duration of Untreated Psychosis **DZ**: Dizygotic (twins) **EAS**: East Asian (ancestry) **ED-DUB-STEP2**: Edinburgh-Dublin-Systematic Treatment Enhancement Program 2 **EEG**: Electroencephalography **EFA**: Exploratory Factor Analysis **EHR**: Electronic Health Records **E/I**: Excitatory-Inhibitory **ENIGMA**: Enhancing Neuro Imaging Genetics through Meta Analysis **EOBD**: Early Onset Bipolar Disorder **EPRS**: Ensemble learning-based Polygenic Risk Score **eQTL**: Expression Quantitative Trait Loci **EUR**: European (ancestry) **FA**: Fractional Anisotropy **FADH2**: Flavin Adenine Dinucleotide (reduced form) ***FADS1***: Fatty Acid Desaturase 1 ***FADS2***: Fatty Acid Desaturase 2 **FDR**: False Discovery Rate **FEP**: First-Episode Psychosis **FFT**: Family-Focused Therapy ***FOX***: Forkhead box (gene family) ***FOXO6***: Forkhead Box O6 **FUMA**: Functional Mapping and Annotation of GWAS ***FURIN***: Furin Paired Basic Amino Acid Cleaving Enzyme **GAF**: Global Assessment of Functioning **GABA**: Gamma-Aminobutyric Acid **GABAAR**: GABA A Receptor ***GABRB1***: Gamma-Aminobutyric Acid Type A Receptor Subunit Beta1 **GBD**: Global Burden of Disease **GBS**: Gene-Based Burden Scores ***GLP2R***: Glucagon-Like Peptide-2 Receptor **GO**: Gene Ontology **GPCRs**: G Protein-Coupled Receptors **GR**: Glucocorticoid Receptor ***GRM7***: Glutamate Metabotropic Receptor 7 **GSA (analysis)**: Gene Set Analysis **GSA (array)**: Global Screening Array **GSEA**: Gene Set Enrichment Analysis ***GSK-3***: Glycogen Synthase Kinase 3 **GWAS**: Genome-Wide

Association Studies **GxE**: Gene-Environment Interactions **HLA**: Human Leukocyte Antigen **HPA**: Hypothalamic-Pituitary-Adrenal **HRC**: Haplotype Reference Consortium **h²snp**: SNP-based heritability **IC**: Intracellular **ICD**: International Classification of Diseases **ICD-9**: International Classification of Diseases, Ninth Revision **ICD-10**: International Classification of Diseases, Tenth Revision **ICD-11**: International Classification of Diseases, Eleventh Revision **IL-6**: Interleukin-6 **IMPase**: Inositol Monophosphatase **IP3**: Inositol Trisphosphate **IPPase**: Inositol Polyphosphate-1-Phosphatase **IPW**: Inverse Probability Weighting **IQR**: Interquartile Range ***ITIH1/3/4***: Inter-Alpha-Trypsin Inhibitor Heavy Chain 1/3/4 ***KCNC2***: Potassium Voltage-Gated Channel Subfamily C Member 2 **KMO**: Kaiser-Meyer-Olkin Criterion **LAT**: Latino (ancestry) **LAVA**: Local Analysis of [Co]Variant Association **LD**: Linkage Disequilibrium **LDSC**: LD Score Regression **lncRNA**: Long Non-Coding RNA **LTP**: Long-Term Potentiation ***MAD1L1***: Mitotic Arrest Deficient 1 Like 1 **MAF**: Minor Allele Frequency **MAGMA**: Multi-marker Analysis of GenoMic Annotation **MAPK**: Mitogen-Activated Protein Kinase **MAR**: Missing at Random ***MARCKS***: Myristoylated Alanine-Rich C Kinase Substrate **maxFDR**: Maximum False Discovery Rate **MDA**: Mean Decrease Accuracy **MDD**: Major Depressive Disorder **MHC**: Major Histocompatibility Complex **MIMIC**: Multiple Indicator Multiple Cause **ML**: Maximum Likelihood **MNAR**: Missing Not at Random **MOFA**: Multi-Omics Factor Analysis **MPS**: Multiple Polygenic Risk Score **MR**: Mendelian Randomization **MRI**: Magnetic Resonance Imaging **mRNA**: Messenger Ribonucleic Acid **MSigDB**: Molecular Signatures Database **MTAG**: Multi-Trait Analysis of GWAS **MVMR**: Multivariable Mendelian Randomization **MZ**: Monozygotic (twins) **NagR2**: Nagelkerke's R-squared **NADH**: Nicotinamide Adenine Dinucleotide + Hydrogen ***NCAN***: Neurocan **Neff**: Effective sample size ***NEK4***: NIMA Related Kinase 4 **NGS**: Next-Generation Sequencing **NIMH**: National Institute of Mental Health **NMDA**: N-methyl-D-aspartate ***NRG1***: Neuregulin 1 ***NT-4***: Neurotrophin-4 **OCD**: Obsessive-Compulsive Disorder **OFC**: Orbitofrontal Cortex **OPCRIT**: Operational Criteria checklist for psychotic illness **OPCs**: Oligodendrocyte Precursor Cells **OR**: Odds Ratio ***PACS1***: Phosphofurin Acidic Cluster Sorting Protein 1 ***PBRM1***: Polybromo 1 **PCA**: Principal Component Analysis **PD**: Panic Disorder **PFC**: Prefrontal Cortex **PGC**: Psychiatric Genomics Consortium **PI**: Phosphatidylinositol **PKA**: Protein Kinase A **PKC**: Protein Kinase C **PPV**: Positive Predictive Value **PRIMED**: Population Architecture using Genomics and Epidemiology **PRS**: Polygenic Risk Score(s) **PRS-CS**: Polygenic Risk Score - Continuous Shrinkage **PRS-CSx**: Polygenic Risk Score - Cross-ancestry **pT**: *P*-value Threshold (PRS method) **PTSD**: Post-Traumatic Stress Disorder **PTVs**: Protein-Truncating Variants **QC**: Quality Control **QEgger**: Cochran's Q for MR-Egger **QoL**: Quality of Life **RC**: Rapid Cycling **RC-BD**: Rapid Cycling Bipolar Disorder **RDoC**: Research Domain Criteria **RF**: Random Forest **rg**: Genetic Correlation **RICE**: Risk Calculation for Integration of Common and Rare variants **RMSE**: Root Mean Squared Error **RMSEA**: Root Mean Square Error of Approximation **RNA-seq**: RNA Sequencing **RO**: Romania(n) **ROC**: Receiver Operating Characteristic **ROS**: Reactive Oxygen Species **SA**: Suicide Attempt **SADS-L**: Schizophrenia and Affective Disorder Schedule-Lifetime **SBayesS**: Summary-data-based BayesS **SCHEMA**: Schizophrenia Exome Meta-analysis ***SCN2A***: Sodium Voltage-Gated Channel Alpha Subunit 2 **SCZ**: Schizophrenia **SCZ3**: Third Schizophrenia GWAS by PGC **scRNA-seq**: Single-Cell RNA-sequencing **SD**: Standard Deviation **s.e.**: Standard Error **s.e.m.**: Standard Error of the Mean **SEM**: Structural Equation Modeling **SES**: Socioeconomic Status **sgACC**: Subgenual Anterior Cingulate Cortex **SI**: Suicidal Ideation ***SLC6A3***: Solute Carrier Family 6 Member 3 **SLEs**: Stressful Life Events **SNP**: Single Nucleotide Polymorphism **SNV**: Single Nucleotide Variant ***SP4***: Sp4 Transcription Factor **SSRI**: Selective Serotonin Reuptake Inhibitor **STEP-UCL**: Systematic Treatment Enhancement Program - University College London **STREGA**: Strengthening the Reporting of Genetic Association studies **SUD**: Substance Use Disorders ***SYN3***: Synapsin III **SZA**: Schizoaffective Disorder (bipolar type) **T3**: Triiodothyronine **TLI**: Tucker-Lewis Index **TMS**: Transcranial Magnetic

Stimulation ***TRANK1***: Tetratricopeptide Repeat and Ankyrin Repeat Containing 1 **TRD**: Treatment-Resistant Depression ***Trk***: Tyrosine Kinase (receptors) **TWAS**: Transcriptome-Wide Association Studies **UK**: United Kingdom **UM**: Unipolar Mania **VIF**: Variance Inflation Factor **VIP**: Vasoactive Intestinal Peptide **VLPFC**: Ventrolateral Prefrontal Cortex **vmPFC**: Ventromedial Prefrontal Cortex **VPA**: Valproic Acid **vPFC**: Ventral Prefrontal Cortex **VTA**: Ventral Tegmental Area **WES**: Whole-Exome Sequencing **WGCNA**: Weighted Gene Co-expression Network Analysis **WGS**: Whole-Genome Sequencing **WHO**: World Health Organization **WLSMV**: Weighted Least Squares Mean and Variance adjusted **WTCCC**: Wellcome Trust Case Control Consortium ***ZNF318***: Zinc Finger Protein 318.

1 Introduction

Bipolar disorder (BD) presents a psychiatric challenge, primarily due to its profound clinical and genetic heterogeneity. This complexity hinders accurate diagnosis, effective patient management, and the elucidation of underlying etiological mechanisms, ultimately complicating the development of individualised therapies.

This thesis presents a comprehensive investigation into the systematic dissection of this heterogeneity. The thesis is structured to first critically review current understanding of BD's diverse presentations, complex genetics, and research limitations (Chapter 1). Subsequent empirical chapters will develop dimensional models of BD psychopathology (Chapter 3); examine the impact of transdiagnostic polygenic risks on clinical outcomes (Chapter 4); delineate distinct and shared genetic architectures of numerous clinical subphenotypes (see section 1.1 below) through large-scale multi-trait analyses (Chapter 5); and evaluate methodological factors, including cohort ascertainment and ancestry, that influence polygenic risk prediction (Chapter 6). Finally, these diverse findings will be synthesised and their broader implications discussed (Chapter 7). Achieving a deeper, more nuanced understanding of these intricate layers is paramount for advancing the field towards the promise of precision psychiatry.

This first chapter, therefore, provides the crucial foundation for this structured inquiry by reviewing current knowledge, identifying research gaps, and culminating in an outline of the specific aims of this thesis, which endeavours to contribute novel and impactful insights into these crucial issues.

1.1 Bipolar Disorder

Bipolar disorder (BD) arises from a combination of genetic factors and environmental influences and exhibit high heritability. Twin studies have indicated heritability rates of 85% to 89%. Specifically, the rate was 85% with a narrow concordance (95% confidence interval [CI], .73-.93) and 89% with a broad concordance (95% CI, .61-1.0)). While research into the genetic basis of BD has advanced, the search for reliable biomarkers for diagnosis and treatment response continues. This endeavour is complicated by evidence that gene variants genetically associated with BD are also implicated in other psychiatric and human diseases. This challenge is notable given the high heritability estimates for BD [1], which have yet to fully translate into readily identifiable biomarkers. The pathophysiology of BD remains largely undetermined. Observed changes in cellular function and brain structure could suggest neurodevelopmental processes and neuroprogression, which may be associated with epigenetic alterations, mitochondrial dysfunction, neurotrophic factors, inflammation, and oxidative stress mechanisms, according to a selective review [2]. Magnetic Resonance Imaging (MRI) studies have corroborated these findings, showing reduced cortical thickness in widespread frontal and parietal regions among BD patients relative to healthy controls. The same study also found that

a longer duration of illness was specifically associated with reduced thickness in medial parietal and occipital regions [3].

Moreover, BD is characterised as both polygenic and pleiotropic, resulting in substantial bidirectional genetic influences with various other human diseases and traits, including cardiovascular disease (CVD), SCZ and intelligence [4-5]. Comorbidity contributes to the disorder's heterogeneity, affecting its clinical presentation, course, and treatment outcomes.

This can confound research results and hamper diagnoses and response to therapeutics. Although some medical comorbidities can be identified through testing, no specific laboratory test currently exists for BD.

Unpacking Bipolar Disorder: The Concept of Subphenotypes

In the context of bipolar disorder, a **subphenotype** refers to a more specific and relatively uniform subgroup of individuals who all share the broader diagnosis but are distinguished by a particular set of clinical features, patterns of illness, or biological markers. This approach acknowledges that bipolar disorder is not a monolithic entity but rather a heterogeneous condition with diverse presentations and underlying causes.

The core idea behind identifying subphenotypes is to move beyond the general diagnostic criteria of bipolar I or bipolar II disorder and delineate more homogeneous patient groups. This refined classification has significant implications for both research and clinical practice, with the ultimate goal of developing more personalized and effective treatments.

Key Characteristics Used to Define Bipolar Subphenotypes:

Researchers are exploring various characteristics to define these subgroups, often integrating clinical observations with genetic and neurobiological data. Some of the key areas of investigation for bipolar disorder subphenotypes include:

- **Clinical Course and Features:** This is one of the most common ways to categorize subphenotypes. Examples include:
    - **Presence or Absence of Psychosis:** Individuals with a history of psychotic symptoms (delusions or hallucinations) during mood episodes may represent a distinct subphenotype compared to those who have never experienced psychosis.
    - **Age of Onset:** Whether the disorder begins in adolescence or adulthood can signify different underlying mechanisms and long-term outcomes.
    - **Rapid Cycling:** Patients who experience four or more mood episodes within a single year fall into this well-established subphenotype, which often presents unique treatment challenges.
    - **Pattern of Inter-episode Remission:** The degree to which an individual returns to their baseline level of functioning between mood episodes can be a defining characteristic.

- **Comorbidity:** The presence of other co-occurring psychiatric conditions is another critical factor. A common example is the subphenotype of bipolar disorder with a comorbid anxiety disorder, which can influence both the presentation of the illness and the response to treatment.
- **Genetic and Familial Factors:** With advancements in genetic research, scientists are identifying specific genetic markers and polygenic risk scores (an individual's overall genetic predisposition) associated with certain clinical features of bipolar disorder. For instance, some subphenotypes may have a stronger genetic link to schizophrenia, while others may share more genetic overlap with major depressive disorder.
- **Neurobiological Markers:** While still largely in the research phase, efforts are underway to identify biological markers, such as specific patterns of brain activity or inflammation, that could help to objectively define different subphenotypes.

The Goal: From Subphenotype to "Endophenotype"

The identification of subphenotypes is a crucial step towards a deeper understanding of the biological underpinnings of bipolar disorder. The ultimate aim for researchers is to define **endophenotypes**. An endophenotype is a subphenotype that is linked to a specific, measurable biological mechanism. By understanding the distinct pathophysiology of these more uniform groups, clinicians can hope to develop targeted therapies that address the root cause of an individual's specific type of bipolar disorder, moving away from a one-size-fits-all approach to treatment.

1.2 BD Comorbidities

A 2024 review of 114 studies, conducted between 1993 and 2022, detailed frequent comorbid BD disorders such as anxiety, substance use disorders (SUD), Attention-Deficit/Hyperactivity Disorder (ADHD), and impulse-control disorders [6], alongside medical conditions including diabetes, metabolic syndrome, and cardiovascular diseases. For a recent review of BD comorbidities, see Oliva *et al*. (2025) [7]. For BD subphenotype prevalences, see literature in Chapter 2 Table 10 and Supplementary Table 58. Comorbidities are consequential as they may influence risk or resilience, affecting how individuals navigate environmental stressors that can provoke BD episodes, response to treatments and impact the disorder's progression [8]. A study on adolescent BD found that lower socioeconomic status (SES) was associated with a higher likelihood of comorbid disruptive behaviour disorders, anxiety disorders, substance use disorders, and a more severe clinical presentation of BD [9]. These factors, alongside comorbid ADHD and obsessive compulsive disorder (OCD) have been associated with a poorer prognosis, including rapid cycling, more severe illness and adverse functional outcomes [7].

While this thesis emphasises genetic risk factors, a variety of environmental influences likely also interact with genetic susceptibilities. Investigating the interplay between BD, its comorbidities, and the environment (Table 1) will be crucial for refining future diagnostic and treatment approaches.

Table 1 Putative Bipolar Disorder Risk Factors and Prodromal Symptoms

Category	Risk Factors	Relationship to Bipolar Disorder
Biological	Family history of bipolar disorder, Neurodevelopmental factors, Temperament, Specific Genes (e.g., *AKAP11*), Brain Structure and Functioning Differences, Neurotransmitter Imbalances	Increase the individual's vulnerability or predisposition to developing bipolar disorder. Research continues to identify specific genes and brain differences that may play a role. Imbalances in neurotransmitters like serotonin, dopamine, and norepinephrine are implicated.
Environmental	Adverse Childhood Experiences (ACEs) such as childhood, trauma, poverty, stress, sexual, physical abuse, neglect, witnessing violence or emotional abuse), Antidepressants, Major life transitions, or Sleep Deprivation	Research suggests a link between ACEs and increased risk. Can act as triggers for episodes or exacerbate the condition. Trauma and significant stress, especially in childhood, can have long-lasting effects. Disruptions in sleep patterns are a significant trigger. Substance misuse and disrupted sleep can also be consequences of environmental factors.
Dimensional	Psychosis, Hypo(mania), Sleep problems, Comorbidity (Anxiety and depressive symptoms, Mood lability, Early-onset Anxiety Disorders (panic disorder, separation anxiety, generalized anxiety), Conduct Problems/Disorder, ADHD, Impulsivity	Often co-occur with bipolar disorder, can be part of the diagnostic presentation, or indicate a greater severity or complexity. These conditions may also represent early clinical risk factors that precede the onset of full bipolar disorder.
Additional	Physical Health Conditions (e.g., thyroid issues, cardiovascular disease, obesity), Substance Use Disorders (as a primary condition), Seasonal Changes, Inflammation, Gut Microbiome	Certain physical health conditions and substance use disorders have a high rate of comorbidity and may influence risk. Seasonal changes can trigger episodes in some individuals. Emerging research is exploring the role of inflammation and the gut microbiome in mental health disorders, including bipolar disorder.

Adapted from Vieta *et al*. 2018, *Early Intervention in Bipolar Disorder* [10]. These categories of factors often interact and influence each other in the development and course of the disorder. For example, genetic predisposition might interact with environmental stressors to increase the risk of BD.

1.3 History And Classification Of BD

Early Differentiation from Schizophrenia and Depression

The classification of bipolar disorder (BD) relies on diagnostic criteria specified in the International Classification of Diseases (ICD) from the World Health Organization (WHO) [11] and the Diagnostic and Statistical Manual of Mental Disorders (DSM) from the American Psychiatric Association (APA) [12]. The origins of BD criteria trace back to Aristaeus of Cappadocia, a 1st-century Greek physician who described mania and melancholia as manifestations of a single disease, a concept later noted by Falret (1851) and Baillarger (1854) [13]. In 1899, psychiatrist Emil Kraepelin introduced the single concept of 'manic-depressive insanity' to describe cyclical mood states, distinguishing it from the chronic , deteriorating course of dementia praecox (now schizophrenia) based on long-term outcomes and episodic recovery patterns [14]. This pivotal contribution established a framework for distinguishing major psychotic disorders by their trajectory and marked a substantial step in psychiatric nosology by integrating various mood disorders into one unifying concept. This observation of remission as a distinguishing feature of BD has evolved, with current evidence indicating many individuals experience incomplete remission due to residual symptoms, often exacerbated by comorbid disorders [15]. This historical overview highlights the evolving understanding of BD, a crucial context for appreciating the heterogeneity this thesis aims to address.

Evolution of Subclassifications of Bipolar Disorders

The understanding of BD extends beyond a simple dichotomy with unipolar depression. Kleist and Leonhard first proposed subclassifying BD in 1957 to better differentiate it. Schizoaffective disorders were later categorized into schizoaffective bipolar type (SZA) and schizoaffective depressive type in the DSM-III-R in 1987. Individuals with schizoaffective bipolar type have a high risk for psychosis, characterized by symptoms such as hallucinations and delusions, alongside manic and depressive mood episodes [16].

Kraepelin's foundational work also informed our understanding of temperaments and mixed states. Akiskal, in 1998, expanded on these ideas, identifying specific temperaments and their associations with mood disorders, including cyclothymia [17]. Mendel first described hypomania in 1881 [18]. Later, Dunner *et al*. (1976) differentiated bipolar disorder type II (BD2) from bipolar disorder type I (BD1), noting that BD2 is characterized by depressive and hypomanic rather than manic episodes [19]. It remains uncertain if a labile-cyclothymic temperament is clinically distinct from BD2.

Griesinger first articulated the concept of rapid switching in 1845 [20] that foreshadowed the concept of rapid cycling. Dunner and Fieve established the formal definition of rapid cycling in the 1970s, describing a BD course involving four or more affective episodes within a year, which was typically unresponsive to lithium monotherapy [21]. In certain instances, rapid-cycling BD may manifest as mood shifts occurring over hours, a phenomenon termed ultrarapid cycling, associated with a more severe, treatment-resistant form of bipolar illness [22]. This can be accompanied by irritability, impulsivity, and suicidal behaviour, presenting diagnostic challenges, as ADHD and borderline personality disorder (BPD), as defined by the Diagnostic and Statistical Manual of Mental Disorders (DSM) , also exhibit similar mood fluctuations [23-24]. Interestingly, cases of prepubertal and early adolescent BD were distinguished from ADHD by mania-specific criteria, though both often displayed ultra-rapid or ultradian cycling [23]. In BD, mixed states refer to the simultaneous experience of manic and depressive symptoms, while rapid cycling describes four or more distinct mood episodes (mania, hypomania, or depression) within a year. However, mixed states may involve rapid-sequence manic and depressive symptoms [7]. For a comprehensive history of bipolar disorder subclassifications, refer to Angst and Marneros 2001 [13].

History of Diagnostic Criteria (ICD, DSM, RDoC)

BD clinical diagnosis relies on the presence, frequency, and severity of hypo(manic) and depressive symptoms. Three primary diagnostic systems are used in psychiatry today: the ICD from the World Health Organisation (WHO) [11], the DSM from the American Psychiatric Association [12], and the Research Domain Criteria (RDoC) from the National Institute of Mental Health (NIMH) [25]. The DSM originated in the United States in 1952, established by the APA to gather psychiatric hospital statistics. Subsequent revisions led to the current edition, DSM-5, published in 2013 [12], with the most recent update being the DSM-5-TR (Text

Revision), published in 2022 [26]. In contrast, the International Classification of Diseases (ICD) evolved as a global initiative for standardizing data across countries and timeframes. Its origins trace back to the 'Bertillon Classification of Causes of Death' (1893), developed by French statistician Jacques Bertillon. The WHO later adopted this classification, which evolved into ICD-10, published in 1992 [27], and the latest version, ICD-11, was adopted in 2019 and implemented in 2022 [28-29].

Differences exist between the ICD-10 and DSM-5. The DSM-5 targets mental disorders specifically, while the ICD-10 encompasses a broader range of physiological conditions. Additionally, they differ in classifying BD, particularly regarding manic episode frequency: the DSM-5 requires at least one hypo(manic) episode, while the ICD-10 specifies two affective disorder episodes, one of which must be hypo(manic). The DSM-5 acknowledges BD2, whereas the ICD-10 did not differentiate this subtype (though ICD-11 does).

Unlike the DSM and ICD initially, RDoC focuses on the required biological factors rather than solely on symptomatology [25, 30] and aims to address heterogeneity and comorbidity within current classifications. Although RDoC provides a valuable framework, it is not intended for clinical diagnosis of BD but seeks to inform future diagnostic criteria. While RDoC has moved research towards a dimensional approach of BD, it is complex and evolving and is yet to meaningfully impact clinical practice.

1.4 BD Classification Criteria And Course Specifiers

Both the DSM-5 and ICD-11 acknowledge BD1 and BD2. The ICD-11 adopts a dimensional symptom assessment approach, retaining the mixed episode diagnosis and subthreshold states eliminated by the DSM-5 [29]. Both systems require at least one hypomanic and one depressive episode for a BD2 diagnosis, defining hypo(manic) episodes by mood elevation or irritability combined with increased activity or other criteria. The three key subtypes recognized in both the ICD-11 and DSM-5-TR are BD1, BD2, and cyclothymic disorder. BD1 is characterized as a manic-depressive disorder potentially including psychotic features, while BD2 is defined by alternating depressive and less severe hypomanic episodes. Cyclothymic disorder features shorter depression and hypomania episodes. Additionally, a Bipolar Disorder Not Otherwise Specified (BD-NOS) category exists, identified by multiple depressive episodes. Diagnostic distinctions between BD1 and BD2 depend on manic and hypomanic episode severity and duration. BD1 is marked by full manic episodes; BD2 by hypomanic and major depressive episodes. BD2 often presents with higher depressive episode frequency compared to BD1, which has higher hospitalisation rates and more extreme mood episodes [7]. Differentiating schizoaffective disorder, bipolar type (SZA) from BD1 is relevant due to worse outcomes in SZA, including prolonged duration of untreated psychosis (DUP), greater illness severity, and poorer Global Assessment of Functioning (GAF) scores [31].

Bipolar Disorder I

Bipolar Disorder I (BD1) is diagnosed following at least one manic or mixed episode, without requiring preceding hypomanic or depressive episodes. The DSM-5 defines a manic episode as a distinct period of persistently elevated or irritable mood with increased activity for at least one week or necessitating hospitalisation. Confirmation requires three or more of the following symptoms (four if irritability is present): 1. inflated self-esteem, 2. reduced need for sleep, 3. excessive talkativeness, 4. racing thoughts, 5. distractibility, 6. increased goal-directed activity or psychomotor agitation, and 7. risky behaviours. These symptoms must disrupt functionality and not be attributable to substance use disorders (SUD) or medications [26].

Bipolar Disorder II

A Bipolar Disorder II (BD2) diagnosis requires at least one hypomanic and one depressive episode, with no history of manic episodes. A hypomanic episode involves a sustained elevated or irritable mood plus increased activity for at least four consecutive days. Similar to BD1, at least three symptoms (four if irritability is involved) must match those for hypomania. This distinct change in functioning should not cause substantial impairment or psychotic features, nor be attributable to substances or medication [26].

Bipolar Disorder Specifiers

Clinical features serving as BD course specifiers were incorporated into the DSM-IV [32] and DSM-5 [12], enhancing diagnostic utility for prognosis and treatment guidance beyond simple categorical diagnoses. Current DSM-5 specifiers include longitudinal course, remission status, severity, anxious distress, mixed features, catatonia, mood-incongruent psychotic features, peripartum onset, seasonal patterns, and rapid cycling.

Other potential clinical variables are suggested but await formal DSM-5 acceptance [7]. For instance, evidence indicates age of onset can influence clinical manifestation, with early-onset cases leading to a more severe illness course, higher suicidality risk, and more comorbidities [33]. Research on BD course specifiers, including the age of the onset BD, psychotic features, comorbidities and rapid cycling (explored in Chapters 3 to 5 of this thesis), indicates potential distinct genetic factors. However, specifiers are likely influenced by a complex interplay between genetic predisposition and environmental factors, such as childhood trauma, which can affect onset timing and illness severity [7].

Differential Diagnoses

Common differential diagnoses for BD include schizophrenia (SCZ), major depressive disorder (MDD), anxiety disorders (ANX), substance use disorders (SUD) and borderline personality disorder (BPD). In children exhibiting early ‘BD’ symptoms, ADHD and oppositional defiant disorder are prevalent concerns [34-35]. Particular attention is needed for children displaying subsyndromal manic symptoms, mood instability, irritability, anxiety, and depression. However, even in this subset, symptom onset and severity remain heterogeneous,

requiring individual risk assessment [10]. Initial physical BD evaluations may include tests ruling out secondary causes, including urine and blood screenings, metabolic panels, and thyroid function and folate level assessments [36].

Bipolar Disorder Diagnostic Challenges

Current classifications mask considerable genetic heterogeneity within BD, which encompasses various psychiatric conditions [37]. Genetic studies highlight genetic overlap with other disorders but do not consistently align with existing classification systems. Approximately 60% of individuals with BD are initially misdiagnosed, often with unipolar depression. In one national survey, more than one-third remain misdiagnosed for 10 years or more [38]. Only 20% may receive a correct diagnosis within the first year of seeking treatment [39]. This diagnostic difficulty is compounded by the genetic overlap BD shares with other psychiatric conditions, potentially contributing to the challenges in identifying specific genetic markers for BD. Diagnosis can be challenging as BD can initially present as depressive episodes [40-41]. This could be further complicated when prior hypo(manic) episodes go unnoticed or unreported [42]. Family studies indicate that polarity at onset may have heritable components [43]. Identifying divergent genetic markers could therefore help clarify disorder boundaries and trajectories, within a continuum of genetic risk for BD and other psychiatric conditions.

Prognosis

BD prognosis is multifactorial, influenced by timely diagnosis, mood episode severity and frequency, comorbid conditions, and individual treatment response. Early intervention, particularly pharmacological and psychoeducational approaches, may enhance functional outcomes [44]. Individuals with early onset, associated with worse outcomes, could be a target group as they showed increased burden for a wider trait spectrum. Predominance of depressed versus hypomanic episodes may also impact subtype distinctions and prognoses [45]. Furthermore, the clinical course is often complicated by persistent cognitive impairment, which can affect memory, attention, and executive function even during periods of euthymia, significantly impacting long-term functional recovery and quality of life [46].

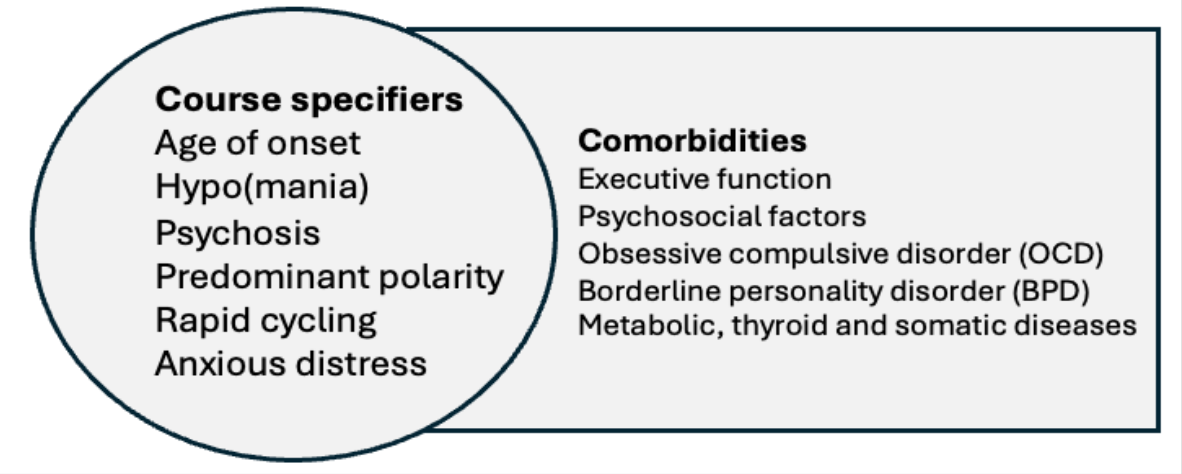


Figure 1 Inter- and Intra-Heterogeneity in Bipolar Disorder.
This illustrates distinct illness trajectories contributing to the clinical heterogeneity characteristic of BD.

A predominantly depressive polarity is frequently associated with an increased risk for depressive illness onset, Bipolar Disorder Type II (BD2), mixed episodes, and suicidality. In contrast, a predominantly manic polarity is often linked to a younger age of onset, a manic or psychotic illness onset, and a higher risk of substance abuse preceding the first mood episode, underscoring the disorder's diverse presentations [47]. A predominantly depressive polarity is associated with increased risk of depressive illness onset, BD2, mixed episodes, and increased suicidality risk. In contrast, a predominantly manic polarity is associated with younger age of onset, manic/psychotic illness onset, and higher pre-first-episode substance abuse risk [47].

Chronicity and comorbid ADHD and ANX are associated with poorer outcomes [48]. Comorbid ADHD-BD subjects had younger BD onset, more depressive episodes, more ANX and substance use/dependency disorders (SUDs), and greater BPD trait and cyclothymic temperament risk [49]. Both mixed states and rapid cycling are associated with a more severe BD form, higher comorbidity, and poor outcomes, potentially leading to inadequate treatment response, higher disability, and greater suicide risk [29]. BD patient mortality risk is elevated, particularly from cardiovascular diseases and suicide, with approximately 30-60% experiencing suicidal ideation and 15-20% completing suicide [50].

Treatments

Treatment typically combines medication and psychotherapy. Medications include mood stabilizers (e.g., lithium, valproic acid, lamotrigine) for managing hypo(manic) and depressive episodes. Antipsychotics (e.g., haloperidol, olanzapine, risperidone) also contribute to mood stabilization. While antidepressants, particularly Selective Serotonin Reuptake Inhibitors (SSRIs), may be used with mood stabilizers, they are contraindicated as standalone treatments and during manic phases due to mania induction risk [51]. Psychotherapy is an important adjunctive therapy to pharmacological BD treatments. While the evidence base has complexities, several modalities such as Cognitive Behavioural Therapy (CBT), Family-Focused Therapy (FFT), and psychoeducation have demonstrated benefits for outcomes such as relapse prevention and medication adherence [52].

Prevalence

Bipolar disorder is a prevalent psychiatric condition, estimates range from 1 to 3% in the general population [53]. According to the latest Global Burden of Disease (GBD) report (2019), around 1 in 150 adults (roughly 40-50 million people globally) are diagnosed with bipolar disorder [54]. Lifetime bipolar spectrum disorder prevalence was estimated at 4.4%, with a 12-month prevalence of 2.8%. Specifically, BD1, BD2, and subthreshold BD prevalence were 1.0%, 1.1%, and 2.4% respectively, with 12-month prevalences of .6%, .8%, and 1.4%. Actual prevalence may be as high as 4 to 6% in outpatient settings when considering subthreshold bipolarity symptoms [53].

The GBD report highlights that BD, similar to SCZ, is highly heritable and shares genetic overlap, maintaining relatively stable worldwide prevalence, although variations occur by income level, birth cohort, and geographical regions. Acute psychotic episodes are associated with the highest disability risk, while depressive and anxiety are among the leading disability causes, elevating severe outcome risks including suicide [54]. The GBD report indicated no sex variation in bipolar disorder burden, aligning with recent comprehensive genetic studies [55-56]. Prior reports suggested greater BD2 prevalence in females; however, current evidence indicates higher bipolar disorder incidence reporting across all forms in females [57]. Notably, about three-quarters of individuals on the bipolar spectrum report a comorbid disorder, with ANX, particularly panic attacks, being most prevalent.

1.5 Clinical Features, Correlates And Functioning

Bipolar disorders are fundamentally characterised by chronic mood instability. The core characteristic of "switching" represents fluctuations between euthymic states, mania, and depression. Episodes can manifest as manic, hypomanic, depressive, or mixed, interspersed with inter-episode periods, with or without subsyndromal symptomatology (Figures 1-4 are referenced generally here, with specific figures detailed below). While BD broadly encompasses symptoms associated with several psychiatric disorders (Figure 2), its distinguishing feature is cycling (Figures 3-4). Therefore, identifying genetic mechanisms of cycling could be key to understanding BD aetiology [58]. Functional impairment is a key distinction between BD1, BD2, and SZA subtypes. BD1's hallmark manic episodes (elevated/irritable mood for at least one week [14, 27] typically cause marked functional impairment. In contrast, BD2's less pronounced hypomanic episodes generally have a lesser immediate functional impact than mania [12]. SZA, combining bipolar disorder and schizophrenia features [59], often results in more severe, persistent functional deficits than BD1 or BD2 due to its combined mood and psychotic symptoms.

Bipolar disorder is distinguished from MDD by hypo(mania) presence. BD1 is characterised by at least one manic episode; BD2 has no manic episodes (Figure 3). Depressive episodes are defined as persistent low moods lasting more than two weeks. Symptoms include loss of interest in typically enjoyed activities, fatigue, insomnia or hypersomnia, hopelessness, suicidal ideation, reduced self-esteem, and social withdrawal [27]. Difficulties differentiating BD1, BD2, and unipolar depression may contribute to up to a 10 year diagnostic and treatment delay [60]. Evidence supports potential unipolar depression misdiagnosis, as 20% of patients developed hypo(mania) within five years in one longitudinal study [61].

Subsyndromal Bipolar Disorder Symptoms

BD can be associated with progressive cognitive deficits, residual symptoms, sleep disturbances, and emotional dysregulation between mood episodes [62]. An estimated 20-50% of patients experienced inter-episodic or chronic subsyndromal symptoms in one review of periods of euthymia [62] (Figure 4).

Early Onset Bipolar Disorder

Early onset bipolar disorder (EOBD) has been proposed as a DSM-5 course specifier. EOBD presence correlates with increased chronicity and comorbidity risk [63]. It is associated with higher comorbid anxiety and SUD instances, more episodes, less euthymia, and greater suicide attempt risk. Most reported EOBD comorbid conditions are ADHD, SUD, and anxiety. In a 983 BD1 adult case study, early-onset BD was associated with more severe illness course, increased suicidality and comorbid psychopathology risk, more episodes, and worse functional outcomes compared to later onset [64]. Childhood onset represented only 5% of cases, 25% adolescence, and 53% at peak ages 15-25 [65]. A recent genetic study (34,658 alcohol use dependency [AUD] and 20,352 BD cases) suggested shared aetiology [66].

The prognosis of BD is shaped by its typical natural history. The age of onset often follows a trimodal distribution, with peaks in adolescence, the early twenties, and around age 40 [67]. EOBD cases between 12-18 years and even earlier are reported (age < 12), most qualifying as early onset (occurring before 17 years of age). Two further onset peaks were reported: 26 and 42 years old [67]. Critically, for many individuals, the illness begins not with mania but with one or more depressive episodes, often leading to initial misdiagnosis and significant treatment delays [41]. Furthermore, a defining feature of the illness is its high rate of recurrence. Seminal longitudinal work, such as Angst's studies of the Zurich cohort, demonstrates that BD is a highly recurrent condition, and full functional recovery between episodes is often incomplete [13, 62]. This pattern of recurrence and residual impairment has direct relevance for the investigation of a chronic illness trajectory in this thesis (Chapter 3).

Rapid Cycling Bipolar Disorder

Rapid cycling (RC-BD) was first noted before available pharmacologic treatments, some potentially worsening switching, suggesting it is not solely a medication artifact. Consistent lithium non-responsiveness evidence also exists [21]. Rapid cycling occurs in approximately 10-20% of BD cases, characterized by four or more episodes per year (RC; ≥ 4 episodes/year). A recent RC-BD systematic review/meta-analysis identified RC-BD in 9.36% of cases (3.74% BD1, 15.2% BD2) [68]. However, another study found higher RC-BD prevalence: in a large 54,257 BD case cross-national community sample (lifetime and 12-month data), approximately 30% met rapid cycling criteria. Rapid cycling may be prognostic for onset, clinical course, and outcomes, associated with increased chronicity and comorbidity risk. It is more often reported early in diagnosis, suggesting the rapid cycling experience may prompt help-seeking behaviour. It is associated with greater severity, chronicity, worse global functioning, and higher suicidal risk [17]. Despite this, no clear treatment consensus exists [69]. One longitudinal study reported rapid cycling often resolved within two years of onset in 4 to 5 cases [70]. While some individuals experience RC-BD temporarily; for others, it is recurring or persistent.

Mixed features

Mixed features and rapid cycling share a similar poor BD trajectory. At least 30-70% of BD patients present with mixed mania or depression [29]. Frequent mixed episodes are associated with a severe, chronic course, comorbid disorders, cognitive impairments, rapid mood swings, and treatment resistance [7].

Bipolar disorder with Psychosis

Similar to EOBD and RC-BD, other specifiers such as psychosis might be better described as dimensional, existing on a severity spectrum (Figures 1-2). While BD and SCZ can involve psychosis, the key difference is mood episode presence and psychotic symptom persistence: in BD, psychosis typically occurs during manic/depressive episodes; in SCZ, psychosis is primary and persistent. In BD, psychosis describes a state of being disconnected from reality, often

involving hallucinations, delusions, and disorganized thoughts or speech, occurring during manic or depressive episodes. Grandiose delusions and paranoia are mania features; however, psychotic symptom presence represents severe BD. The prevalence of psychotic symptoms varies across the different phases of the illness. A systematic review highlighted that such symptoms are significantly more common during manic and mixed episodes compared to depressive phases, underscoring the strong link between psychosis and elevated mood states in BD (See Chapter 4 [39]).

Psychosis is classified as either mood congruent (symptoms align with current mood state) or mood incongruent (symptoms do not correspond). The distinction between mood-congruent and mood-incongruent psychosis has significant diagnostic implications, particularly at the boundary between bipolar disorder and schizophrenia. The presence of mood-incongruent psychotic symptoms, especially when persistent, raises the diagnostic possibility of schizoaffective disorder, bipolar type (SZA). The work of researchers such as Andreasen *et al*. (1987); Akiskal and Pinto (1999), has been central to debating these diagnostic boundaries, highlighting the challenge of classifying patients who present with a mix of severe mood and psychotic features [71-72]. This classification quandary is not merely academic; as demonstrated in a machine learning analysis of the Northwick Park functional psychosis trial, these symptom dimensions can help separate affective psychoses from schizophrenia [73-74]. This has direct relevance for the genetic analyses in this thesis, where polygenic risk for schizophrenia is used to probe the biological basis of psychotic features within BD (Chapter 4).

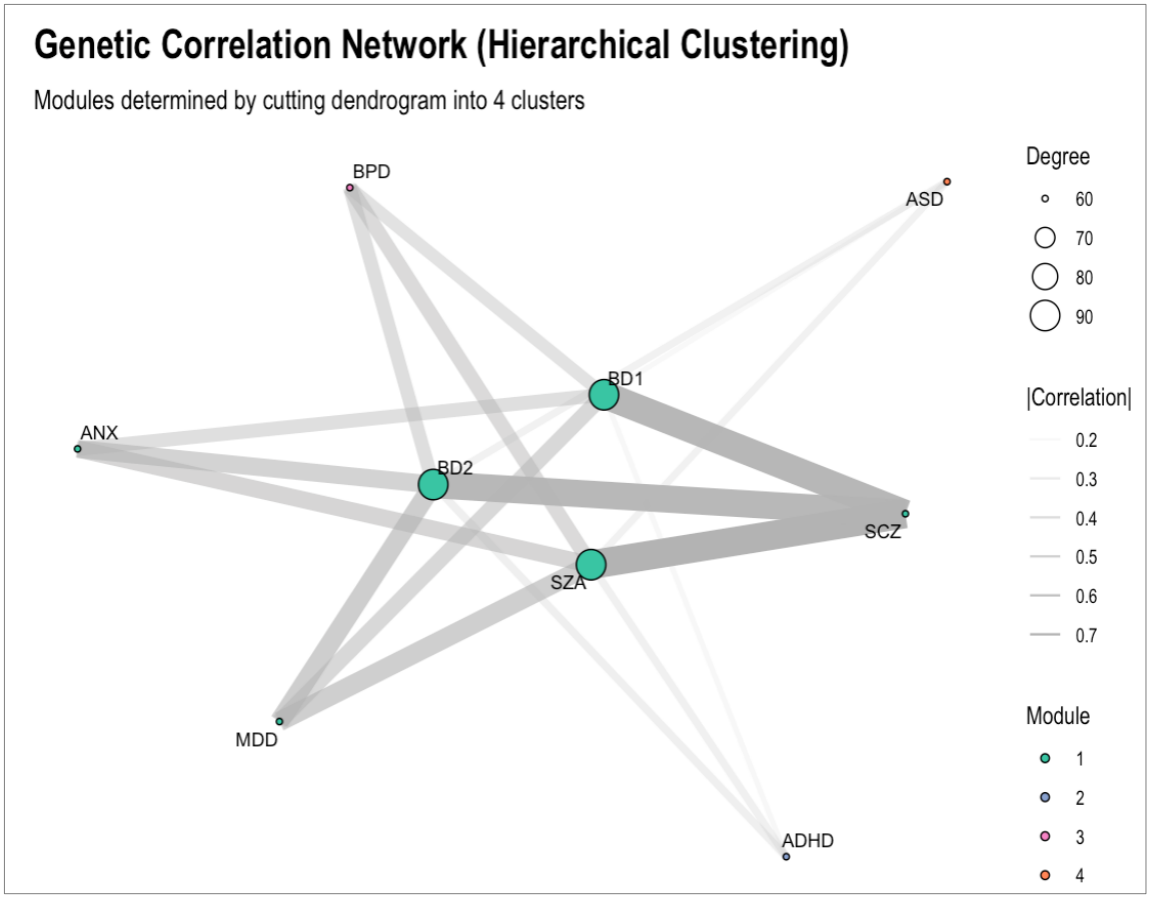


Figure 2 Shared Phenotypic and Genetic Correlations.
Adapted from Gordovez and McMahon (2020), *The genetics of bipolar disorder* [58]. The genetics of bipolar disorder, which used estimated genetic correlation (*rG*) extracted from an atlas of genetic correlations (see Chapter 2 [24]). This figure displays instead the genetic correlation (*rG*) generated in this thesis, Chapter 5. This network reveals psychiatric trait (modules) genetic correlations. Each node represents a specific psychiatric trait: BD1, BD2, autism spectrum disorder (ASD), SZA, BPD, ADHD, MDD, and ANX. Node size is proportional to its degree (number of other traits with genetic correlations), with larger nodes indicating a more widespread influence on the overall genetic correlation structure. Node colour represents the trait's module (cluster) determined by unsupervised hierarchical clustering, where same-coloured traits exhibit stronger genetic interconnectedness patterns, helping identify broader shared genetic underpinnings across different disorders. Edges (lines) connect trait pairs with reported genetic correlation (*rG* value), with edge opacity and width reflecting the strength (absolute value) of this relationship; thicker, darker edges indicate stronger genetic associations.

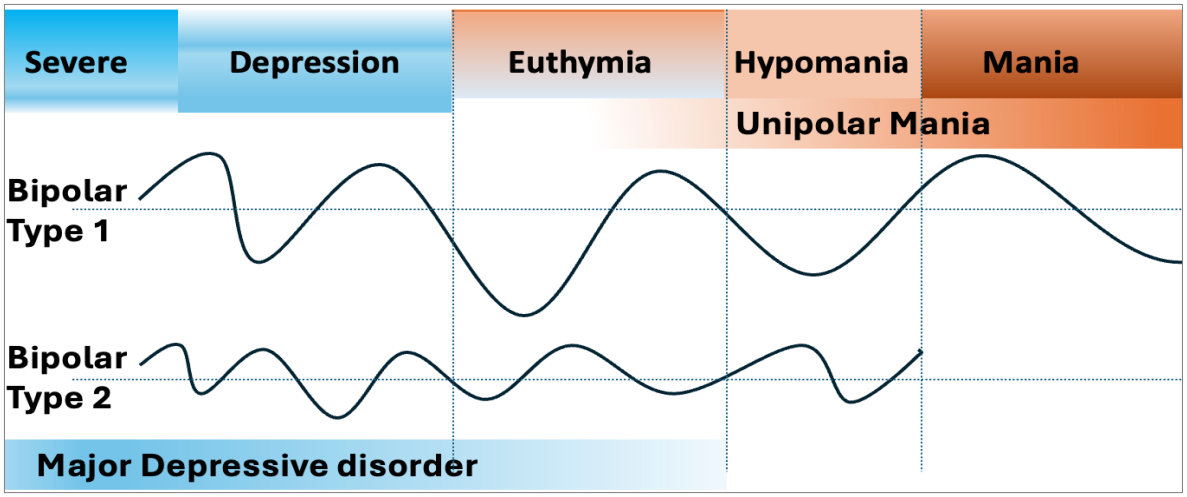


Figure 3 Mood Frequencies Across BD and Depression.
Adapted, from O'Connell and Coombes (2021), *Genetic contributions of bipolar disorder: current status and future direction* [75]. The figure represents a comparison of polarity and mood switch frequency across BD1, BD2, and unipolar mania (UM) and depression.

Subsyndromal Bipolar Disorder Symptoms

BD is associated with progressive cognitive deficits, residual symptoms, sleep disturbances, and emotional dysregulation between mood episodes. An estimated 20-50% of patients experience inter-episodic or chronic subsyndromal symptoms [62] (Figure 4).

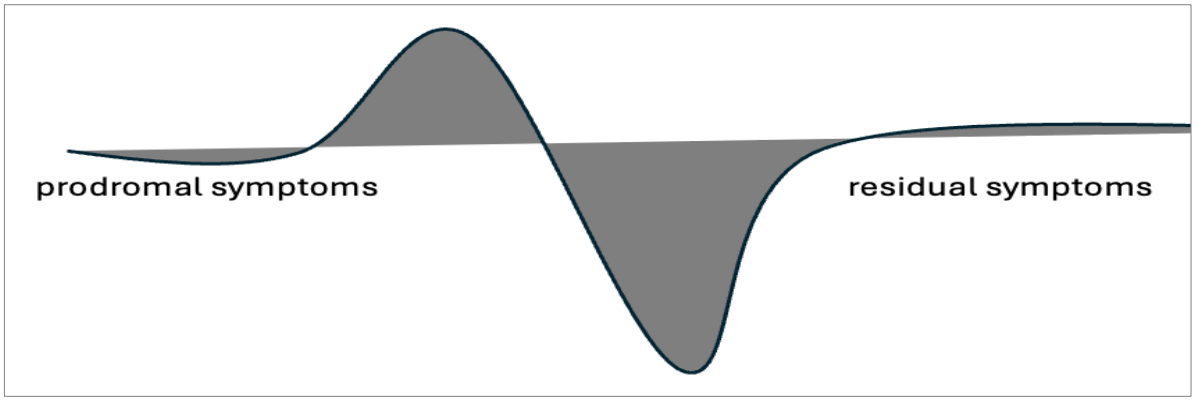


Figure 4 Subsyndromal Symptoms in Bipolar Spectrum Disorders.
Adapted from Grunze and Born (2020), *The Impact of Subsyndromal Bipolar Symptoms on Patient's Functionality and Quality of Life* [62]. Prodromal symptoms may occur before full condition onset, influencing functionality, QoL, and increasing relapse risk. Residual symptoms may persist after an episode (e.g., cyclothymia [low-level depression, mildly elevated mood/irritability], sleep disturbances, and difficulty concentrating).

1.6 Bipolar Disorder Aetiology

This section reviews known genetic, biomarker, and environmental BD contributors. Heritability in BD aetiology is multifaceted; genetics is a pivotal risk factor. Children of BD parents are eight to 10 times more likely to develop BD, though most do not [30]. These offspring, however, have a heightened risk for other psychiatric disorder risk [76]. A longitudinal study found preschool ADHD children with early-onset BD parents had higher BD development risk than community controls [77].

A Neurodevelopmental Model of BD

Neurodevelopmental disorders stem from early brain abnormalities due to genetic/environmental neurodevelopmental influences. Evidence suggests BD may develop early, leading to adverse adult conditions [78]. This premise is debatable, possibly applying more to psychosis than BD phenotypes [79]. Kloiber *et al*. (2020) provides a comprehensive review discussing neurodevelopmental abnormality evidence in early-onset BD-linked psychotic symptoms [80]. Neurodevelopmental evidence may be too subtle for pre-onset BD detection or distinguishing some early-stage psychiatric diseases [81]. Shared pathogenic mechanism evidence with other neurodevelopmental disorders (intellectual disability, ASD and ADHD) led some researchers to propose BD exists on a neurodevelopmental continuum with these early-onset disorders [82]. This prompted increased focus on adolescent/young adult longitudinal studies, as BD symptoms often emerge then. Typically, depression presents first, often during or before puberty [83] while manic episodes usually manifest post-puberty [84]. Most commonly, BD onset is in young adulthood, when brain development slows and synaptic pruning increases, enhancing efficiency by eliminating redundant neural connections [85]. In contrast, during adolescence, BD individuals may experience grey matter and neuron loss without typical white matter connection increase seen in unaffected adolescents [3, 86], particularly in prefrontal cortex and insula (MRI studies). Imaging research revealed BD patient hyper- and hypoactivation differences compared to healthy controls. Amygdala, prefrontal cortex, and visual system hyperactivation may be critical in emotional dysfunction. Anterior cingulate cortex (ACC) hypoactivation could contribute to cognitive deficits in BD patients younger than 18 years [87].

When compared with the neurodevelopmental model for SCZ, the trajectory for BD appears distinct. Landmark longitudinal studies, such as the Dunedin cohort, suggest that SCZ is often preceded by subtle motor and cognitive deficits in early childhood [88]. In contrast, the major functional and structural brain abnormalities in BD typically emerge later, during adolescence and young adulthood, often coinciding with the onset of the first mood episode [80, 83]. Within this framework, mania can be seen as the fulcrum that differentiates the BD subtypes. The emergence of a full manic episode, often linked to more pronounced disruptions in prefrontal cortical development, defines the transition to BD1. In contrast, the absence of mania in BD2 may suggest a different, possibly less severe, neurodevelopmental impact [89]. This distinction is critical for understanding the different long-term outcomes and treatment needs associated with the BD1 and BD2 diagnoses.

Neural Substrates in Bipolar Disorder

Mechanisms underlying BD symptoms are complex. Specific brain region structural abnormalities may correlate with emotional dysregulation and cognitive dysfunction. Cortical thickness/surface area show high heritability, are associated with various genetic influences [90-91] and may be affected by largely distinct gene sets [92-93]. BD structural changes are documented across multiple brain regions (prefrontal/temporal cortices). Factors considered include genetics, comorbid disorders, and accelerated aging [94]. Neuroimaging studies identified neuroanatomical alterations (cortical thickness/surface area changes, and grey matter volume changes) corresponding to BD-associated cognitive/behavioural functional impairments [3]. Cross-sectional studies highlighted BD-specific structural abnormalities primarily in prefrontal/temporal cortex, cingulate gyrus, subcortical regions, and insula. Notably, amygdala, hippocampus, and thalamus subcortical alterations are documented in BD patients [94].

Manic episodes consistently correlate with cortical volume or thickness reductions, especially prefrontal [3]. Additional findings indicate psychotic history BD patients demonstrate thinner frontal, temporal, and parietal cortical grey matter (both hemispheres), alongside reduced cortical surface area [94]. The large-scale Enhancing Neuro Imaging Genetics through Meta Analysis (ENIGMA) project further evidenced thinner frontal/temporal cortices in BD patients [3, 95]. The ventrolateral prefrontal cortex (VLPFC), crucial for emotional regulation/reward processing, shows greatest cortical thickness depletion [96]. Investigating regional cortical thickness/surface area discrepancies may facilitate identifying meaningful biomarkers for different BD subtypes and course specifiers. However, a critical challenge in current neuroimaging biomarker identification efforts, is main BD phenotype heterogeneity [97], potentially exacerbated in large-scale consortium studies by intra- and inter-cohort differences.

Progressive Deterioration in Brain Structures

Research indicates repeated manic episodes may contribute to structural changes (particularly prefrontal cortex), with observed correlation between episode frequency and illness severity [94]. Psychosis presence and type (mood-congruent/incongruent) in first-episode mania were suggested to have different phenotypic markers [98]. This ‘neuro-progressive’ model, which posits that mood episodes themselves may have a neurotoxic effect, is supported by some longitudinal evidence. Large-scale collaborative studies from the ENIGMA Bipolar Disorder Working Group have demonstrated correlations between a higher number of manic episodes and accelerated cortical thinning over time, particularly in prefrontal regions [3]. Such findings could bolster the rationale for early and sustained intervention to mitigate potential long-term structural brain changes. While some interpret this correlation within a ‘neuro-progressive’ model where episodes may have a neurotoxic effect, it is important to note that much of the evidence is cross-sectional. Such study designs cannot definitively distinguish between illness progression and pre-existing vulnerabilities. Robust longitudinal studies are needed to confirm a causal relationship and rule out other confounders such as medication effects or comorbid conditions.

Chapter 4 investigates the potential of using SCZ PRS to identify individuals with BD at higher risk of psychosis, which aligns with the need for early detection highlighted by this neuro-progressive model. Understanding neurobiological underpinnings is crucial for episode prevention to mitigate further damage. However, exact brain volume reduction mechanisms are not fully understood. Neuroinflammation, altered neurotransmitter activity, and disrupted brain connectivity are associated with psychosis [99]. Full episode structural changes may be caused by neuroinflammatory/oxidative stress [100-101], dysregulated hypothalamic-pituitary-adrenal (HPA) system hormonal release, and neurotrophic factor secretion defects [102]. These alterations can persist during euthymia. MRI scans show BD-associated focal demyelination and axon/nerve fibre loss, observed in children and adolescents at rates similar to unipolar depression and schizophrenia [103]. A notable lack of longitudinal studies tracks neuroanatomical changes across the lifespan. Untreated patient research is scarce; many studies focus on euthymic patients for methodological reasons, limiting understanding of manic episode functional and structural changes. Nonetheless, one study controlling for confounders still identified BD cognitive impairments [104]. Similarly, another found untreated bipolar patients had smaller left anterior cingulate volumes than healthy controls [105]. Lithium-treated bipolar patient comparisons suggested lithium might influence cingulate volumes, possibly via neuroprotective effects [105]. These findings emphasize identifying BD genetic factors and biological mechanisms, as this knowledge could help predict early signs and facilitate targeted interventions potentially preventing full-blown bipolar disorder episodes.

Cognitive Deficits in Bipolar Disorder

Besides disentangling grey matter volume loss contributors, understanding how mania-related changes translate to symptomatology, such as social and cognitive functioning, is crucial for treatment. Cognitive impairment is a central BD feature, affecting memory, attention, and executive function, impacting recovery, work ability, and quality of life (QoL). Frontal, subcortical, and limbic structure functional abnormalities are broadly implicated in mood disorder pathophysiology, where BD neuropathology involves mood, cognition, and behaviour dysregulation.

Premorbid BD cognitive deficit studies report lower risk compared to SCZ [104,106]. Accordingly, psychotic history BD patients had greater impairment in several cognitive domains. However, effect size differences between BD subjects with and without psychosis were moderate, potentially representing a severity spectrum rather than a qualitative distinction [107]. Orbitofrontal cortex (OFC) subregion activation neuroimaging studies reported decreased activity during manic episodes and in depressed bipolar subjects [108]. While OFC abnormalities are reported across psychiatric disorders, in BD the OFC mediates executive function, including inappropriate response control, decision-making, and behavioural flexibility [109].

These cognitive deficits are not limited to acute episodes but are also observed during euthymic phases [46], where they may be complicated by medication effects. BD1 cognitive impairment is reportedly more severe and widespread across cognitive measures than BD2 [110]. However,

the neural correlates that could explain these cognitive changes over time are largely unknown. One five-year BD neurocognitive trajectory longitudinal study found a positive association between the number of hypo(manic) episodes and a greater decline in cognitive measures such as working memory [111]. Comorbidity also complicates treatment; for instance, BD comorbid with ADHD may exacerbate anxiety and mood dysregulation risk, suggesting a hierarchical treatment plan is necessary [112].

The frequent co-occurrence of cognitive deficits and personality traits in bipolar disorder, along with the need to understand their underlying genetic links, provides a key impetus for the dimensional and cross-trait analyses explored in this thesis. This need for a deeper biological understanding is critical, as currently no medications specifically improve BD cognitive functional outcomes. Moreover, common bipolar disorder medication side effects (affecting concentration, memory, processing speed difficulties, and impaired executive function) may exacerbate symptoms.

Anterior Insula in BD Symptomatology

Converging evidence indicates early subcortical, caudal, and especially ventral prefrontal cortex (vPFC) and insula dysfunction in BD, and between these brain region interconnections. Altered functioning within these regions may be implicated in specific BD symptoms: interoception (insular cortex), motor changes (precentral gyrus), and cognition (prefrontal cortex). The insula is a hub for saliency, cognitive control (inhibitory control, behavioural regulation), and interoceptive (internal bodily) awareness. The anterior insula (AI) together with the anterior cingulate cortex (ACC) integrates external/internal bodily information to guide goal-directed behaviour. The insula has substantial DLPFC connections, especially from the AI, influencing attention, working memory, and decision-making. Early childhood BD manifestations include inattention, hyperactivity, and disruptive behaviours [113]. AI and frontoparietal executive control/saliency network functional connectivity is reported as a differential biomarker between BD and unipolar depression, a potential BD therapeutic target [114] and an early indicator of the disorder [86]. A BD neurological model proposes emotion circuitry area activity imbalance, disrupting emotion regulation. BD occurs when the ventral system (regulates emotion perception in the amygdala, insula, ACC, and prefrontal cortex) is overactivated. Conversely, the dorsal system (regulates emotion in hippocampus, dorsal ACC [dACC], and prefrontal regions) is under activated [115].

Biological Pathways in Bipolar Disorder

In summary, BD is a complex trait influenced by multiple genetic variants across various biological pathways. Key pathways include signalling mechanisms, epigenetic processes, and neurotransmitter systems. Specific signalling pathways involved are Gamma-Aminobutyric Acid (GABA), glutamate, and calcium signalling, alongside neurotransmitter systems (serotonergic, noradrenergic, and dopaminergic). Additional affected biological functions encompass neuroinflammation, oxidative stress, mitochondrial dysfunction, impaired neuroplasticity, and circadian rhythm dysregulation. These factors contribute to cellular

changes crucial to BD pathophysiology. Notably, BD dysregulation is associated with intracellular calcium level disturbances, interneuron deficits, and glial cell abnormalities. While cellular changes are evident across brain regions, prefrontal cortex and hippocampus may be especially impacted.

Excitatory/Inhibitory Balance in BD

Research indicates excitatory and inhibitory (E/I) neuronal activity imbalance might be crucial in BD. Post-mortem BD individual studies show neurotransmission changes involving glutamate (excitatory) and GABA (inhibitory) signalling. The E/I balance concept (excitation/inhibition ratio), initially an autism spectrum disorder (ASD) model [116], is now associated with various neurodevelopmental and neuropsychiatric conditions, such as intellectual disability [117-18] and schizophrenia [119]. Rebalancing E/I ratio is suggested as part of lithium's therapeutic effect by promoting inhibition [120]. Accumulating evidence suggests glutamate is involved in BD aetiology. Post-mortem analyses uncovered prefrontal cortex (PFC) excitotoxicity [121], ACC glutamatergic function and synaptic connection abnormalities [122], and broader glutamatergic system disruptions. Glial cells are crucial in glutamate metabolism management; astrocytes are vital in synaptic cleft glutamate uptake [123]. Both post-mortem studies and in vivo Transcranial Magnetic Stimulation (TMS) research revealed impaired BD cortical inhibition [124]. Elevated glutamate levels are associated with executive dysfunction [125]. Excessive glutamate can activate ionotropic receptors in extra-synaptic locations, leading to neurotoxicity via calcium influx and free radical (e.g., nitric oxide) production. Persistent [123] or repeated mood episode glutamate elevation may contribute to BD neuro-progressive pathogenesis.

Glutamate and GABA network disruptions could lead to BD-associated neurotransmission and neuronal plasticity irregularities. Further research highlighted GABA neurotransmission's role in BD mood-regulating brain region interneuron synapses [126]. Various studies show GABA level alterations in brain, cerebrospinal fluid, and blood [127]. Glutamic acid decarboxylase (enzyme essential for GABA synthesis) decreased activity is associated with depressed patients, potentially diminishing GABAergic activity [126]. GABAergic system changes could also be associated with BD cognitive deficits [128]. GABAergic interneurons help integrate information to synchronise neural networks. Post-mortem studies found reduced cortical interneuron densities in BD1 parahippocampal tissue, resembling SCZ patterns compared to healthy controls [129]. However, confounders such as comorbid panic disorders [130], anxiety [131], and alcohol dependence [132] could complicate the assumed direct neurotransmission and BD association. Recent proton magnetic resonance spectroscopy (1H-MRS) studies indicated obsessive-compulsive disorder (OCD) individuals exhibit higher Anterior Cingulate Cortex (ACC) glutamate and lower GABA levels than those without OCD [133]. Future research must fully clarify glutamate/GABA effects in BD, distinguishing them from BD comorbid disorder effects. (See Chapter 5, which demonstrates differential glutamate and GABA gene set and cell type expression across BD subtypes, specifiers, and comorbid disorders).

Neurotransmission

Neurotransmitters (brain chemical messengers) play a key BD role. Substantial data support BD neurotransmitter system dysfunction which is frequently investigated as potential therapeutic targets. BD is associated with imbalances in serotonin, dopamine, norepinephrine, and GABA. Glutamate and GABA abnormalities are consistently identified in BD literature; these two amino acids are the most abundant in brain excitation/inhibition controlling neurotransmitters [127]. Brain catecholamines (dopamine, norepinephrine, and epinephrine) are relatively low compared to other neurotransmitters but are crucial in regulating brain functions and are vital therapeutic targets.

The Dopamine Hypothesis in BD

Neurotransmitters such as dopamine, norepinephrine (noradrenaline), and serotonin are abnormally regulated in BD. Limbic system dysfunction impacts sleep, alertness, and emotion regulation [134]. The 'cholinergic-adrenergic balance' hypothesis initially explained different BD affective states [135]. However, in a neuroimaging, neuropharmacological, and genetic study review, it was suggested there is stronger evidence of the 'catecholaminergic-cholinergic balance' hypothesis. Nevertheless, these neurotransmitter system interplays do not fully account for mania/depression cycling [136]. Dopamine is singled out as a key player in BD core symptoms, especially depression/mania transition [137]. Signalling primarily involves G protein-coupled receptors (GPCRs), modulating fast synaptic transmission in glutamatergic/GABAergic neurons. These receptors are crucial for various physiological and cellular processes [138], making GPCRs pivotal therapeutic targets (many medications aim to modify their activity). The BD dopamine hypothesis posits dopamine transport and receptor availability dysregulation may explain the disorder's depressive/manic phases [139]. Evidence indicates heightened dopamine transmission (particularly in the mesolimbic region, associated with reward and motivation) is linked to manic episodes, possibly with increased D2/3 receptor availability and a hyper-responsive reward system. Conversely, reduced dopamine activity (possibly due to elevated dopamine transporter [DAT] levels) is linked to depressive episodes. Extensive support for dopamine's BD role exists, with robust research accumulating since the 1970s (post-mortem, pharmacological, functional magnetic resonance, and molecular imaging) (see Ashok *et al*. 2017 [139]). Post-mortem analyses noted DLPFC D2/3 receptor upregulation in BD patients; however, studies do not specify illness phase [140-41]. This systematic review of BD dopamine effects [139] highlights converging pharmacological and imaging study results, indicating elevated D2/3 receptor availability and reward processing network hyperactivity contribute to mania. DAT level imbalances are observed in bipolar depression, however other dopaminergic functioning aspects yielded inconsistent results.

More recent multimodal imaging work continues to refine this model. For instance, a 2025 study by Jauhar and colleagues in *JAMA Psychiatry* provided further evidence linking striatal dopamine function not just to psychosis, but to the interaction between psychosis and mood severity in affective disorders [142]. Their findings suggest that the dopaminergic dysregulation in bipolar psychosis may be distinct from that seen in SCZ, potentially being

more closely tied to the affective state, which has significant implications for diagnosis and the development of state-specific therapeutics. Across all patient groups, higher dopamine synthesis in the associative region of the striatum was linked with greater severity of positive psychotic symptoms (e.g., hallucinations, delusions), regardless of the specific mood disorder diagnosis. The study also found however, that dopamine dysregulation was not uniform. Patients with manic psychosis showed higher dopamine synthesis, particularly in the brain's limbic region, compared to those with psychosis and depression. The results suggest that the biological basis of psychosis does not perfectly align with traditional diagnostic categories. This implies that antipsychotic drugs, which modulate the dopamine system, could be beneficial for treating psychotic symptoms across a wider range of mood disorders than is current practice.

Noradrenaline

Noradrenaline levels were low in bipolar disorder and depression; however, greater noradrenaline metabolite levels were detected during manic episodes [143], suggested to be due to low inhibitory alpha2-adrenaline receptor sensitivity, and also observed in panic disorder (PD) [144], a common BD comorbidity.

Serotonin

Serotonin studies have associated serotonin with other commonly BD-comorbid disorders [145]. In contrast, small BD patient studies yielded inconclusive results [146]. Several research efforts associated cerebrospinal fluid (CSF) 5-hydroxyindoleacetic acid (5-HIAA) (serotonin metabolite) concentrations with impulsivity, aggression, and unipolar depression suicide risk [147]. However, 5-HIAA CSF level differences were not clearly distinguishable between manic depressive episode patients and unipolar depression patients [148-149].

Intracellular Signalling

BD pathophysiology research includes intracellular signalling cascade network studies, searching for new mood disorder treatments. Complex signalling networks support cell communication involving mood/wakefulness-related targets (glucocorticoids, thyroid/gonadal hormones). Intracellular (IC) signal transduction system changes are a focal point [146]. Various intermediaries are associated with BD; post-mortem studies and pharmacological evidence implicate lithium in the phosphatidylinositol (PI) pathway. Functional changes occur as neurotransmitters and neuromodulators bind GPCRs. Cyclic adenosine monophosphate (cAMP) and diacylglycerol (DAG) impact protein kinase A (PKA) and protein kinase C (PKC), regulating metabolism and transcription factors [150]. Lithium's bidirectional cAMP impact suggests broad therapeutic effects across manic and depressive phases [151]. Lithium impacts PI pathway by depleting myo-inositol levels, reducing intracellular transmission via targeted key protein downregulation (Figures 5, 6) [152], including PKC phosphoprotein substrate myristoylated alanine-rich C kinase substrate (MARCKS)[152]. Lithium inhibition is proposed

to mitigate increased intracellular calcium levels which is expected to be hyperactive in BD patients [153-155].

Neuroplasticity and Neurotrophic Signalling

Neuroplasticity and neuroprotection effects are associated with BD pathophysiology. Neuroplasticity may explain decreased cellular plasticity and damage in BD, possibly associated with functional deficits increasing mood episode severity. Neurotrophic factors are involved in key processes: (proteins regulating neuronal cell survival/growth, synapse formation, and neuroplasticity processes [synapse remodelling, long-term potentiation (LTP), axonal growth, synaptogenesis and neurogenesis]) [156]. Neurotrophins modulate central nervous system (CNS) via tyrosine kinase (Trk) receptors, activating mitogen-activated protein kinase (MAPK) pathway, increasing neuroprotective proteins such as Bcl-2 [157] (see Figure 6). Brain-derived neurotrophic factor (BDNF) is frequently implicated in BD. BDNF and neurotrophin-4 (NT-4) bind TrkB receptor. Several studies report decreased BDNF/TrkB levels in blood/brain of medicated/unmedicated BD patients [158]. Antidepressant/mood stabiliser action mechanisms are also associated with BDNF levels. Transcription factor cAMP Response Element-Binding Protein (CREB) influences BDNF function; increased levels of both are reported in antidepressant-treated patients [146]. Additionally, glycogen synthase kinase 3 (GSK-3) enzyme, which promotes apoptosis, is inhibited by BDNF [159], lithium, and valproate [160]. PKA's role recently gained focus, *AKAP11* gene (deficiencies can inhibit PKA-activation of GSK-3) which was identified in the largest BD Whole-Exome Sequencing (WES) study to date [161] (see below).

Mitochondrial Dysfunction and Oxidative Stress

Accumulating evidence indicates mitochondrial dysfunction and reactive oxygen species (ROS) production in BD pathogenesis. Brain mitochondria (organelles) are critical for cell survival [162], producing energy (adenosine triphosphate [ATP]) for neuronal function; dysfunction can contribute to neuronal degeneration. Mitochondrial function is essential for synaptic plasticity (crucial for learning and memory)[163]. BD patient brain studies report glycolytic shift, indicating mitochondrial dysfunction related to neuronal sodium (Na+)/potassium (K+)-ATPase activity. This dysfunction may promote neurodegeneration via glutamate excitotoxicity/neuronal apoptosis [164], potentially inducing hyperexcitable state (mania) or inhibiting neurotransmitter release (depression)[165]. Other energy metabolism mechanisms are implicated in BD, including nuclear messenger ribonucleic acid (mRNA) product downregulation (Krebs cycle involvement). This suggests decreased nicotinamide adenine dinucleotide + hydrogen (NADH) and flavin adenine dinucleotide (FADH2) oxidation in BD might increase ROS production, causing oxidative stress. Excessive ROS can impair cognitive functions (learning, memory, and executive function)[166].

This link between mitochondrial function and BD is further substantiated by pharmacogenomic research. For example, studies using Induced Pluripotent Stem Cells (iPSCs) derived from patients with BD have shown that lithium responders exhibit a rescue of mitochondrial deficits

that are not seen in non-responders. Specifically, lithium treatment in cells from responders has been shown to normalize deficits in energy metabolism and reduce oxidative stress. Much of this research has pointed towards dysfunction in mitochondrial complex I (MC1) as a key pathological hub, with evidence showing altered MC1 activity and expression in patient-derived cells, a deficit that may be directly modulated by lithium's therapeutic action [167].

Immune-Inflammatory Imbalance and Kynurenine Pathway

Dysfunctional kynurenine metabolism can result from inflammatory response, reported in mood disorders, potentially contributing to BD patient volume loss [168]. Pro-inflammatory cytokines such as interleukin-6 (IL-6) are associated with BD, suggesting a potential direct immune dysfunction role in BD pathogenesis. Several mechanisms are proposed: blood-brain barrier integrity, genetic factors, gut-brain axis dysfunction, and kynurenine pathway involvement. Kynurenine metabolites are associated with neurotoxicity/impaired neurotransmission [145]. Unmedicated BD patient research found dendritic atrophy may correlate with amygdala and hippocampus volume loss, potentially due to protective kynurenine metabolite loss [169]. However, translating these associations into effective treatments has proven challenging. While observational data suggest a role for immune dysfunction, the clinical evidence for anti-inflammatory interventions remains equivocal. Notably, a large-scale randomized controlled trial of the anti-inflammatory agent minocycline as an adjunctive treatment for bipolar depression failed to find a significant benefit over placebo, highlighting the complexity of targeting this pathway in BD [170].

Circadian Rhythm

Circadian rhythm disruptions may be associated with sleep disturbances (often reported in BD during acute and inter-episode periods) and can influence body temperature and hormone secretion (melatonin and cortisol levels typically follow a circadian pattern). Studies show BD patient cortisol secretion is higher than controls (regardless of circadian phase), aligning with increased hippocampal/amygdala glucocorticoid receptor (GR) mRNA levels in BD [171-72]. Sleep deprivation is known to trigger BD manic episodes [173]. Interestingly, some research indicates short-term antidepressant effects in some bipolar depression individuals [174], possibly due to rapid BDNF level increase, resembling antidepressant actions [175]. Genetic studies revealing numerous associations with circadian rhythm regulating genes, bolster these consistent findings [176]. Identifying exact BD pathophysiology mechanisms is challenging, which relies largely on isolating individual functional mechanisms or postmortem brain tissue use. Such studies may not accurately reflect active, holistic brain physiology and their results are likely influenced by lifetime medication use. A recent living donor fresh brain tissue study highlighted potential postmortem and live tissue discrepancies [177]. Also, cell metabolism results from intricate genetic and environmental factor interactions. While genetic predispositions are substantial, BD pathophysiology remains dynamic; accumulated psychosocial stress and sleep deprivation can instigate mood episodes independently of genetic factors. Genetic predisposition may interact with various environmental factors including early-life adversities, leading to epigenetic, endocrine, and inflammatory alterations [7].

Neuroendocrine System

Numerous studies report Hypothalamic-Pituitary-Adrenal (HPA) axis hyperactivity even in unmedicated depressed/bipolar depressed patients. The HPA axis includes the hypothalamus, pituitary, and adrenal glands. Elevated HPA activity is associated less with manic episodes than mixed episodes/bipolar depression. Bipolar disorder studies associate increased HPA activity with mixed manic states/depression, and less consistently with classical manic episodes [178]. HPA dysregulation is also linked to bipolar disorder clinical course outcomes, increasing cognitive deterioration and risk for relapse [179]. Robust evidence suggests manic episodes may be preceded by heightened cortisol/adrenocorticotropic hormone (ACTH) levels [180]. HPA alterations are tied to familial risk; bipolar disorder patient first-degree relatives reportedly exhibit elevated baseline cortisol [181]. A thyroid hormone/mood disorder relationship in BD is well-supported. Thyroid hormones have neurotrophic effects; thyroxine/triiodothyronine (T3) treatments for treatment-resistant depression (TRD) or bipolar disorder increase intracellular CREB [182]. Gonadal hormone influence on mood disorders is well-documented. Oestrogen modulates serotonin's antidepressant effects via neurotransmitters (noradrenaline, dopamine, GABA), and influences neuroplasticity via intracellular PKC signalling [146]. Progesterone and testosterone, recognised primarily for reproductive functions, also substantially affect mood and mental health; imbalances can potentially lead to anxiety, depression, and mood swings.

Epigenetic Mechanisms

Epigenetic mechanisms (DNA methylation, histone modifications, chromatin remodelling) influence BD physiology by modulating gene expression. These long-term gene function modifications occur responding to environmental factors. DNA methylation may dysregulate BD gene expression; abnormal DNA methylation is observed in known BD risk genes such as BDNF, this suggested this was more affected in BD2 than BD1 [183]. Epigenetics is implicated in the main BD phenotype, psychosis, and suicide risk. Twin studies report serotonin transporter gene *SLC6A4* hypermethylation and lower *KCNQ3* gene methylation (associated with BD via neuronal hyperactivity regulation role). Candidate Plasticity Gene 2 (CpG2) (*SYNE1* splice variant) methylation status predicts previous mood episode number and suicide attempts. However, epigenetic mechanism associations with BD are relatively limited compared to**,** for example borderline personality disorder (BPD) [184]. Epigenetic alterations may influence BD development of risk and resilience (captured in BD parent offspring who developed the disorder) [185-86]. Conversely, early-life trauma-induced alterations destabilising epigenetic methylation, can increase later adult psychopathology risk [187-88]. DNA methylation related to gene expression repression, may become dysregulated from early childhood adversities, continuing into adulthood via sustained adult prefrontal cortex BDNF gene expression depletion. Recent hypotheses suggest 5-hydroxytryptamine 3A receptor (5-HT3AR) methylation could mediate early adversity effects on adult psychopathology, potentially modulating the risk for developing BD, BPD and ADHD. Additionally, *RELN* and *GAD67* gene downregulation (involved in GABA synthesis and secretion) was studied, this revealed respective promoter CpG island (CGI) hypermethylation evidence in BD and SCZ

patients. Post-translational histone modifications modulating transcription (CREB histone acetylation, *H3K4* trimethylation in synapsin genes) are also associated with BD, based on BD patient post-mortem brain sample studies [184].

Environmental Risk

Several environmental factors are proposed to trigger BD in biologically vulnerable individuals. The "Developmental Risk Factor" Model may clarify BD/SCZ similarities beyond shared genetic liabilities [189-90]. The model posits psychosis genetic predisposition combined with early-life experiences contributes to disorder development [191-92]. Identified risk factors include obstetric complications, peripartum asphyxia, low birth weight, maternal pregnancy stress, and perinatal infections. Maternal stress is also associated with SCZ, depression, anxiety, and a range of ADHD symptoms [193]. Research indicates concordant associations between obstetric complications, peripartum asphyxia and BD. Peripartum asphyxia (newborn oxygen deprivation) can lead to potential brain damage, resulting from complications [194] such as premature birth, prolonged labour, or cord suppression. Notably a brain MRI study found perinatal asphyxia/severe obstetric setbacks correlated with smaller amygdala and hippocampal volumes later [195]. A Finnish birth cohort study noted maternal smoking also increased BD risk [196]. However, a recent epidemiological twin study reinforced maternal stress poses a higher risk compared to smoking or alcohol consumption. Maternal stress was associated with subclinical hypomania, elevated mood, irritability symptoms in BD-risk youths and young adults [197]. Current research aims to identify potential epigenetic mechanisms illustrating gene-environment interaction risk influence in longitudinal studies. More understanding could be essential for effective prevention strategies.

Pharmacogenetics

Available BD medications include mood stabilisers, antipsychotics, antidepressants, and anti-anxiety medications. While pathophysiology and drug action understanding has grown, the exact BD medication combination depends on individual symptoms. Antidepressants are not recommended alone without a mood stabiliser (especially in BD1), as evidence suggests antidepressants may induce hypo(manic) episodes or increase mood switching. Antidepressants are also not advised if mixed features are present (coinciding depressive/hypomanic symptoms, are often indicated by irritability)[198]. Antidepressant BD mania trigger mechanisms are also unknown; one study identified opposing lithium (mood stabiliser) and fluoxetine (antidepressant), their therapeutic effects both converging in the PI pathway [199]. Antidepressants also affect serotonin and norepinephrine systems (SSRIs), and dopamine. Combined with a mood stabiliser (lithium, valproic acid [VPA], lamotrigine) to regulate mood, antipsychotics (olanzapine, risperidone) can also address psychosis symptoms. Medication responses vary widely; some patients cycle through medications before finding effective treatment with minimal side effects. Pharmacogenomic studies aim to leverage genetics to help predict treatment responses. A pivotal BD pharmacogenomics challenge is in measuring treatment response, restricted by follow-up duration, medication adherence, and multi-drug strategy confounders. The Alda lithium response scale [200], a systematic, high

inter-rater reliability rating system was developed to quantify BD clinical improvement during treatment, accounting for response confounders [201]. However, study design and sample heterogeneity yielded limited replication. While not yet replicable or presenting inconsistent results, promising BD pharmacogenomic findings were summarised in a recent review, see Gordovez and McMahon (2020) [58]. One Swedish/UK patient genetic study implicated an intronic SNP on chromosome 2q31.2 related to *SESTD1* (spectrin repeat containing nuclear envelope protein 1), a phospholipid regulation gene; intriguing as phospholipids are strongly associated with lithium targets [202]. Subsequently, a chromosome 21 locus was identified, involving long non-coding RNA (lncRNA) genes *AL157359.3 / AL157359.4* , which are crucial for CNS gene expression regulators [203]. An early identified genome-wide association (below genome-wide significance) implicated the gene *GRIA2* (glutamate ionotropic receptor AMPA type subunit 2) [204]. A notable Han Chinese patient study implicated the *GADL1* (glutamate decarboxylase-like protein 1) gene [205] that was not replicated in European [206] or Asian [207-208] samples. A recent meta-analysis (6,300 BD cases previously analysed for lithium responsiveness) replicated the *ADCY1* (adenylate cyclase 1) protein-coding gene association [202, 209-10]. *ADCY1* plays essential roles in regulatory processes implicating neuroplasticity, dopamine D4 receptors, sleep disturbances, and circadian rhythm dysfunction.

Most prior pharmacogenomic studies focused on lithium response. However, more recent studies explored the genetic associations with anti-epileptic mood stabiliser response, providing insights into two SNP-level associations (*THSD7A*, *SLC35F3*), and two gene-level associations (*ABCC1, DISP1*) [211]. Recent genetic findings illuminating probable drug targets associated with bipolar disorder (BD) suggest a potential for repurposing existing pharmacological agents. For instance, calcium channel blockers (CCBs), traditionally utilised in the treatment of hypertension and cardiovascular conditions, have garnered renewed interest as a therapeutic avenue for BD [212]. This resurgence is largely due to the widespread implication of the *CACNA1C* gene (calcium voltage-gated channel subunit alpha1 C), a locus consistently identified as one of the strongest genetic associations with BD [58]. Research further indicates that CCBs may exert neuroprotective effects [213] and influence neuroplasticity [214], although the potential for these agents to exacerbate certain symptoms, particularly cognitive deficits in bipolar disorder, warrants careful consideration, see section ‘Cognitive Deficits in Bipolar Disorder’ above.

Lithium in the PI Pathway and Calcium Signalling

Lithium is believed to exert therapeutic effects by modulating E/I balance, assuming a hyperactive excitatory system is key to BD pathogenesis (chronic animal model treatment reduced mGluR5-PKC signalling) [120]. Lithium is the first-line BD treatment reducing episode frequency and may also reduce suicide risk [215]. The exact mode of lithium actions is not well-understood; the pharmacological studies above suggest lithium targets multiple signalling pathways and regulatory network mechanisms.

Inositol depletion hypothesis

IMPase inhibition has the most support for lithium action mechanisms (shown to target many enzymes, often via Mg2+ co-factor) [199]. Similar structure enzymes also targeted (GSK3, β-arrestin-2-Akt complex) [216]) (Figure 6). Lithium inhibits two key mechanisms: inositol monophosphate (IMPase) and inositol polyphosphate-1-phosphatase (IPPase), depleting the available inositol cycle which supports downstream IC calcium-signalling. G protein activation by phospholipases via Trk receptors initiates phosphatidylinositol 4,5-bisphosphate hydrolysis, subsequently initiating the PI signal transduction cascade. Lithium IMPase inhibition prevents inositol availability for downstream targets (required in Inositol trisPhosphate (IP3), decreasing intracellular calcium release, preventing DAG/PKC activity. This downregulation and lithium's transduction cascade mechanism effects are well-supported. Altered PI signalling is reported in BD [215]. In vivo studies support lithium treatment reduces magnetic resonance (MR) spectroscopy myo-inositol [217]. Although inositol depletion is yet to be refuted, inconsistent findings exist, prompting researchers to explore alternative mechanisms.

Dopamine

Mood stabilising drugs lithium/sodium valproate also impact dopamine signalling. Valproate reportedly increased DAT gene expression via Sp transcription factor family interaction [218]. L-dopa (dopamine precursor) treats Parkinson's hallucinations. Certain antipsychotics for example, Haloperidol reduce mania by decreasing dopaminergic transmission via D2 receptor blockade. Observed therapeutic effects via dopamine transmission suggest D2/3 receptor blockers could be beneficial in BD depression [139]. Mood stabilising drugs lithium and sodium valproate also impact dopamine signalling. Valproate reportedly increased DAT gene expression via Sp transcription factor family interaction [218]. This corresponds with in-vivo neuroimaging findings; for example, a positron emission tomography (PET) study by Yatham and colleagues demonstrated that treatment with valproate was associated with a reduction in dopamine synthesis capacity in patients with mania, suggesting a direct modulatory effect on the presynaptic dopamine system [219].

GABA

Long-term BD patient mood stabilizer treatment reportedly upregulated PFC/hippocampus GABA receptors, simultaneously downregulating hypothalamus GABA receptors. In another study, lithium and valproate administration reinforced GABA importance. Serum GABA (reportedly to be low in depressed patients) increased with manic patient valproate treatment [146]. Despite therapeutic advances, many patients remain nonresponsive [220] or noncompliant (partly due to side-effect burden). In one lithium acute mania treatment meta-analysis, only 47% of BD patients had a good response [221].

Ketamine

Several N-methyl-D-aspartate (NMDA)-receptor antagonists recently gained attention for BD depression antidepressant effects [222]. Ketamine (anaesthetic drug) acts on NMDA receptor antagonist to target glutamate [223]. However, ketamine can have side effects including dissociation, increased blood pressure, nausea, and short-term cognitive changes. However, one review suggested there is still no clear treatment consensus. A recent case study suggested intranasal ketamine efficacy, demonstrating affective symptom stabilization at an 18-month follow-up. Future epigenetic therapeutics may include BD epigenetic effect study identified targets to address neuroprogression (e.g., histone methyltransferase inhibitor use has been suggested) [184]. This has precedence: lithium and antidepressants exert therapeutic effects via neurotrophic effects, maintaining adult CNS neuroplasticity. Neuroplasticity modulates mood, cognition, and behaviour sustaining mechanisms (including dendritic function, synaptic remodelling, long-term potentiation (LTP), axonal/neurite growth, synaptogenesis, and neurogenesis). Neuroprogression, potential BD patient brain volume loss with repeated mood episodes, may also be associated with lower treatment responsiveness, especially lithium and CBT [224-25]. This reiterates the importance of early therapeutic intervention to potentially preserve mechanisms, especially in those with a genetic predisposition to a chronic BD trajectory.

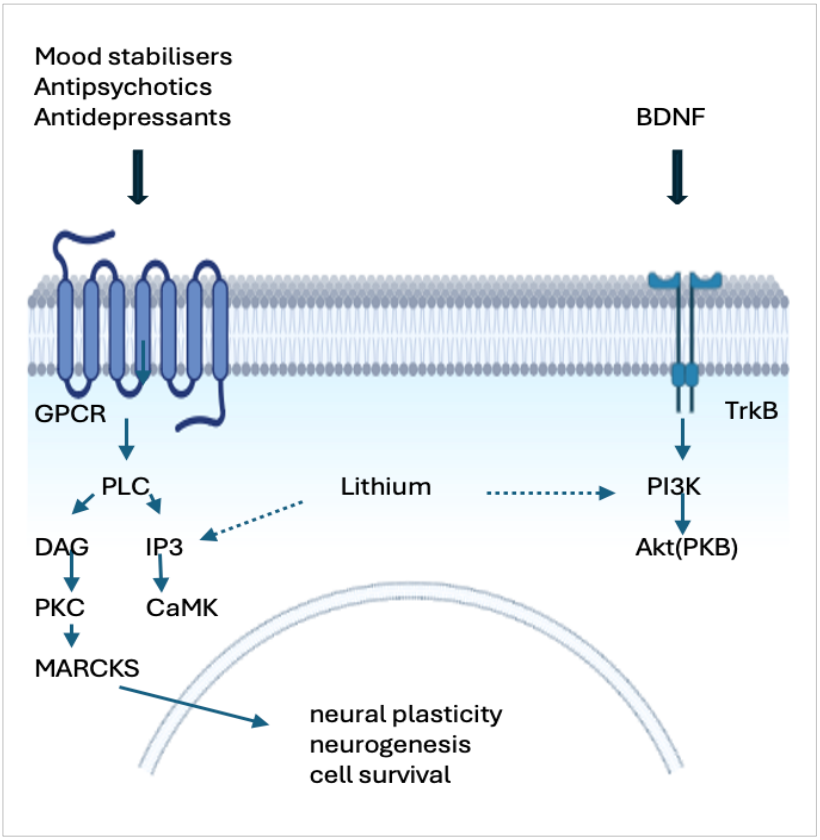


Figure 5 Intracellular Mechanisms of Therapeutics.
Adapted from Lee *et al*. (2022) [146]. *Neuromolecular Aetiology of Bipolar Disorder: Possible Therapeutic Effects of Mood Stabilisers.*

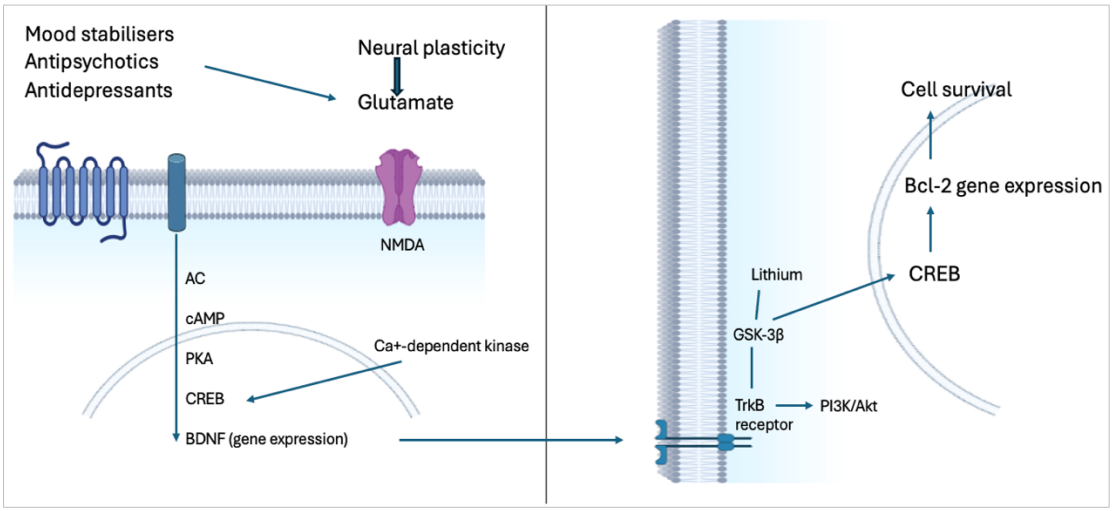


Figure 6 Neuroplasticity Effects of Therapeutics.
Adapted from Lee *et al*. (2022) [146]. *Neuromolecular Aetiology of Bipolar Disorder: Possible Therapeutic Effects of Mood Stabilisers*.

1.7 Genetics Of Bipolar Disorder

Family, twin, and adoption studies established BD is highly heritable, suggesting a strong genetic contribution. Instead of single dominant genes, research indicates multiple genes are involved, similar to other complex disorders. Recent BD genetic studies have revealed a complex inheritance pattern with multiple modest-effect genes contributing to risk, including common and rare genetic variants, and substantial overlap with other psychiatric disorders.

Twin and Adoption Studies

BD familial aggregation does not solely reflect genetic contributions; environmental factors also play a role. Adoption and twin studies help clarify genetics' impact. BD shows the highest psychiatric/behavioural disorder heritability [226] (estimates 59-87%). In Monozygotic (MZ) twins, this is approximately 38.5 to 58% concordance; dizygotic (DZ) twins: eight to 20% [227]. Adoption studies are inconclusive [228] but support largely genetic aetiology (BD risk was elevated only in biological parents). Despite this strong genetic contribution, no Mendelian inheritance pattern has been found, suggesting a complex inheritance which may also be influenced by assortative mating and shared environments [229]. This reiterates BD pathogenesis as multifactorial, involving genetics, social factors, trauma, as well as stress.

BD Familial Burden Studies

A large body of familial studies since 1960 indicates a strong BD genetic component (particularly BD1). A Swedish family-based study estimated BD risk is 7.9, 3.3, and 1.6 times higher for first-, second-, and third-degree BD patient relatives, respectively, compared to unaffected families. BD heritability was estimated at 58%. First-degree relatives' BD risk may be approximately 9% (nearly ten times greater than the general population) [230-31]. Evidence also suggests partial BD subtype genetic segregation (BD1 to BD2 *rG* is approximately .88 which is less than 1) (Figure 7). BD2 risk is higher among relatives of BD2 diagnosed individuals than for those with BD1 [232-34]. BD2 is considered more heterogeneous, positioned between BD1 and MDD [235]. In a large Swedish cohort, other psychiatric disorder genetic risks were 9.7-22.9 for BD individuals, and 1.7-2.8 for full siblings [236]. Similarly, in a Danish cohort study, a first-degree relative with mental illness increases SZA (bipolar type) relative risk to 2.76. Risks varied for related conditions: 2.57 for SCZ, 3.23 for BD, and 1.92 for SZA [237].

Familial BD Psychiatric Burden Studies

Studying multimorbidity in multiplex families (which feature a high concentration of individuals with BD) allows for delving further into BD genetic aetiology. In one bipolar multiplex family study, they reported that familial BD cases and unaffected members both exhibited higher genetic risk for BD, SCZ, and MDD [238] relative to unaffected families. SCZ, Autism Spectrum Disorder (ASD), and depression show particularly strong familial BD correlations [75]. MDD risk is greater than BD risk in BD families, with relatives of BD

diagnosed individuals more likely to be affected by this more prevalent MDD [236]. Familial BD risk also correlates with increased familial ADHD and personality disorders [30, 237, 239]. Studies consistently indicate BD relatives are more likely to have ADHD, suggesting shared familial and genetic predisposition. A familial genetic study meta-analysis revealed higher ADHD prevalence among BD relatives, with greater BD1 prevalence among ADHD relatives [240]. Borderline personality disorder (BPD) and antisocial personality disorder (ASPD) are also frequently observed in BD families. BPD may be more common in BD2 risk individuals. Research also indicates a substantial percentage of BD individuals also meet BPD criteria, leading some experts to propose BPD may reside on the BD spectrum. Although BPD and BD are both distinct, and separately diagnosable, approximately 20% of BD2 individuals and 10% with BD1 also qualify for a BPD diagnosis [241]. BD/BPD comorbidity is associated with more severe outcomes (increased psychosocial deficits, impulsivity, aggression, anxiety including OCD, post-traumatic stress disorder (PTSD), somatoform disorders, earlier mood symptom onset, hospitalisation, and worse treatment response) [242]. ADHD individuals show elevated BPD risk (nearly 20 times more likely than those without ADHD) [243].

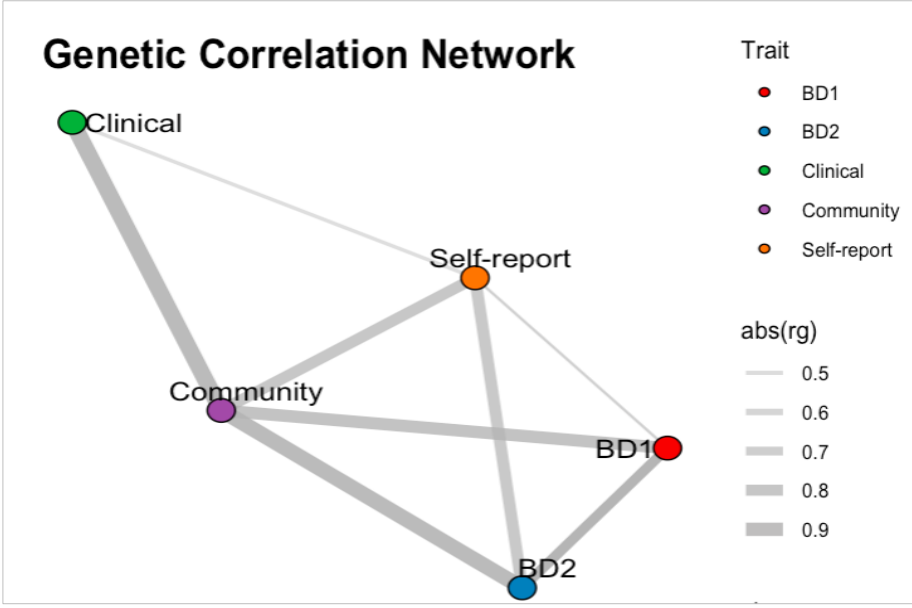


Figure 7 Genetic Correlation Between BD1 and BD2 Stratified by Ascertainment.
This figure is adapted from O'Connell, Koromina, van der Veen *et al*., 2025, *Genomics Yields Biological and Phenotypic Insights into Bipolar Disorder* [55]. Each node (circle) represents a specific BD phenotype (trait), indicated by its label and colour. Colours correspond to different ascertainment traits (Clinical, Community, Self-report, BD1 and BD2), as per the legend. Edges (lines) connecting nodes represent genetic correlations (*rG*) strength between traits. Thicker, darker edges indicate stronger genetic correlations. Node spatial arrangement by force-directed layout algorithm (Fruchterman-Reingold) positions nodes with stronger associations closer, identifying correlated traits.

1.8 Molecular Genetic Studies

Linkage Studies

Linkage studies indicated various BD-associated genomic locations by analysing family genetic marker inheritance. Several chromosomal regions are implicated, helping identify higher risk-associated candidate genes (e.g., a notable rare Copy Number Variant [CNV] on chromosome 16p11.2). Some region findings have replicated, however BD linkage study consistency remains elusive. To enhance statistical power, multi-study result meta-analyses were conducted. In one instance, 7 studies, found a potential 13q/22q loci [244]. In another, 18 studies, identified no linkage candidates, though several regions (9p22.3-21.1, 10q11.21-22.1, 14q24.1-32.12, and chromosome 18) showed evidence of linkage [245]. However, data in these meta-analyses tend to be less robust than direct combined data analyses [246]. Family linkage research has also focused on family subgroups (exploring greater genetic homogeneity, particularly concerning psychosis). Various studies found psychotic subtype linkage support. Noteworthy loci include: 1q42 near *DISC1* (scaffold protein) associated with SZA bipolar type [247]; 9q31/8p21 associations for BD with psychosis [248 Park]; and 5q26/18q12-q21 in a combined BD/SCZ analysis [249]. Overall, while linkage studies implicated certain chromosomal regions, results remained inconsistent. Methods have evolved with the growing evidence of polygenicity in complex disorders including BD. This complexity contrasts sharply with straightforward Mendelian disorder (e.g., cystic fibrosis) linkage analyses [250]. Researchers pivoted to GWAS to discover BD-associated genetic variants, adapting to traditional linkage study limitations.

Candidate Gene Studies

Research transitioned from gene linkage to BD-related gene associations investigated via candidate gene studies (examining specific "candidate gene"/BD development risk relationships by analysing genetic variations in affected/unaffected individuals). The initial focus established BD mechanisms involving neurotransmitter systems (dopamine, serotonin, and norepinephrine). Notable genes findings included *MAOA*, *COMT*, serotonin transporter, and circadian rhythm-associated clock genes (an extract of BD GWAS discoveries can be found in Appendix 9.1). The SCZ-related *DISC1* gene was examined but showed no consistent BD associations. In contrast, *DAOA* (D-amino acid oxidase activator)/G30 locus (chromosome 13q) variations are associated with BD susceptibility. Neuregulin 1 (*NRG1*) was also associated with BD (crucial in neurogenesis, synaptic transmission, and myelination roles). BDNF research highlighted larger sample size requirements. BDNF (brain derived neurotrophic factor superfamily member, chromosome 11p13) is vital for axonal development and neuronal population survival [251]. Initial studies suggested a *BDNF* polymorphism/BD association; later research yielded inconsistent findings [252]. Other genes (dopamine receptor D4 [*DRD4*], solute carrier family 6, member 3 [*SLC6A3*]) provided modest study support. Many early investigations faced type I error issues, further emphasising larger sample sizes' importance for identifying BD-associated genes.

To tackle the challenge of small sample sizes, Genome-Wide Associations Studies (GWAS) compared unrelated affected and unaffected individuals across many cohorts' using meta-analysis, eventually replicating the robust *BDNF* finding in BD [253]. Despite more than a decade of GWAS findings however, common variants alone evidently do not account for most twin study observed heritability, the ("missing heritability" problem). GWAS has been essential in robustly identifying hundreds of common genetic variants associated with complex traits including BD, thereby providing unprecedented insights into BD underlying biological pathways and polygenic architecture.

Rare Variants

Combined common SNP-heritability (h^2snp) is far lower than family/twin study estimates. Twin studies suggest BD heritability of approximately 60-90% [228]. SNP-heritability (h^2snp) (GWAS-derived estimate) is approximately 17-23% (based on a liability scale, .5-2% population prevalence) [55, 254-255]. Despite many common variants identified, none seem to substantially increase disease risk; many are also associated with SCZ and MDD. This raises the possibility that rare variants contribute to the heritability gap (structural variants [CNVs], rare SNVs, short indels), with potential gene-gene/gene-environment interactions. High costs meant BD whole-genome and exome sequencing (WGS/WES) studies typically focused on large family lineages (as they expected higher rare variant prevalence). Due to reduce costs for WES/WGS through innovation, studies began to provide evidence of increased rare deleterious variant burden in BD. BD1 diagnosed individuals exhibit greater rare deleterious SNV/rare CNV burden [256]. High disruptive variant burden is associated with BD age of onset [257]. In familial studies, up to 378 rare, non-synonymous, possibly functional variants were identified, indicating rare BD variant genetic overlap with ASD/SCZ [258-260]. A large BD exome sequencing study (Bipolar Exome [BipEx] consortium; including approximately 14,000 cases and 14,000 controls) by Palmer *et al*. (2022) [161] found excess ultra-rare protein-truncating variants (PTVs) in evolutionarily constrained BD patient genes. These PTVs were notably enriched within previously SCZ-implicated genes (Schizophrenia Exome Meta-analysis [SCHEMA] consortium). Combining their results with SCHEMA [309] data, *AKAP11* (A-kinase anchoring protein 11) was identified as a definitive shared BD and SCZ risk gene (odds ratio [OR]=7.06). Functionally, *AKAP11* interacts with GSK3β (a hypothesised lithium target). These findings support BD's polygenicity, re-establish rare coding variation role in BD aetiology, and underscore shared BD and SCZ genetic risk.

Copy Number Variants

Copy Number Variants (CNVs) have been investigated in BD. CNVs are stretches of deoxyribonucleic acid (DNA) that result in an individual having one (a deletion), three (a duplication), or more copies of a particular chromosomal region, instead of the typical two copies found in a diploid human genome. While large, rare CNVs have been associated with an increased risk for SCZ, they appear to play a smaller role in BD compared to their frequency and impact in SCZ or other neurodevelopmental disorders. Notably, a duplication on chromosome 16p11.2, initially detected in SCZ [261], has demonstrated the most robust

association with BD and has also been linked to an early age of onset for BD [262]. The chromosomal regions 1q21.1 and 3q29 have also been associated with BD, although these findings were below genome-wide significance [263]. Both the 16p11.2 and 3q29 regions are also implicated in SCZ, ASD, and intellectual disability [264]. One study has suggested that the contribution of rare CNVs to BD may be limited to cases of schizoaffective bipolar type disorder (SZA) [265].

Overall, the burden of CNVs has not been found to be substantially different between BD patients and controls. However, when stratified by subtype, an increased risk from CNVs was observed for SZA but not for other BD subtypes. This aligns with stronger evidence for CNV associations in SCZ and SZA compared to other BD subtypes [266]. These findings highlight a shared genetic architecture between BD and other psychiatric or neurodevelopmental disorders that extends beyond common variants, while also suggesting potential differential mechanisms underlying each condition. The comparatively reduced CNV burden in BD might be related to cognitive function, as CNVs identified in SCZ have been associated with cognitive dysfunction [265]. However, it is possible that smaller CNVs (less than approximately 30 kilobytes (kb) in size) are associated with BD but are more difficult to detect with current technologies [58].

De Novo Variants

Evidence suggests a potential impact of rare genetic variants, particularly de novo mutations (DNMs), on the genetic architecture of BD. Indeed, studies have identified de novo CNVs in individuals with BD, highlighting the role of these mutations. For instance, the first trio-based exome sequencing study in BD identified 71 de novo point mutations and one de novo copy-number mutation, many of which were predicted to be loss-of-function or protein-altering [267]. Certain BD subtypes, such as early-onset BD, exhibit a higher frequency of CNVs, including de novo ones [263], emphasizing the importance of these genetic variations in stratified risk assessment. Two studies specifically indicate that de novo CNVs contribute to the likelihood of early-age BD development [268]. A total of 107 de novo variants affecting protein-coding genes have been identified, showing enrichment in genes associated with the post-synaptic density and in phosphoinositide-linked pathways, which may be relevant to lithium's therapeutic effects [269]. The occurrence of de novo point mutations has been found to correlate with paternal age [270], and older fathers have an increased risk of having offspring with BD [271-72]. This finding aligns with similar associations observed in SCZ and ASD, reinforcing the relevance of paternal age to genetic risk for psychiatric disorders [273].

Single Nucleotide Variants

Rare Single Nucleotide Variants (SNVs) and small insertions/deletions (indels) are typically not captured by Genome-Wide Association Study (GWAS) Single Nucleotide Polymorphism (SNP) arrays. However, the advent of next-generation sequencing (NGS) technology has enabled several studies to identify rare, transmitted risk variants in both living BD patients and

post-mortem brain tissue [259, 274]. Large-scale WES efforts, such as the BipEx study which identified *AKAP11* as a risk gene [161] (as detailed above), have markedly advanced the understanding of rare coding variants in BD. A study by Ament *et al*. (2015) [274] involving the sequencing of 201 individual genomes, found that risk variants were predominantly noncoding with predicted regulatory effects. This research implicated rare variant associations with several genes (including *ANK3*, *CACNA1B*, *CACNA1C, CACNA1D, CACNG2*, *CAMK2A*, and *NGF*) that have also been associated with BD in common variant studies. These variants were found to be enriched in neuronal excitability pathways, such as those involving GABA and voltage-gated calcium channels. Indeed, SNV risk in BD families and case-control cohorts has been most strongly associated with risk variants in these ion channel receptor subunits.

RNA sequencing (RNA-seq) of post-mortem BD brain tissue has been used as a method to identify relevant BD-specific gene expression changes compared to those in ASD or SCZ [275]. Such brain transcriptome analyses using RNA-seq have detected potential dysfunctions in neuroplasticity, circadian rhythms, and GTPase binding, all of which are processes implicated in BD. However, sample sizes in next generation sequencing (NGS) studies remain relatively low, and collaborative efforts are likely needed to uncover variants that current studies lack the statistical power to detect [58]. Furthermore, recent evidence suggests a fundamental limitation in the use of postmortem tissue, as a study on cell type-specific transcriptional differences identified discrepancies between living and postmortem human brain tissue. Specifically, cell type proportion estimation was found to be more accurate in samples from living individuals compared to postmortem samples [276].

Genetic Interactions

Additional components of the still unaccounted for missing variance explained in BD may reside in genetic effects arising from gene-environment interactions (GxE) or gene-gene interactions (epistasis). For example, one study found that an interplay between a history of childhood trauma and reduced BDNF mRNA levels was associated with psychosis risk, potentially by impacting neurogenesis and leading to lower hippocampal volumes [277]. The *COMT* (Val158Met) polymorphism (rs4680) affects a catecholamine-degrading enzyme (which metabolises neurotransmitters such as dopamine), particularly in the prefrontal cortex (PFC). The Val allele leads to higher enzyme activity (faster dopamine breakdown) compared to the Met allele. Stressful life events (SLEs) are known to impact these same catecholamine systems. Therefore, it is plausible that an individual's *COMT* genotype could moderate their response to stress, thereby influencing BD susceptibility or its course [278]. A subsequent investigation in a sample of patients with First Episode Psychosis (FEP) found that the *COMT* (Val158Val) genotype moderated the association between severe SLEs and depressive symptoms, with Val/Val patients experiencing SLEs reporting the highest levels of depressive symptoms. It is important to note that FEP cohorts can include individuals who later develop BD; however, this study was not conducted in an exclusively BD-diagnosed cohort and specifically examined depressive symptoms within the FEP context [279].

1.9 Genome-Wide Association Studies

Genome-Wide Association Studies (GWAS) compares unrelated individual genomes with and without disease, to identify phenotype-associated genetic markers. Variation discovery helps identify potentially disease-contributing genes and pathways. Understanding these provides insights into underlying disease biological processes. GWAS feasibility improved through technological advancements, for example, the HapMap Project [280], and the 1000 Genomes Project Consortium [281]. Cost-effective technologies also facilitated 500,000 to 2 million SNP genotyping. Large-scale collaborations have emphasised combining cohort sample sizes in GWAS meta-analyses to boost statistical power for detecting small common variant effect sizes.

Early GWAS Insights

Early research in BD genetics centred on family and twin studies to identify genetic factors. While linkage and candidate gene studies hinted at chromosomal regions and specific genes, their outcomes were often not definitive, even with enlarged sample sizes. GWAS allow for the examination of genetic variant associations in much larger groups of BD cases and healthy controls, without requiring harder to recruit related individuals. In the past decade, many BD-associated loci have been reported and subsequently confirmed in meta-analyses, thereby consolidating our understanding of BD aetiology (for an overview of BD GWAS findings see 9.1).

In 2008, Baum *et al*. conducted the first BD GWAS (550,000 single nucleotide polymorphisms [SNPs]), uncovering a SNP association in the diacylglycerol kinase eta (*DGKH)* gene (critical for lithium-sensitive PI pathway) [282]. Although the BD risk effect was modest, suggesting BD complex polygenic disorder, follow-ups confirmed *DGKH* association in Han-Chinese [283] samples, and a nominal Japanese association [284]. To minimise type 1 errors, as millions of SNPs are tested, stringent multiple testing thresholds are used (typically a Bonferroni correction of $P < 5.0\text{x}10\text{-}^{8}$).

Cichon *et al*. (2011) proposed the neurocan (*NCAN*) gene as a potential BD susceptibility candidate [285]. *NCAN* is implicated in other mood disorders, suggesting overlapping genetic risk with ADHD, depression [286], and dyslexia [287]. The gene crucial for cellular adhesion/migration, is associated with brain volume and structure measures. The *NCAN* risk allele is associated with BD, depression, and SCZ patient mania. Ncan-deficient mice display hyperactive behaviour and impaired inhibition (a potential link to BD-reported cognitive deficits) [288]. *NCAN* loss could lead to cognitive impairments, and reduced brain volumes. BD brain imaging has indicated decreased cortical thickness, lower subcortical volume, and disrupted white matter integrity (described above).

In a further meta-analysis, Ferreria *et al*. (2008) evaluated the WTCCC, STEP-UCL, and ED-DUB-STEP2 cohorts. They identified strong associations with Ankyrin-G (*ANK3*) gene and L-

type voltage-gated calcium channel α-1C subunit gene (*CACNA1C*), both involved in ion channel functionality [252]. Ankyrin-G influences cell motility, activation, proliferation, and modulates neuronal sodium channel activity, which suggests ion channel dysfunction involvement in BD pathogenesis. These genes and ion channel dysfunction have subsequently been shown as robust BD associations, e.g., most recently in O'Connell *et al*. (2025) [55]. Not all associations replicated independently (e.g., suggested chromosome 16q12 locus, produced mixed results across samples). In the WTCCC study, locus in gene-rich high disequilibrium chromosome 16q12 region associated with BD [289]. This was not replicated in an independent reference study; other studies have found 16p12 linkage signal evidence in BD [290] and in psychosis [291]. Among the findings from the Wellcome Trust Case Control Consortium (WTCCC) in 2009 was an association with the potassium voltage-gated channel subfamily C member 2 (*KCNC2*) gene, which encodes the Kv3.2 subunit, although this association was below genome-wide significance [289]. This finding potentially implicated alterations in neuronal excitability in the mood episodes characteristic of BD. Support also existed for the previously identified involvement of the GABA and glutamate systems [292]. For example, gamma-aminobutyric acid type A receptor subunit beta-1 (*GABRB1*), which encodes the GABAAR β1 subunit, showed a high-ranking association in the WTCCC data, along with *SYN3* (synapsin III) [289]. These findings were strongly associated with SZA in a follow-up study by Craddock *et al*. (2010), which also included additional associations with *GABAAR* α4, α5, and β3 subunits [293]. The WTCCC study also identified a genetic association with BD in the region of gene *GRM7* (glutamate metabotropic receptor 7) [289]. The mGlu7 receptor, encoded by *GRM7*, is a presynaptic G-protein coupled receptor (GPCR) that modulates neurotransmission. Mutations or reduced expression of this gene have been associated with neurodevelopmental disorders and were previously linked to BD [294] and BD-related personality traits [295]. The potential role of this gene received additional support from findings showing a BD association with a rare CNV at the *GRM7* locus [296]. Following these initial discoveries, numerous GWAS have confirmed early findings and highlighted novel loci associated with BD.

Advances with Larger GWAS and Meta-Analyses

As GWAS sample sizes began to exceed 10,000 participants with BD, the number of genome-wide discoveries for BD increased substantially. Similar to many common traits, a large proportion of these variants are in non-coding regions of the genome and often have small effect sizes, with odds ratios (ORs) typically ranging from 1.1 to 1.3. The highest standard for validating these genetic associations is the replication of findings in independent cohorts. Several genes have been consistently associated with BD across multiple studies, including *ANK3*, *NCAN*, CACNA1C, fatty acid desaturase 2 (*FADS2*), mitotic arrest deficient 1 like 1 (*MAD1L1*), and tetratricopeptide repeat and ankyrin repeat containing 1 (*TRANK1*). However, limitations in sample size can still hinder the detection of variants with weaker effects. Collaborative meta-analyses are essential for increasing statistical power in genetic association studies. Nevertheless, they can introduce challenges, such as heterogeneity between study cohorts, which may impact the overall power to detect associations. This is compounded by

the inherent heterogeneity of the main BD phenotype itself. A more comprehensive list of key genes implicated in bipolar disorder can be found in Table 68, Appendix 9.2.

Advances in GWAS and Insights from the Psychiatric Genomics Consortium (PGC)

Several large-scale Genome-Wide Association Studies (GWAS) for bipolar disorder (BD) have been conducted by the Psychiatric Genomics Consortium (PGC) since 2011. The PGC's Bipolar Disorder Working Group (BDWG) has been instrumental in leading many of the genetic discoveries in the field. An early PGC-led GWAS (Sklar *et al*. 2011), which included 11,974 BD cases and 51,792 controls, reaffirmed the previously observed association between BD and the *CACNA1C* gene. Furthermore, a combined GWAS of BD and Schizophrenia (SCZ) uncovered strong SNP associations in the *CACNA1C* as well as in NIMA related kinase 4 (*NEK4*)-inter-alpha-trypsin inhibitor heavy chain 1/3/4 (*ITIH1/3/4*) region [297]. A subsequent BD GWAS conducted by Stahl *et al*. (2019) analysed SNP data from 29,764 BD patients and 169,118 controls, identifying 30 genome-wide loci associations [8]. This study again highlighted the roles of ion channels, neurotransmitter transporters, and synaptic components in the aetiology of BD. The strong association with *CACNA1C* was replicated, as were associations with *NCAN* and *ANK3*. Notably, fatty acid desaturase 1 (*FADS1*) and adenylate cyclase 2 (*ADCY2*) were among the newly associated genes. *FADS1* is associated with diacylglycerol lipase alpha (*DAGLA*), an enzyme crucial in the production of the endocannabinoid 2-arachidonoylglycerol (2-AG), which is involved in lithium's mechanism of action, retrograde synaptic signalling, axonal growth, and adult neurogenesis. The *ADCY2* gene had also been previously implicated as a BD risk gene (Mühleisen *et al*. 2014 [298]). However, the strongest association at the T*RANK1* locus reported in some earlier studies was not replicated in all follow-ups at that time.

A more recent GWAS (Mullins *et al*. 2021 [256]), encompassing 41,917 individuals with BD and 371,549 controls from more than 50 clinical cohorts, identified 64 independent loci [41]. This study successfully replicated 28 out of the 30 loci reported by Stahl *et al*., including the *TRANK1* association. The top association in the Mullins *et al*. study was also at the *TRANK1* locus on chromosome 3. Expression quantitative trait loci (eQTL) analyses suggested stronger, correlated expression regulation of doublecortin like kinase 3 (*DCLK3*), located upstream of *TRANK1*. BD was also associated with decreased expression of the furin paired basic amino acid cleaving enzyme (*FURIN*) gene, which has been implicated in neurodevelopmental disorders and in a 2019 SCZ GWAS [299]. The study further found that BD associations were enriched in gene sets related to neuronal compartments and synaptic signalling. BD risk alleles were particularly enriched in genes expressed in neurons and known to be targets for antipsychotics, calcium channel blockers (CCBs), and antiepileptic medications [256].

Investigating Homogeneous Subgroups in BD Genetics

The analysis of BD more homogenous subgroups within GWAS can enhance statistical power, especially for identifying genetic variants with smaller effects. This approach reduces

heterogeneity that enables more precise genetic analyses [300]. The largest genetic study of BD to date, conducted by O'Connell *et al*. (2025), involved a multi-ancestry meta-analysis of GWAS data from 79 cohorts, comprising 158,036 individuals with BD and 2,796,499 controls [55]. In this study, cases were stratified by ascertainment type (clinical, community, or self-report) and by ancestry (detailed in Chapter 6). This approach led to the discovery of 298 genome-wide loci, representing a fourfold increase in known associations, with 267 of these loci being novel to BD. The analyses highlighted the importance of specific cell types, notably GABAergic interneurons and medium spiny neurons, in the pathophysiology of BD. Common variants associated with BD were found to be particularly enriched in synaptic regions, as well as in prefrontal cortex and hippocampal interneurons, and hippocampal pyramidal neurons. Gene and gene set analyses indicated enrichments for targets of anticonvulsant, antipsychotic, and anxiolytic medications. The genetic architecture of BD was observed to vary among the ascertainment-stratified subtypes. This suggests that creating more homogeneous subgroups can help unravel the genetic basis of heterogeneous phenotypes, a crucial consideration for future BD genetic studies (evidenced in Chapter 5).

BD Pathway, Tissue, and Cell-type Enrichment Analyses

Secondary post-GWAS analyses are crucial for assessing functional enrichment in specific tissues or cell types, fine-mapping loci to identify credible causal variants, and potentially applying findings to risk prediction through individualised Polygenic Risk Scores (PRS). Genetic studies in BD have successfully identified specific biological pathways implicated in the disorder, including the regulation of insulin secretion, retrograde endocannabinoid signalling, glutamate receptor activity, and calcium channel activity.

Gene and Gene Set Pathway Analysis

Increasing sample sizes in BD GWAS have enabled robust pathway enrichment analyses, leading to numerous findings that pinpoint biological pathways associated with vulnerability to BD. Calcium signalling has been repeatedly implicated in BD (detailed above), and intracellular calcium signalling has been hypothesised as a key mechanism of lithium's therapeutic action [301-302]. Calcium is a ubiquitous signalling molecule that modulates critical neuronal processes such as neurotransmitter release, synaptic plasticity, and neurite outgrowth [303].

Several studies have also implicated *CACNA1C* in other psychiatric disorders, including SCZ and MDD. The PGC's Sklar *et al*. pathway analysis, which aimed to detect Gene Ontology (GO) term enrichment among the top 34 independent BD GWAS SNPs, identified an enriched pathway involving calcium channel subunits. This included three L-type calcium channel family members: calcium voltage-gated channel auxiliary subunits *CACNA1C*, *CACNA1D*, and *CACNB3* [297]. Research suggests that L-type calcium channels influence neuronal firing and regulate neuronal excitability, potentially contributing to the mood instability characteristic of BD [304].

Subsequently, Stahl *et al*. (2019) tested for enrichment in curated biological pathways from multiple sources, using competitive gene-set tests performed with MAGMA (Multi-marker Analysis of GenoMic Annotation) on GWAS data [305]. These analyses controlled for biases related to SNP and gene density, as well as gene size. Their findings reaffirmed the earlier associations of *CACNA1C* and other voltage-gated calcium channel genes with BD. Moreover, this work highlighted ion transport, neurotransmitter receptors, insulin secretion, and endocannabinoid signalling as containing potential novel therapeutic targets. The endocannabinoid system had previously been implicated in the pathophysiology of SCZ [306-308]. The gene set enrichment analysis by O'Connell *et al*. (2025) [55] identified six gene sets related to synapse function and transcription factor activity that were associated with brain gene expression and with early-to-mid-prenatal development. Consistent with a recent SCZ study suggesting that common and rare variants can converge on the same genes and biological pathways [309], O'Connell *et al*. (2025) found that 71 genes mapped to putatively causal SNPs were enriched for ultra-rare (defined as five or fewer minor allele counts) damaging missense or protein-truncating variants (PTVs) reported in the BipEx [310] or SCHEMA [309] datasets.

Fine-mapping Genes and Pathways

GWAS fine-mapping aims to identify the specific candidate genes within broader genomic regions that are most likely to be causally associated with BD. Analysing these gene associations seeks to enhance the understanding of genetic regulatory mechanisms implicated in BD. O'Connell et al. utilised transcriptome-wide association studies (TWAS) [311-12]. which explore correlations between gene-expression data and their associations with GWAS SNPs (using a Bonferroni-corrected $P < .05$). This approach, combined with six other fine-mapping strategies, confirmed the roles of 36 pathobiology-implicated genes in BD. The SP4 (Sp4 transcription factor) gene was highlighted by six of these analyses. SP4 is known to have regulatory influences on *GABAAR* subunit genes and astrocytes. Among the 36 studied genes, eight were mapped to presynaptic or postsynaptic compartments, with *CACNA1B* being identified solely in presynaptic compartments.

Single-cell Gene Expression Insights

Recent advancements in single-cell gene expression analysis have unveiled specific gene expression patterns that suggest potential neuronal dysfunctions associated with BD. Mullins *et al*. (2021) [256] analysed single-cell RNA-sequencing (scRNA-seq) data from adult human and murine brain tissue. They discovered enrichment for genes associated with both excitatory and inhibitory neurons, particularly within the cortex and specifically in the hippocampus. These findings indicated crucial activities in hippocampal pyramidal neurons and interneurons, as well as in the prefrontal cortex, with cell-type specificity being consistently observed across their analyses. Similarly, O'Connell *et al.* (2025) [55] implicated specific cell types, especially GABAergic interneurons and medium spiny neurons, in BD pathophysiology. Enrichment was also noted in dopamine-associated and calcium-associated biological processes, which are more often strongly associated with BD.

Other smaller studies using single-cell RNA-sequencing (scRNA-seq) data have allowed for a deeper examination of cell-specific transcriptional characteristics in key brain regions such as the dorsolateral prefrontal cortex (DLPFC) and subgenual anterior cingulate cortex (sgACC). Disruptions in specific classes of excitatory and inhibitory neurons were found to correlate with BD development. BD dysregulation was associated with two inhibitory cell clusters, specifically involving vasoactive intestinal peptide (VIP) GABAergic interneurons. Gene expression was dysregulated in two excitatory and inhibitory cell clusters, which included VIP GABAergic inhibitory interneurons [313]. Earlier studies had already suggested that VIP cells, which release GABA and inhibit other neurons, may play a role in BD [314].

RNA-seq analysis of postmortem DLPFC tissue has demonstrated differentially expressed (DE) genes and transcripts across various psychiatric disorders. These findings have implicated widespread dysregulation of biological processes, including neuroplasticity, circadian rhythms, and GTPase binding, in psychiatric illnesses [315].

It is likely that both neurons and glial cells are affected in BD. Some studies have suggested a potential decrease in cortical interneuron density in BD [316]. However, how these neurons and glial cells are altered structurally and functionally remains largely unknown. Post-mortem studies also suggest that a stoichiometric imbalance in gene expression, where the relative expression levels of certain genes are imbalanced, might be a key feature in BD development [317]. Stahl et al. (2019) [255] found that BD-associated genomic signals were enriched in neurons and oligodendrocyte precursor cells (OPCs). Analysis of transcriptomes from post-mortem BD brain samples of the sgACC and amygdala, when compared to neurotypical controls, suggested transcriptional changes in genes associated with the immune response, inflammation, and the post-synaptic membrane. These data converged on sodium voltage-gated channel alpha subunit 2) (*SCN2A*) and glutamate ionotropic receptor NMDA type subunit 2A (*GRIN2A*). Enrichment for neuroimmune and synaptic pathway genes, as well as microglia-specific genes, was found to be downregulated in BD [318]. Microglia downregulation has also been reported in PsychENCODE BD brain samples [319].

The exact role of glial-neuronal interactions in BD requires further investigation. Some studies have proposed an association between neuroinflammation and BD pathophysiology, possibly through processes that modulate brain structure and support cognitive and behavioural functioning. These processes likely involve synaptic plasticity, neurotransmission, neurogenesis, neuronal survival, and apoptosis [320].

The multiplicity of pathways implicated in BD pathophysiology may reflect the high genetic heterogeneity among individuals with BD. BD subtypes and other homogeneous subgroups have demonstrated some distinct familial patterns. Delineating BD cases into more refined subsets could enhance the power of discovery and help identify differential functional pathways that are currently masked by the heterogeneity of the main BD phenotype. Divergent genetic architectures for different BD subtypes have been reported, suggesting that increased genomic insights can be gained from stratifying cases by BD subtype or other homogeneous

subgroups, such as those defined by DSM-5 course specifiers, which may reflect differences in individual and familial trajectories.

1.10 BD Subtypes

Genetic Distinctions and Overlaps

GWAS specific to BD subtypes have been conducted, with secondary analyses often focusing on three prominent subtypes: BD1, BD2 and SZA. While clinical divergence exists between these BD subtypes, genetic analyses provide evidence of substantial overlap, particularly between BD1 and BD2 (as illustrated in Figure 7). However, when stratified by subtype, unique mechanisms and overlaps between the subtypes and other psychiatric disorders, previously obscured by analyses of the collective BD phenotype, become clearer.

Stahl et al. (2019) [255] identified 14 loci specific to BD1. In contrast, smaller analyses for BD2 and SZA in that study yielded no loci exceeding the threshold for genome-wide significance ($P < 5.0x10\text{-}8$). SNP-based heritability (h^2snp) (estimate of proportion of a trait's variance explained in a single sample by looking at a specific, common set of genetic markers (SNPs) across the entire genome) was found to differ across subtypes: BD1 and SZA showed similarly high heritability, while BD2 had a lower SNP-heritability (BD1 $h^2snp = .25$, s.e.m. = .014; BD2 $h^2snp = .11$, s.e.m. = .028; SZA $h^2snp = .25$, s.e.m. = .10). Linkage Disequilibrium Score Regression (LDSC), used for estimating h^2snp and genetic correlations between traits, revealed divergent genetic correlations for the subtypes. A stronger genetic correlation was evident between SCZ and BD1 (rG=.71, s.e.m. = .025) compared to SCZ and BD2 (rG=.51, s.e.m. = .072). Conversely, a stronger genetic relationship was found between MDD and BD2 (rG=.69, s.e.m. = .093) versus MDD and BD1 (rG=.30, s.e.m. = .028).

Polygenic Risk Scores (PRS) have increasingly facilitated the genetic risk stratification of BD subtypes. PRS are a statistical estimate of an individual's genetic predisposition to develop a complex disease by summing up the effects of thousands of common genetic variants across their genome. Stahl *et al*. (2019) used PRS analyses to differentiate BD subtypes and psychotic cases based on their genetic burden for SCZ and MDD (as shown in Figure 8). PRS for SCZ risk alleles were higher in BD1 cases than in BD2 cases, and higher in psychotic cases versus non-psychotic cases. Conversely, PRS for MDD risk alleles were elevated in BD2 cases compared to BD1 cases. This suggests that MDD risk alleles contribute more to case-control differences in BD2, while SCZ risk alleles are more predictive of variance in BD1 and psychosis.

Mullins *et al*. (2021) [256] also conducted stratified GWAS analyses for BD1 and BD2 separately, which increased the discovery of BD1-specific loci to 44, with 13 of these being unique to BD1. The strongest signal among these 13 unique BD1 loci was in a region associated with the *HTR6* (5-hydroxytryptamine receptor 6) gene. This gene encodes a G-protein coupled

receptor (GPCR) that is a target for various antidepressant and antipsychotic medications. For BD2, a single association was identified with the slit guidance ligand 3 (*SLIT3*) gene, which is involved in axon guidance, cell migration, proliferation, and differentiation. Genetic correlation analyses indicated that BD1 and BD2 are overlapping yet partially distinct phenotypes, with a correlation between them ranging from .85 to .88 (s.e.=.05), as illustrated in Figure 7.

O'Connell *et al*. (2025) [55] further explored the genetic architecture of BD subtypes. They identified heterogeneity related to ascertainment type among 25,060 BD1 cases and 6,781 BD2 cases, and stratified their analyses by ascertainment source (clinical, community, or self-report), as shown in Figure 9 (and detailed in Chapter 6). A substantial proportion of BD1 cases were clinically or community-reported, whereas self-reported cases had a higher representation of BD2. This observation was supported by genetic correlations with other psychiatric disorders: Schizophrenia (SCZ) showed a stronger correlation with BD1, while self-reported BD indicated higher genetic correlations with major depressive disorder (MDD), anxiety-related obsessive-compulsive disorder (OCD), attention-deficit/hyperactivity disorder (ADHD), and borderline personality disorder (BPD). Divergent patterns for BD1 and BD2 had been previously characterized, with more genetic overlaps observed between BD2 and other psychiatric conditions compared to BD1 [321]. Specifically, for BD1, the genetic correlations were: MDD (.34, s.e.m. = .023), anxiety-related OCD (.29, s.e.m. = .067), and ADHD (.14, s.e.m. = .032). This contrasts with the higher genetic correlations found for BD2 with these conditions: MDD (.65, s.e.m. = .048), anxiety-related OCD (.50, s.e.m. = .113), and ADHD (.42, s.e.m. = .049) [55]. Figure 9 illustrates this same pattern of association signals. This discrepancy could indicate the overdiagnosis of BD in outpatient settings, particularly among individuals with comorbidities [322-23], which may confound the observed differences in subtype genetic architecture [324]. Despite these subtype differences, the inclusion of additional multi-ancestry data risk alleles improved both the discovery of subtype-specific associations and polygenic prediction for these subtypes [55] (as detailed in Chapter 6).

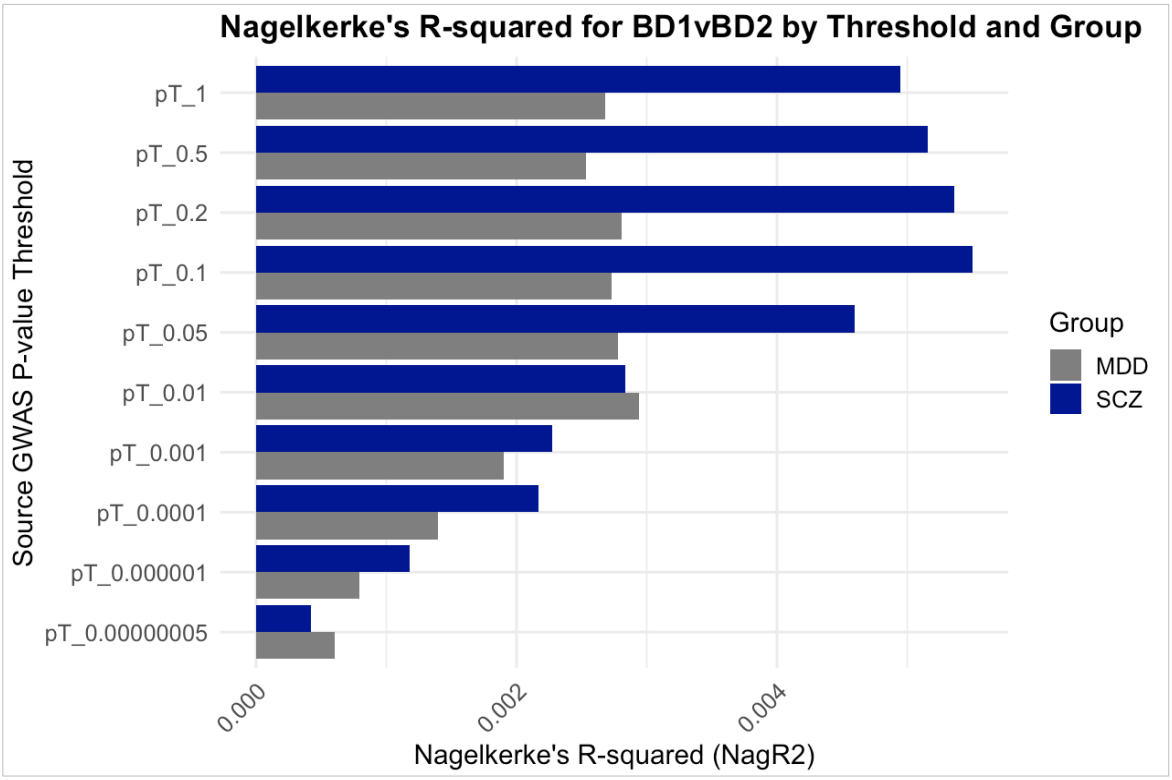


Figure 8 PRS of MDD or SCZ in BD1 and BD2.
Adapted from Stahl *et al*. (2019), *Genome-wide association study identifies 30 loci associated with bipolar disorder* [255]. Nagelkerke's R-squared (NagR2) for BD1 versus BD2 case status association with PRS, stratified by source GWAS *P*-value threshold (pT) across disorder (SCZ: schizophrenia, MDD: Major Depressive Disorder). Bars represent NagR2 value (subtype variance [BD1 vs BD2] explained by PRS at each threshold, within SCZ or MDD group). This illustrates PRS BD1 and BD2 subtype specific polygenic signals across varying PRS inclusion thresholds (stringency), separately for individuals using SCZ or MDD GWAS (colour) as variant discovery (SNPs) set. Bar length indicates association strength (NagR2).

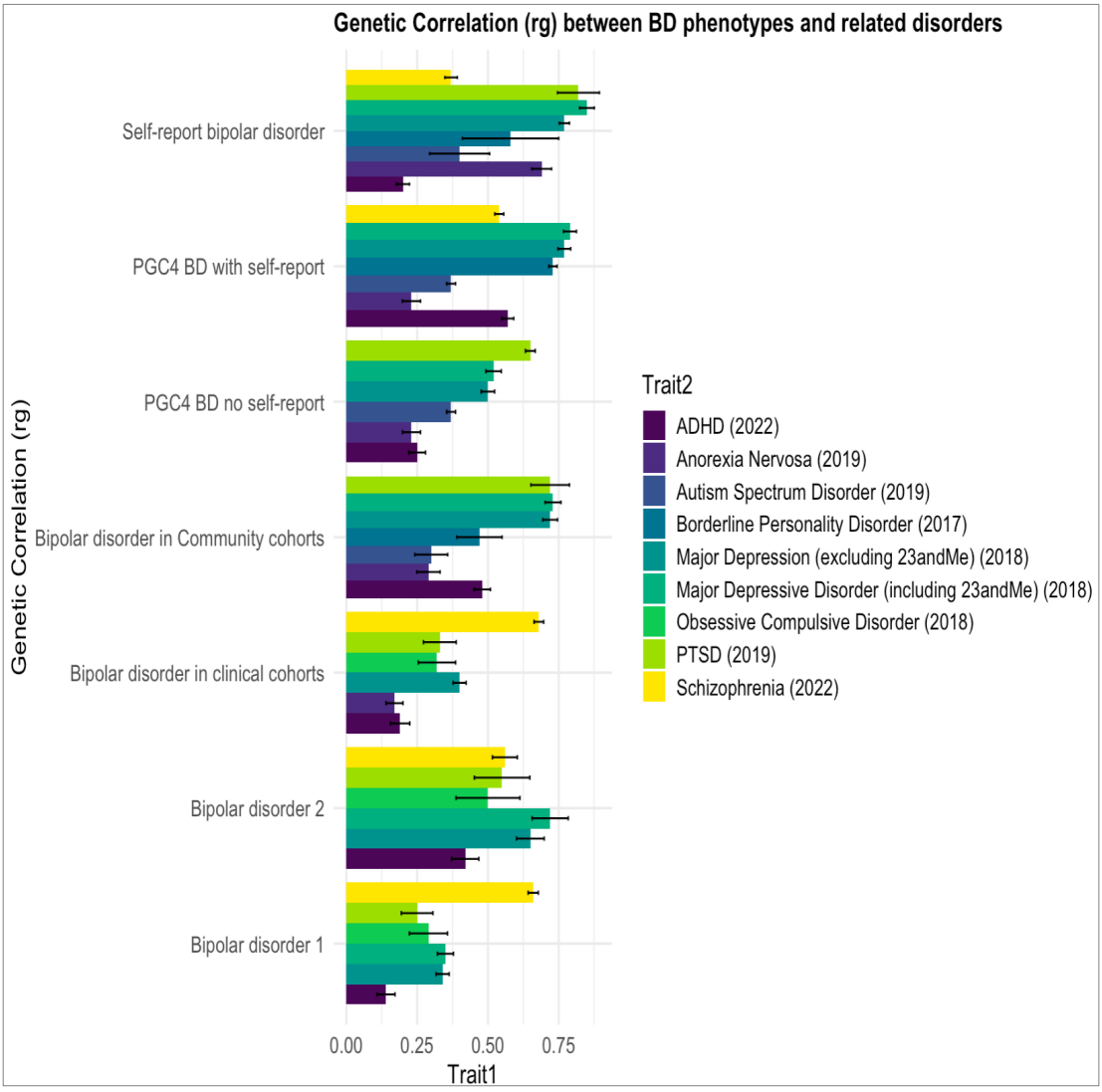


Figure 9 Genetic Correlations of BD1 and BD2 by Ascertainment and Related Traits.
This is adapted from O'Connell, Koromina, van der Veen *et al*. (2025), *Genomics Yields Biological and Phenotypic Insights into Bipolar Disorder* [55]. Genetic correlations (*rG*) and standard errors (s.e.) between primary BD phenotype (Trait1, y-axis) and related disorders (Trait2, bar colour). Horizontal bars: estimated positive *rG*. Black error bars: ±1 SE around the *rG* estimate. This illustrates shared and differential genetic burdens in BD subtypes.

1.11 Polygenic Risk Scores

BD's high pleiotropy and polygenicity contribute to substantial heterogeneity, which complicates the identification of genetic causes through GWAS that focus on the primary BD phenotype. Besides examining genetic correlations, Polygenic Risk Scores (PRS) are utilised to evaluate an individual's genetic burden associated with a specific trait and to assess overlaps in genetic burden with other relevant traits. PRS can help reduce heterogeneity by classifying individuals based on their genetic burden signatures relative to numerous psychiatric conditions and human diseases. Although no single PRS method currently accounts for a substantial portion of the variation in the main BD subtypes, PRS for BD and other related traits are employed in research to gauge genetic burdens and explore their clinical implications, such as in illness onset and progression (as detailed in Chapter 3 and 4). While PRS have demonstrated reliable but modest predictive power across complex phenotypes, further diversification of these methods is needed. This is required to address issues related to the missing variance explained, including the biases introduced by the predominance of European ancestry, sole focus on common variants (SNPs) and between-cohort heterogeneity in large consortia genetic studies. However, it remains uncertain whether BD PRS will ever become robust enough for future clinical classifications, such as enhancing diagnostic procedures or potentially providing more personalised treatment approaches.

The PRS for an individual *j* is typically calculated using the following equation:

$$PRS\_j = \Sigma_i (\beta_i * dosage_{ij})$$

where:

- **PRS_j** is the polygenic risk score for individual *j*.
- Σ_i represents the summation across all SNPs *i* that have passed the clumping and the current P-value threshold.
- β_i is the **effect size estimate** of the *i*-th SNP on the trait of interest, as reported in the GWAS summary statistics (e.g., beta coefficient for quantitative traits, log odds ratio for binary traits).
- $dosage_{ij}$ is the number of risk alleles carried by individual *j* at the *i*-th SNP (typically coded as 0, 1, or 2, representing the number of copies of the risk allele).

For each SNP selected at a given P-value threshold after clumping, the weight contributing to the individual's PRS is the effect size observed for that SNP in the original GWAS. Individuals with more risk alleles for SNPs with larger effect sizes will have a higher PRS. The equation used to calculate weights in PRSice, uses the beta coefficients (or log odds ratios) from the input (discovery) GWAS summary statistics for the SNPs that pass the clumping and P-value thresholding.

Figure 10 *P*-value thresholds approach.
pT+clump approach in PRSice software adapted from Euesden *et al*. (2015), *Polygenic Risk Score Software* [325]. PRS is estimated using a two-step redundancy minimizing process: 1. Clumping selects independent SNPs, and 2. *P*-value filtering retains most significant SNPs for PRS calculation.

The goal is to model the relationship between the phenotype (y) and a large number of SNPs (X):

$$y = X\beta + \varepsilon$$

where:

- **y** is the vector of phenotypes.
- **X** is the genotype matrix.
- **β** is the vector of SNP effect sizes (weights we want to estimate)
- **ε** is the error term.

PRS-CS places a **continuous shrinkage** (CS) prior on the **SNP effect sizes (β)**, which induces sparsity. This prior is crucial for inducing sparsity, which means it assumes most SNPs have small or zero effects, with few larger effects. The degree of overall shrinkage is controlled by the **global shrinkage parameter (phi - φ)**. **PRS-CS-auto avoids the need for a validation set by using a fully Bayesian approach to automatically learn the optimal value of φ directly from the GWAS summary statistics, placing** a **standard half-Cauchy prior** on φ to inform about the level of sparsity:

$$\varphi \sim \text{Cauchy}+(0, 1)$$

In contrast to PRSice, the weights in PRS-CS-auto, are dependent on the Bayesian model and the Gibbs sampling algorithm used for posterior inference, rather than a single equation for each weight. The final weight for each SNP is a data-driven estimate of its true effect size, shrunk towards zero based on the overall genetic architecture and the LD patterns.

Figure 11 PRS-CS Continuous Shrinkage Method.
Adapted from Ge *et al*. (2019), *Polygenic Prediction via Bayesian Regression and Continuous Shrinkage Priors* [326]. This method uses a Bayesian regression framework with continuous shrinkage priors applied to SNP effect sizes, learned directly from the data, to improve risk prediction accuracy by better handling LD structure.

Polygenic Risk Scoring Methods

PRS methods are essential for understanding the genetic underpinnings of complex traits such as BD. GWAS have identified numerous variants associated with BD; however, these often have small individual effects and collectively represent only a portion of the overall heritability. Yang *et al*. (2010) illustrated that most of the heritability for traits such as height can be explained by aggregating the effects of thousands of SNPs [327]. The classic PRS calculation involves summing an individual's risk alleles, with each allele (SNP) weighted by its effect size on the phenotype as determined by a GWAS. Large-scale GWAS from consortia such as the Psychiatric Genomics Consortium (PGC) offer robust discovery datasets due to their size, leading to more accurate individual scores. Typically, common biallelic alleles, defined by a minor allele frequency (MAF) greater than 1%, are considered in PRS construction, although variants with lower MAF, which are rarer, have been incorporated in more recent PRS studies [257, 263].

Different PRS methods are employed to select and weight genetic variants from GWAS, addressing the challenges inherent in robust polygenic score analysis. A common issue with some PRS approaches is the potential for overfitting, particularly when variant selection is based solely on *P*-value thresholds. Shrinkage methods, which can reduce genetic effect estimates, are used to enhance model generalisability.

When comparing PRS methodologies, several have been analysed for their predictive utility. These include traditional methods including *P*-value thresholding and clumping (pT+clump) (Figure 10), as well as more modern approaches such as Polygenic Risk Score - Continuous Shrinkage (PRS-CS)-auto [326] (Figure 11), which utilises continuous shrinkage (CS) priors. A recent study suggested that methods such as PRS-CS-auto can outperform classic pT+clump techniques in terms of predictive accuracy [328]. Chapter 4 of this thesis replicated this improvement using PRS-CS-auto, which was subsequently utilised in the work for Chapters 3-6. Unlike methods such as pT+clump (e.g., as implemented in PRSice [325], PRS-CS-auto utilizes a Bayesian Regression Framework. This framework incorporates a continuous shrinkage (CS) prior and facilitates automatic learning of the global shrinkage parameter (phi). A key difference is that PRS-CS-auto does not rely on fixed *P*-value thresholds for SNP selection. Instead, it estimates the posterior effect size for each SNP simultaneously, with most of these effect sizes being shrunk towards zero. The framework's use of a CS prior for SNPs allows for improved handling of linkage disequilibrium (LD) and SNP effect sizes. This, in turn, enhances local LD modelling and improves the prediction of genetic liability for complex traits. To calculate a comprehensive PRS for a target cohort, PRS-CS-auto requires PLINK 2.0 [325] to weight SNPs by their respective effect sizes to estimate individual's risk scores.

Polygenic Risk Burden in BD

A higher bipolar disorder (BD) Polygenic Risk Score (PRS) is associated with an increased risk of BD in offspring (Hiser and Koenigs 2017 [329]). It has also been associated with risks for other traits, including other psychiatric disorders, variations in brain structures, differences in cognitive abilities, and various clinical outcomes. BD PRS reflects a genetic predisposition that is crucial in the familial transmission of BD. These scores are typically higher in parents with BD and their offspring compared to unaffected individuals, even beyond the consideration of parental diagnosis [330]. Notably, in a recent PRS analysis of new-onset cases among high-risk offspring, BD-PRS predicted person-level BD, particularly in the offspring of parents with an earlier age of onset who also presented with anxiety (ANX) or depression symptoms [331]. An increased BD PRS has been associated with increased odds of developing psychotic symptoms [332]. Furthermore, a higher BD PRS correlates with an increased likelihood of developing both BD1 and BD2, with a particularly strong association for BD1 [55, 255-256] and could be a function of symptom severity [324].

While BD2 was previously considered a milder version of BD1, suggested by evidence of lower BD genetic burden, recent research has challenged this notion, based on the evidence of potential increase cross-disorder burden. A comparison of clinical differences between BD1

and BD2 in multiplex families revealed a continuum of severity, where BD1 was associated with a higher BD PRS, which in turn predicts more severe manic and depressive symptoms [324]. Conversely, BD2 was found to be correlated with an increased genetic risk for comorbidities, potentially predisposing individuals to chronic illness. The burden of depression, ADHD, and anxiety is reported to be greater in individuals with BD2 than in those with BD1 [321].

PRS for BD and other traits have been utilised to help explain common comorbidities observed in BD. A higher BD PRS also predicts suicidal ideation in BD multiplex families. While a PRS for suicide has been associated with suicide itself [333], it is not always predictive of such outcomes [334]. Suicidality is understood to be influenced by a combination of genetic, environmental, and clinical factors. Suicide attempts in individuals with BD have been associated with a higher genetic liability for depression [335] and trauma-related outcomes [336]. PRS for ADHD [337], MDD, and ANX [334] have also been associated with suicidal behaviour.

Research has also demonstrated an association between a higher BD polygenic burden and potential endophenotypes. For example, an association was recorded between a thinner ventromedial prefrontal cortex (vmPFC), a brain region critical for social and affective functioning (including emotional regulation, decision-making, and social recognition) and a BD diagnosis [329]. A higher BD PRS has also been associated with lower fractional anisotropy (FA), indicative of reduced widespread white matter integrity. It has been suggested that distinct bipolar subtypes may reflect varying degrees of disease expression, with an observed increase in white matter microstructure disruption from BD2 to BD1 [338]. Associations with brain structure changes are likely age-dependent and may be influenced by the number of mood episodes experienced and the neuro-progressive effects of medication. Ongoing research seeks to clarify the exact mechanisms and extent of these relationships. Notably, in a randomised trial examining brain structure changes in youth with BD, alterations in pretreatment neuroanatomic features were found to predict treatment outcomes, with these features later improving with treatment [339].

Comorbid Polygenic Burden in BD

Investigations of multiplex BD families have shown a higher genetic burden for common SCZ and MDD variants in these families [239]. Patients with psychotic features experienced a higher genetic risk profile, as indicated by SCZ and BD PRS, which explained 9% and 2% of the variance in psychosis, respectively [340]. A genetic overlap is also apparent between BD and ADHD, though this overlap is less pronounced with other childhood psychopathologies [341]. Individuals with a childhood history of ADHD who later develop BD have shown increased ADHD genetic liability, an earlier onset of BD, and higher chances of comorbid ANX and SUD [336, 342-343]. An interplay between the genetic burdens for ADHD, SCZ, and BD has been observed to increase the risk for alcohol and nicotine dependence [344-347].

Some studies suggest that BD may have a neurodevelopmental basis [348], with early signs likely preceding major mood episodes, a theory supported by twin studies [349]. Longitudinal research has also associated BD PRS with childhood conduct and oppositional defiant difficulties [113], while elevated ADHD and ANX PRS have been associated with a higher risk for rapid cycling BD [332, 350]. Individuals with rapid cycling BD typically experience an earlier onset of the disorder and have an increased suicide risk compared to those with non-rapid cycling BD [68]. A transdiagnostic PRS approach has also shown promise in improving predictions of lithium response in BD patients [351], as heightened genetic liability for depression and SCZ correlates with poorer responses to lithium [352-353].

Clinical Dimensions

Personality traits are often represented dimensionally, existing on continuous distributions rather than as distinct categories, which allows for a more nuanced understanding of individual differences. Similarly, there is increasing recognition of the dimensionality of BD symptoms. Twenty-four clinical variables related to BD have been stratified to demonstrate shared and differential genetic burdens between BD and SCZ. Psychosis showed a high polygenic risk from both BD and SCZ PRS, whereas mania was better predicted by BD PRS specifically [190]. This aligns with recent research suggesting that mania, depression, and psychosis can be considered distinct dimensions of bipolar disorder, each with potentially unique underlying causes and outcome patterns [354]. This recent study of BD patients found that MDD PRS was most strongly associated with the depression dimension, while BD PRS best predicted the mania dimension. The psychosis dimension, in turn, was most strongly associated with SCZ genetic burden (replicated in this thesis Chapter 3). Another transdiagnostic dimensional study reported that BD PRS was negatively associated with the depression dimension [355]. Genetic signatures for ADHD and ANX have also been implicated in BD pathophysiology, especially in rapid cycling BD, suggesting that this three-factor model (mania, depression, psychosis) should be extended to include these additional genetic burdens [351] (as evidenced in Chapter 3).

Beyond identifying individuals at higher risk for complex diseases, PRS offer the promise of clinical risk stratification, advancing personalised medicine, facilitating early intervention, and potentially informing therapeutic decisions. However this methodology has constraints. PRS do not consider environmental confounders. Environmental factors can interact with genetic predispositions, for example, by altering gene expression and impacting BD development, which in turn can alter the utility and interpretation of polygenic prediction. Epigenetic factors have been proposed to exert neurobiological consequences in BD, as well as in the context of childhood trauma, psychotic disorders, rapid cycling BD, and particularly ADHD [329, 350, 356-358].

Sex Differences

Sex differences in BD are primarily observed at the phenotypic level, affecting symptoms, course, and outcomes, rather than at the genotypic level. These differences may stem from environmental influences or subtle genetic interactions that current studies may not fully capture. Sex differences exist in the presentation and progression of BD: females tend to experience more depressive episodes, mixed features, rapid cycling, and report higher rates of suicide attempts [359]. Males, on the other hand, often report mania more frequently and show a higher prevalence of SUD. Comorbid conditions such as thyroid disease, migraines, obesity, and ANX are more common in females.

Despite these observable differences in symptoms, there is limited evidence to indicate that sex affects the response to mood stabiliser treatment [360]. The general response to lithium does not appear to be sex-dependent but rather may be in part driven by individual differences in transdiagnostic genetic burdens, although some studies suggest that females might experience more adverse pharmaceutical side effects, including hypothyroidism [361]. Historically, BD has been categorized with psychiatric conditions that show no gender difference in lifetime prevalence within the general population, which contrasts with MDD's consistent higher prevalence in females [359]. The first large-scale sex-stratified GWAS of BD indicated a largely overlapping genetic architecture between sexes, an overlap that was even more similar when focusing solely on BD1. This suggests that observable sex differences in BD might be predominantly associated with the risk architecture of BD2 and its overlap with MDD (Yang *et al*. 2023 [56]).

A recent multivariate analysis of the genetic architecture of eight psychiatric disorders, which identified three primary factors (psychotic, neurodevelopmental, and internalizing), found that problematic alcohol use and PTSD loaded more on the internalising factor for females. Additionally, four phenotypes (educational attainment, insomnia, smoking, and deprivation) demonstrated some, albeit small, sex-differentiated associations with the psychotic factor [362]. Future research in BD aimed at exploring sex differences and specific subgroups will necessitate larger sample sizes to effectively investigate sex differences at the subphenotype level. Moreover, this underscores the importance of thoroughly understanding the complex interplay of genetic and environmental factors that potentially contribute to sex differences in BD.

Predictive Utility of BD PRS

While PRS could enhance early detection and risk stratification for BD, their predictive power is presently limited and insufficient for routine clinical application. Integrating clinical data with PRS can boost predictive accuracy, especially in specific populations, such as the offspring of patients with early-onset mood disorders (including anxiety and depression) [331], or BD1 cases with psychosis (evidenced in this thesis Chapter 4).

Challenges remain in estimating the broad-sense heritability of the BD phenotype due to the complexity of genetic interactions. Most PRS models rely on narrow-sense heritability, which is the proportion of variance attributable only to additive variant effects and are further limited to variants that are high-quality genotyped or well-imputed, leading to only incremental gains in predictive utility. The performance of PRS derived from large multi-ancestry GWAS has explained only approximately 9% of the phenotypic variance in European cohorts. Individuals in the top quintile (20%) for BD risk, as determined by these PRS, had an odds ratio (OR) of 7.06 (95% CI = 3.9-10.4) for a BD diagnosis (detailed in Chapter 6). Moreover, the ability of BD PRS to explain phenotypic variance in European cohorts is still relatively low; however, accuracy is expected to increase with the inclusion of larger samples from non-European ancestries to address current disparities.

Chapter 5 PRS results were competitive with, and for several BD subphenotypes, exceed the results reported in several, recent, larger-scale PGC studies [256, 55] for a broadly defined BD phenotype. This suggests that the subphenotypic approach can enhance predictive power by leveraging more specific genetic signals, even when individual cohort sample sizes for subphenotypes might be smaller.

Ongoing research (as detailed in Chapters 1 to 7) indicates that BD shares both unique and overlapping genetic mechanisms with multiple disorders. Transitioning towards diagnoses based on biological pathology will likely require the use of multiple, cross-trait, broader genetic architectures than only polygenic prediction measures. This approach may enhance patient assessment and therapeutic interventions. This is particularly important given the critical association between delayed treatment and adverse outcomes in BD. BD is often misdiagnosed, leading to delays in treatment or the administration of inappropriate treatment, thereby lengthening periods of distress, disability and potentially increasing morbidity and mortality; it remains a leading cause of lost life years for individuals aged 15-44 [54].

PRS Methodological Advancements

Methodological advancements, such as focusing on genetically distinct subgroups and incorporating rare variants, could improve the predictive power of PRS. The ongoing development of frameworks that combine common and rare variants [363] is leading to better predictive accuracy. Rare variants, typically those with a minor allele frequency (MAF) of less than 1%, are often excluded from PRS calculations due to low statistical power. Williams *et al*. [363] recently developed a new PRS framework that calculates separate PRS for common and rare variants. Analysis of real data using this framework showed an improved predictive accuracy by an average of 25.7% when compared to leading PRS methods that use only common variants. Multiple polygenic risk score (MPS) approaches and machine learning techniques show promise in improving diagnostic precision and predictions of treatment response. Krapohl *et al*. [364] demonstrated a 10-fold increase in the variance explained for developmental outcomes by using an MPS approach that incorporated data from 81 well-powered GWAS. Craig *et al*. [365] derived PRS based on a multi-trait analysis using GWAS data for glaucoma and its endophenotypes. The multi-trait PRS demonstrated better prediction

ability than PRS based on any single input trait. A novel strategy involving first stratifying patients genetically by their BD and SCZ risk using PRS and then training machine-learning models with clinical predictors, led to large improvements in predicting lithium response (Cearns *et al*. 2022 [6]). Hansen *et al*. (2025) [366] showed that the detection of SCZ progression is achievable by applying machine learning algorithms to clinical data from electronic health records (EHR), potentially facilitating a reduction in diagnostic delay. PRS combined with clinical data was most predictive of outcomes in BD1 in Chapter 4. Overall, while PRS currently have limitations, future developments in study design and methodology hold the potential for further the methodology to advance our understanding and treatment of BD.

Summary

Investigating pleiotropy (one gene affecting multiple traits) and polygenicity (multiple genes contributing to a trait) in bipolar disorder is vital for understanding its genetic basis, which can enhance diagnosis, treatment, and prevention strategies, in addition to help elucidate BD genetic architecture. Bipolar disorder is complex and heritable, but specific involved genes are not fully known. GWAS have identified several BD-associated genetic variants, yet these explain only a portion of its heritability ("missing heritability") while Subphenotyping may help in identifying some of the "hidden heritability". Examining pleiotropy and polygenicity can help uncover more missing heritability by identifying non-European ancestry variants and shared genetic influences. Pleiotropy and polygenicity findings thus far suggest substantial genetic overlaps between bipolar disorder and other psychiatric and human disease traits. Studying these overlaps can reveal biological mechanisms and potential treatment targets. The primary objective of this thesis is precisely this: to explore these genetic overlaps to better account for patient heterogeneity and understand the shared biological mechanisms in bipolar disorder, to potentially identify functional genomics targets which could lead to the development of new therapeutic targets.

AIMS OF THESIS

Overarching Aim: The overarching aim of this thesis is to deconstruct the clinical and genetic heterogeneity of bipolar disorder (BD) to provide a more biologically grounded understanding of the illness. An enhanced understanding of the mechanisms underlying BD subgroups is essential for predicting illness course, improving treatment response, and developing personalized therapies.

To achieve this, a series of integrated specific aims were established:

1. To critically synthesize the existing literature on BD's nosology, risk factors, and the limitations of current research paradigms, thereby establishing the foundation for this investigation (Chapter 1).
2. To develop and validate a novel dimensional framework for BD that incorporates premorbid factors, moving beyond traditional diagnostic categories to identify the 'Adverse Chronic Trajectory' (ACT) and other genetically informative subgroups (Chapter 3).
3. To assess the transdiagnostic utility of schizophrenia polygenic risk for predicting severe outcomes, such as psychosis and age of onset, in high-risk BD1 patients and to identify associated biological pathways (Chapter 4).
4. To perform a large-scale, multi-trait analysis across 11 clinical subphenotypes to delineate their shared and distinct genetic foundations and identify novel biologically-based dimensions, such as 'Severe Illness' and 'Comorbidity' dimensions (Chapter 5).
5. To rigorously evaluate the methodological factors critical for genetic discovery, specifically by examining the impact of patient ascertainment strategies and genetic ancestry on the accuracy of polygenic prediction (Chapter 6).
6. To synthesize these empirical findings into a more nuanced model of BD's genetic architecture and to outline limitations and key future directions for research and clinical practice (Chapter 7).

2 General Methods

2.1 Study Population and Phenotypic Data

Cohort Ascertainment and Characteristics

The research presented in this thesis utilized data from multiple large-scale, international collaborations, encompassing a wide range of participant cohorts with diverse ascertainment strategies. Chapter 3 analyses included 2590 BD cases at the University College London (UCL) recruited via the National Health Service (NHS), United Kingdom (UK). The analyses in Chapter 4 utilized a combined European cohort of 1878 BD1 cases and 2751 controls from Romania (RO) and the UK (UCL). All participants were of European ancestry and provided written informed consent under ethically approved protocols. A detailed breakdown of the clinical characteristics and the traits for these samples is provided in Tables 2-6 and 17-18 below.

- Romanian (RO) Cohort: Unrelated BD1 patients (*N*=574) were recruited from the Obregia Psychiatric Hospital in Bucharest. Genealogical information was collected to ensure a homogeneous genetic sample. Diagnosis was confirmed using the Diagnostic Interview for Genetic Studies (DIGS) [1] based on DSM-IV [2] criteria, supplemented by medical records and information from relatives. Population-based controls (*N*=534) were screened with the DIGS to exclude major psychiatric history.
- United Kingdom (UK) Cohort: The UK sample included 1304 BD1 subjects who fulfilled Research Diagnostic Criteria for BD1. Clinical data were collected using the Schizophrenia and Affective Disorder Schedule-Lifetime (SADS-L) [3] and the OPCRIT [4] checklist. The UK controls (*N*=2217) were primarily population-based and screened with the SADS-L. Several of the UK controls consisted of random blood donors who were not screened for psychiatric disorders.
- As the mean and median Age of Onset (AO) differed significantly between the Romanian and UK samples, the AO data was normalized for the combined analysis using a rank-based inverse-normal transformation.

The total dataset in Chapter 6 for the multi-ancestry meta-analysis included up to 158,036 cases with bipolar disorder and 2,796,499 controls from 79 distinct cohorts. The effective sample size (Neff) was 535,720, with participants primarily of European (EUR) ancestry (82.3%), followed by Latino (LAT) (9.1%), African (AFR) (4.4%), and East Asian (EAS) (4.2%) ancestry. A subset of these 79 cohorts (all EUR) with available subphenotype data were included in Chapter 5 analyses.

The cohorts were broadly categorized into three ascertainment types:

- Clinical Cohorts: Participants were assessed using semi-structured or structured clinical interviews, such as the Diagnostic Interview for Genetic Studies (DIGS) or the Schizophrenia and Affective Disorder Schedule-Lifetime (SADS-L).
- Community Cohorts: Participants were assessed using data from medical records, national registries, and detailed questionnaires.
- Self-Reported Cohorts: Participants were classified as cases if they self-reported having received a clinical diagnosis or treatment for bipolar disorder in response to web-based surveys.

Specific analyses within this thesis drew upon different subsets of these larger cohorts. In Chapter 6, individual-level genotype and phenotype data were available for 53 'internal' cohorts, with the remaining 26 'external' cohorts contributing summary statistics. The large-scale multi-trait analysis of eleven clinical subphenotypes presented in Chapter **5** drew upon a sample of up to 23,819 BD cases and 163,839 controls from 56 of the 79 distinct cohorts.

2.1.1 Specific Cohort Characteristics by Analysis

Sample for Dimensionality Analysis (Chapter 3)

The dimensional analysis detailed in Chapter 3 utilized a sample of 2590 individuals with a DSM-IV bipolar disorder diagnosis and 2402 healthy controls. The characteristics of the cases, stratified by subtype, are shown in Table 2.

Table 2 Participant characteristics stratified by bipolar disorder subtypes

Characteristic	Overall	SZA	BD1	BD2	BD-NOS	*P*-Value
	***N*=2590**[a]	***N*=332**	***N*=1475**	***N*=387**	***N*=204**	
Psychosis						$1.70x10^{-77}$
No	38% (983/2590)	14% (47/332)	36% (525/1475)	79% (307/387)	51% (104/204)	
Yes	66% (1711[b]/2590)	86% (285/332)	64% (950/1475)	21% (80/387)	49% (100/204)	
Rapid cycling						$6.80x10^{-6}$
No	58% (1506/2590)	76% (252/332)	69% (1024/1475)	59% (230/387)	[N/A]	
Yes	27% (688/2590)	24% (80/332)	31% (451/1475)	41% (157/387)	[N/A]	
BD Age Onset	28 (11)	25 (9)	28 (11)	28 (11)	[N/A]	$2.10x10^{-4}$
Sex						$2.60x10^{-1}$
Female	61.3% (1588/2590)	59% (197/332)	62% (913/1475)	58% (223/387)	N/A]	
Male	38.7% (1002/2590)	41% (135/332)	38% (562/1475)	42% (164/387)	N/A]	
Age interviewed	46 (12)	49 (13)	46 (12)	51 (14)	[N/A]	$3.10x10^{-7}$

Abbreviations: SZA, schizoaffective disorder; BD1, bipolar disorder I; BD2, bipolar disorder II; BD-NOS, bipolar disorder not otherwise specified, N/A, data unavailable. Kruskal-Wallis rank sum test, Pearson's Chi-squared test. [a] Subtype information was missing for 192 participants. [b] Occurrence of psychosis information was missing for 296 participants. *N (%); Median (IQR).*

Sample for SCZ-PRS Analysis in BD1 (Chapter 4)

The investigation into schizophrenia-derived polygenic risk in Chapter 4 focused on a well-characterized European cohort of BD1 cases and controls from Romania (RO) and the United Kingdom (UK). The clinical characteristics of the BD1 cases are compared across the two recruitment sites in Table 3.

Table 3 Comparison of clinical traits in BD1 cases across samples

Variable	Overall	RO	UK
	***N*=1878**	***N*=574**	***N*=1304**
Sex (Male)	38% (718/1878)	38% (216/574)	38% (502/1304)
Age-at-interview (Mean (SD))	47 (13)	42 (13)	49 (13)
Age-of-onset BD1 (Mean (SD))	25 (10)	27 (10)	25 (10)
Psychosis (Yes)	70% (1331/1878)	84% (482/574)	65% (849/1,304)
Rapid cycling (Yes)	16% (309/1878)	10% (55/574)	19% (254/1304)
Irritable mania (Yes)	19.5% (366/1878)	59% (341/574)	2% (25/1304)
Family history psychoses (Yes)	27% (499/1878)	60% (343/574)	12% (156/1304)
Note: Missing data exists for some variables in the UK sample.			

Sample for Multi-Trait Subphenotype Analysis (Chapter 5)

The large-scale multi-trait analysis of 11 clinical subphenotypes in Chapter 5 drew upon a EUR-only sample from 56 international cohorts. Clinical characteristics are stratified by BD subtype and by key homogeneous subphenotype groups in Tables 4-5.

Table 4 Clinical Characteristics Stratified by BD Subtype

Characteristic	SZA	BD1	BD2	NOS
	***N*=1449**	***N*=11553**	***N*=2401**	***N*=405**
Psychosis	593 (96%)	6473 (68%)	476 (25%)	86 (56%)
Rapid Cycling	58 (39%)	1505 (29%)	586 (45%)	27 (31%)
Suicide attempt	139 (49%)	2852 (41%)	464 (39%)	20 (57%)
AlcSUD	122 (35%)	2339 (27%)	449 (26%)	38 (25%)
Age onset BD	21 (17, 27)	22 (16, 29)	22 (16, 31)	23 (18, 33)
N (%); Median (IQR). Sample sizes for specific characteristics may be smaller than the total cohort size (23,819 BD cases) due to missing data.				

Table 5 Clinical Characteristics Stratified by Homogenous Groups

Characteristic	Psychosis (No)	Psychosis (Yes)	Rapid Cycling (No)	Rapid Cycling (Yes)
	***N*=5186**	***N*=8476**	***N*=5617**	***N*=2373**
Suicide attempt	1292 (25%)	2218 (26%)	1533 (27%)	703 (30%)
AlcSUD	1105 (21%)	1892 (22%)	1139 (20%)	635 (27%)
Subtype BD1	2995 (58%)	6473 (76%)	3741 (67%)	1505 (63%)
Subtype BD2	1445 (28%)	476 (5.6%)	704 (13%)	586 (25%)
Age onset BD	23 (16, 31)	22 (17, 29)	24 (19, 32)	20 (15, 29)
N (%); Median (IQR). Sample sizes for specific characteristics may be smaller than the total cohort size (23, 819 BD cases) due to missing data.				

Sample for PRS Optimization Analysis (Chapter 6)

The study on optimizing PRS prediction in Chapter 6 involved the largest multi-ancestry meta-analysis from the PGC, utilizing data from mostly EUR, however additionally also AFR, and EAS ancestry cohorts. The characteristics of the target cohorts used for PRS testing are summarized in Table 6.

Table 6 Target Cohorts for PRS Optimization Analysis for Chapter 6

Ancestry Group	Number of Cohorts	Total Cases	Total Controls
European (EUR)	55	40,992	80,215
African (AFR)	1	347	669
East Asian (EAS)	3	4473	65,923
Note: These numbers represent the target cohorts in which PRS performance was evaluated			

2.2 Phenotypic Assessment and Diagnosis

Diagnosis and phenotypic characterization across the participating cohorts were established using internationally recognized criteria and comprehensive assessment tools. Diagnoses were made according to various versions of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III, DSM-IV, DSM-IV-TR, DSM-5) [5-7] and the International Classification of Diseases (ICD-9, ICD-10) [8-9]. In many cohorts, a consensus best-estimate diagnostic procedure was employed, integrating all available clinical information to ensure diagnostic accuracy.

To gather detailed clinical and symptomatic information, researchers utilized a range of semi-structured and structured interviews. The most frequently used instruments across the cohorts included the Schedule for Affective Disorders and Schizophrenia-Lifetime Version (SADS-L), the Diagnostic Interview for Genetic Studies (DIGS), and the Structured Clinical Interview for DSM (SCID) [10].

A key instrument for the detailed documentation of symptoms, premorbid functioning, and longitudinal course was the 90-item Operational Criteria checklist for psychotic illness (OPCRIT). The OPCRIT was used to systematically assess a wide array of psychopathological features, providing the foundational data for the dimensional analyses presented in Chapter 3. Inter-rater reliability for the OPCRIT assessments was formally assessed and found to be high (mean κ Statistic = .85).

Detailed Cohort Descriptions

Detailed descriptions of the specific diagnostic criteria and assessment procedures for each of the 79 cohorts contributing to the analyses in Chapters 3-6 are provided in the Appendix, 9.4.

Subphenotype Definitions

Across the various analyses in this thesis, the following clinical subtypes and subphenotypes were defined and investigated to deconstruct the heterogeneity of bipolar disorder. These were selected based on their clinical relevance and evidence for clustering within families, suggesting more genetically homogeneous subgroups. The definitions are as follows:

- Bipolar Disorder I (BD1): Characterized by the occurrence of at least one lifetime manic episode.
- Bipolar Disorder II (BD2): Characterized by at least one hypomanic episode and one major depressive episode, with no history of manic episodes.
- Bipolar Disorder Not Otherwise Specified (BD-NOS): A category for individuals who do not meet the full criteria for BD1 or BD2 but exhibit clear bipolar features, often identified by multiple depressive episodes.
- Schizoaffective Disorder, Bipolar Type (SZA): A diagnosis that includes symptoms of both a major mood episode (manic or major depressive) and the active-phase symptoms of schizophrenia, with at least a two-week period of delusions or hallucinations in the absence of a major mood episode.
- Psychotic Features: The presence of hallucinations or delusions during a manic or major depressive episode.
- Rapid Cycling (RC): The occurrence of four or more distinct mood episodes (manic, hypomanic, or depressive) within a 12-month period.
- Unipolar Mania (UM): Characterized by recurrent manic episodes without any history of major depressive episodes.
- Alcohol or Substance Use Disorder (AlcSUD): A comorbid diagnosis of an alcohol or substance abuse or dependence disorder.
- Obsessive-Compulsive Disorder (OCD): A comorbid diagnosis of OCD characterized by obsessions and/or compulsions.
- Panic Disorder (PD): A comorbid diagnosis of panic disorder characterized by recurrent unexpected panic attacks.
- Suicide Attempt (SA): A lifetime history of one or more suicide attempts.
- Suicidal Ideation (SI): A lifetime history of thoughts of harming oneself.
- Age at Onset (AOO): The age at which the individual first met criteria for any primary mood episode (manic, mixed, or major depressive).
- Age of onset of depression (AO_depr): The age at which the individual first met criteria for a major depressive episode.
- Age of onset of mania or mixed episodes (AO_Man/Mix): The age at which the individual first met criteria for a manic or mixed episode.

Note: The main source of these descriptions is the DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition) or its revision, the DSM-IV-TR.

2.3 Genotyping, Imputation, and Quality Control

Genetic data for the analyses in this thesis were processed through rigorous, state-of-the-art pipelines to ensure high quality and accuracy. While the core principles of quality control (QC), phasing, and imputation were consistent across all studies, specific parameters and reference panels were tailored to the requirements of each analysis, from the focused European cohort studies to the large-scale multi-ancestry meta-analyses of the Psychiatric Genomics Consortium (PGC). This section outlines the general procedures, with specific details for each major analysis presented in Table 9.

Genotyping Platforms

A variety of high-density single nucleotide polymorphism (SNP) genotyping arrays were used across the 79 contributing cohorts, reflecting the collaborative and multi-stage nature of the research. The most common platforms included the Illumina Global Screening Array (GSA), Illumina PsychArray, Illumina Omni Express, and the Affymetrix Gene Chip 500k Assay. The specific platform breakdowns for the focused analyses in Chapters 3 and 4 are detailed below (Tables 7-9).

For the SCZ-PRS study (Chapter 4), specific post-imputation filtering steps were applied to handle correlated SNPs, with different approaches tailored to the analyses for each cohort. In the Romanian sample, this was followed by Linkage Disequilibrium (LD) clumping (parameters: 500 SNP window, 100 SNP overlap, r^2 threshold of .05). This procedure creates a set of approximately independent genetic variants, which is a necessary prerequisite for methods such as the traditional 'clumping and thresholding' polygenic scoring, as it prevents the same genetic signal from being counted multiple times and inflating the score.

In the UK sample, a locus-definition approach was used, where the most significant SNPs were retained within a physical distance of less than 250 kb and an r^2 greater than .1. This is the standard method for identifying distinct, independent genetic loci from a GWAS. It ensures that multiple correlated SNPs in the same genomic region, which are likely tagging the same underlying causal variant, are correctly treated as a single genetic signal rather than multiple independent discoveries.

For the dimensionality analysis (Chapter 3), the sample was genotyped across three primary platforms, with the following distribution of participants and post-QC imputed SNPs:

Table 7 Genotyping Array Frequencies for Chapter 3

Sample	Affymetrix Gene Chip 500k	Illumina Global Screening Array	Illumina PsychChip	*N*
Post-QC'd Sample size	491 cases, 495 controls	416 cases, 533 controls	1683 cases, 1374 controls	4992
Percent %	.197	.190	.613	
Post-QC'd Imputed SNPs	3,080,075	3,164,648	3,443,778	*Mdn*=3,164,648

For the SCZ-PRS analysis (Chapter 4), the Romanian and UK cohorts were genotyped on a different combination of platforms:

Table 8 Genotyping Array Frequencies for Chapter 4

Sample	Affymetrix Gene Chip 500k	Illumina Global Screening Array	Illumina Omni Express	Illumina PsychChip	*N*
Romanian (RO)	0	309	799	0	1108
UK	840	533	0	2148	3521
Percent %	18.15	18.19	17.26	46.40	
Post-QC Imputed SNPs					
RO		3,930,194	3,917,108		
UK	3,080,075	3,164,648		3,443,778	

For the large-scale multi-trait and multi-ancestry meta-analyses presented in Chapters 5 and 6, a wider array of genotyping platforms was used across the numerous contributing cohorts. A detailed breakdown of the specific platforms used for each of the 56 cohorts in the subphenotype analysis and the 79 cohorts in the PRS optimization analysis is provided in the Appendix, 9.4.

GWAS Quality Control (QC)

All datasets underwent stringent QC procedures aligned with PGC standards to remove low-quality variants and samples before imputation. This involved standardized thresholds for both variant-level and sample-level metrics. See Table 9 below.

Variant-level QC typically excluded SNPs with low call rates, significant deviation from Hardy-Weinberg Equilibrium (HWE), and low minor allele frequency (MAF). Sample-level QC removed individuals with low call rates, excessive heterozygosity (FHET), sex discrepancies, and cryptic relatedness to other individuals in the sample.

Genotype Imputation

To standardize genotypes across different array platforms and increase genomic coverage, all datasets were imputed to a common reference panel. For most analyses, the Haplotype Reference Consortium (HRC) [11] panel was used, which provides a high-quality reference for individuals of European ancestry. The standard procedure involved a pre-phasing step using EAGLE2 [12] followed by imputation using Minimac3 [13]. For the Romanian cohort analysis in Chapter 4, the 1000 Genomes Project panel [14] was used to suit the specific sample characteristics.

Post-Imputation Filtering

Following imputation, variants were filtered based on imputation quality scores (INFO or R2) to ensure that only accurately imputed SNPs were included in the downstream association analyses. A common threshold was an INFO score > .8, although more stringent filters were applied in some analyses to maximize confidence in the results.

Table 9 Genotyping and QC Parameters for Analyses

Parameter	Dimensionality Study (Chapter 3)	SCZ-PRS Study (Chapter 4)	Multi-Trait Subphenotype Study (Chapter 5)	PGC4 PRS Optimization Study (Chapter 6)
QC Pipeline/Standard	Standard PGC Protocols	Standard PGC & local protocols	Standard PGC Protocols	Standard PGC Protocols
Sample-Level QC				
Subject Missingness	< 2%	< 2% (UK) / < 5% (RO)	< 2%	< 2%
Heterozygosity (FHET)	Outside +/- 0.20	> 1 SD from mean (RO) / |Fhet| < .2 (UK)	Outside +/- 0.20	Outside +/- 0.20
Relatedness (pi_hat)	> 0.2	> 0.2	Not specified in main text	> 0.2
Sex mismatch	Mismatches between pedigree and genetically determined sex were removed based on the *F* statistic of X chromosome homozygosity (female $F < 0.2$ and male $F > 0.8$).			
Variant-Level QC				
SNP Missingness	< 5%	< 5%	< 5%	< 5%*
Minor Allele Freq. (MAF)	> 1%	> 0.1% (RO) / > 1% (UK)	> 1%	> 1%
HWE *P*-value (Controls)	$< 1 \times 10^{-6}$	$< 1 \times 10^{-6}$	$< 1 \times 10^{-6}$	$< 1 \times 10^{-6}$
HWE *P*-value (Cases)	$< 1 \times 10^{-10}$	$< 1 \times 10^{-10}$	$< 1 \times 10^{-10}$	$< 1 \times 10^{-10}$
Imputation				
Imputation Reference	HRC Panel	1000 Genomes (RO) / HRC (UK)	HRC panel (r1.1 2016)	HRC panel (v1.0)
Phasing / Imputation Software	Eagle / Minimac3	Eagle / Minimac3	Eagle / Minimac3	Eagle / Minimac3
Post-Imputation Filter	INFO score > 0.8 (standard)	Rsq > 0.8 (RO) / INFO > 0.9 (UK)	INFO score > 0.8	Filtered SNPs in < 75% of total Neff
No. of SNPs in Primary GWAS (Million)	3.16	3.08-3.93	4.57-7.40	3.97-9.74
Note: "Standard PGC Protocols" refers to the established quality control and analysis standards developed by the Psychiatric Genomics Consortium, which are implemented in the pipeline RICOPILI. * < 0.05 (before sample removal) and < 0.02 (after sample removal). Case-Control Missingness Difference: < 0.02.				

2.4 Statistical and Genetic Analysis

This section outlines the statistical and genetic methods used across the chapters. Primary Genome-Wide Association Studies (GWAS) were conducted for the analyses in Chapters 5 and 6, while existing external GWAS summary statistics were utilized for polygenic scoring in Chapters 3 and 4.

2.4.1 Sample Size and Prevalence Parameters

Population and Sample Prevalences

The transformation of SNP-based heritability (h^2snp) and Polygenic Risk Score (PRS) variance explained (R^2) from the observed 0-1 scale to the unobserved continuous liability scale is a critical step for interpreting results for binary traits. This transformation requires specifying an estimate of the trait's prevalence in the general population. The specific prevalences used varied across the thesis analyses to match the context of each study:

- For the dimensional analysis which included all BD subtypes in Chapter 3, a population prevalence of 2% was used.
- For the SCZ-PRS study in Chapter 4, a Bipolar Disorder I (BD1) population prevalence of 1% was assumed.
- For the PGC-BD PRS optimization study in Chapter 6 (and in Chapter 5), both 1% and 2% population prevalences were used for comparison.

For the additional multi-trait subphenotype analyses in Chapter 5, population prevalences were estimated from major epidemiological studies. For course specifiers that do not have direct population estimates, the prevalence was calculated by multiplying the general prevalence of bipolar disorder by the proportion of individuals with BD who exhibit that feature. For example, the prevalence for the psychosis subphenotype was estimated by multiplying the ~1% lifetime prevalence of Bipolar Disorder (e.g., Merikangas *et al*., 2007)[15] by the ~50% proportion of BD individuals who experience psychosis (e.g., Perälä *et al*., 2007)[16], resulting in an estimated population prevalence of .5%. For comorbid disorders, the direct lifetime prevalence was taken from the literature. The specific values and primary sources used are detailed in Table 10.

For all analyses, population prevalence estimates were based on a review of the relevant scientific literature. The choice of population prevalence was intentionally tailored to the specific scientific goal of each analysis, following a 'fit-for-purpose' strategy. For broad assessments of a general Bipolar Disorder PRS across diverse cohorts (Chapters 3, 4 and 6), standard 1-2% prevalences were used to ensure the results were comparable with the wider field and major genomic consortia. In contrast, for more granular genetic architecture analyses of specific subphenotypes (Chapter 5), the most precise, subtype-specific epidemiological estimates were used. This approach maximizes the accuracy of the heritability calculations for

those specific traits by ensuring the assumptions of each analysis were best aligned with its scientific question.

Table 10 Population Prevalences Literature Sources

Subphenotype	Population Prevalence Used (%)	Primary Source(s)
Psychosis	.005	Merikangas *et al*. (2007)[15]; Perälä *et al*. (2007)[16]
Rapid Cycling	.0025	Merikangas *et al*. (2007); Tondo *et al*. (2003) [17]
BD1	.006	Merikangas *et al*. (2007)
BD2	.004	Merikangas *et al*. (2007)
SZA	.003	Perälä *et al*. (2007)
Panic Disorder	.027	Kessler *et al*. (2005) [18]
OCD	.023	Ruscio *et al*. (2010) [19]
AlcSUD	.139	Grant *et al*. (2015) [20]
Suicide Attempt	.042	Nock *et al*. (2008) [21]
Unipolar Mania	.060	Angst & Marneros (2001) [22]

A key exception was for the LD Score Regression (LDSC) analyses; where the effective sample size was provided as input, the sample prevalence was accordingly set to .5 as per standard methodological practice.

Use of Total (N) vs. Effective (Neff) Sample Size

Both total sample size (*N*; the actual count of cases and controls) and effective sample size (Neff; calculated to account for case-control imbalance) were utilized for distinct purposes throughout the analyses.

- The effective sample size (Neff) was used when the statistical power of a case-control sample was the most relevant metric. Its applications included:
    - Quality control in the large-scale meta-analyses (Chapters 5 and 6), where SNPs had to be present in a minimum percentage of the total Neff to be included.
    - Weighting results in the meta-analyses of Polygenic Risk Score (PRS) performance across cohorts (Chapters 3-6).
    - As the sample size input for all major post-GWAS summary-statistic-based analyses, including LDSC [23-24], MTAG [25], Multi-marker Analysis of GenoMic Annotation (MAGMA) [26] (within Functional Mapping and Annotation of GWAS [FUMA] [27]), Transcriptome-Wide Association Studies (TWAS) [28], Local Analysis of [Co]variant Association (LAVA) [29], and Summary-data-based BayesS (SBayesS) [30].
- The total sample size (N) was required for statistical models that explicitly use the number of cases and controls as parameters. Its primary use was:
    - As input for the liability scale conversion of PRS R2 (e.g., using the Lee *et al*., 2012 formula) [31], which mathematically requires the number of cases and controls in the sample.

2.5 Primary GWAS Association Analyses

The foundational step in identifying genetic variants associated with bipolar disorder and its subphenotypes was to conduct a Genome-Wide Association Study (GWAS) on each cohort, followed by a meta-analysis to increase statistical power by combining the results.

- GWAS and Meta-Analysis: To form the basis of analyses in Chapters 5 and 6, GWAS were run in each cohort using an additive logistic regression model in PLINK(v.190) [32], with the first five principal components as covariates. For these primary meta-analyses, a genome-wide significant locus was defined as the region around a lead SNP ($P < 5.0 \times 10^{-8}$) including all variants in Linkage Disequilibrium (LD) at $r^2 > .1$ within a 3,000-kb window, based on the ancestry-matched HRC reference panel. The DENTIST [33] tool was used for quality control to detect and filter problematic variants. Cohort-level summary statistics were then combined using an inverse-variance-weighted fixed-effect model in METAL [34]. To ensure robustness, all SNPs present in less than 75% of the total effective sample size were removed from the meta-analyses. For the subphenotype GWAS (Chapter 5), Linkage Disequilibrium Score Regression (LDSC) confirmed that confounding from population stratification was minimal, with a median intercept of 1.015. The attenuation ratio, an estimate of the proportion of the GWAS signal due to confounding, had a median of .183, which is in line with values reported for similar large-scale psychiatric analyses. (See Watanabe *et al*. [53] and Chapter 5 [43] for comparison).
- Multi-Trait Analysis of GWAS (MTAG): As detailed in Chapter 5, MTAG was used to boost statistical power by integrating the primary subphenotype GWAS with large external GWAS for Bipolar Disorder (BD) and Schizophrenia (SCZ). This was performed only for a subgroup of subphenotypes showing a strong median initial genetic correlation (*rG*> .70) with the external study. The reliability of these analyses was confirmed by low median maximum False Discovery Rate (maxFDR) [25] values (BD-only: < .00014; BD+SCZ: < .00013).

The MTAG method was chosen specifically to enhance statistical power for subphenotype analyses. Its application was contingent on a strong initial genetic between the primary subphenotype GWAS and the external study, ensuring a valid basis for integration. The reliability of the MTAG results was confirmed by low median maximum False Discovery Rate (maxFDR) values, suggesting a reliable synthesis of signals rather than distortion from the larger external GWAS. Phenotypes that exhibited a higher maxFDR, including suicide ideation and the age of onset variables, were excluded from downstream MTAG analyses to ensure the robustness of the findings.

2.6 Polygenic Risk Scoring (PRS)

To move beyond single-variant associations and capture the cumulative genetic risk for a disorder, polygenic risk scores (PRS) were constructed. This approach aggregates the small effects of thousands of genetic variants into a single, quantitative score representing an individual's genetic liability. For the PRS in Chapters 3, 4, 5, and 6, a leave-one-cohort-out approach was used. This method ensures that the PRS for a given target cohort is not biased by its inclusion in the discovery GWAS by creating a unique set of summary statistics for each target cohort that excludes its own data.

2.6.1 PRS Construction, Performance and Evaluation

For analyses across Chapters 3, 4, 5, and 6, PRS were constructed from large, external GWAS summary statistics using two primary methods:

- Clumping and Thresholding (pT+clump): Implemented in PRSice (v.2.3.3) [35], this method only used in Chapter 4, selected approximately independent SNPs based on a clumping threshold of R2< .1 within a 250 kb window, retaining SNPs that passed specific *P*-value thresholds (pT). For the analysis in Chapter 4, this was implemented in PRSice v2.3.3 using its default settings, and scores were generated for eight *P*-value thresholds ranging from 5×10^{-8} to .05.
- PRS-CS-auto: For the analysis in Chapter 4 to 6, the auto-setting was used to learn the global shrinkage parameter. The PRS-CS-auto [36] setting uses a fully Bayesian approach to learn the optimal global shrinkage parameter directly from the discovery GWAS data, avoiding the need for a separate validation set. Raw scores were generated using the PLINK v2.0 score function from the posterior SNP effect means. Power analysis for the PRS was conducted using the AVENGME package in R [37-38] and G*Power 3 [39] was used for calculating the sample size and power for the statistical tests (*F, t*, $\chi 2$, *Z*). The following parameter were used for the power calculation; number of independent SNPs produced by the PRS-CS package; sample size training sample (sample size of the schizophrenia GWAS used); heritability and prevalence for bipolar disorder were obtained from Table 1 in Wray *et al*., 2010 [40] and the proportion of null markers were set as 90%.
- Chapter 5 evaluated the predictive performance of PRS for BD subphenotypes. The core methodology involved developing subphenotype-specific PRS using MTAG. These MTAG-derived effect sizes were then used for PRS construction in target cohorts via PRS-CS-auto, employing a leave-one-cohort-out approach. Within each target cohort, PRS were standardized and their association with phenotype status was assessed using logistic regression, adjusted for the first five PCs. Nagelkerke's R^2 was converted to R2 on the liability scale (R2-liability) using the method by Lee *et al*. (2012). A formal random-effects (RE) meta-analysis of the per-cohort R^2-liability values was additionally conducted to model between-cohort heterogeneity.
- LD Reference Panels: The selection of the Linkage Disequilibrium (LD) reference panel was tailored to the specific analysis and ancestry of the samples. For the PRS-

CS analyses in the dimensionality study (Chapter 3) and the SCZ-PRS study (Chapter 4), the 1000 Genomes Project European LD reference panel was used. For the primary GWAS meta-analysis and subsequent post-GWAS analyses (including FUMA, SBayesS, and LAVA) in the multi-trait subphenotype study (Chapter 5), the ancestry-matched Haplotype Reference Consortium (HRC) panel was used. For the TWAS analysis in this chapter, the European 1000 Genomes Project LD panel was used. For the main PRS-CS analyses in the PRS optimization study (Chapter 6), the UK Biobank European LD reference panel, as provided by the PRS-CS developers, was used.

PRS Performance Evaluation

Raw scores were standardized to z-scores (mean=0, SD=1) to make them comparable across individuals. The predictive power of the PRS was assessed in linear and logistic regression models using Nagelkerke's R2 (converted to liability scale R2) and the Area Under the Curve (AUC). To summarize performance across cohorts in Chapters 5 and 6, per-cohort R2 estimates on the liability scale were pooled via a random-effects meta-analysis, and heterogeneity was assessed with the I2 and Cochran's Q statistics. The variance explained by PRS was calculated as Nagelkerke's pseudo-R^2 using the fmsb [41] package in R, while the Area Under the Curve (AUC) was calculated using the pROC [42] package.

2.6.2 PRS Performance Evaluation and Metrics

The performance of the Polygenic Risk Scores (PRS) was assessed using a comprehensive suite of metrics. The specific metrics reported were tailored to the primary aims of each analysis.

The PRS-CS-auto method was selected for most analyses because it demonstrated superior predictive accuracy compared to the traditional pT + clump method in Chapter 4, explaining nearly 2% more variance on the liability scale in comparative tests (Chapter 4, Table 26).

Metrics for the Multi-Trait Subphenotype Study (Chapter 5)

Individual-level Polygenic Risk Scores (PRS) were constructed for participants in European target cohorts using effect sizes from discovery meta-analyses which combined subphenotype-specific GWAS with data from bipolar disorder (BD) cases lacking subphenotype information while systematically excluding each target cohort from its respective discovery dataset. These PRS were adjusted for population stratification. As detailed in the external xlsx Supplementary Table 58, the reported metrics included:

- Cohort: An identifier for the specific study or dataset.
- Sample.Size_N: The total number of individuals (cases plus controls) in the analysed sample for that cohort.
- N_eff_half: Half of the effective sample size, calculated as 2×Ncases×Ncontrols /(Ncases+Ncontrols), accounting for case-control imbalances.

- Proportion_Cases_P: The proportion of cases within the cohort's analysed sample.
- Cases_NCA and Controls_NCO: The number of cases and number of controls, respectively.
- NagelkerkeR2_obs: Nagelkerke's pseudo R2 value, measuring variance explained by the logistic regression model on the observed scale.
- LiabilityR2_adj: The R^2 value on the liability scale, estimating the proportion of variance in underlying disease liability explained by the PRS, adjusted for population prevalence (K) and sample case proportion (P).
- PerCohort_Weighted_LiabR2_pct: The cohort's R2-Liability multiplied by its relative effective sample size, expressed as a percentage, indicating its weighted contribution to an overall average.
- PRS_PVal_adj_wCovars: The *P*-value for the overall statistical significance of the PRS model including covariates.
- Coef_PRS: The regression coefficient (beta) for the standardized PRS from the logistic regression, representing the log-odds change per standard deviation increase in PRS.
- Coef_SE_PRS: The standard error of the PRS coefficient.
- CoefL_PRS and CoefH_PRS: The lower and upper bounds of the 95% confidence interval for the PRS coefficient.
- Z_value_PRS: The Z-statistic for the PRS coefficient.
- AUC (Area Under the ROC Curve): Measures the PRS model's ability to discriminate between cases and controls.
- AUC_Low and AUC_High: The 95% confidence interval for the AUC.
- Absolute Risk:
  Estimated absolute risks are given for the AbsRisk_Quintile_Top (top PRS quintile),
  AbsRisk_Quintile_Bottom (bottom PRS quintile),
  AbsRisk_Top1pct (top 1% of PRS distribution), and
  AbsRisk_Top10pct (top 10% of PRS distribution).

Metrics for the PRS Optimization Study (Chapter 6)

For the study focused on comparing different ascertainment strategies and ancestries, reporting was focused on the primary metrics of predictive power. The key metrics included:

- Variance Explained (Liability Scale R2): This was the main outcome measure used to compare the performance of the different discovery GWASs.
- Risk Stratification (Odds Ratio): To assess clinical potential, the odds ratio (OR) was calculated for individuals in the top quintile (top 20%) of the PRS distribution compared to those in the middle quintile.
- Discriminative Ability (AUC): The AUC was reported to measure overall discriminative accuracy. To isolate the predictive value added by the PRS itself, the AUC gain was also calculated by subtracting the median AUC of a model containing only covariates from the median AUC of the full model.

In Chapter 5, a sensitivity analysis using the Slope-Hunter [43] method in R to adjust for potential index event bias was also explored but was not used for the final results as it was found to inflate the test statistics. As a sensitivity analysis, the Slope-Hunter method was explored to assess the potential impact of index event bias on the effect sizes from the primary, single-subphenotype GWAS analyses. However, this correction was not carried forward to the final Polygenic Risk Score (PRS) analyses for two key reasons. First, initial tests on the single-subphenotype GWAS indicated that, the Slope-Hunter adjustments were inflating the test statistics, suggesting that the underlying model assumptions of the tool were not a good fit for the data. Second, applying this correction to the downstream MTAG-derived summary statistics would be methodologically invalid, as MTAG results represent a complex mixture of effect sizes from cohorts with different ascertainment strategies. Therefore, a random-effects meta-analysis was chosen as the more appropriate and robust method to account for heterogeneity in the final PRS performance estimates.

External GWAS Summary Statistics for PRS Analyses

The discovery GWASs used as a basis for PRS construction were selected as they were the largest and most recent available at the time of analysis. They included:

Table 11 External GWAS Summary Statistics for Analyses

Disorder/Trait	Study	Sample Size (N)	Used in Chapter(s)
Psychiatric Disorders			
Bipolar disorder	O'Connell *et al*., 2025 [44]	840,309	3, 4, 5, 6
Schizophrenia (SCZ)	Trubetskoy *et al*., 2022 [45]	130,644	3, 4, 5
Major depressive disorder (MDD)	Howard *et al*., 2019 [46]	500,199	3, 5
Attention deficit and hyperactivity disorder (ADHD)	Demontis *et al*., 2023 [47]	225,534	3, 5
Anxiety (ANX)	Purves *et al*., 2020 [48]	114,091	3, 5
Autism spectrum disorder (ASD)	Grove *et al*., 2019 [49]	46,350	5
Mood swings (MOOD)	Neale Lab UKBB, 2018 [50]	604,063	5
Post traumatic stress disorder (PTSD)	Nievergelt *et al*., 2019 [51]	174,659	5
Borderline personality disorder (BPD)	Witt *et al*., 2017 [52]	2543	5
Insomnia (INS)	Watanabe *et al*., 2022 [53]	386,888	5
Cognitive Traits			
Intelligence (INTEL)	Savage *et al*., 2019 [54]	269,867	5
Matrix	de la Fuente *et al*., 2020 [55]	11,356	5
Memory	de la Fuente *et al*., 2020	331,679	5
Trail Making Test B (TMTB)	de la Fuente *et al*., 2020	78,547	5
Tower	de la Fuente *et al*., 2020	11,263	5
Symbol and digit (SymDig)	de la Fuente *et al*., 2020	87,741	5
VNR	de la Fuente *et al*., 2020	171,304	5
Reaction time (RT)	de la Fuente *et al*., 2020	330,024	5

Note: To align phenotypes, only GWAS summary statistics without 23andMe self-report data were included. Matrix = Matrix Pattern Completion task; Memory = Memory – Pairs Matching Test; RT = Reaction Time; Symbol Digit = Symbol Digit Substitution Task; Trails-B = Trail Making Test – B; Tower = Tower Rearranging Task; VNR = Verbal Numerical Reasoning Test. Phenotype data was scaled before analyses and higher scores aligned to indicate better cognitive performance. See Figure 26, a presentation of the global genetic correlations presented in Supplementary table 59. To ensure comparability, only versions of the discovery sets which excluded 23andMe self-reported data were included for analysis (Chapter 3-5). Chapter 6 modelled PRS for datasets with and without 23andMe data.

2.7 Covariate and Bias Control

The handling of covariates and bias differs across chapters due to the distinct goals and methodologies of each analysis. The choice of which covariates to control for and how to control for them is driven by the specific research question being addressed and the potential sources of bias inherent in that particular study design.

Covariate and Bias Control

- Population Stratification (Principal Component Analysis [PCA]): This method was applied consistently across all analyses because the entire study population shares the same fundamental potential confounder: genetic ancestry. PCA is a standard practice in genetic studies to control for population structure, which can create spurious associations between genetic markers and traits if not accounted for. By including the first five to ten principal components as covariates, the regression models were adjusted for this systematic bias, ensuring that any observed associations were not simply a result of shared ancestry.
- Covariate Adjustment (Residualisation): This approach was specifically used for the Polygenic Risk Score (PRS) analyses in Chapters 3 and 4. PRS is a score derived from genetic data to predict an individual's risk for a specific trait or disease. To ensure the PRS itself was the primary variable of interest and that its predictive power wasn't inflated by other factors, the scores were "residualised." This means that the effects of non-genetic factors, including age, sex, and genotyping batch, were statistically removed. The resulting residuals represent the portion of the PRS that is independent of these covariates, allowing for a cleaner and more accurate assessment of the PRS's direct association with the outcome.
- Ascertainment Bias (IPW): This method was exclusively applied to the dimensional analysis in Chapter 3 when modelling within-case severity. Ascertainment bias occurs when the method of selecting a study sample systematically favours certain individuals, potentially distorting the results. In this case, the analysis was performed on a case-control sample, which is inherently biased because individuals were selected based on their disease status. Inverse Probability Weighting (IPW) [56-57] was used to correct for this. By weighting the cases and controls based on their probability of being selected, the method effectively rebalances the sample to be more representative of the source population, thereby mitigating the bias introduced by the case-control sampling design.

2.7.1 Chapter 3: A Four-Dimensional Genetic Model of Bipolar Disorder

This chapter's primary goal was to investigate the dimensional structure of Bipolar Disorder (BD) using a Multiple Indicators and Multiple Causes (MIMIC) model [58]. The methods to control for bias were comprehensive. To mitigate confounders and ascertainment bias, including collider bias which can be an issue in case-only studies, this study adopted a case-

control design. To further address potential selection bias, stabilised weights were implemented for the Inverse Probability Weighting (IPW).

- Population Stratification: To control for confounding due to genetic ancestry, a Principal Component Analysis (PCA) was conducted using the EIGENSTRAT (v6.1.4) [59] software. The first ten ancestry-specific principal components were then included as covariates in all statistical models to adjust for population structure.
- Ascertainment Bias: Because the analysis used a case-control sample, Inverse Probability Weighting (IPW) using propensity scores was applied to mitigate potential selection bias and adjust for imbalances between groups. This procedure was implemented using the R statistical environment.
- Covariate Adjustment (Residualisation): For the PRS analyses, scores were fully adjusted before being used in the final models. The effects of covariates including age, sex, the first ten principal components, and genotyping batch/platform, were regressed out of the PRS. The resulting standardized residuals were used as the final PRS predictor, a method known as residualisation, which was performed in R.

The following table demonstrates the effect of the multi-step covariate correction applied to the BD PRS in the dimensionality study (Chapter 3). The 'pre-correction' result reflects the strong, unadjusted association between the PRS and case-control status. The 'post-correction' result illustrates a known statistical artifact that occurs when controlling for a variable that is a proxy for the outcome itself. Such a complete elimination of the signal is the expected outcome when adjusting for a factor like illness severity in both cases and controls, as this statistically removes the core difference between the groups. In this analysis, a multi-stage correction was applied where standard covariates (age, sex, PCs, array) were controlled for, followed by an IPW adjustment for severity (hospitalization [see Table 18, OCPRIT 01 'Source of rating']) in the cases only.

Table 12 Correction for Covariates for Chapter 3

BD PRS	Analysis of Variance (ANOVA)		
	F	*df*	*P*
Pre-correction	85.22	2, 4989	$P < .001$
Post-correction	.092	2, 4989	.912

2.7.2 Chapter 4: Schizophrenia-Derived Polygenic Risk

This chapter's analysis of a combined cohort from two different sites required specific corrections for batch and site effects in addition to standard covariate control.

- Population Stratification: To control for genetic ancestry differences, PCA was performed using the EIGENSTRAT (v6.1.4) software package for both the Romanian and UK samples. The first ten ancestry-specific principal components were included as covariates to control for population structure.
- Batch and Site Effects: A specific two-step correction was implemented to address non-genetic variance in the PRS calculations. First, batch effects due to different genotyping platforms within each cohort were regressed out. Second, site effects between the Romanian (RO) and United Kingdom (UK) cohorts were corrected to mitigate potential bias.
- Covariate Adjustment (Residualisation): The final PRS predictor was a standardized residual. The effects of age, sex, and the first 10 principal components were regressed out of the PRS scores prior to their use in regression and Random Forest (RF) models. All analyses were conducted in R.

Correction for Batch and Site Effects: In the analyses for Chapter 4, the Romanian (RO) and United Kingdom (UK) samples were genotyped on different platforms (Table 8). These between-platform and between-cohort differences can introduce batch effects into the PRS calculations. To address this, batch effects due to platform differences were first regressed out of the PRS for each cohort separately in Chapter 4. Subsequent corrections were then made for site effects between the two cohorts.

The following tables demonstrate the successful data harmonization process applied to the SCZ-PRS in the study described in Chapter 4. The 'pre-correction' results show that significant, systematic differences in the mean PRS existed due to technical factors, specifically the genotyping platform (batch effects) and recruitment site (site effects). The 'post-correction' results show that the two-step correction method successfully removed this non-biological variance, as indicated by the drop from highly significant to non-significant statistics. This essential harmonization step created a clean PRS variable, ensuring that the main downstream analyses were not biased by these technical confounders.

Table 13 PRS Batch Effects (Genotype Array) Correction for Chapter 4

Romanian	*T*-test	
	t	*df*
Pre-correction	13.578	1,1107
Post-correction	0.116	1,1107
UK	*ANOVA*	
	F	*df*
Pre-correction	85.22	2, 3518
Post-correction	0.092	2, 3518

Table 14 PRS Sample (Site) Correction for Chapter 4

RO/UK	*T*-test	
	t	*df*
Pre-correction	6.861	2,4627
Post-correction	0.180	2,4627

2.7.3 Chapter 5: Multi-Trait Analysis of Eleven Clinical BD Subphenotypes

This large-scale meta-analytic approach relied on including covariates directly within the statistical models rather than using pre-adjusted residualised scores; the same approach was adopted in Chapter 6.

- Population Stratification: Standard Genome-Wide Association Study (GWAS) procedures, which include PCA, were conducted using the RICOPILI [60] automated pipeline. The first five to ten principal components of ancestry were included as covariates in all primary GWAS and downstream regression models. The Linkage Disequilibrium Score Regression (LDSC) intercept was also monitored to confirm that confounding from uncorrected population stratification was minimal, i.e., close to 1.
- Ascertainment Cohort Heterogeneity: The analysis addressed between-cohort heterogeneity in a multi-step process. First, the DENTIST tool was used on the GWAS summary statistics to identify and remove problematic SNPs that showed significant heterogeneity across the different cohorts. In the subsequent Polygenic Risk Score (PRS) analysis, random-effects models were employed to directly measure and model the remaining heterogeneity in the prediction estimates. This provides a more robust and generalizable estimate that properly reflects the variability observed in the underlying data.

- Covariate Adjustment: Unlike the analyses in the preceding chapters, residualisation was not performed. Instead, PRS and other relevant covariates were included as independent variables directly in the regression models to assess their associations with the outcomes (as noted above).

2.7.4 Chapter 6: Optimising BD Polygenic Risk Prediction

This chapter's primary goal was to investigate the impact of ascertainment and ancestry, using stratification as the main analytical method.

- Population Stratification: PCA was conducted for each cohort using EIGENSTRAT (v6.1.4). The first five principal components were included as covariates in the logistic regression models.
- Covariate Adjustment: The GWAS for each cohort was conducted using PLINK (v1.90), which directly included principal components as covariates in the regression model. The final PRS performance was assessed using the `glm()` function in R, which also included sex and principal components as covariates alongside the standardized PRS. The PRS was not residualised beforehand.
- Ascertainment Bias: This was the central focus of the investigation rather than a factor to be corrected statistically. The analysis addressed ascertainment by stratifying cohorts based on their recruitment method (clinical, community, and self-report) and comparing PRS performance across these distinct groups.

2.8 Post-GWAS Functional and Genetic Architecture Analyses

Individual-level pathway analysis was applied in Chapter 4 using PRSet to explore the genetic architecture of specific clinical features. The analysis used PRSet in the PRSice package, which provides an individual-level representation of genetic burden within a gene-set, in contrast to population-level methods like Multi-marker Analysis of GenoMic Annotation (MAGMA). The analysis was applied to 1878 cases and 2751 controls in the combined RO/UK sample. After excluding 893 SNP regions not present in both the SCZ3-GWAS summary statistics and the target genotypes, a total of 31,937 gene regions from the Molecular Signatures Database (MsigDB) [61-62] database were included for a hypothesis-free analysis of psychosis and individual-level subtype risk. PRSet was run with all SNPs included (P-value threshold < 1) and performed two types of gene-set analysis: a 'self-contained' analysis to test if a gene set is associated with the phenotype, and a 'competitive' analysis to test if the gene set is more associated than a random set of genes with similar properties. The method was restricted to SNPs within a 10-kilobase window around each gene, and SNPs were clumped independently for each pathway (R^2 threshold = 0.1, P-value threshold = 1, 2-megabase window). The 'competitive' enrichment P-value was derived from 10,000 permutations, with significance set at $P < .05$.

To translate genetic associations from the MTAG analyses in Chapter 5 into biological insights, the following suite of post-GWAS methods was employed:

- **LD Score Regression (LDSC):** To estimate SNP-based heritability (h^2snp) and genetic correlations from summary statistics while distinguishing true polygenicity from confounding, Linkage Disequilibrium Score Regression (LDSC) was used. A low median LDSC intercept of 1.015 confirmed minimal inflation from uncorrected population stratification.
- **Functional Mapping and Annotation of GWAS (FUMA)**: FUMA (v1.8.0/v1.5.2) was used to functionally map and annotate genetic associations using GWAS summary statistics aligned to the GRCh37 (hg19) reference. The SNP2GENE and GENE2FUNC functions were used to identify independent genomic loci and annotate putative causal genes, with significance based on a Bonferroni correction across 19,139 genes ($P<2.61\times10\text{-}^6$). For reference datasets see Table 15 below.
    - Locus Definition: Standard clumping was applied in FUMA ($r^2 = .1$, 250 kb window) using the 1000 Genomes Project European-ancestry reference panel.
    - Genomic risk loci were defined by identifying independent significant SNPs ($P\leq5\times10\text{-}8$, r2< .6) which were then clumped at a stricter threshold (r2< .1) to define lead SNPs. Loci were formed by merging LD blocks of independent SNPs within a 250 kb distance. Loci were classified as "novel" if situated more than 500 kb from loci previously reported in the GWAS Catalog for BD or SCZ.
    - Gene Mapping: Three strategies were used to link SNPs to genes:
        - Positional Mapping: SNPs within a 10 kb window of a gene's boundaries (based on ANNOVAR) were mapped to that gene.
        - eQTL Mapping: SNPs were mapped to genes if they were significant cis-eQTLs in any of the brain tissue types considering pairs up to 1Mb apart.
        - Chromatin Interaction Mapping: SNPs were mapped to genes via long-range Hi-C data from tissue/cell types, including adult and foetal brain samples (e.g., Giusti-Rodriguez *et al*., 2019; PsychENCODE)[63-64]. A mapping was established if a SNP's region interacted with a gene's promoter (250 base pairs [bp] upstream to 500bp downstream of the transcription start site).
    - Functional Annotation: Combined Annotation Dependent Depletion (CADD) scores were used to predict the deleteriousness of genetic variants. A CADD score exceeding the widely accepted threshold of 12.37 is considered indicative of a potentially deleterious genetic variant [65].
- Gene-Set Analysis (MAGMA): MAGMA (v1.10) performed a competitive gene-set analysis to identify enriched biological pathways. SNPs within a window of 35 kb upstream and 10 kb downstream of a gene were assigned to it. The analysis tested 17,023 gene sets (including "Canonical pathways" and "GO terms") from MsigDB (v2023.1Hs), with significance at a Bonferroni-corrected threshold of $P<2.94\times10\text{-}^6$.
- Cell-Type Specificity Analysis: To identify the specific brain cell types where the genetic risk for a subphenotype is concentrated, a gene-property analysis was

performed. A gene-property analysis using MAGMA tested for enrichment across 226 unique cell types from 31 public single-cell RNA sequencing datasets from the human brain (including data from Wang *et al*., 2018; Hodge *et al*., 2019; Habib *et al*., 2017; La Manno *et al*., 2016; and Hochgerner *et al*., 2017) [66-70], [Table 15].

The analysis used 16,830 genes and conditioned on covariates (e.g., gene size, density). A 3-step workflow identified specific associations: (1) a per-dataset analysis, (2) a within-dataset conditional analysis, and (3) a cross-dataset conditional analysis. Significance was determined using the Benjamini-Hochberg (BH) [71] method, with a final Bonferroni-corrected threshold of $P \leq 2.2 \times 10\text{-}^{5}$.

- **Transcriptome-Wide Association Studies (TWAS)**: While FUMA annotates and maps risk variants to genes, TWAS provides a formal statistical test to identify which of those genes are likely causal by testing if their genetically predicted expression level is directly associated with the trait. To help prioritize potentially causal genes at GWAS loci by testing whether genetic risk is mediated through gene expression, a Transcriptome-Wide Association Study (TWAS) was conducted. The analysis was implemented using the FUSION [28] software (within the GenomicSEM T-SEM module) [72] and utilized precomputed functional weights from large-scale eQTL datasets. These included 15 brain tissues from the GTEx Consortium (v8) and CommonMind Consortium (CMC), Table 15. To mitigate confounding from the highly complex Major Histocompatibility Complex (MHC) region, all primary analyses were conducted both with and without the MHC region, defined as coordinates chr6:28,477,797-33,448,354 (GRCh37/hg19).
- The analysis was restricted to genes with significant evidence of cis-heritable expression ($P < .01$), and transcriptome-wide significance was set at a Bonferroni-corrected threshold ($P \leq 5.54 \times 10^{-7}$). To distinguish true causal effects from associations driven by linkage disequilibrium (LD) with other nearby genes, a conditional analysis was also performed within the FUSION framework. This analysis tests whether a gene's association with a subphenotype remains significant after statistically accounting for the effects of all other associated genes within the same locus. This step is crucial for dissecting complex GWAS loci where multiple genes may show a TWAS signal, helping to pinpoint which gene has the most direct, independent effect on the trait.
- While this conditional analysis helps to identify independent signals, it is distinct from more advanced fine-mapping methods such as FOCUS [73]**.** FOCUS goes a step further by using the information from all genes in a locus to calculate a posterior probability that each specific gene is the true causal gene. The conditional analysis performed here provides an essential intermediate step, giving stronger evidence for a gene's independent role and increasing confidence in its prioritization for further biological investigation, but does not provide a formal probabilistic estimate of causality that FOCUS does.

- **Local Genetic Correlation (LAVA)**: Local Analysis of [Co]variant Association (LAVA) was used to estimate local genetic correlations and identify specific genomic regions with shared risk between subphenotypes.
- Genetic Architecture Analysis (SBayesS): This summary-level Bayesian model was used to estimate SNP-based heritability (h^2snp), polygenicity, and negative selection (*S*) for subphenotypes. Heritability was transformed to the liability scale using a shrunk LD matrix from GCTA (available at https://yanglab.westlake.edu.cn/software/gcta). Model convergence was confirmed by the Gelman and Rubin statistic (R^<1.2).
- **Credible Gene Set Prioritization**: A "credible" gene was defined as one meeting two criteria: (1) a significant association in the conditional TWAS analysis, and (2) implication by at least one of three mapping strategies in FUMA. To prioritize a high-confidence set of risk genes, a gene was defined as "credible" if it was significant in the conditional TWAS analysis and was also implicated by at least one of the three FUMA mapping strategies (positional, eQTL, or chromatin interaction). The statistical validity of this credible gene set was then confirmed by testing for enrichment of established rare-variant risk genes using a one-sided Fisher's exact test.
- **Validation with Rare-Variant Data**: The credible gene sets were tested for enrichment of established rare-variant risk genes from the Schizophrenia Exome Meta-analysis (SCHEMA) [74] and Bipolar Exome (BipEx) consortia [75]. The enrichment was assessed using a one-sided Fisher's exact test ($P < .0125$).

Table 15 Reference Datasets and Publications for FUMA Analysis Modules and TWAS

FUMA Module	Category	Specific Dataset/Tool	Reference
Available at https://fuma.ctglab.nl/links			
Cell Type	scRNA-seq	Adult Human Brain	Siletti *et al*. (2023). *Science*. 382(6667).
		GSE168408	Herring *et al*. (2022). *Cell*. 185, 4428-4447.
		Allen Brain Atlas (Human MTG)	Hodge *et al*. (2018). *bioRxiv*. doi: 10.1101/384826.
		DroNc	Habib et al. (2017). *Nat. Methods*. 14, 955-958.
		GSE76381	La Manno *et al*. (2016). *Cell*. 167, 556-580.
		GSE101601	Hochgerner *et al*. (2017). *Sci. Rep*. 7: 16327.
		GSE104276	Zhong et al. (2018). *Nature*. 555, 524-528.
		GSE67835	Darmanis et al. (2015). *Proc. Natl. Acad. Sci. USA*. 112, 7285-90.
SNP2GENE	LD Reference Panel	1000 Genomes Project Phase 3	The 1000 Genomes Project Consortium. (2015). *Nature*. 526, 68–74.
		UK Biobank	Bycroft *et al.* (2018). *Nature*. 562(7726), 203–209.
SNP2GENE	Gene Expression (MAGMA)	BrainSpan	Kang *et al*. (2011). *Nature*. 478, 483-489.
SNP2GENE	eQTL Mapping	Blood eQTL Browser	Westra *et al.* (2013). *Nat. Genet*. 45, 1238-1243.
		BIOS QTL Browser	Zhernakova *et al*. (2017). *Nat. Genet*. 49, 139-145.
		BRAINEAC	Ramasamy *et al*. (2014). *Nat. Neurosci*. 17, 1418-1428.
		CommonMind Consortium	Fromer et al. (2016). *Nat. Neurosci*. 16, 1442-1453.
		MuTHER	Grundberg et al. (2012). *Nat. Genet*. 44, 1084-1089.
		xQTLServer	Ng *et al*. (2017). *Nat. Neurosci*. 20, 1418-1426.
		eQTLGen	Vosa *et al*. (2018). *bioRxiv*. doi: 10.1101/447367.
		DICE	Schmiedel *et al*. (2018). *Cell*. 175, 1701-1715.e16.
		van der Wijst et al. scRNA eQTLs	van der Wijst *et al*. (2018). *Nat. Genet*. 50, 493-497.
		eQTL Catalogue	Kerimov *et al*. (2021). *Nucleic Acids Res*. 49(D1), D997-D1003.
		EyeGEx	Ratnapriya *et al*. (2019). *Nat. Genet*. 51(4), 615-624.
		InsPIRE	Viñuela *et al*. (2020). *Cell Reports*. 31(10), 107727.
		TIGER	Alonso *et al*. (2021). *Cell Reports*. 37(13), 110167.
SNP2GENE	Chromatin Interaction	Hi-C (GSE87112)	Schmitt *et al*. (2016). *Cell Rep*. 17, 2042-2059.
		Hi-C (Giusti-Rodriguez et al.)	Giusti-Rodriguez *et al*. (2019). *bioRxiv*. doi: 10.1101/406330.
		FANTOM5	Andersson *et al*. (2014). *Nature*. 507, 455-461.
GENE2FUNC	Gene Expression	BrainSpan	Kang *et al.* (2011). *Nature*. 478, 483-489.
GENE2FUNC	Gene Set Enrichment	WikiPathways	Kutmon *et al*. (2016). *Nucleic Acids Res*. 44, 488-494.
		DrugBank	Wishart *et al.* (2008). *Nucleic Acis Res*. 36, D901-6.

FUMA Module	Category	Specific Dataset/Tool	Reference
All Modules	Core Tool	PLINK	Purcell *et al.* (2007). *Am. J. Hum. Genet*. 81, 559-575.
		MAGMA	de Leeuw *et al*. (2015). *PLoS Comput. Biol*. 11, e1004219.
	Annotation Tool	ANNOVAR	Wang *et al*. (2010). *Nucleic Acids Res*. 38:e164.
	Annotation Score	CADD	Kircher *et al*. (2014). *Nat. Genet*. 46, 310-315.
		RegulomeDB	Boyle *et al.* (2012). *Genome Res*. 22, 1790-7.
	Annotation Data	15-core chromatin state (ChromHMM)	Roadmap Epigenomics Consortium. (2015). *Nature*. 518, 317-330.
		GTEx	The GTEx Consortium. (2020). *Science*. 369(6509), 1318-1330.
		PsychENCODE	Wang *et al*. (2018). *Science*. 362, eaat8464.
	Gene Score	pLI (from ExAC)	Lek *et al*. (2016). *Nature*. 536, 285-291.
		ncRVIS	Petrovski *et al*. (2015). *PLOS Genet*. 11, e1005492.
	Gene Set Enrichment	MSigDB	Liberzon *et al*. (2011). *Bioinformatics*. 27, 1739-40.
		GWAS Catalog	MacArthur *et al.* (2016). *Nucleic Acids Res*. pii:gkw1133.
TWAS Module	Download at http://gusevlab.org/projects/fusion		
TWAS Module	T-SEM (FUSION software in GenomicSEM module)	CommonMind Consortium (CMC)	Brain (DLPFC) - RNA-seq, Brain (DLPFC) - RNA-seq splicing
		GTEx v8	Amygdala, Anterior cingulate cortex (BA24), Caudate (basal ganglia), Cerebellar Hemisphere, Cerebellum, Cortex, Frontal Cortex (BA9), Hippocampus, Hypothalamus, Nucleus accumbens (basal ganglia), Putamen (basal ganglia), Spinal cord (cervical c-1), Substantia nigra
		Foetal	O'Brien, Heath E., *et al*. "Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders." *Genome biology* 19.1 (2018): 194.

2.9 Psychometric and Predictive Modelling

Dimensional Structure Analysis

To investigate the underlying structure of clinical symptoms in the bipolar disorder sample, a multi-stage approach was employed in Chapter 3 using both Exploratory and Confirmatory Factor Analysis (EFA/CFA). The initial data-driven exploration used EFA to uncover latent dimensions of psychopathology from 77 clinical items from the Operational Criteria (OPCRIT) checklist. Based on multiple criteria including parallel analysis, scree plots, and model fit indices (Table 16, and Chapter 3 Table 19, Figures 12-13), a four-factor structure was identified as the most robust and clinically relevant.

Table 16 Factor Model fit indices description for Chapter 3 & 5

Chi-Square Test	Represents the difference between the observed and expected covariance matrices. A non-significant *p*-value indicates a good model fit, although this test can be sensitive to sample size.
Comparative Fit Index (CFI)	Compares the fit of the specified model to a baseline (often a null model). Values more than .90-.95 indicate a good fit.
Root Mean Square Error of Approximation (RMSEA)	Measures the error of approximation in the population. Values that are lower than .05 are considered a good fit, values below .08 are considered acceptable.
Tucker-Lewis Index (TLI)	Like CFI, the value accounts for model complexity. A TLI above .90 suggests the model has a good fit.

Data for 77 clinical symptoms from the OPCRIT with adequate sample sizes were included for analysis. Items with zero- or near-zero variance were removed to enable model convergence. The clinical sample was partitioned into balanced 60/40 splits for the exploratory (calibration) and confirmatory (validation) phases using the createDataPartition function in the Caret [76] package. The analysis was conducted on a calibration subsample of 1554 BD patients. The 'WLSMV' estimator was used for the ordinal categorical items, and Geomin rotation was applied to allow the latent factors to correlate. Items with very low frequencies were analysed separately via regression. For items with a high pairwise correlation (.7 or above), the item with the least missingness and most clinical relevance was retained to ensure a parsimonious model.

To prepare the data for factor analysis, several steps were taken in R. Redundancy between clinical items was assessed using the hetcor function in the polycor [77] package. Missing data, which was low at 8%, was assessed using the var_miss function in the Naniar [78] package, and its pattern was confirmed to be Missing At Random (MAR) using the missing_compare function in the finalfit [79] package. To avoid potential overfitting, imputation was not performed. The clinical sample was partitioned into balanced 60/40 splits

for exploration and confirmation using the createDataPartition function in the caret package. To further confirm the data's suitability for structure detection, the Kaiser-Meyer-Olkin (KMO) [80] test was used to measure sampling adequacy, and Bartlett's test of sphericity [81] was used to test for significant interrelatedness between variables. For preliminary Principal Component Analysis (PCA) on the clinical data, scales were normalized before computing components with the 'prcomp' function in R.

The number of factors to retain was determined using multiple criteria, including parallel analysis with the fa.parallel function in the psych [82] package and scree plots generated with the fviz_eig function in the factoextra [83] package. Exploratory and Confirmatory Factor Analyses were conducted using the efa and cfa functions, respectively, from the lavaan [84] package. The final path diagrams were visualized using the lavaanPlot [85] package.

This structure was then formally tested and validated using CFA on an independent subsample. A parsimonious model consisting of the 20 core OPCRIT symptoms that loaded most strongly and consistently onto the four factors was developed. The items comprising this final four-factor model are detailed in Table 17 below. A complete list of all 77 items included in the initial exploratory analysis, along with their full factor loadings, can be found in Chapter 3, Table 22. The selection of the 20 symptoms for the CFA was based on a median factor loading above .6 to ensure a parsimonious and reliable model, which is a widely accepted practice for retaining meaningful indicators. The initial threshold of .4 was used to interpret the EFA factor loadings. This threshold was chosen to align with criteria used in previous factor analyses of bipolar disorder symptoms (Allardyce *et al*., 2023) [86].

The fit of the final factor models was evaluated using multiple fit indices in Chapter 3 and 5, though the Standardised Root Mean Squared Residual (SRMR) was not used due to evidence of bias in binary data.

Further checks for multicollinearity (using eigenvalues) and the linearity assumption (using bivariate scatterplots of predicted factor scores) were also performed.

To assess the genetic contributions to these latent factors, the CFA was extended into a Multiple Indicator Multiple Cause (MIMIC) model, a special case of Structural Equation Modeling (SEM). This was implemented with the 'sem' function in lavaan, and an *a priori* power analysis for the model was conducted using the semPower [87] package in R. This integrated model was preferable to separate multiple regressions as it allows for the simultaneous modelling of both the factor-level and item-level associations with genetic load.

2.10 Validation and Sensitivity Analyses

Sensitivity analyses for dimensional modelling for Chapter 3

To test the robustness of the final CFA model and the independent contribution of each of the 20 core clinical items in Chapter 3, a multi-stage validation process was used. Two primary sensitivity analyses were performed.

First, individual-level factor scores for each of the four dimensions were estimated using a ”leave-one-out” approach, where each item was omitted from the CFA model in turn. The resulting factor scores were then used in regression models to confirm that a dimension’s score was the best predictor for its own constituent symptoms.

Second, to ensure the genetic associations identified at the global SEM level held for individual items, regression analyses were conducted using the individual-level Polygenic Risk Scores (PRS) for each of the five psychiatric disorders to predict the presence of each of the 20 core symptoms. For both of these sensitivity analyses, participants were dichotomized into the top 10% of scores versus the remaining 90% to assess the increased risk (Odds Ratio) for reporting a given symptom.

Finally, a third set of post-hoc regression analyses was performed for two primary reasons. The first was to explore the association between the newly identified chronicity dimension and other clinically important variables known to be associated with poorer outcomes in BD. The second, more specific reason, was to investigate variables like rapid cycling, suicide thoughts, and substance use. These variables were clinically expected to correlate with the chronicity dimension but did not meet the strict statistical cutoff (factor loading > .6) for inclusion in the final, parsimonious 20-item CFA model. This post-hoc approach allowed these crucial relationships to be investigated without degrading the primary model.

Table 17 OPCRIT Variables for Analyses in Chapters 3 & 4

OPCRIT Item No.	Variable Name	Full OPCRIT Definition & Original Coding	Use in Thesis (Chapter & Analysis)	Analytical Coding
Part I: Core Items for Dimensional Analysis (Chapter 3)				
24	Slowed activity	Obvious slowing of movement, reaction time, and speech. (0, 1, 2)	Chapter 3: EFA/CFA	Treated as Ordinal
42	Excessive self-reproach	Pathological feelings of guilt, including self-blame and remorse which is persistent, inappropriate or out of proportion. (0, 1, 2)	Chapter 3: EFA/CFA	Treated as Ordinal
39	Loss of pleasure (anhedonia)	A pervasive loss of interest or pleasure in all or almost all of the patient's usual activities. (0, 1, 2)	Chapter 3: EFA/CFA	Treated as Ordinal
25	Loss of energy/tiredness	Subjective experience of tiredness, weariness or loss of energy. (0, 1, 2)	Chapter 3: EFA/CFA	Treated as Ordinal
37	Dysphoria	An unpleasant mood state with features of depression, anxiety and/or irritability. (0, 1, 2)	Chapter 3: EFA/CFA	Treated as Ordinal
31	Racing thoughts	Thoughts that are so rapid the patient cannot 'keep up with them'. (0, 1, 2)	Chapter 3: EFA/CFA	Treated as Ordinal
30	Pressured speech	An increase in the amount and/or speed of speech which is difficult for the interviewer to interrupt. (0, 1, 2)	Chapter 3: EFA/CFA	Treated as Ordinal
22	Reduced need for sleep	Patient feels rested and full of energy after only a few hours sleep (e.g. 3 hours less than usual). (0, 1, 2)	Chapter 3: EFA/CFA	Treated as Ordinal
19	Excess activity	An increase in the level of activity, e.g. at work, socially or sexually. (0, 1, 2)	Chapter 3: EFA/CFA	Treated as Ordinal
35	Elevated mood	A sustained feeling of wellbeing, cheerfulness, or elation, which is not in keeping with the patient's circumstances. (0, 1, 2)	Chapter 3: EFA/CFA	Treated as Ordinal
58	Delusions of influence	The belief that one's feelings, impulses, thoughts, or actions are not one's own, but are imposed by some external force. (0, 1)	Chapter 3: EFA/CFA	Treated as Ordinal
54	Persecutory/jealous delusions	A delusion of being persecuted (e.g. being followed, harassed, conspired against), or of the infidelity of one's spouse or partner. (0, 1)	Chapter 3: EFA/CFA	Treated as Ordinal
67	Thought withdrawal	The experience of thoughts being removed from one's mind by an outside agency. (0, 1)	Chapter 3: EFA/CFA	Treated as Ordinal
66	Thought insertion	The experience of thoughts, which are not one's own, being inserted into one's mind. (0, 1)	Chapter 3: EFA/CFA	Treated as Ordinal
68	Thought broadcast	The experience of one's thoughts being broadcast or escaping from one's mind so that others can hear them. (0, 1)	Chapter 3: EFA/CFA	Treated as Ordinal
10	Premorbid poor social adjustment	Patient found difficulty entering or maintaining normal social relationships, showed persistent social isolation, withdrawal or maintained solitary interests prior to onset of psychotic symptoms. (0, 1)	Chapter 3: EFA/CFA & Post-hoc Regressions	Treated as Ordinal / Binarised
11	Premorbid personality disorder	Evidence of inadequate/schizoid/schizotypal/paranoid/cyclothymic/psychopathic/sociopathic personality disorder present since adolescence and prior to the onset of psychotic symptoms. (0, 1)	Chapter 3: EFA/CFA & Post-hoc Regressions	Treated as Ordinal / Binarised
9	Premorbid poor work adjustment	Refers to work history before onset of illness. Scored if the patient was unable to keep any job for more than 6 months, had a history of frequent	Chapter 3: EFA/CFA	Treated as Ordinal

		changes of job or was only able to sustain a job well below that expected. (0, 1)		
88	Inter-episode remission (subsyndromal)	Deterioration from premorbid level of functioning: Patient does not regain his premorbid social, occupational or emotional functioning after an acute episode of illness. (0, 1)	Chapter 3: EFA/CFA	Treated as Ordinal
90	Course of disorder (chronic)	Course of disorder: 1 = Single episode with good recovery, 2 = Multiple episodes with good recovery between, 3 = Multiple episodes with partial recovery between, 4 = Continuous chronic illness, 5 = Continuous chronic illness with deterioration. (1, 2, 3, 4, 5)	Chapter 3: EFA/CFA	Treated as Ordinal
Part II: Key Variables for Transdiagnostic & Predictive Analyses (Chapters 3 & 4)				
4	Age of Onset	The age at which the proband first met criteria for a manic, mixed or major depressive episode.	Chapter 4: Regression/RF Models	Continuous (Age in years)
36	Irritable mood	A mood state characterized by a pervasive feeling of irritability. (0, 1, 2)	Chapter 4: Regression/RF Models	Binarised (0=No, 1=Yes)
43	Suicidal thoughts	Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide. (0, 1)	Chapter 3: Post-hoc Regressions	Binarised (0=No, 1=Yes)
80	Other substance abuse/ dependence	A lifetime diagnosis of abuse of or dependence on any other specified substance. (0, 1)	Chapter 3: Post-hoc Regressions	Binarised (0=No, 1=Yes)
87	Impairment/ incapacity during disorder	0 = No impairment, 1 = Subjective impairment, 2 = Impairment in major life role, 3 = No function at all in major life role.	Chapter 3: EFA	Treated as Ordinal
N/A	Rapid Cycling	A derived variable based on the OPCRIT assessment: “Four or more mood disturbances in one year?”	Chapters 3 & 4: Regressions/RF Models	Binarised (0=No, 1=Yes)
N/A	Psychosis (Overall)	A composite variable defined by the presence of any OPCRIT item related to delusions or hallucinations.	Chapter 4: Regression/RF Models	Binarised (0=No, 1=Yes)

Statistical Learning and Predictive Models for Chapter 4

To extend the predictive analyses beyond standard regression and account for non-linear relationships and interactions between predictors, Random Forest (RF) models were employed, implemented via the cforest function in the caret package in R. As detailed in Chapter 4, these models were used to evaluate the predictive performance of the SCZ3-PRS alone and in combination with other clinical variables for several BD1 subphenotypes. The key OPCRIT-derived variables used in these analyses are defined in Table 17 above.

The RF models utilized a conditional inference framework (cforest) [88] to reduce the risk of overfitting in data with correlated predictors. Model performance for binary outcomes (e.g., psychosis) was assessed using 10-fold cross-validation to calculate the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC). For continuous outcomes (e.g., age of onset), performance was assessed with Root Mean Squared Error (RMSE) and R-squared (R2).

The relative importance of each predictor in the models was determined using the Mean Decrease Accuracy (MDA) score. For comparison, penalized regression models (elastic net) were also used to assess variable importance using cv.glmnet [89] in R.

To formally compare the predictive performance of different models (e.g., a model with clinical variables versus a model with both clinical and genetic variables), pairwise Bonferroni-corrected one-sample t-tests were performed on the performance metrics (AUC or R2) generated during cross-validation. For interpretation, a model was considered to have clinical utility if the AUC and Positive Predictive Value (PPV) reached at least .8. An AUC value between .71 and .79 was considered moderately discriminative, while an AUC ≥ .79 was considered strongly discriminative.

The statistical significance of each predictor's importance score (Mean Decrease Accuracy, MDA) was determined using a permuted, cross-validated *P*-value implemented in the Vita [90] R package. To account for potential bias from correlated predictors, conditional permutation importance (CPI)[was calculated using the Permimp [91] R package to establish the final variable importance rankings.

To ensure the reliability of the prediction of BD1 traits and to handle correlated clinical variables, this thesis used a non-parametric algorithm in addition to standard regressions. This approach can detect non-linear relationships and was implemented using penalized 'elastic net' modelling and Conditional Random Forest ('cforest') functions within the 'caret' R package, which uses a conditional inference framework to reduce the risk of overfitting.

Multivariate Regression Models: For the elastic net penalty regression models, individuals with BD1 were randomly allocated to training, validation, or testing sets (70%:15%:15%). Ten-fold cross-validations were implemented to further avoid overfitting, with classification statistics calculated in the 'cvAUC' [92] package in R. These models served as a robustness check for comparison with the Random Forest models' variable importance rankings.

Non-parametric Random Forest Models: For the RF models, individuals with BD1 were also randomly allocated to training, validation, and testing sets (70%:15%:15%). RF predictions rely on bootstrapping 1000 decision trees, and tuning parameters (mtry = 2, 4, 7, 10) were used to optimize the models. Predictive performance for binary outcomes was determined using ten-fold cross-validated models to calculate the Receiver Operating Characteristic (ROC) curve, Area-Under-the-Curve (AUC), sensitivity, specificity, and accuracy. For continuous outcomes, accuracy was assessed with Mean Absolute Error (MAE) and the more stringent Root Mean Squared Error (RMSE) along with R-squared (R2).

Ranking Variable Importance: The importance of variables in predicting psychosis and its subtypes was compared between the penalized elastic net regression and the conditional random forest models. For regressions, the effect size is reported as the log odds ratio (LogOR). For random forest, variables are ranked based on their Mean Decrease Accuracy

(MDA) score; higher scores represent more accuracy loss when the variable is excluded from the model.

2.11 Derivation of Genetic-Clinical Dimensions from Subphenotypes

To provide an empirical framework for the clinical heterogeneity of bipolar disorder, a multi-step analysis was performed in Chapter 5. The primary goal was to identify underlying latent factors that could group the subphenotypes into broader, more genetically coherent dimensions.

To empirically deconstruct the clinical heterogeneity of bipolar disorder for the analysis in Chapter 5, a multi-stage factor analysis was performed on 11 clinical subphenotypes in a sample of 18,800 BD cases. The suitability of the data for this analysis was first confirmed with the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and a significant Bartlett's test of sphericity. Initial exploratory techniques included Principal Component Analysis (PCA) and Factor Analysis of Mixed Data (FAMD), which was implemented in the FactoMineR [93] R package to visualize the main components. Additionally, hierarchical clusters were investigated using the 'iclust' algorithm from the psych package, where subphenotypes were merged into composite scales based on an increase in coefficients alpha and beta. A parallel analysis, conducted using the psych package in R, provided statistical support for a four-factor model, which was then formally tested and validated using Confirmatory Factor Analysis (CFA) in the lavaan package. The final four-factor clinical model was selected after demonstrating a superior fit compared to more parsimonious models with fewer factors. Finally, to validate the clinical structure with genetic data, a separate Principal Component Analysis (PCA) was performed on the genome-wide significant MTAG loci. This analysis was conducted using the FactoMineR package for computation and the factoextra package for visualization. The statistical reliability of the resulting genetic-clinical dimensions was then confirmed with a one-way ANOVA using independent results from the LAVA analyses.

To assess for phenotypic heterogeneity before pooling data for the meta-analyses, generalized linear mixed-effects models (GLMMs) were performed with geographic region included as a random effect. The random effect was consistently non-significant across the models, confirming a high degree of phenotypic homogeneity across recruitment sites and supporting the validity of the combined analysis.

Table 18 Additional OPCRIT Variables for Analyses in Chapters 3 & 4

OPCRIT Item No.	Variable Name	Full OPCRIT Definition & Original Coding	Use in Thesis (Chapter & Analysis)	Analytical Coding
1	Source of rating	The source of the patient data – 1 = Hospital case notes (charts). 2 = Structured interview with subject. 3 = Prepared abstract. 4 = Interview with informant. 5 = Combined sources including structured interview. 6 = Combined sources not including structured interview	Chapters 3 & 4: Regressions/RF Models	Nominal
4	Age of Onset	The age at which the proband first met criteria for a manic, mixed or major depressive episode.	Chapter 4: Regression/RF Models	Continuous (Age in years)
36	Irritable mood	A mood state characterized by a pervasive feeling of irritability. (0, 1, 2)	Chapter 4: Regression/RF Models	Binarised (0=No, 1=Yes)
N/A	Rapid Cycling	A derived variable based on the OPCRIT assessment: "Four or more mood disturbances in one year?"	Chapters 3 & 4: Regressions/RF Models	Binarised (0=No, 1=Yes)
N/A	Psychosis (Overall)	A composite variable defined by the presence of any OPCRIT item related to delusions or hallucinations.	Chapter 4: Regression/RF Models	Binarised (0=No, 1=Yes)
N/A	Congruent/ Incongruent Psychosis	A composite variable indicating psychotic symptoms consistent/inconsistent with the patient's mood state. (Available for RO cohort only).	Chapter 4: Regression/RF Models	Binarised (0=No, 1=Yes)

All statistical analyses were carried out in R version 4.4.2 [3] on data stored securely on computer clusters supported by University College London (London, UK).

3 Bipolar Disorder Dimensionality

A preprint version of the research in this chapter is available on *medRxiv* at
doi: https://doi.org/10.1101/2025.05.17.25327825

3.1 Abstract

Background: Bipolar disorder (BD) factor models offer limited dimensional understanding due to incomplete integration of chronic deficits, long-term outcomes, and transdiagnostic genetics, thus restricting personalised interventions. This study aimed to provide a holistic understanding of BD psychopathology, overcoming this limitation.

Aims: In this study I aimed to develop and validate a novel dimensional model of bipolar disorder (BD) that integrates premorbid factors, and to investigate the transdiagnostic genetic architecture of its dimensions using polygenic risk scores. The study hypothesized that a distinct dimension of bipolar disorder exists that links premorbid factors to a poor long-term illness course. Furthermore, it was predicted that this adverse trajectory would be genetically associated with a higher risk for ADHD and anxiety.

Methods: Exploratory Factor Analysis of 77 OPCRIT items revealed four psychopathological dimensions, and Confirmatory Factor Analysis validated a 20-item, four-factor BD model. Polygenic Risk Scores for five relevant disorders were calculated, and Structural Equation Modelling analysed the genetic contributions to this dimensional model. The study applied Inverse Probability Weighting to address biases in a sample of 4992 participants.

Results: Confirmatory Factor Analysis revealed a novel Adverse Chronic Trajectory (ACT) dimension, characterised by the co-occurrence of premorbid deficits, reduced inter-episode remission and poorer long-term outcomes in individuals with BD. Structural Equation Modelling further showed distinct patterns of genetic liability: BD PRS for mania, Schizophrenia (SCZ) PRS for psychosis, and Major Depressive Disorder (MDD) PRS for depression. Notably, the ACT dimension exhibited a positive association with Attention-Deficit/Hyperactivity Disorder (ADHD) and anxiety PRSs, and an inverse relationship with BD PRS.

Conclusions: This study offers a novel and clinically relevant dimensional model of BD by identifying the ACT dimension, which uniquely integrates crucial premorbid factors and outcomes. The identified direct genetic link between ADHD and anxiety with ACT (a trajectory associated with poorer BD outcomes) provides important new insight into a

challenging illness course. This potentially enables earlier identification and facilitates targeted interventions to reduce risk for a chronic outcome and overall quality of life in BD.

3.2 Introduction

Limitations of Categorical Diagnosis

Bipolar disorder (BD) shows diverse outcomes influenced by genetics beyond current subtyping (BD1, BD2). Traditional classifications often overlook the critical impact of premorbid factors on long-term outcomes. While course specifiers aim to improve treatment alignment [1-2], this study proposes a novel dimensional approach for a more nuanced understanding of BDs inherent heterogeneity beyond categorical diagnoses.

Bipolar Disorder: A Symptom Continuum

Many BD patients experience continuous symptoms beyond discrete episodes: cognitive deficits in remission are reported, with prevalence as high as 70% [3-5]; 20-50% experience inter-episodic symptoms [6], highlighting limitations of episodic models. Even during euthymia, executive dysfunction and anxiety persist, indicating vulnerability [7-8]. Personality traits also influence BDs onset, progression, and course [9].

Dimensional Frameworks in Bipolar Disorder

Dimensional approaches dissect BDs heterogeneity, allowing researchers to identify potentially more genetically similar subgroups based on specific symptom profiles. Acknowledging this heterogeneity, research increasingly focuses on genetic differences within more homogeneous subgroups [10] to understand genetic contributions to diverse presentations. While specific BD course specifiers show familiality [11], and genetic liabilities for subphenotypes are being identified, single regression models can complicate interpretation [12-24]. A dimensional framework offers a powerful alternative by examining psychopathology along continuous axes, enabling nuanced analysis of specifier interrelations and combined genetic liabilities for a holistic understanding of BD heterogeneity.

Impact of Premorbid Factors on Bipolar Disorder

Cognitive and functional deficits, not fully recognised specifiers [1, 25], contribute to BD variability and impair quality of life, even during mood stability [4, 26]. These deficits exist on a spectrum, negatively impacting relationships and productivity, often leading to social withdrawal [27] and affecting 30-60% of adults with BD [28]. Early onset of these deficits links to worse outcomes including anxiety, substance use, and suicidality, with increased childhood risk [29-30]. Recognising genetic predisposition could potentially reduce diagnostic delays [31] and suicide rates in BD [32]. Examining these deficits within a broader psychopathological spectrum may also clarify connections to other disorders. While research on psychosis [12] explored premorbid risk factors, their specific impact on long-term BD outcomes remains less understood.

Genetic Contributions and Polygenic Risk Scores

Genetic factors substantially contribute to BD comorbidity [33], with approximately 35-65% of individuals with BD meeting criteria for another psychiatric condition [34], indicating complex psychopathology interplay. This high comorbidity suggests single-disorder analyses might miss critical genetic factors contributing to this broader spectrum of co-occurring conditions. Symptoms often begin early and persist, and may be worsened by environmental factors [35-39].

Polygenic risk scores (PRSs) are valuable tools for investigating the genetic basis of BD heterogeneity and comorbidities [40-41]. Analysing symptom clusters may reveal stronger genetic associations than isolated disorder analyses. For example, higher SCZ PRS is linked to mood-incongruent psychotic symptoms and earlier BD onset [16, 22]. Higher ADHD and anxiety risk correlates with rapid cycling [19, 42]. ADHD increases multimorbidity risk, worsening symptom severity and functional impairment [44-45]. Polygenic ADHD burden has been linked to earlier BD onset and lithium resistance, while lithium response can be influenced by family history and absence of anxiety or rapid cycling [18, 46-50]. Factor analysis can simplify complex relationships between symptoms and disorders, revealing underlying factors and their genetic contributions within a spectrum framework.

3.3 Aims

Introducing Adverse Chronic Trajectory (ACT)

This study aimed to develop a novel dimensional model of bipolar disorder by integrating premorbid factors with other clinical symptoms. Building on prior BD modelling using OPCRIT items, this study introduces a novel four-factor model. By combining OPCRIT items and PRS, I identified an Adverse Chronic Trajectory (ACT) dimension, demonstrating correlations between premorbid deficits and adverse BD outcomes, with shared genetic burdens for ADHD and anxiety prominently associated with APT, thus emphasising its role and genetic links for advancing BD understanding, classification, and intervention.

3.4 Methods

The underlying structure of 77 OPCRIT [4] items was investigated using Exploratory and Confirmatory Factor Analysis. The genetic architecture of the resulting dimensions was then explored by integrating five transdiagnostic Polygenic Risk Scores into a Structural Equation Model (MIMIC) [72]. Central to this chapter's investigation of long-term outcomes were two key OPCRIT variables used to define the illness trajectory. The "Reduced inter-episode remission" (item 88) was coded as a binary measure to capture whether a patient returned to their premorbid baseline after an acute episode. To complement this, the "Course of disorder"

(item 90) was treated as an ordinal variable, allowing us to model the full spectrum of outcomes from complete recovery to a chronic, deteriorating course (for a full description, see Table 23).

A complete description of the participant cohorts, the OPCRIT instrument, factor analysis procedures, and PRS calculations is provided in Methods (Chapter 2).

3.5 Results

(i) Clinical Characteristics

77 clinical symptoms were examined and five psychiatric disorder PRS estimated in 2590 individuals with BD and 2402 healthy controls. The overall sample consisted of 61% females and 39% males, with no sex distribution differences across BD subtypes. A difference in age of onset within cases was found across BD subtypes, specifically between SZA and BD2 (Chapter 2, Table 2).

(ii) EFA

Initially 77 clinical symptoms were evaluated in a calibration sample of 1554 BD patients (60%). Seventy-six symptoms loaded ($P < .05$) across four factors; Family history of schizophrenia (OPCRIT 13) was the exception. Symptoms exceeding .4 were visualised (Figure 14**,** Table 22). A four-factor EFA model fit best ($\chi^2 = 304$, RMSEA = .033 [90% Confidence Intervals [CI] .024–.037], CFI = .989, and TLI = .986). Four factors were retained based on the lower RMSEA, parallel analysis, and scree plot (see Table 19**,** Figures 12-13 below).

3.5.1.1 Exploratory factor analysis (EFA) models fit indices

Table 19 Exploratory factor analysis (EFA) models fit indices

Model	Parameters	Chi.square	RMSEA
1-factor	77.00	877.00	.05
2-factor	153.00	678.00	.04
3-factor	318.00	552.00	.04
4-factor	304.00	462.00	.03

Note. The data was extracted from the 1-factor to 4-factor EFA models using BD clinical symptoms.

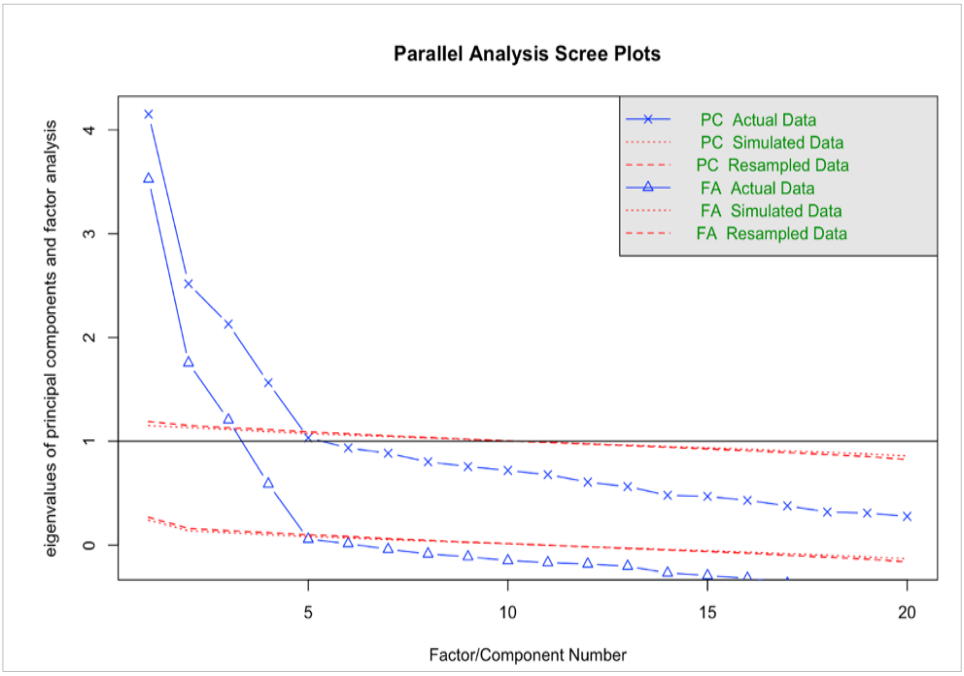


Figure 12 Parallel Analyses for exploratory factor analysis.
This figure illustrates the results of the parallel analysis conducted to determine the number of factors to retain in the exploratory factor analysis (EFA). The plot displays the eigenvalues obtained from the actual data (blue line) compared to the eigenvalues from random, uncorrelated data (red line). The intersection of the eigenvalues or the point where the real data eigenvalues drop below the random data eigenvalues typically suggests the appropriate number of underlying factors. In this specific analysis, the real data eigenvalues remain above the simulated data eigenvalues for four factors, suggesting that a four-factor model is appropriate.

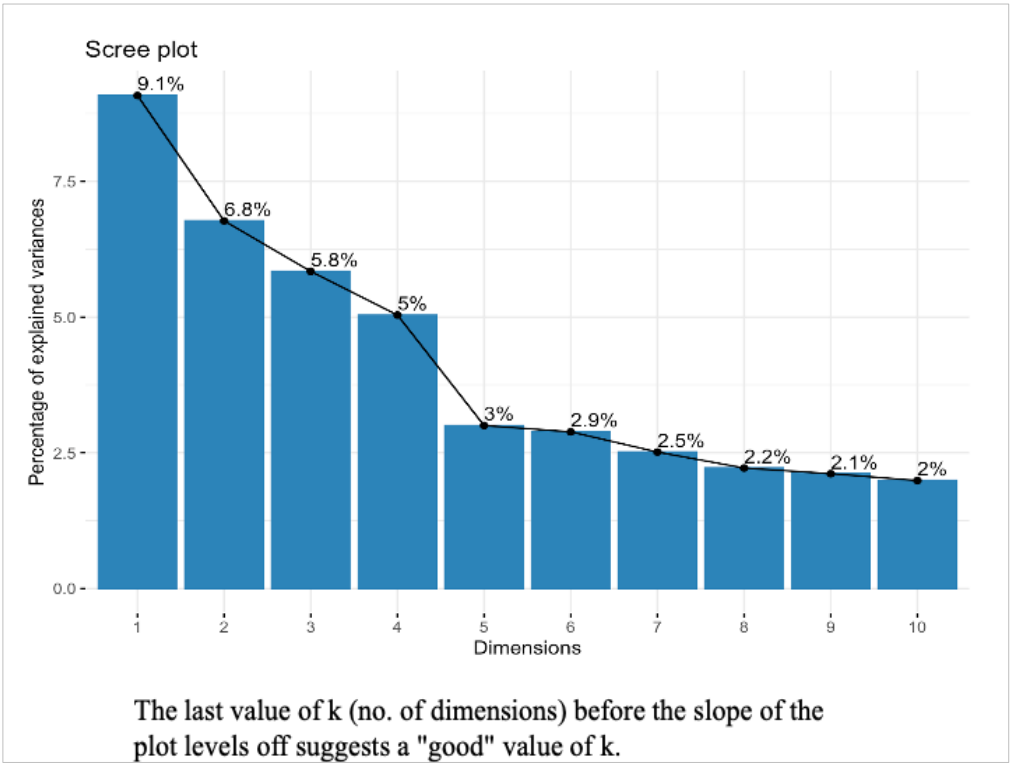


Figure 13 Scree plot for exploratory factor analysis.
This figure illustrates the scree plot, which is a graph of the eigenvalues of the factors plotted against the factor number. The shape of the plot helps to determine the number of factors to retain in EFA. The "elbow" or point of inflection in the scree plot typically indicates where the amount of variance explained by subsequent factors starts to diminish, suggesting an optimal number of factors before the "scree" begins. In this scree plot, the elbow is observed at the fourth factor, suggesting the retention of four factors is appropriate.

Table 19 presents the fit indices for Exploratory Factor Analysis (EFA) models with one to four factors, tested on a calibration subsample of bipolar disorder participants (*N*=1554). The fit indices included are Chi-Square, Root Mean Square Error of Approximation (RMSEA) with its 90% Confidence Intervals (CI), Comparative Fit Index (CFI), and Tucker-Lewis Index (TLI). These indices were used to evaluate the model fit for each number of factors to determine the optimal factor structure for the OPCRIT data. Lower RMSEA values and higher CFI and TLI values (typically above .90-.95) generally indicate a better model fit. A four-factor EFA model fit best (χ2=304, RMSEA = .033 [90% Confidence Intervals [CI] .024–.037], CFI = .989, and TLI = .986).

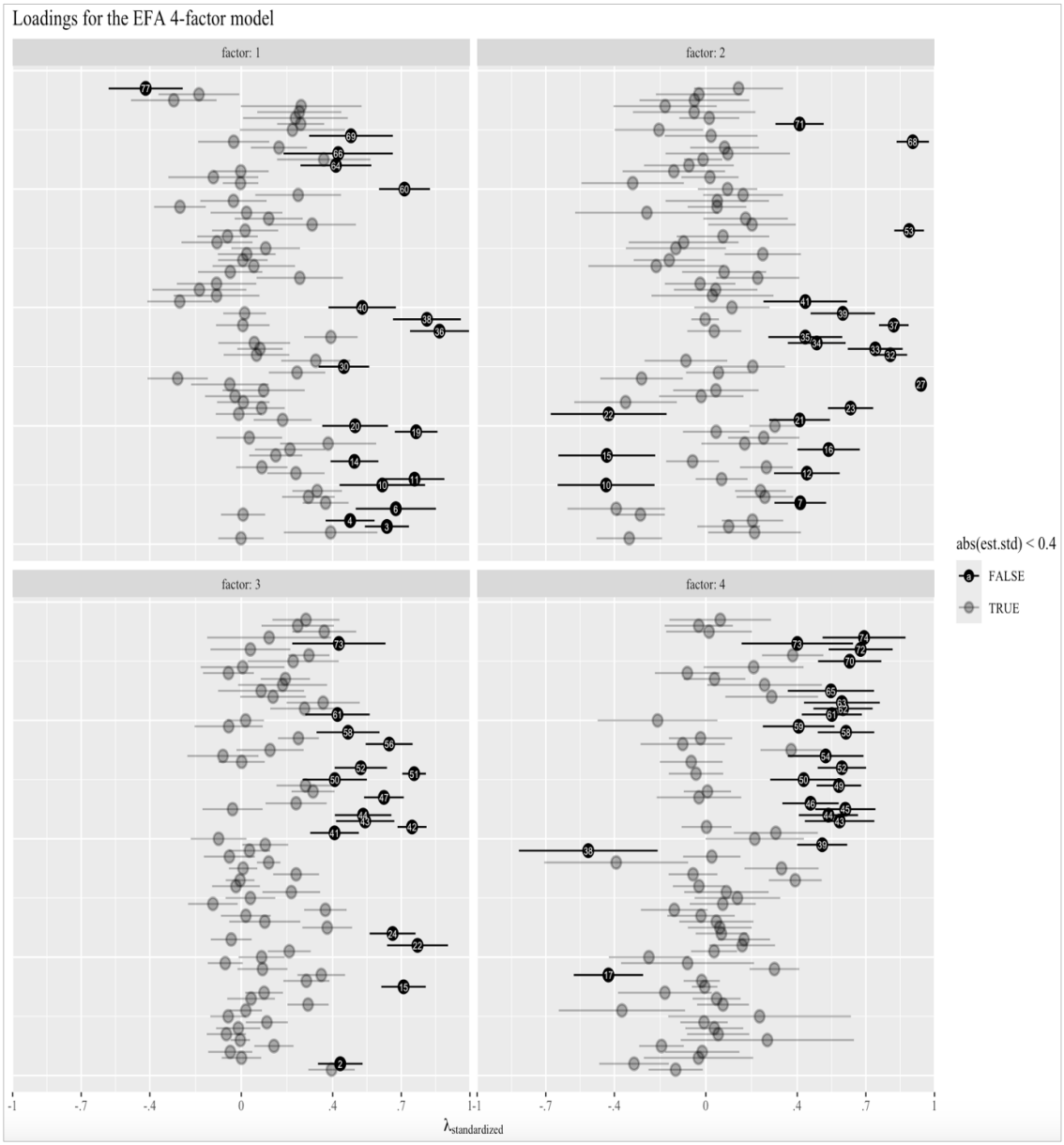


Figure 14 Exploratory factor analysis of 77 OPCRIT for Chapter 3.
This figure visualises the standardised factor loadings (lambda values) and their 90% confidence intervals (CIs) for 77 OPCRIT items derived from Exploratory Factor Analysis. Each item is represented by a circle on the x-axis according to its factor loading. Circles are color-coded to indicate loadings above (black, ≥ .4) or below (grey, <.4) a threshold of .4. The plot reveals four distinct factors, labelled as: Factor 1 - Depression, Factor 2 - Mania, Factor 3 - Adverse Chronic Trajectory (ACT), and Factor 4 - Psychosis.

(iii) **CFA**

Items were identified with a median threshold of .6 for EFA factor loadings to each of four dimensions, to ensure a parsimonious CFA model with literature-relevant items. Twenty core symptoms formed a four-factor model validated by CFA. The 4-factor CFA model using 20 clinical symptoms indicated a good fit (χ^2 = 505.88, RMSEA = .03 [90% CI .03–.04], CFI =

.99, TLI = .99; Figure 15 below). Factors, defined by highest EFA loadings, showed robust associations ($P < .05$) with all 20 symptoms. CFA generated a four-factor model with interrelated mania, psychosis, depression, and ACT symptom dimensions. Lower covariances between dimensions compared to dimension items indicated distinct structures with minimal overlap.

Model	Chi.sq	RMSEA	CI.lower	CI.upper	CFI	TLI
4-factor	505.88	.03	0.03	0.04	.99	.99

Note. The data was extracted from the 4-factor CFA model using 20 clinical symptoms.

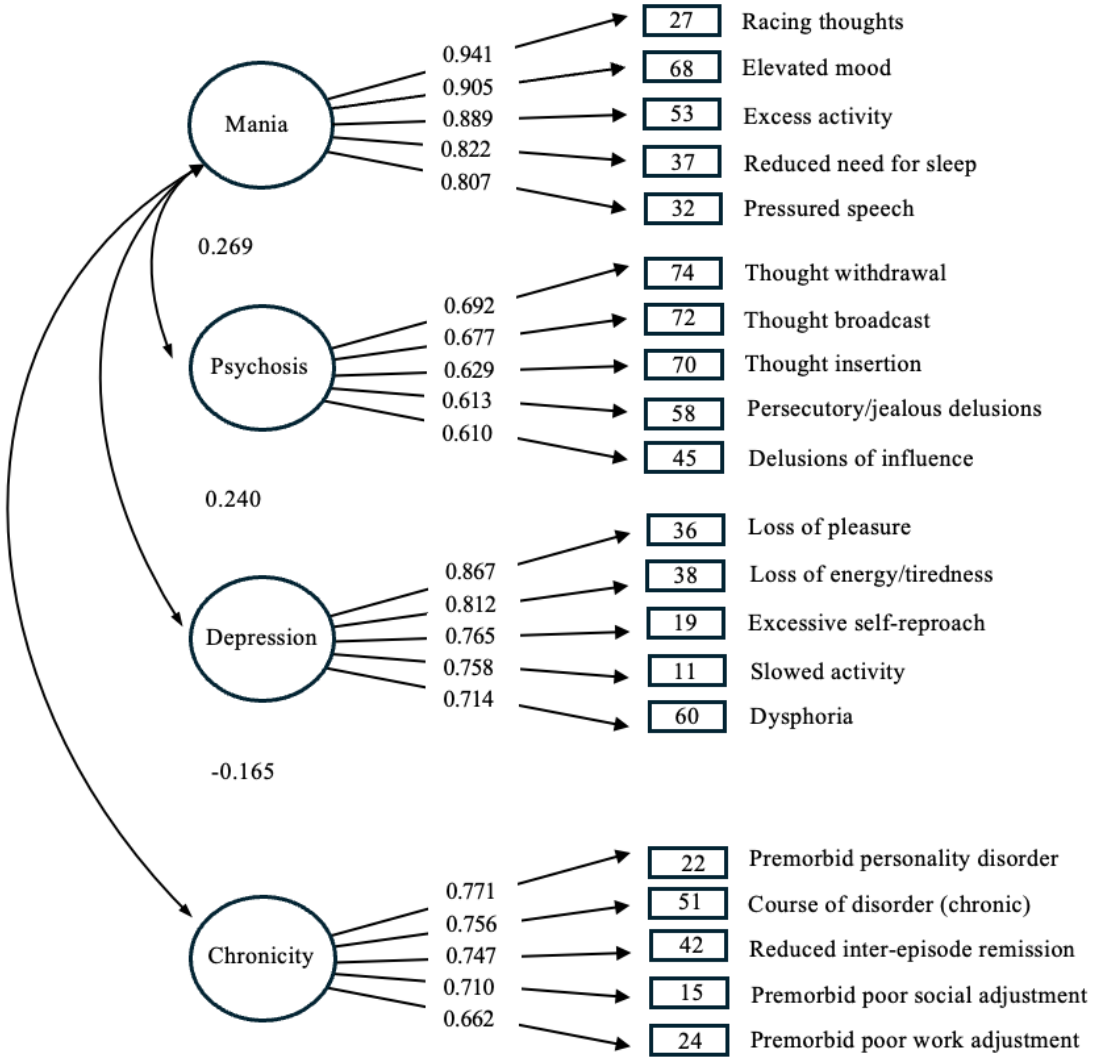


Figure 15 Confirmatory four-factor analysis (CFA) and fit indices.
This figure displays the path diagram for the four-factor Confirmatory Factor Analysis (CFA) model. The circles represent the four latent symptom dimensions: Mania, Psychosis, Depression, and Adverse Chronic Trajectory (ACT). Arrows on the the left hand side represent covariances between mania and the other dimensions.

The squares indicate the 20 core OPCRIT items that load onto these dimensions. The square boxes illustrate the factor loadings of each item onto its respective dimension (circle), while arrows on the right hand side also show the covariances between the latent dimensions. The model demonstrated good fit to the data ($\chi 2$=505.88, RMSEA = .03 [90% CI .03–.04], CFI = .99, TLI = .99).

3.5.1.2 Confirmatory factor analysis (CFA) loadings for 20 core OPCRIT items

Table 20 Confirmatory factor analysis (CFA) loadings for 20 core OPCRIT items

Predictor	Target	Item Number	Coefficient/ Loading
Mania	Racing thoughts	27	0.941
Mania	Elevated mood	68	0.905
Mania	Excess activity	53	0.889
Mania	Reduced need for sleep	37	0.822
Mania	Pressured speech	32	0.807
Psychosis	Thought withdrawal	47	0.692
Psychosis	Thought broadcast	72	0.677
Psychosis	Thought insertion	70	0.629
Psychosis	Persecutory/jealous delusions	58	0.613
Psychosis	Delusions of influence	45	0.610
Depression	Loss of pleasure	36	0.867
Depression	Loss of energy/tiredness	38	0.812
Depression	Excessive self-reproach	19	0.765
Depression	Slowed activity	11	0.758
Depression	Dysphoria	60	0.714
Chronicity	Premorbid personality disorder	22	0.771
Chronicity	Course of disorder (chronic)	51	0.756
Chronicity	Reduced inter-episode remission	42	0.747
Chronicity	Premorbid poor social adjustment	15	0.710
Chronicity	Premorbid poor work adjustment	24	0.662

This table presents the standardised factor loadings of the 20 core OPCRIT items on their respective latent dimensions (Mania, Psychosis, Depression, and Adverse Chronic Trajectory) derived from the confirmatory factor analysis. Significance levels for the adjusted Bonferroni *P*-values are also indicated to show the strength of the relationship between each item and its assigned dimension. This table supports the validity and internal consistency of the four-factor model. Factors, defined by highest EFA loadings, all showed robust associations ($P < .05$) with their respective OPRCIT items

(iv) **SEM** Multiple Indicator Multiple Cause **(MIMIC) model**

The final MIMIC model indicated distinct genetic liabilities across the four clinical dimensions (see Figure 16, Table 21 below). The statistical significance of the 20 path coefficients from the five PRS to the four latent dimensions was assessed against a Bonferroni-corrected alpha threshold of $P<0.0025$ (0.05 / 20 tests) to account for multiple testing. The mania dimension, strongest associated with BD PRS, associated positively with psychosis and depression, and inversely with ACT symptoms which correlated with worse outcomes. The PRS correlated strongest with their symptom dimensions; SCZ with psychosis, BD with mania, MDD with depression, and ADHD and anxiety with ACT. The MIMIC model fit acceptably ($\chi^2 = 348.45$, RMSEA = .04 [90% CI .04–.04], CFI = .92, TLI = 0.90) but with less reliability than the CFA, likely due to the additional complexity.

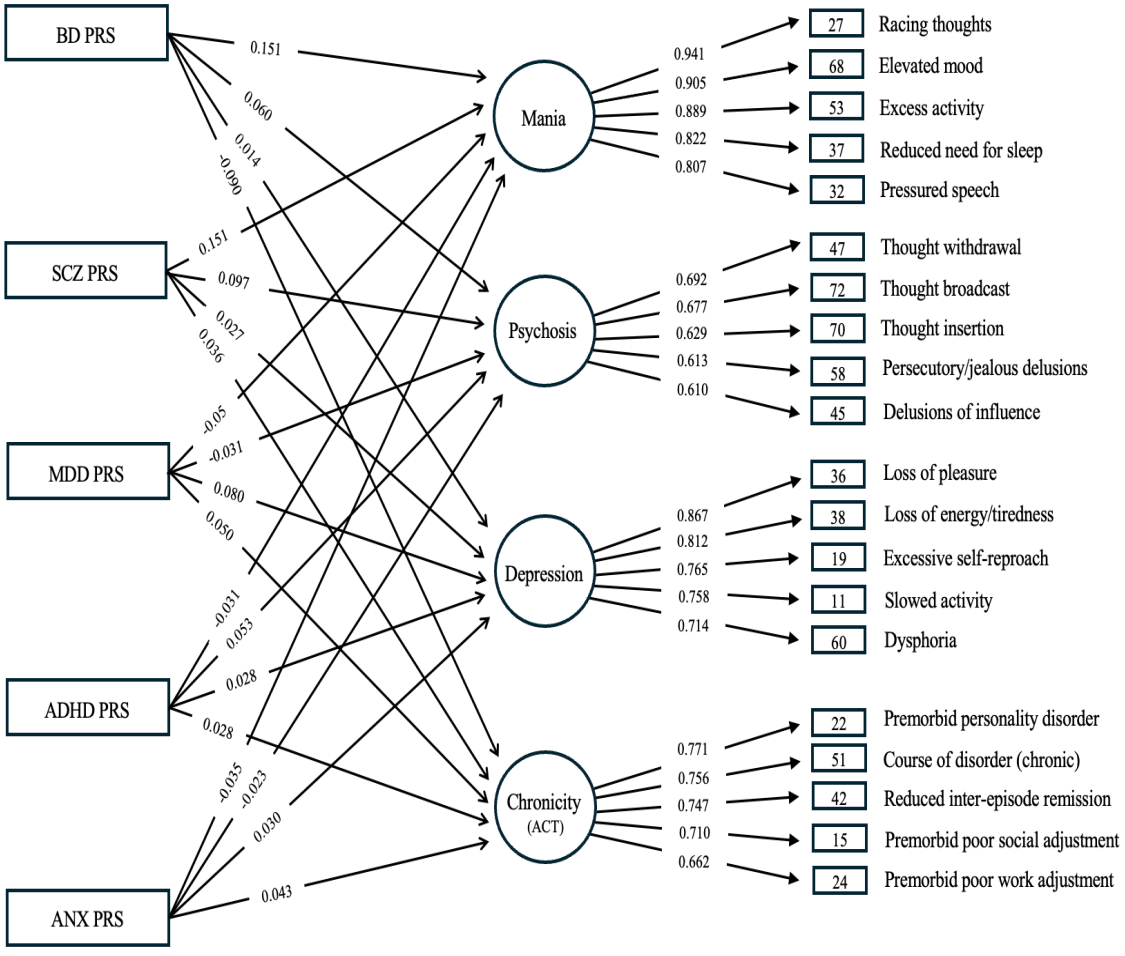


Figure 16 Structural equation (MIMIC) models (SEM) fit indices.
This figure illustrates the results of the Structural Equation Model (SEM) using the Multiple Indicator Multiple Cause (MIMIC) approach. Rectangles on the left represent the five Polygenic Risk Scores (PRSs) used as predictors: Bipolar Disorder (BD), Schizophrenia (SCZ), Major Depressive Disorder (MDD), ADHD, and Anxiety (ANX). The central circles represent the four latent symptom dimensions derived from the factor analysis: Mania, Psychosis, Depression, and the Adverse Chronic Trajectory

(ACT). Arrows originating from the PRSs indicate the path coefficients predicting each latent dimension; the specific values for these paths are detailed in Table 21. Arrows pointing from the latent dimensions to the boxes on the right represent the factor loadings on the 20 core OPCRIT items.

3.5.1.3 Estimates for SEM (MIMIC) of 20 OPCRIT items and five genetic covariates

Table 21 Estimates for SEM (MIMIC) of 20 OPCRIT items and five genetic covariates

Dimension	PRS	Estimate	Odds Ratio (OR)	Std.Error	Z-value	*P*Bonferroni	*P*Bonf.signif
Depression	ANX	0.030	1.030	0.028	1.071	0.041	*
Depression	BD	0.014	1.014	0.030	0.467	0.043	*
Depression	ADHD	0.028	1.028	0.029	0.966	0.333	ns
Depression	MDD	0.080	1.083	0.029	2.759	0.007	**
Depression	SCZ	0.027	1.027	0.028	0.964	0.0189	*
Mania	ANX	-0.035	0.966	0.027	-1.296	0.046	*
Mania	BD	0.151	1.163	0.030	5.033	$5.44x10^{-7}$	****
Mania	ADHD	-0.043	0.958	0.028	-1.536	0.013	*
Mania	MDD	-0.050	0.951	0.028	-1.786	0.0086	**
Mania	SCZ	0.054	1.055	0.028	1.929	0.0035	**
Psychosis	ANX	-0.023	0.977	0.030	-0.767	0.431	ns
Psychosis	BD	0.060	1.062	0.031	1.935	0.005	**
Psychosis	ADHD	0.053	1.054	0.030	1.767	0.08	ns
Psychosis	MDD	-0.031	0.969	0.030	-1.033	0.306	ns
Psychosis	SCZ	0.097	1.102	0.029	3.345	$3.0x10^{-5}$	****
Chronicity	ANX	0.043	1.044	0.033	1.303	0.003	**
Chronicity	BD	-0.090	0.914	0.034	-2.647	$8.0x10^{-5}$	****
Chronicity	ADHD	0.071	1.074	0.033	2.152	$3.0x10^{-4}$	***
Chronicity	MDD	0.050	1.051	0.033	1.515	0.003	**
Chronicity	SCZ	0.036	1.037	0.032	1.125	0.026	*

This table displays the results of the Structural Equation Model (SEM) using the Multiple Indicator Multiple Cause (MIMIC) approach. It shows the path coefficients indicating the strength and direction of the relationships between the five genetic covariates (PRSs for BD, SCZ, MDD, ADHD, and ANX) and both the latent symptom dimensions and the individual 20 core OPCRIT items. Significance levels for the adjusted Bonferroni *P*-values are also included. This table illustrates the distinct genetic liabilities associated with each of the identified symptom dimensions. *(Significance levels of adjusted Bonferroni *P*-value, < .0001 ****, < .001 ***, <.01 **, < .05).

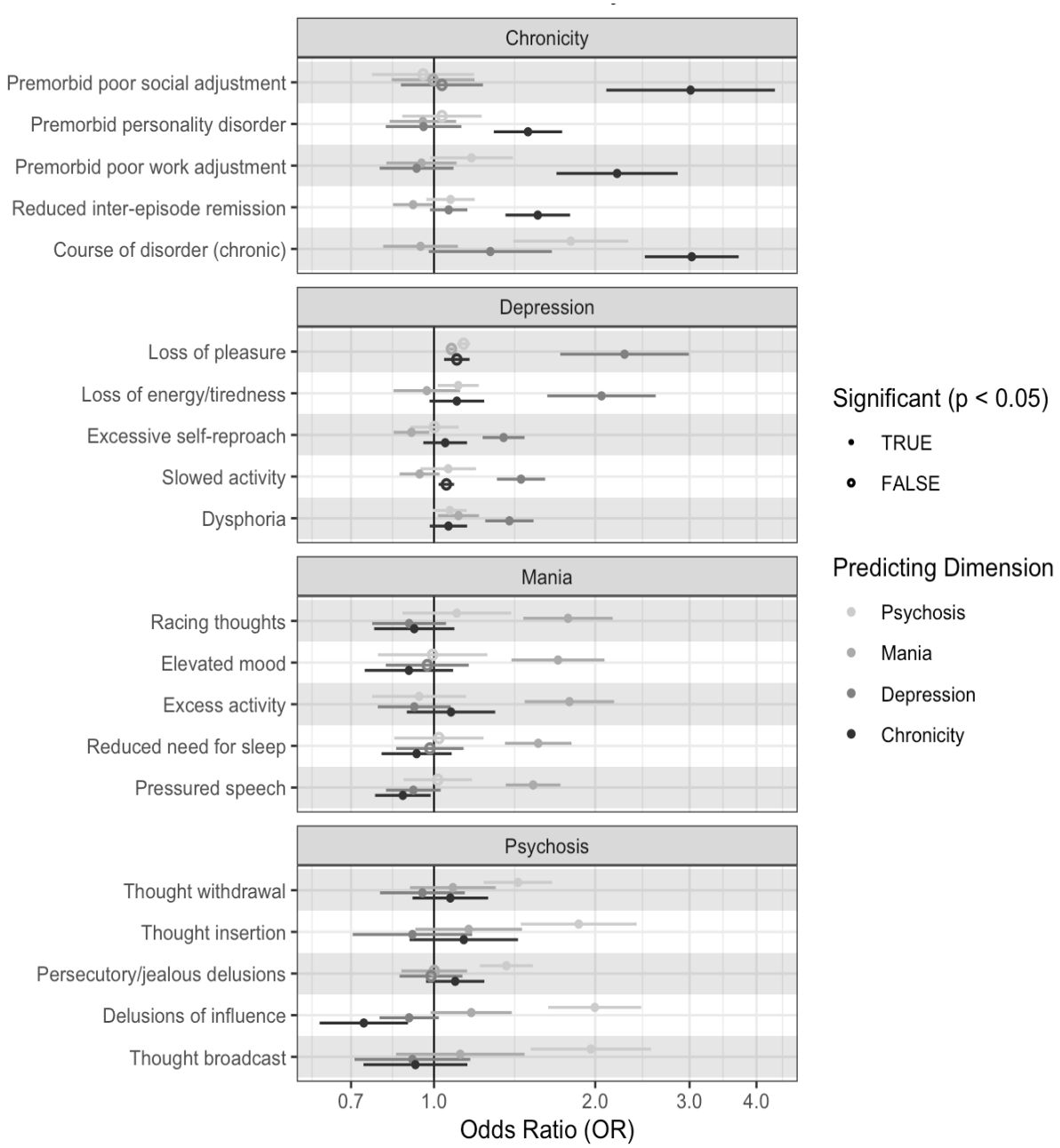


Figure 17 Core items associations using individuals' leave-one-out factor scores.
This forest plot displays the odds ratios (ORs) and 95% confidence intervals from a "leave-one-out" validation analysis. The y-axis lists the 20 core clinical symptoms. The x-axis represents the OR for reporting a symptom, comparing individuals in the top 10% of a given factor score distribution to the remaining 90%. Each point is the result of a separate logistic regression, where a symptom was predicted by the factor scores derived from a model in which that symptom was excluded. The greyscale colour of the points denotes the predicting factor dimension. A solid point indicates a statistically significant association after Bonferroni correction, while a hollow point is not significant.

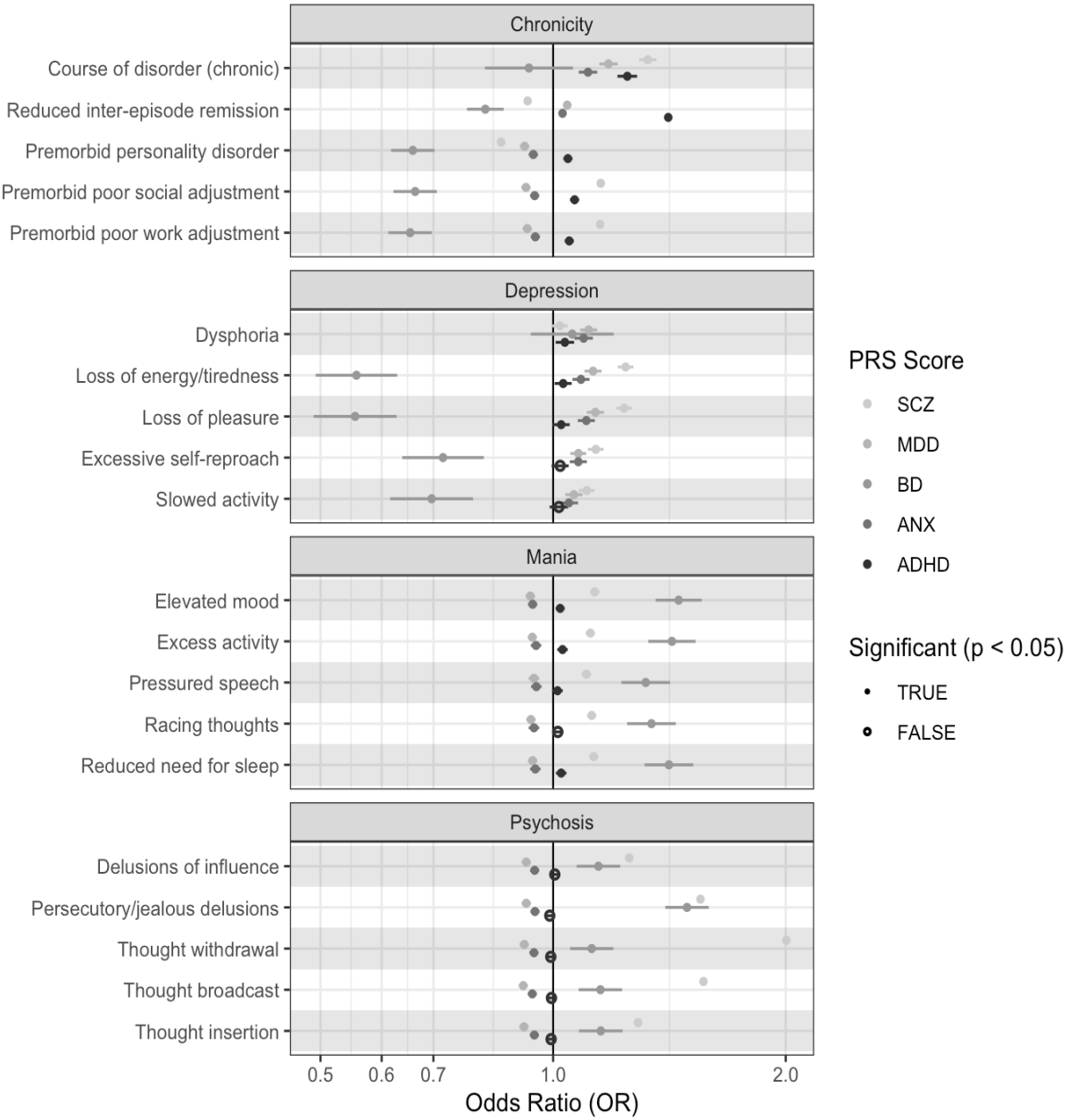


Figure 18 Core predictions using five transdiagnostic individual-level PRS scores.
This forest plot displays the odds ratios (ORs) from separate logistic regression analyses testing the direct association between transdiagnostic genetic risk and individual symptoms. The y-axis lists the 20 core clinical symptoms. The x-axis represents the OR for reporting a symptom, comparing individuals in the top 10% of a specific PRS distribution to the remaining 90%. Each point represents a single model where one symptom was predicted by one PRS (e.g., "Racing thoughts" predicted by BD PRS). The greyscale colour of the points denotes the predicting PRS. A solid point indicates a statistically significant association after Bonferroni correction, while a hollow point is not significant.

(v) Sensitivity analyses

For the sensitivity analyses predicting the 20 core clinical symptoms from the five individual PRS scores, a Bonferroni correction was also applied within each set of tests. Specifically, the significance threshold was adjusted for the 20 symptoms tested against each PRS (alpha = 0.05 / 20 = 0.0025) to control for the family-wise error rate.

Individual-level factor (dimension) scores

To analyse each of the 20 core items independent contribution to the CFA model, regression analyses was performed using individual-level factor scores. Factor scores were estimated using leave-one-out CFA analyses. The median RMSEA remained relatively stable within the full CFA models confidence intervals, indicating a robust model. The 20 items were each predicted by one of the individual-level factor scores for each dimension. Participants in the top 10% of scores were more likely to report the symptom compared to those in the lower 90%. Their factor score for the symptom-related dimension was a better predictor than scores from other dimensions. The odds ratio (OR) of reporting symptoms was increased for participants in the top 10% of factor scores. The specificity of these factor scores is illustrated in Figure 17 and Table 24. For example, when predicting the symptom 'Racing thoughts,' the Mania factor score (shown as a dark grey point) had a significantly higher odds ratio (OR > 2.5) than the scores for the Psychosis, Depression, or Chronicity dimensions (all with ORs near 1.0). This pattern, consistent across the 20 core items, confirms that each factor score is the most potent predictor of its own constituent symptoms.

Individuals PRS Scores

To ensure the five genetic contributions at the global SEM level held for each dimension item, I performed regression analyses using each item and individual-level PRS scores in turn. Participants with the top 10% compared to the lower 90% of scores for the respective dimension, were associated with a higher risk (OR) for dimension-related symptoms. Separation of global effects revealed a mixture of effect directions related to ANX and SCZ PRS for the ACT dimensions (Figure 18 and Table 25).

(vi) ***Post hoc*** **Regression Analyses**

Rapid cycling (RC), considered a chronic form of BD [42-43], positively associated with the ACT dimension in EFA and inversely with mania. RC also showed a positive association with premorbid social adjustment (OPCRIT 10) (OR 1.185, $P = 1.04 \times 10^{-10}$) and personality disorders (OR 1.391, $P = 1.288 \times 10^{-10}$). Premorbid personality disorder was associated with substance abuse (OPCRIT 80) (OR 1.160, $P = 4.98 \times 10^{-10}$) and suicidal ideation (OPCRIT 43) (OR 1.140, $P = 1.28 \times 10^{-10}$). Both personality disorder and RC were associated with a higher ADHD PRS (1.325 and OR 1.209, respectively, both $P < 5 \times 10^{-5}$) and an earlier onset of BD ($F = -3.782$, $P = 6.46 \times 10^{-6}$ and $F = -3.026$, $P = 4.821 \times 10^{-5}$, respectively).

3.6 Discussion

A Novel Four-Factor Model and the Adverse Chronic Trajectory Dimension

The current research specifically addresses an under-examined association between premorbid deficits and a chronic BD course in genetic studies. The fourth dimension, which was termed the 'Adverse Chronic Trajectory' (ACT), empirically captures a clinically crucial aspect of bipolar disorder that goes beyond acute mood symptoms. The core items loading onto this factor, poor premorbid work and social adjustment, personality disorder, and a chronic illness course with reduced inter-episode remission, may represent the long-term, cumulative burden of the illness. This dimension aligns with extensive research suggesting that a substantial portion of the disability in BD may stem not just from acute episodes, but from a persistent course characterized by enduring behavioural deficits [27].

Conceptually, the dimensions of Mania, Depression, and Psychosis are well-established clinical constructs that form the core of bipolar disorder psychopathology and have been identified in previous factor-analytic studies [21]. The current Chapter 3 model was designed to first confirm their foundational three-factor structure within the current independent dataset. Empirically, the decision to retain four factors was strongly supported by our statistical analyses as well as their supplementary materials. Both the current parallel analysis and scree plot tests clearly indicated that a four-factor solution provided the optimal fit for the data, explaining significantly more variance than a three-factor model without overfitting. The fourth dimension, termed the 'Adverse Chronic Trajectory' (ACT), emerged directly from the exploratory factor analysis as a distinct and coherent construct, and was confirmed in the confirmatory stage.

3.6.1 ACT Dimension and Long-Term Outcomes

One interpretation is that the ACT dimension represents a neurodevelopmental factor within bipolar disorder. The items loading onto this factor, poor premorbid social and occupational adjustment, personality difficulties, and a chronic course, are consistent with an illness trajectory rooted in early developmental processes. This aligns with a neurodevelopmental model where early-life abnormalities may contribute to long-term functional deficits (Chapter 1[82]). It could identify a subgroup of patients whose illness is defined not just by mood episodes, but by a persistent trajectory of functional decline rooted in cognitive and behavioural deficits. This distinction is critical, as it suggests that the genetic liabilities contributing to the ACT factor may be linked to the mechanisms that govern long-term illness progression and cognitive outcomes in bipolar disorder, rather than just the risk for acute mood states. However, a key limitation of this interpretation is that the OPCRIT checklist, while detailed, was not designed to capture the full spectrum of neurodevelopmental traits, such as those associated with Autism Spectrum Disorder (ASD). Therefore, while the ACT factor points towards a developmental trajectory, its characterization is constrained by the scope of the measurement tool used. Similarly, the personality disorder item in ACT also lacks specificity.

ACT Links To Cognitive and Behavioural Deficits

The elements of the ACT factor are strongly linked in the literature to underlying cognitive and behavioural impairments. Cognitive impairments, affecting memory, attention, and executive function, is considered a central feature of bipolar disorder. A chronic course with subsyndromal symptoms, a core feature of the ACT, is associated with these persistent cognitive deficits, which are observed even during stable, euthymic phases of the illness (Chapter 1[46]). Furthermore, a greater number of mood episodes has been longitudinally associated with a greater decline in cognitive measures, including working memory (Chapter 1[111]). These cognitive deficits may manifest behaviourally as difficulty maintaining employment, social withdrawal, and an overall failure to return to the previous level of functioning, thereby negatively impacting relationships and productivity [27].

ACT Genetic Links To ADHD and Anxiety

This model uniquely links a genetically influenced ACT dimension, connecting premorbid deficits and adverse long-term outcomes to genetic risk for ADHD and anxiety, highlighting a distinct pathway to illness severity and their contribution to social functioning, work, personality, and a less stable BD course [39]. Identifying this ACT dimension and its genetic links offers a new understanding of challenges beyond BD mood episodes, suggesting a biological basis emphasising transdiagnostic risks in BDs spectrum and variable outcomes.

The finding that a higher ADHD PRS is associated with a more adverse chronic trajectory (ACT) in bipolar disorder aligns with evidence from Agnew-Blais *et al*. (2021) [73], who demonstrated that higher ADHD genetic risk is associated with a more persistent course of ADHD into young adulthood. Supporting this, Duffy (2012) [74] also suggests that childhood ADHD may be linked to a subtype of BD with a more severe course and poorer treatment response [48].

Parental BD elevates child ADHD risk [75] and early chronic challenges. These factors and inherited genetic predisposition may heighten suicidality risk [76].

The ACT dimension and its genetic links provide a new framework for understanding diverse BD clinical presentations. This highlights the need for integrated assessment and treatment, particularly when addressing co-occurring ADHD and anxiety to improve long-term chronic outcomes. For BD individuals with chronic/cognitive deficits, clinicians could tailor integrated treatment plans for optimal outcomes [77].

Dimensional Assessment and Early Intervention Implications

These findings also support a dimensional assessment in BD. Evaluating an individuals chronic trajectory and genetic risk for associated conditions could inform more comprehensive, personalised treatment plans, suggesting earlier identification of individuals predisposed to a more challenging BD course. This could enable preventative or early intervention strategies focused on bolstering cognitive and functional deficits [28, 78]. The strong genetic associations

with the ACT dimension, particularly with ADHD and anxiety [74], further underscore the potential for early intervention, as these often present in childhood and adolescence.

Four versus a Three-Factor Model

Building on dimensional approaches, this study increased the number of OPCRIT measures included than prior studies, yielding a fourth ACT dimension, validated in both this and one other study (eResults 4) [21]. Here, three clinical dimensions (mania, depression, psychosis) and their genetic associations replicated their findings. Importantly, in the current analysis, the additional measures and PRSs loaded strongest to the novel ACT dimension, thus only a four-factor model could account for the genetic signatures in BD course specifiers. Additionally, the inclusion here of the propensity scores, likely provided more accurate assessment, adjusted for potentially inflated effect sizes commonly reported when analysing Electronic Health Records (EHR) data.

Predictive Utility of the PRS

Sensitivity analyses confirmed symptom strength independent of global dimensions. Factor and PRS scores better predicted risk for the dimension symptoms within than across dimensions in unseen data. PRS provided incremental predictive value to clinical data, with a median positive predictive value (PPV) at a .8 clinical utility threshold [79].

Factor Loading Thresholds

A .6 factor loading threshold for OPCRIT items was used to maintain clinical relevance and parsimony, though prior analyses have used lower thresholds [12, 21, 78, 70]. Here, EFA robustness at .4 suggests future studies could use a lower threshold. While model fit and sensitivity were adequate, more items do not guarantee better accuracy and risk overfitting, reducing generalisability [70].

Genetics of the ACT Dimension

The novel ACT dimension showed distinct genetic signatures. Higher BD burden indicated resilience against premorbid deficits and chronic illness progression, predicting higher functioning in an independent BD dataset [23] and an inverse relationship with rapid cycling [19]. Similarly, the mania dimension positively associated with BD PRS was inversely related to the ACT dimension.

Genetics of the ACT Dimension Symptoms

Symptoms associated with BD and SCZ PRS linked to higher inter-episode remission, in contrast to MDD, ANX, and especially ADHD, which positively associated with reduced inter-episode remission. Higher BD PRS predicted inter-episode remission and reduced anxiety in an independent BD dataset [23]. Depression, anxiety, and cognitive issues are often early BD symptoms [5, 80]. Rapid cycling (RC) correlated with higher ANX or ADHD PRS but

inversely with BD in a prior study [19]. Co-occurring ADHD and anxiety elevate the risk for BD onset [18, 81-82], suggesting a less favourable trajectory.

Severe incapacity (OPCRIT 87) linked primarily to mania and psychotic features, less to ACT symptoms, and least to depressive symptoms. BD PRS correlated with increased symptom severity and lower depression polygenic burden in multiplex BD families [83]. Here, strongest associations existed between premorbid occupational (OPCRIT 9) and social adjustment (OCPRIT 10), and ADHD or SCZ PRS. The observed negative association between ACT (including premorbid adjustment) and mania aligns with the inverse relationship found by Allardyce *et al*. (2007) [12]. Novel in the current study however, is the inverse association between genetic liability to BD PRS and ACT, capturing the association of illness chronicity and personality within the ACT dimension.

Longitudinal data suggests enduring chronic and cognitive deficits in BD [5]. ACT impairments affect (30-60)% of adults with BD [28], especially with comorbid anxiety and ADHD [34]. Sensitivity analysis showed ADHD PRS consistently positive with the ACT dimension symptoms, while ANX and SCZ PRS effects were more complex across indicators, suggesting nuanced relationships needing further granular investigation.

The higher BD1 proportion of cases here, linked to lower anxiety, might have limited ANX PRS and ACT dimension item-level associations. Prior factor analyses found the largest BD subgroup to be characterised by affective stability with low anxiety and low risk for ADHD-like behaviours, supporting this chapter's genetic findings [15].

ADHD PRS uniquely correlated here with a higher risk for premorbid personality disorders (OPCRIT 11) and other ACT dimension symptoms. ADHD and BD comorbidity increases the risk for personality disorder and more frequent episodes, leading to poorer functioning. Childhood ADHD is associated with higher borderline personality disorder (BPD) risk [84].

Future Studies

Future research should focus on validating the four-factor models reproducibility across independent ancestral datasets, ideally utilising the same OPCRIT items to ensure comparability. Furthermore, the collection and analysis of longitudinal data will be essential for further understanding the temporal dynamics between genetic risk, the emergence of premorbid factors, and the subsequent longitudinal course of bipolar disorder. By tracking individuals over extended periods, future studies can help to establish the precise temporal order of these events and to identify potential causal pathways. Longitudinal data incorporating detailed symptom scales could also be invaluable in identifying specific temporal links and triggers for mood episodes, especially targeting those individuals at elevated risk of suicidality.

While an individuals underlying genetic code remains relatively stable throughout their lifespan, environmental factors can influence how these genes are expressed (through epigenetic mechanisms) and interact with one another to either increase or decrease the

likelihood of developing bipolar disorder. Investigating the specific mechanisms through which ADHD and anxiety might trigger or exacerbate mood episodes in BD, especially in rapid cycling, could be a logical next step, potentially involving neuroimaging or neurochemical studies to explore underlying brain circuitry.

Clinical Practice

This study suggests early identification of chronic difficulties in individuals with higher genetic burden for ADHD and anxiety offers a crucial opportunity for interventions to improve long-term BD outcomes. The findings underscore the potential utility of incorporating comprehensive assessments for premorbid functioning and any co-occurring symptoms of ADHD and anxiety in individuals with or at risk for BD. This more holistic approach could facilitate the earlier identification of those individuals who may be on a more adverse chronic trajectory, allowing for the implementation of proactive and personalised interventions that may ultimately improve the overall course of their illness and their quality of life.

3.7 Limitations

The sample, while large, primarily comprised individuals recruited through clinical settings, potentially overrepresenting those with more severe or chronic forms of BD who are more likely to seek and remain in treatment. While Inverse Probability Weighting (IPW) was applied to mitigate ascertainment, bias related to hospitalisation and symptom severity, the generalisability of the findings to community-based populations or individuals with milder presentations of BD warrants further investigation. The decision to exclude OPCRIT items with low frequency (less than 8% missingness) could potentially limit the generalisability of these findings to individuals presenting with rarer symptoms. The cross-sectional nature of the data limits the ability to infer the temporal relationships between genetic risk, premorbid factors, and the longitudinal course of BD. It is important to note that current PRSs for complex psychiatric disorders, including BD, explain a modest proportion of the overall variance in these conditions, and the findings, while informative at a group level, reflect trends rather than definitive individual-level predictions [40]. Further research efforts, including larger genome-wide association studies and the inclusion of more diverse ancestral populations, are needed to enhance the predictive power of PRSs for clinical applications.

3.8 Conclusions

M analysis indicates a broader transdiagnostic genetic signature, beyond traditional mood disorders, contributes to a more adverse BD trajectory, potentially worsening long-term outcomes due to chronic and cognitive deficits, notably linked to higher ADHD and anxiety polygenic burden. The MIMIC model revealed a complex interplay between mania and the novel ACT dimension. While ADHD PRS showed a consistent positive association with ACT, ANX and SCZ PRS effects on ACT items were more nuanced, requiring further research.

These findings underscore the importance of considering transdiagnostic genetic risks in understanding BD heterogeneity, linked to predicting its trajectory.

3.9 Supplementary Materials

Exploratory Factor Analysis (EFA)

Table 22 Exploratory factor analysis (EFA) loadings of 77 OPCRIT items

Note: In Exploratory Factor Analysis, communalities represent the proportion of each symptom's variance that the shared factors can explain. Unique variance is the proportion that is not explained by the factors and is unique to the symptom itself. These 77 items were selected for analysis due to adequate sample size and less than 8% missingness.

Item No.	Item Description	Depres (f1)	Mania (f2)	Chronicity (f3)	Psychosis (f4)	Unique.var	Communalities
1	Rapid cycling	-0.001	-0.335	0.395	-0.132	0.817	0.183
2	Weight loss	0.391	0.213	0.433	-0.314	0.435	0.565
3	Diminished libido	0.637	0.1	0.001	-0.032	0.530	0.470
4	Diurnal variation	0.476	0.204	-0.048	-0.016	0.647	0.353
5	Mode of onset	0.008	-0.286	0.143	-0.194	0.868	0.132
6	Weight gain	0.676	-0.392	-0.004	0.269	0.263	0.737
7	Early morning waking	0.369	0.413	-0.065	0.054	0.551	0.449
8	Middle insomnia	0.294	0.258	-0.013	0.037	0.775	0.225
9	Initial insomnia	0.332	0.239	0.112	-0.008	0.742	0.258
10	Increased appetite	0.617	-0.436	-0.057	0.235	0.224	0.776
11	Slowed activity	0.758	0.069	0.02	-0.367	0.300	0.700
12	Agitated activity	0.239	0.442	0.292	0.075	0.532	0.468
13	Stressor prior to onset	0.09	0.265	0.043	0.047	0.893	0.107
14	Excessive sleep	0.495	-0.058	0.1	-0.179	0.752	0.248
15	Poor premorbid social adjustment	0.151	-0.433	0.71	-0.005	0.616	0.384
16	Distractibility	0.213	0.537	0.285	-0.018	0.452	0.548
17	Poor appetite	0.38	0.17	0.35	-0.426	0.477	0.523
18	Inappropriate affect	0.035	0.253	0.093	0.3	0.822	0.178
19	Excessive self reproach	0.765	0.045	-0.07	-0.08	0.389	0.611
20	Poor concentration	0.498	0.302	0.089	-0.249	0.475	0.525
21	Irritable mood	0.181	0.41	0.21	0.036	0.668	0.332
22	Premorbid personality disorder	-0.011	-0.425	0.771	0.159	0.504	0.496
23	Increased sociability	0.089	0.633	-0.044	0.167	0.512	0.488
24	Poor premorbid work adjustment	0.009	-0.351	0.662	0.068	0.590	0.410
25	Alcohol/drug abuse within one year of onset	-0.027	-0.02	0.376	0.061	0.857	0.143
26	Family history of schizophrenia	0.098	0.044	0.103	0.046	0.969	0.031
27	Thoughts racing	-0.05	0.941	0.02	-0.021	0.150	0.850
28	Unemployed	-0.278	-0.281	0.368	-0.138	0.635	0.365
29	Family history of other psychiatric disorder	0.244	0.055	-0.124	0.074	0.904	0.096
30	Restricted affect	0.449	0.205	0.04	0.138	0.635	0.365
31	Blunted affect	0.326	-0.087	0.219	0.09	0.846	0.154
32	Pressured speech	0.067	0.807	-0.023	-0.03	0.298	0.702
33	Increased self esteem	0.082	0.741	-0.005	0.391	0.298	0.702
34	Reckless activity	0.057	0.485	0.240	-0.056	0.657	0.343
35	Relationship psychotic/affective symptoms	0.392	0.435	0.008	0.331	0.364	0.636
36	Loss of pleasure	0.867	0.037	0.12	-0.392	0.143	0.857
37	Reduced need for sleep	0.007	0.822	-0.052	0.026	0.321	0.679
38	Loss of energy/tiredness	0.812	-0.003	0.036	-0.514	0.169	0.831
39	Grandiose Delusions	0.015	0.599	0.105	0.509	0.351	0.649

Item No.	Item Description	Depres (f1)	Mania (f2)	Chronicity (f3)	Psychosis (f4)	Unique.var	Communalities
40	Negative formal thought disorder	0.529	0.114	-0.099	0.214	0.577	0.423
41	Positive formal thought disorder	-0.269	0.435	0.408	0.306	0.587	0.413
42	Reduced inter-episode remission	-0.108	0.029	0.747	0.003	0.414	0.586
43	Widespread Delusions	-0.183	0.043	0.542	0.585	0.373	0.627
44	Well organised delusions	-0.108	-0.025	0.533	0.536	0.440	0.560
45	Delusions of influence	0.256	0.227	-0.038	0.610	0.435	0.565
46	Other (non affective) auditory hallucinations	-0.048	0.08	0.24	0.458	0.732	0.268
47	Life time diagnosis of cannabis abuse/depend	0.055	-0.217	0.624	-0.03	0.398	0.602
48	Single (v married)	0.007	-0.16	0.314	0.007	0.884	0.116
49	Persecutory Delusions	0.024	0.249	0.281	0.582	0.502	0.498
50	Abusive/accusatory/persecutory voices	0.107	-0.131	0.409	0.428	0.631	0.369
51	Course of disorder (chronic)	-0.106	-0.097	0.756	-0.043	0.553	0.447
52	Delusions & hallucinations last for one week	-0.06	0.075	0.523	0.595	0.378	0.622
53	Excessive activity	0.017	0.889	0.002	-0.064	0.191	0.809
54	Primary delusional perception	0.31	0.202	-0.08	0.524	0.492	0.508
55	Non-affective hallucination in any modality	0.12	0.174	0.126	0.373	0.767	0.233
56	Life time diagnosis of other abuse/depend	0.023	-0.258	0.647	-0.101	0.397	0.603
58	Persecutory/jealous delusions & hallucinations	-0.034	0.05	0.467	0.613	0.412	0.588
59	Other primary delusions	0.249	0.163	-0.055	0.406	0.685	0.315
60	Dysphoria	0.714	0.095	0.019	-0.211	0.408	0.592
61	Third person auditory hallucinations	-0.003	-0.320	0.421	0.551	0.444	0.556
62	Bizarre Delusions	-0.122	0.018	0.277	0.600	0.572	0.428
63	Running commentary voices	-0.002	-0.14	0.358	0.595	0.511	0.489
64	Delusions of guilt	0.414	-0.074	0.139	0.288	0.718	0.282
65	Delusions of passivity	0.361	-0.012	0.087	0.547	0.520	0.480
66	Nihilistic Delusions	0.474	0.096	0.181	0.257	0.643	0.357
67	Life time diagnosis of alcohol abuse/depend	0.165	0.082	0.193	0.038	0.909	0.091
68	Elevated mood	-0.033	0.905	-0.056	-0.081	0.205	0.795
69	Delusions of poverty	0.480	0.023	0.006	0.209	0.692	0.308
70	Thought insertion	0.224	-0.205	0.227	0.629	0.476	0.524
71	Impairment/incapacity during disorder	0.26	0.41	0.296	0.38	0.668	0.332
72	Thought broadcast	0.237	0.015	0.04	0.677	0.444	0.556
73	Thought echo	0.254	-0.051	0.427	0.400	0.577	0.423
74	Thought withdrawal	0.262	-0.178	0.122	0.692	0.409	0.591
75	Rapport difficult	-0.295	-0.05	0.363	0.014	0.779	0.221
76	Information not credible	-0.185	-0.03	0.247	-0.031	0.902	0.098

Item No.	Item Description	Depres (f1)	Mania (f2)	Chronicity (f3)	Psychosis (f4)	Unique.var	Communalities
77	Lack of insight	-0.418	0.143	0.283	0.063	0.786	0.214

3.8.1.1 Item Definitions

Table 23 OPCRIT Items for Adverse Chronic Trajectory Dimension

Item No.	Description
9 'Chronicity (1)'	Poor work adjustment: Refers to work history before onset of illness. It should be scored if the patient was unable to keep any job for more than 6 months, had a history of frequent changes of job or was only able to sustain a job well below that expected by his educational level or training at time of first psychiatric contact. Also, score positively for a persistently very poor standard of housework (housewives) and badly failing to keep up with studies (students). (0, 1)
10 'Chronicity (2)'	Poor premorbid social adjustment: Patient found difficulty entering or maintaining normal social relationships, showed persistent social isolation, withdrawal or maintained solitary interests prior to onset of psychotic symptoms. (0, 1)
11 'Chronicity (3)'	Premorbid personality disorder: Evidence of inadequate/ schizoid/ schizotypal/ paranoid/ cyclothymic/ psychopathic/ sociopathic personality disorder present since adolescence and prior to the onset of psychotic symptoms.(0, 1)
87 'BD outcome (4) Symptom severity'	Impairment/incapacity during disorder: 0 = No impairment 1 = Subjective impairment at work, school, or in social functioning 2 = Impairment in major life role with definite reduction in productivity and/or criticism has been received 3 = No function at all in major life role for more than 2 days or inpatient treatment has been required or active psychotic symptoms such as delusions or hallucinations have occurred
88 'BD outcome (5) Inter-episode remission'	Deterioration from premorbid level of functioning: Patient does not regain his premorbid social, occupational or emotional functioning after an acute episode of illness. (0, 1)
90 'BD outcome (6) Illness recovery to chronic course'	Course of disorder: 1 = Single episode with good recovery 2 = Multiple episodes with good recovery between 3 = Multiple episodes with partial recovery between 4 = Continuous chronic illness 5 = Continuous chronic illness with deterioration (nb score this item in hierarchical fashion, e.g. if patient's course in past rated '2',but for the time-period now being considered it rates '4', then the correct rating is '4'.) (1, 2, 3, 4, 5,)

Item No.	Description
Note:	19. OPCRIT (version 4) (Chapter 2 [4]) includes 90 items of psychopathology, premorbid functioning, personal and family history information. Expansion of OPCRIT necessitated an increase in the number of items comprising the OPCRIT checklist beyond the original 74-item checklist. Lifetime occurrence was assessed for each patient. Inter-rater reliability was formally assessed using 20 randomly selected cases (mean κ Statistic = .85).

3.8.1.2 Confirmatory Factor Analysis (EFA)

Table 24 Coefficients of 20 core OPCRIT items with four individual factor scores

Dimension	OPCRIT Item	PRS Score	Estimate	Odds Ratio (OR)	Std.Error	*P*Bonferroni	*P*Bonf.signif
Depression	Dysphoria	ADHD	0.035	1.036	0.014	0.010	**
Depression	Dysphoria	ANX	0.091	1.095	0.014	<0.0001	****
Depression	Dysphoria	BD	0.057	1.059	0.063	<0.0001	****
Depression	Dysphoria	MDD	0.106	1.112	0.013	<0.0001	****
Depression	Dysphoria	SCZ	0.020	1.020	0.012	<0.0001	****
Depression	Loss of energy/tiredness	ADHD	0.030	1.030	0.013	0.027	****
Depression	Loss of energy/tiredness	ANX	0.083	1.087	0.013	<0.0001	****
Depression	Loss of energy/tiredness	BD	-0.586	0.557	0.062	<0.0001	****
Depression	Loss of energy/tiredness	MDD	0.119	1.126	0.013	<0.0001	****
Depression	Loss of energy/tiredness	SCZ	0.216	1.241	0.012	<0.0001	****
Depression	Loss of pleasure	ADHD	0.024	1.024	0.013	0.008	**
Depression	Loss of pleasure	ANX	0.099	1.104	0.013	<0.0001	****
Depression	Loss of pleasure	BD	-0.590	0.554	0.063	<0.0001	****
Depression	Loss of pleasure	MDD	0.126	1.134	0.013	<0.0001	****
Depression	Loss of pleasure	SCZ	0.212	1.236	0.012	<0.0001	****
Depression	Self-reproach	ADHD	0.021	1.021	0.013	0.128	ns
Depression	Self-reproach	ANX	0.075	1.078	0.013	<0.0001	****
Depression	Self-reproach	BD	-0.328	0.720	0.062	<0.0001	****
Depression	Self-reproach	MDD	0.075	1.078	0.012	<0.0001	****
Depression	Self-reproach	SCZ	0.127	1.135	0.012	<0.0001	****
Depression	Slowed activity	ADHD	0.017	1.017	0.014	0.208	ns
Depression	Slowed activity	ANX	0.047	1.048	0.014	0.001	***
Depression	Slowed activity	BD	-0.362	0.696	0.063	<0.0001	****
Depression	Slowed activity	MDD	0.062	1.064	0.013	<0.0001	****
Depression	Slowed activity	SCZ	0.100	1.105	0.012	<0.0001	****
Mania	Elevated mood	ADHD	0.021	1.021	0.007	0.007	**
Mania	Elevated mood	ANX	-0.061	0.941	0.007	<0.0001	****
Mania	Elevated mood	BD	0.374	1.454	0.035	<0.0001	****

Dimension	OPCRIT Item	PRS Score	Estimate	Odds Ratio (OR)	Std.Error	*P*Bonferroni	*P*Bonf.signif
Mania	Elevated mood	MDD	-0.068	0.934	0.007	<0.0001	****
Mania	Elevated mood	SCZ	0.124	1.132	0.007	<0.0001	****
Mania	Excess activity	ADHD	0.028	1.028	0.008	<0.0001	****
Mania	Excess activity	ANX	-0.051	0.950	0.008	<0.0001	****
Mania	Excess activity	BD	0.354	1.425	0.036	<0.0001	****
Mania	Excess activity	MDD	-0.062	0.940	0.007	<0.0001	****
Mania	Excess activity	SCZ	0.111	1.117	0.007	<0.0001	****
Mania	Pressured speech	ADHD	0.013	1.013	0.008	0.001	***
Mania	Pressured speech	ANX	-0.050	0.951	0.008	<0.0001	****
Mania	Pressured speech	BD	0.276	1.318	0.037	<0.0001	****
Mania	Pressured speech	MDD	-0.057	0.945	0.008	<0.0001	****
Mania	Pressured speech	SCZ	0.099	1.104	0.007	<0.0001	****
Mania	Racing thoughts	ADHD	0.014	1.014	0.008	0.089	ns
Mania	Racing thoughts	ANX	-0.057	0.945	0.008	<0.0001	****
Mania	Racing thoughts	BD	0.293	1.340	0.037	<0.0001	****
Mania	Racing thoughts	MDD	-0.066	0.936	0.007	<0.0001	****
Mania	Racing thoughts	SCZ	0.115	1.122	0.007	<0.0001	****
Mania	Reduced need for sleep	ADHD	0.024	1.024	0.008	0.003	**
Mania	Reduced need for sleep	ANX	-0.053	0.948	0.008	<0.0001	****
Mania	Reduced need for sleep	BD	0.345	1.412	0.037	<0.0001	****
Mania	Reduced need for sleep	MDD	-0.061	0.941	0.007	<0.0001	****
Mania	Reduced need for sleep	SCZ	0.121	1.129	0.007	<0.0001	****
Psychosis	Delusions of influence	ADHD	0.005	1.005	0.007	0.505	ns
Psychosis	Delusions of influence	ANX	-0.055	0.946	0.007	<0.0001	****
Psychosis	Delusions of influence	BD	0.135	1.145	0.033	<0.0001	****
Psychosis	Delusions of influence	MDD	-0.080	0.923	0.007	<0.0001	****
Psychosis	Delusions of influence	SCZ	0.227	1.255	0.006	<0.0001	****
Psychosis	Persecutory/jealous delusions	ADHD	-0.010	0.990	0.007	0.189	ns
Psychosis	Persecutory/jealous delusions	ANX	-0.054	0.947	0.007	<0.0001	****
Psychosis	Persecutory/jealous delusions	BD	0.399	1.490	0.033	<0.0001	****
Psychosis	Persecutory/jealous delusions	MDD	-0.080	0.923	0.007	<0.0001	****
Psychosis	Persecutory/jealous delusions	SCZ	0.439	1.551	0.006	<0.0001	****
Psychosis	Thought withdrawal	ADHD	-0.007	0.993	0.007	0.320	ns
Psychosis	Thought withdrawal	ANX	-0.057	0.945	0.007	<0.0001	****
Psychosis	Thought withdrawal	BD	0.115	1.122	0.033	<0.0001	****
Psychosis	Thought withdrawal	MDD	-0.086	0.918	0.007	<0.0001	****
Psychosis	Thought withdrawal	SCZ	0.695	2.004	0.006	<0.0001	****
Psychosis	Thought broadcast	ADHD	-0.005	0.995	0.007	0.530	ns
Psychosis	Thought broadcast	ANX	-0.062	0.940	0.007	<0.0001	****
Psychosis	Thought broadcast	BD	0.141	1.151	0.033	<0.0001	****
Psychosis	Thought broadcast	MDD	-0.089	0.915	0.007	<0.0001	****
Psychosis	Thought broadcast	SCZ	0.448	1.565	0.006	<0.0001	****

Dimension	OPCRIT Item	PRS Score	Estimate	Odds Ratio (OR)	Std.Error	*P*Bonferroni	*P*Bonf.signif
Psychosis	Thought insertion	ADHD	-0.006	0.994	0.007	0.429	ns
Psychosis	Thought insertion	ANX	-0.056	0.946	0.007	<0.0001	****
Psychosis	Thought insertion	BD	0.142	1.153	0.033	<0.0001	****
Psychosis	Thought insertion	MDD	-0.087	0.917	0.007	<0.0001	****
Psychosis	Thought insertion	SCZ	0.253	1.288	0.006	<0.0001	****
Chronicity	Course of disorder (chronic)	ADHD	0.221	1.247	0.015	0.005	**
Chronicity	Course of disorder (chronic)	ANX	0.104	1.110	0.014	<0.0001	****
Chronicity	Course of disorder (chronic)	BD	-0.072	0.931	0.067	<0.0001	****
Chronicity	Course of disorder (chronic)	MDD	0.165	1.180	0.014	<0.0001	****
Chronicity	Course of disorder (chronic)	SCZ	0.282	1.326	0.013	<0.0001	****
Chronicity	Reduced inter-episode remission	ADHD	0.343	1.409	0.006	0.007	**
Chronicity	Reduced inter-episode remission	ANX	0.028	1.028	0.006	<0.0001	****
Chronicity	Reduced inter-episode remission	BD	-0.202	0.817	0.028	<0.0001	****
Chronicity	Reduced inter-episode remission	MDD	0.042	1.043	0.006	<0.0001	****
Chronicity	Reduced inter-episode remission	SCZ	-0.076	0.927	0.005	<0.0001	****
Chronicity	Premorbid personality disorder	ADHD	0.044	1.045	0.007	0.006	**
Chronicity	Premorbid personality disorder	ANX	-0.059	0.943	0.007	<0.0001	****
Chronicity	Premorbid personality disorder	BD	-0.418	0.658	0.033	<0.0001	****
Chronicity	Premorbid personality disorder	MDD	-0.085	0.919	0.007	<0.0001	****
Chronicity	Premorbid personality disorder	SCZ	-0.155	0.856	0.006	<0.0001	****
Chronicity	Premorbid poor social adjustment	ADHD	0.064	1.066	0.007	0.004	**
Chronicity	Premorbid poor social adjustment	ANX	-0.055	0.946	0.007	<0.0001	****
Chronicity	Premorbid poor social adjustment	BD	-0.411	0.663	0.033	<0.0001	****
Chronicity	Premorbid poor social adjustment	MDD	-0.081	0.922	0.007	<0.0001	****
Chronicity	Premorbid poor social adjustment	SCZ	0.142	1.153	0.006	<0.0001	****
Chronicity	Premorbid poor work adjustment	ADHD	0.048	1.049	0.007	0.007	**
Chronicity	Premorbid poor work adjustment	ANX	-0.053	0.948	0.007	<0.0001	****
Chronicity	Premorbid poor work adjustment	BD	-0.426	0.653	0.033	<0.0001	****
Chronicity	Premorbid poor work adjustment	MDD	-0.077	0.926	0.007	<0.0001	****
Chronicity	Premorbid poor work adjustment	SCZ	0.140	1.150	0.006	<0.0001	****

This table presents the coefficients and their significance levels from the regression analyses where each of the 20 core OPCRIT items was predicted by the individual factor scores for the four latent dimensions (Mania, Psychosis, Depression, and Adverse Chronic Trajectory) in a 'leave-one-out' cross-validation approach. These results demonstrate the predictive ability of the factor scores for their respective symptoms in each of the four dimensions. *(Significance levels of adjusted Bonferroni *P*-value, < .0001 ****, < .001 ***, <.01 **, < .05 *).*

Table 25 Coefficients of 20 core OPCRIT items with five individual PRS scores

Dimension	OPCRIT Item	PRS Score	Estimate	Odds Ratio (OR)	Std.Error	*P*Bonferroni	*P*Bonf.signif
Depression	Dysphoria	ADHD	0.035	1.036	0.014	0.010	**
Depression	Dysphoria	ANX	0.091	1.095	0.014	<0.0001	****
Depression	Dysphoria	BD	0.057	1.059	0.063	<0.0001	****
Depression	Dysphoria	MDD	0.106	1.112	0.013	<0.0001	****
Depression	Dysphoria	SCZ	0.020	1.020	0.012	<0.0001	****
Depression	Loss of energy/tiredness	ADHD	0.030	1.030	0.013	0.027	****
Depression	Loss of energy/tiredness	ANX	0.083	1.087	0.013	<0.0001	****
Depression	Loss of energy/tiredness	BD	-0.586	0.557	0.062	<0.0001	****
Depression	Loss of energy/tiredness	MDD	0.119	1.126	0.013	<0.0001	****
Depression	Loss of energy/tiredness	SCZ	0.216	1.241	0.012	<0.0001	****
Depression	Loss of pleasure	ADHD	0.024	1.024	0.013	0.008	**
Depression	Loss of pleasure	ANX	0.099	1.104	0.013	<0.0001	****
Depression	Loss of pleasure	BD	-0.590	0.554	0.063	<0.0001	****
Depression	Loss of pleasure	MDD	0.126	1.134	0.013	<0.0001	****
Depression	Loss of pleasure	SCZ	0.212	1.236	0.012	<0.0001	****
Depression	Self-reproach	ADHD	0.021	1.021	0.013	0.128	ns
Depression	Self-reproach	ANX	0.075	1.078	0.013	<0.0001	****
Depression	Self-reproach	BD	-0.328	0.720	0.062	<0.0001	****
Depression	Self-reproach	MDD	0.075	1.078	0.012	<0.0001	****
Depression	Self-reproach	SCZ	0.127	1.135	0.012	<0.0001	****
Depression	Slowed activity	ADHD	0.017	1.017	0.014	0.208	ns
Depression	Slowed activity	ANX	0.047	1.048	0.014	0.001	***
Depression	Slowed activity	BD	-0.362	0.696	0.063	<0.0001	****
Depression	Slowed activity	MDD	0.062	1.064	0.013	<0.0001	****
Depression	Slowed activity	SCZ	0.100	1.105	0.012	<0.0001	****
Mania	Elevated mood	ADHD	0.021	1.021	0.007	0.007	**
Mania	Elevated mood	ANX	-0.061	0.941	0.007	<0.0001	****
Mania	Elevated mood	BD	0.374	1.454	0.035	<0.0001	****
Mania	Elevated mood	MDD	-0.068	0.934	0.007	<0.0001	****
Mania	Elevated mood	SCZ	0.124	1.132	0.007	<0.0001	****
Mania	Excess activity	ADHD	0.028	1.028	0.008	<0.0001	****
Mania	Excess activity	ANX	-0.051	0.950	0.008	<0.0001	****
Mania	Excess activity	BD	0.354	1.425	0.036	<0.0001	****
Mania	Excess activity	MDD	-0.062	0.940	0.007	<0.0001	****
Mania	Excess activity	SCZ	0.111	1.117	0.007	<0.0001	****
Mania	Pressured speech	ADHD	0.013	1.013	0.008	0.001	***
Mania	Pressured speech	ANX	-0.050	0.951	0.008	<0.0001	****
Mania	Pressured speech	BD	0.276	1.318	0.037	<0.0001	****
Mania	Pressured speech	MDD	-0.057	0.945	0.008	<0.0001	****
Mania	Pressured speech	SCZ	0.099	1.104	0.007	<0.0001	****

Dimension	OPCRIT Item	PRS Score	Estimate	Odds Ratio (OR)	Std.Error	*P*Bonferroni	*P*Bonf.signif
Mania	Racing thoughts	ADHD	0.014	1.014	0.008	0.089	ns
Mania	Racing thoughts	ANX	-0.057	0.945	0.008	<0.0001	****
Mania	Racing thoughts	BD	0.293	1.340	0.037	<0.0001	****
Mania	Racing thoughts	MDD	-0.066	0.936	0.007	<0.0001	****
Mania	Racing thoughts	SCZ	0.115	1.122	0.007	<0.0001	****
Mania	Reduced need for sleep	ADHD	0.024	1.024	0.008	0.003	**
Mania	Reduced need for sleep	ANX	-0.053	0.948	0.008	<0.0001	****
Mania	Reduced need for sleep	BD	0.345	1.412	0.037	<0.0001	****
Mania	Reduced need for sleep	MDD	-0.061	0.941	0.007	<0.0001	****
Mania	Reduced need for sleep	SCZ	0.121	1.129	0.007	<0.0001	****
Psychosis	Delusions of influence	ADHD	0.005	1.005	0.007	0.505	ns
Psychosis	Delusions of influence	ANX	-0.055	0.946	0.007	<0.0001	****
Psychosis	Delusions of influence	BD	0.135	1.145	0.033	<0.0001	****
Psychosis	Delusions of influence	MDD	-0.080	0.923	0.007	<0.0001	****
Psychosis	Delusions of influence	SCZ	0.227	1.255	0.006	<0.0001	****
Psychosis	Persecutory/jealous delusions	ADHD	-0.010	0.990	0.007	0.189	ns
Psychosis	Persecutory/jealous delusions	ANX	-0.054	0.947	0.007	<0.0001	****
Psychosis	Persecutory/jealous delusions	BD	0.399	1.490	0.033	<0.0001	****
Psychosis	Persecutory/jealous delusions	MDD	-0.080	0.923	0.007	<0.0001	****
Psychosis	Persecutory/jealous delusions	SCZ	0.439	1.551	0.006	<0.0001	****
Psychosis	Thought withdrawal	ADHD	-0.007	0.993	0.007	0.320	ns
Psychosis	Thought withdrawal	ANX	-0.057	0.945	0.007	<0.0001	****
Psychosis	Thought withdrawal	BD	0.115	1.122	0.033	<0.0001	****
Psychosis	Thought withdrawal	MDD	-0.086	0.918	0.007	<0.0001	****
Psychosis	Thought withdrawal	SCZ	0.695	2.004	0.006	<0.0001	****
Psychosis	Thought broadcast	ADHD	-0.005	0.995	0.007	0.530	ns
Psychosis	Thought broadcast	ANX	-0.062	0.940	0.007	<0.0001	****
Psychosis	Thought broadcast	BD	0.141	1.151	0.033	<0.0001	****
Psychosis	Thought broadcast	MDD	-0.089	0.915	0.007	<0.0001	****
Psychosis	Thought broadcast	SCZ	0.448	1.565	0.006	<0.0001	****
Psychosis	Thought insertion	ADHD	-0.006	0.994	0.007	0.429	ns
Psychosis	Thought insertion	ANX	-0.056	0.946	0.007	<0.0001	****
Psychosis	Thought insertion	BD	0.142	1.153	0.033	<0.0001	****
Psychosis	Thought insertion	MDD	-0.087	0.917	0.007	<0.0001	****
Psychosis	Thought insertion	SCZ	0.253	1.288	0.006	<0.0001	****
Chronicity	Course of disorder (chronic)	ADHD	0.221	1.247	0.015	0.005	**
Chronicity	Course of disorder (chronic)	ANX	0.104	1.110	0.014	<0.0001	****
Chronicity	Course of disorder (chronic)	BD	-0.072	0.931	0.067	<0.0001	****
Chronicity	Course of disorder (chronic)	MDD	0.165	1.180	0.014	<0.0001	****
Chronicity	Course of disorder (chronic)	SCZ	0.282	1.326	0.013	<0.0001	****
Chronicity	Reduced inter-episode remission	ADHD	0.343	1.409	0.006	0.007	**
Chronicity	Reduced inter-episode remission	ANX	0.028	1.028	0.006	<0.0001	****

Dimension	OPCRIT Item	PRS Score	Estimate	Odds Ratio (OR)	Std.Error	*P*Bonferroni	*P*Bonf.signif
Chronicity	Reduced inter-episode remission	BD	-0.202	0.817	0.028	<0.0001	****
Chronicity	Reduced inter-episode remission	MDD	0.042	1.043	0.006	<0.0001	****
Chronicity	Reduced inter-episode remission	SCZ	-0.076	0.927	0.005	<0.0001	****
Chronicity	Premorbid personality disorder	ADHD	0.044	1.045	0.007	0.006	**
Chronicity	Premorbid personality disorder	ANX	-0.059	0.943	0.007	<0.0001	****
Chronicity	Premorbid personality disorder	BD	-0.418	0.658	0.033	<0.0001	****
Chronicity	Premorbid personality disorder	MDD	-0.085	0.919	0.007	<0.0001	****
Chronicity	Premorbid personality disorder	SCZ	-0.155	0.856	0.006	<0.0001	****
Chronicity	Premorbid poor social adjustment	ADHD	0.064	1.066	0.007	0.004	**
Chronicity	Premorbid poor social adjustment	ANX	-0.055	0.946	0.007	<0.0001	****
Chronicity	Premorbid poor social adjustment	BD	-0.411	0.663	0.033	<0.0001	****
Chronicity	Premorbid poor social adjustment	MDD	-0.081	0.922	0.007	<0.0001	****
Chronicity	Premorbid poor social adjustment	SCZ	0.142	1.153	0.006	<0.0001	****
Chronicity	Premorbid poor work adjustment	ADHD	0.048	1.049	0.007	0.007	**
Chronicity	Premorbid poor work adjustment	ANX	-0.053	0.948	0.007	<0.0001	****
Chronicity	Premorbid poor work adjustment	BD	-0.426	0.653	0.033	<0.0001	****
Chronicity	Premorbid poor work adjustment	MDD	-0.077	0.926	0.007	<0.0001	****
Chronicity	Premorbid poor work adjustment	SCZ	0.140	1.150	0.006	<0.0001	****

This table shows the coefficients and their significance levels from the regression analyses where each of the 20 core OPCRIT items was predicted by the five individual polygenic risk scores (BD, SCZ, MDD, ADHD, and ANX). These results illustrate the relationship between the genetic burden for each disorder and the individual clinical symptoms in each of the four dimensions. *(Significance levels of adjusted Bonferroni P-value, < .0001 ****, < .001 ***, <.01 **, < .05).

Chapter 3 established a dimensional model of bipolar disorder, revealing distinct genetic influences on dimensions such as mania, psychosis, depression, and a novel chronicity factor. While this provides a broader understanding of BD's structure, the significant genetic overlap between BD and schizophrenia (SCZ), particularly concerning the severe psychotic features often prominent in bipolar disorder I (BD1), warrants more focused investigation. Therefore, building on the utility of Polygenic Risk Scores (PRS) in dissecting heterogeneity, Chapter 4 will focus on this critical area of transdiagnostic overlap to explore its clinical application. This chapter aims to specifically evaluate the utility of the SCZ3-PRS in predicting key clinical features of BD1, namely the presence and severity of psychosis and age of onset, while also exploring the associated biological pathways to identify potential biomarkers.

---

4 PRS-SCZ3 and BD1

A published version of the research in this chapter is available in the *Journal of Affective Disorders* (2024) at DOI: 10.1016/j.jad.2024.04.066

4.1 Abstract

Background: Schizophrenia (SCZ) and bipolar disorder (BD) exhibit shared genetic liability. This study aimed to investigate the predictive value of polygenic risk scores (PRS) derived from the most recent schizophrenia genome-wide association study (GWAS) (SCZ3) for phenotypic traits of bipolar disorder type 1 (BD1).

Aims: To determine the predictive power of SCZ3-PRS, alone and in combination with clinical variables, for various BD1 subphenotypes, including age of onset (general, depression, mania), psychosis (overall, congruent, incongruent), and rapid cycling, in a European BD1 case-control cohort, with validation in an independent cohort. Additionally, the aim was to identify biological pathways associated with psychosis in BD1 using individual-level gene set pathway analysis.

Methods: SCZ3-PRS was computed using PRSice-v2.3.3 (clumping and thresholding) and PRS-CS (Continuous Shrinkage) in 1878 BD1 cases and 2751 controls from Romania (RO) and the United Kingdom (UK). Univariate linear and logistic regressions assessed the predictive power of SCZ3-PRS for BD1 subphenotypes. Random forest (RF) models evaluated the predictive performance of SCZ3-PRS alone and in combination with nine clinical variables. Pathway analysis using PRSet explored gene sets associated with psychosis.

Results: SCZ3-PRS predicted psychosis (overall and incongruent), general age-of-onset of BD1, age-of-onset of depression and mania, and rapid cycling in univariate analyses. An inverse relationship was observed between SCZ3-SNP loading and rapid cycling, potentially suggesting different underlying genetic mechanisms. A negative correlation was observed between the number of depressive episodes and psychosis (mainly incongruent). RF models showed that combinations of SCZ3-PRS-CS and clinical variables provided the best predictions for BD1 subphenotypes, closely followed by models using only clinical variables. The most important clinical variables in predicting psychosis alongside SCZ3-PRS were family history and irritable mania. Gene set pathway analysis identified 22 pathways underlying psychosis in BD1.

Conclusions: These findings suggest that SCZ3-PRS has a modest clinical utility in predicting phenotypic traits of BD1. Its predictive performance is enhanced when combined with clinical variables. These results highlight the shared genetic underpinnings of SCZ and BD1 while also emphasising the importance of considering clinical information for improved prediction of BD1 subphenotypes.

4.2 Introduction

Developmental psychopathology, family studies, and genetic investigations have consistently indicated a shared vulnerability between schizophrenia (SCZ) and bipolar disorder (BD), alongside specific features that distinguish these diagnoses. Genome-wide association studies (GWAS) have corroborated this overlap, revealing a substantial genetic correlation (*rG*) of .70 between SCZ and BD based on common single nucleotide polymorphisms (SNPs) [1]. A broader analysis across eleven major psychiatric disorders further identified a "psychotic factor" encompassing shared genetic variants for both BD and SCZ [2]. Despite this shared genetic architecture, subsets of SNPs exert differential effects in SCZ and BD [3], potentially contributing to their distinct clinical presentations. Earlier research explored the association between SCZ-derived SNP sets [4] and certain BD phenotypic traits, such as age of onset (AO) [3, 5], and psychosis [6-8]. While some associations were reported, replication across studies has been inconsistent [9]. Notably, studies have suggested a higher loading of SCZ polygenic risk scores (PRS) in bipolar disorder I (BD1) compared to BD2 [8], and even higher in BD1 with psychosis compared to those without [10]. Furthermore, SCZ-PRS have been associated with treatment response within BD [11-12]. A recognised challenge in large-scale psychiatric GWAS is the inherent phenotypic heterogeneity arising from diverse populations, varying diagnostic criteria, recruitment settings, and the inclusion of different BD subtypes, including schizoaffective disorder [8-9]. To address this issue, the current study focused on investigating the predictive value of PRS derived from the most recent and largest schizophrenia GWAS (SCZ3) for phenotypic traits of BD1 within two phenotypically more homogeneous and well-characterised samples - a Romanian sample with detailed genealogical data to ensure genetic homogeneity and a well-phenotyped UK sample with comprehensive clinical assessments.

4.3 Aims

The aim was to investigate the predictive value of polygenic risk scores (PRS) derived from the Psychiatric Genomics Consortium (PGC) Schizophrenia GWAS 2022 (SCZ3) [13] for phenotypic traits of BD1 in two phenotypically homogeneous and well-characterised samples: a RO sample and a UK sample. More specifically, the aims were to determine the predictive power of SCZ3-PRS alone and in combination with clinical variables for several BD1 subphenotypes, i.e., clinical features. These included: General age of onset (AO) of BD1, age of onset of the first depressive episode, age of onset of the first manic/mixed episode, presence of psychosis (overall, as well as congruent and incongruent types), and rapid cycling. Additionally, the aim was to conduct pathway analysis using SCZ3-PRS to identify biological pathways associated with psychosis in BD1.

4.4 Methods

The predictive utility of schizophrenia-derived polygenic risk scores (SCZ3-PRS) for key clinical features of Bipolar Disorder I (BD1) was assessed in a European case-control cohort.

PRS were computed using both clumping-and-thresholding and continuous shrinkage methods. The performance of the SCZ3-PRS, alone and combined with clinical variables, was evaluated using regularised regression analyses and Random Forest (RF) models. Individual-level gene set pathway analysis was performed using PRSet to identify biological pathways associated with psychosis.

Full details regarding the cohorts, genotyping, PRS computation, and all statistical models are available in the General Methods (Chapter 2).

4.5 Results

The predictive power was analysed of SCZ3-PRS-CS (Table 27) and of eight P-thresholds (pT) containing SNPs associated with SCZ (P = 5 x 10^{-8} to P = .05) (Tables 31-35) for several phenotypic traits of BD1 (general AO, AO-first depression, psychosis and incongruent psychosis) in the RO, the UK and the combined RO-UK samples.

Differentiation of cases and controls

First, the ability of the SCZ3-PRS-CS and of the eight SCZ3-PRS-pTs was tested to differentiate the cases from controls in each national sample and in the combined RO-UK sample. All versions of SCZ3-PRS distinguished the cases from controls with high certainty, see corrected *P*-values (Table 27), demonstrating a clear separation based on PRS. The detailed results for the separate Romanian and UK samples are provided in Table 38 and Table 39, respectively. More variance was explained when using the superior PRS-CS-auto methodology rather than the existing pT threshold method (Table 26).

Table 26 Comparison of two methods for calculating SCZ3 individual-level PRS

PRSice (pT)	NgR2 (liability scale %)
1e-8	.838
1e-7	.933
1e-6	1.279
1e-5	1.943
1e-4	2.813
.001	4.094
.01	5.031
.05	6.044
C+T (pT *Mdn*)	**2.378**
PRS-CS	**4.261**

Table 27 PRS-SCZ3 BD1 and subphenotypes in RO-UK samples

PRS-CS-auto SCZ3 score (predictor)				Effect size		Variance %	AUC	P	FDR-P
Sample	**Outcome**	**Beta**	**SE**		**95%CI**				
	BD1 vs Ctrl			**OR**		**R^2N[a]**			
RO-UK	BDI	0.533	0.033	1.705	1.71; 1.82	4.261	0.643	2.84E-64	4.83E-63
RO	BDI	0.638	0.068	1.893	1.89; 2.17	5.901	0.669	3.55E-23	1.21E-22
UK	BDI	0.514	0.038	1.672	1.67; 1.80	3.963	0.639	1.43E-44	1.22E-43
	Psychosis								
RO-UK	Psychosis	0.463	0.035	1.589	1.48; 1.70	3.264	0.624	1.53E-42	2.14E-41
RO	Psychosis	0.576	0.068	1.778	1.56; 2.03	4.915	0.654	2.89E-19	1.35E-18
UK	Psychosis	0.443	0.042	1.557	1.43; 1.69	2.981	0.619	4.45E-27	3.12E-26
				OR		**R^2N[a]**			
RO-UK	Incongruent	0.464	0.058	1.591	1.42; 1.79	3.285	0.626	3.30E-16	1.16E-15
RO-UK	Congruent	0.381	0.057	1.464	1.31; 1.64	2.243	0.603	1.03E-11	2.88E-11
RO-UK	Rapid cycling	−0.451	0.382	0.637	0.61; 0.69	2.014	0.583	2.13E-09	2.43E-08
				t-test		**Adj. R2**			
RO-UK	AO general	−0.944	0.256	−3.68	−1.45;−0.44	0.728	–	2.38E-04	3.33E-04
RO	AO general	−1.88	0.431	−4.38	−2.73;−1.04	0.935	–	1.44E-05	3.36E-05
UK	AO general	−0.535	0.315	−1.69	−1.15; 0.08	0.472	–	0.0904	9.04E-02
				t-test		**Adj. R2**			
RO-UK	AO depress	−1.2	0.313	−3.84	−1.81; 0.59	0.98	–	1.26E-04	1.96E-04
RO	AO depress	−1.92	0.495	−3.88	−2.89; 0.95	1.197	–	1.17E-04	1.96E-04
UK	AO depress	−0.86	0.396	−2.17	−1.64; 0.08	0.526	–	3.01E-02	3.24E-02
				t-test		**Adj. R2**			
RO-UK	AO mania	−1.34	0.316	−4.24	−1.96; 0.72	0.731	–	2.41E-05	4.82E-05
RO	AO mania	−1.83	0.504	−3.62	−2.82; 0.84	0.815	–	3.16E-04	4.02E-04
UK	AO mania	−1.07	0.403	−2.65	−1.86; 0.28	0.363	–	8.30E-03	9.68E-03

Abbreviations. AO general – BD1 age of onset; AO depres – Age of onset of depression; Incongruent – mood incongruent psychosis.[a] Nagelkerke pseudo R-squared on the liability scale.

Phenotypic traits of bipolar disorder I (BD1)

Subsequently, several phenotypic traits of BD1 were analysed: the general age of onset (AO) of BD1 irrespective of polarity at onset, AO of the first depressive episode, AO of the first manic/mixed episode, presence of psychosis, presence of incongruent and congruent psychosis, and rapid cycling.

General age of onset of BD1

In the combined sample the general AO was predicted by SCZ3-PRS-CS (median $P = 3.33 \times 10^{-4}$) and by all eight pTs (pT+clump method) (FDR-corrected *P*-values) (Table 27). In the RO sample the general AO was predicted by SCZ3-PRS-CS ($P = 3.36 \times 10^{-5}$), while in the UK sample just a trend was visible ($P = .090$) (Table 27). In both national samples the regression coefficients were negative indicating that a higher SCZ3-SNP loading was associated with a younger AO in BD1 patients.

Age of onset of the first depressive episode

Both in the combined RO-UK sample (median $P = 1.96 \times 10^{-4}$) and in the separate national samples the age of onset of the first depressive episode was predicted by the SCZ3-PRS-CS ($P = 1.96 \times 10^{-4}$ for RO; $P = 3.24 \times 10^{-2}$ for UK) (Table 27). Similarly, all eight SCZ3-pTs computed through the pT+clump method predicted the AO of depression with significant *P*-values and with negative regression coefficients in the RO-UK sample (Table 32) and in the national samples (data not shown) indicating a negative effect of SCZ3-PRS on AO of depression. The AO-depression was younger in psychotic patients than in non-psychotic patients in the RO-UK sample (AO-depression in psychotic cases mean = 25.90, SD = 10.38, AO-depression in non-psychotics =27.04, SD = 11.44; $t = 2.89$; $P = .004$), as well the general AO of BD1 (AO in psychotic cases mean = 25.05; SD = 9.54; AO in non-psychotic cases mean = 26.35; SD =11.16 ($t = 2.49$; $P = .013$).

Age of onset of the first manic episode

Age of onset of the first manic episode was predicted by the SCZ3-PRS-CS both in the combined sample ($P = 4.82 \times 10^{-5}$) and in the national samples ($P = 4.02 \times 10^{-4}$ for RO; $P = 9.68 \times 10^{-3}$ for UK) (Table 27). The pT+clump method did not predict the AO of the first manic episode either in the national or in the combined samples (data not shown). In the samples there was a significant difference in AO of depression between female and male cases (RO sample mean AO-depression: males mean = 30.27 years (SD = 10.60); females mean = 27.11 years (SD = 9.77; t = 3.364, $df = 1/497$, $P = .00082$; UK sample: males mean = 26.45 (SD = 11.33; females Mean = 24.26 (SD = 10.41; $t = 2.95$, $df = 1/884$, $P = .003$); RO-UK sample AO-depression; males mean = 27.76, SD = 11.23; females mean = 25.31, SD = 10.27, $df = 1/1381$; $t = 4.15$, $P = 1.7 \times 10^{-5}$). Linear regressions for AO-mania were performed with sex as a covariate in the combined sample (data not shown).

Presence of lifetime psychosis and incongruent psychosis

Similar to other BD samples [39] the prevalence of psychosis (congruent and incongruent) reached 71% in the RO-UK BD1 sample. Both PRS computation methods (CS and pT+clump) yielded highly significant *P*-values for the prediction of psychosis irrespective of type and for the mood incongruent psychosis in the combined sample (Tables 27, 33-34) and the national samples (data not shown). A novel finding indicated a negative correlation between the number

of depressive episodes and psychosis. This finding was confirmed by a multivariate logistic regression (β = -.407; SD = .143, Wald = 8.113; OR = .666, 95%CI = .503-.881, *P* = .004) including six clinical variables, in regularised regressions (RF) (Table 30). The same negative correlation was valid for the incongruent psychosis, but not for congruent psychosis. The mood congruent psychosis was predicted only by the SCZ3-PRS-CS in the combined sample, but not by the pT +clump method.

Rapid cycling

Both the SCZ3-PRS-CS method (Table 27) and the pT +clump SCZ3-PRS method with five pTs (Table 35) predicted the rapid cycling trait. But the ORs were below 1 and the regression coefficients were negative suggesting that rapid cycling and SCZ3-PRS loading have an inverse relationship. This could suggest a different underlying genetic architecture for rapid cycling compared to other psychosis-related features.

Family history for major psychoses

Major psychoses (schizophrenia, schizoaffective disorders, bipolar disorder, unipolar major depression) was nominally predicted by three SCZ3-pTs (Table 30) indicating that only specific SNPs and genes are involved in familial inheritance.

Table 28 PRS-SCZ3 prediction of BD-traits (10-fold cross-validated RF classification)

Outcome	Sample	Accuracy	Accuracy CI (95%)	Pos Pred Value (PPV)	F1	AUC	AUC CI (95%)
Psychosis	RO/UK	0.765	0.737-0.799	0.777	0.852	0.785	0.744-0.824
Incongruent	RO/UK	0.805	0.769-0.836	0.819	0.884	0.787	0.746-0.829
Congruent	RO/UK	0.724	0.685-0.760	0.74	0.803	0.761	0.719-0.803
		RMSE	R-squared	MAE	RMSE-SD	R-squared SD	MAE-SD
AO BPI	RO/UK	5.591	0.733	3.716	0.647	0.057	0.357
AO Depression	RO/UK	3.884	0.874	2.105	0.62	0.048	0.265
AO Mania	RO/UK	5.082	0.774	2.801	0.698	0.063	0.314

Abbreviations: Adj.R2,adjusted R2; RMSE, Root Mean Squared Error; MAE, Mean Absolute Error; AUC, Area Under the Curve (AUCROC)

All models were constructed using conditional inference Random Forest (RF) to reduce risk of overfiitted models in data with correlated predictor variables

Performance metrics from a 10-fold cross-validated Random Forest (RF) model, detailing the accuracy of the SCZ3-PRS in predicting various continuous and binary BD traits.

Table 29 Random forest 10-fold cross-validated predictions

Model	Accuracy	Accuracy CI (95%)	AUC	AUC 95% CI	Adj. P-value[a]
Psychosis					
Clinical + genetic	0.765	0.737-0.799	0.785	0.744-0.824	-
Clinical	0.719	0.704-0.794	0.761	0.722-0.817	7.32E-03
Genetic (SCZ3-PRS)	0.711	0.685-0.733	0.625	0.593- 0.657	
Incongruent psychosis					
Clinical + genetic	0.805	0.769-0.836	0.787	0.746-0.829	-
Clinical	0.786	0.752- 0.779	0.753	0.644-0.724	4.59E-04
Genetic (SCZ3-PRS)	0.738	0.714-0.761	0.606	0.575-0.638	
Congruent psychosis					
Clinical + genetic	0.724	0.685-0.760	0.761	0.719-0.803	-
Clinical	0.696	0.65-0.706	0.693	0.643-0.714	0.096
Genetic (SCZ3-PRS)	0.685	0.659-0.710	0.601	0.568-0.634	
	RMSE	Adj.R-squared			
AO BPI					
Clinical + genetic	5.591	0.733	-	-	-
Clinical	6.279	0.714	-	-	< 2.2E-16
Genetic (SCZ3-PRS)	10.651	0.016			
AO depression					
Clinical + genetic	3.884	0.874	-	-	-
Clinical	5.037	0.768	-	-	8.00E-03
Genetic (SCZ3-PRS)	9.874	0.017			
AO mania					
Clinical + genetic	5.082	0.774	-	-	-
Clinical	6.143	0.681	-	-	0.078
Genetic (SCZ3-PRS)	10.766	0.015			

Abbreviations: Adj.R-squared, adjusted R-squared; RMSE = Root Mean Squared Error; MAE, Mean Absolute Error; AUC, Area Under the Curve (AUCROC)

a. Bonferroni corrected P-value for pairwise comparisons between model fit statistics

This analysis compared the predictive performance of models using either clinical predictors only, genetic predictors only (SCZ3-PRS-CS), and a combination of both in the RO/UK samples.

Table 30 Comparison between variable importance

Outcome	Predictor	LogOR[c]	SE	CLow	CLhigh	P-value	Sig.	MDA[d]	SE	Sig.
Psychosis	Fam hist maj psychoses	1.769	0.138	1.48	2.076	<0.001	***	0.0216	0.011	**
Psychosis	SCZ3 PRS	0.194	0.051	0.1	0.293	<0.001	***	0.003	0.009	*
Psychosis	Nr depres epis	-0.132	0.066	-0.244	-0.011	0.028	**	0.002	0.002	*
Psychosis	Nr mania epis	0.29	0.109	0.232	0.627	<0.001	***	0.002	0.031	*
Psychosis	Age onset mania	-0.014	0.013	-0.039	0.011	0.204	n.s.	0.0002	0.017	*
Psychosis	Rapid cycling	-0.746	0.325	-1.293	-0.02	<0.001	***	0.002	0.027	*
Psychosis	Irritable mania	2.479	0.225	2.053	2.954	<0.001	***	0.002	0.052	
Psychosis	Nr epis total	0.108	0.026	0.015	0.12	0.012	**	0.001	0.023	*
Psychosis	Age onset depress	-0.037	0.015	-0.065	-0.009	0.016	*	0.0001	0.024	*
Psychosis	Age onset BPI	-0.069	0.017	-0.108	-0.041	<0.001	***	0.0003	0.029	*
MIP	SCZ3 PRS	0.261	0.047	0.211	0.404	<0.001	***	0.008	0.012	*
MIP	Fam hist maj psychoses	2.325	0.142	2.158	2.744	<0.001	***	0.006	0.009	**
MIP	Irritable mania	1.739	0.226	1.295	2.165	<0.001	***	0.001	0.003	*
MIP	Age onset depress	-0.041	0.017	-0.067	-0.002	0.02	*	0.0005	0.021	*
MIP	Age onset mania	-0.009	0.013	-0.032	0.018	0.456		0.001	0.021	*
MIP	Nr epis mania	0.197	0.039	0.13	0.268	0.004	**	0.001	0.028	*
MIP	Age onset BPI	-0.086	0.013	-0.117	-0.064	<0.001	***	0.001	0.027	*
MIP	Nr epis depres	0.067	0.111	-0.313	0.04	0.4		0.003	0.033	*
MIP	Rapid cycling	-0.083	0.197	-0.412	0.39	0.776		0.0005	0.009	**
MIP	Nr epis total	0.109	0.019	0.077	0.158	<0.001	***	0.0003	0.04	*
MIP	Fam hist aff dis	1.85	0.119	1.575	2.033	<0.001	***	0.01	0.013	***
MIP	SCZ3 PRS	0.173	0.051	0.068	0.265	<0.001	***	0.005	0.009	**
MIP	Nr epis depres	0.27	0.053	0.192	0.367	0.004	**	0.001	0.024	*
MIP	Nr epis mania	0.384	0.042	0.307	0.478	<0.001	***	0.001	0.024	*
MIP	Age onset mania	-0.075	0.014	-0.12	-0.056	<0.001	***	0.001	0.026	*
MIP	Rapid cycling	-0.277	0.305	-0.427	-0.122	<0.001	***	0.0002	0.004	**
MIP	Irritable mania	1.68	0.186	1.358	2.121	<0.001	***	0.002	0.022	*
MIP	Nr epis total	0.2	0.022	0.158	0.249	<0.001	***	0.0004	0.034	*
MIP	Age onset depres	-0.11	0.019	-0.178	-0.106	<0.001	***	0.001	0.027	*
MIP	Age onset BPI	-0.118	0.016	-0.171	-0.108	<0.001	***	0.001	0.043	*

Notes: a. MIP – Mood-Incongruent Psychosis. A Penalised (Elastic Net) logistic regression used for measuring outcome associations with constrained, correlated predictor variables. Bootstrapped Std.Error, CI, Confidence Intervals (95%) and P-values. Significant at * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$ b. Random (conditional) forest modelling for nonlinear approximation of relationships between outcome and the predictor variables. Variable importance measure used Conditional Permutation Importance c. LogOR, absolute log odds ratio (OR) per standard deviation (SD) d. **MDA**, **Mean Decrease Accuracy** expresses how much accuracy the model losses by excluding each variable.

This analysis compared the variable importance rankings derived from Random Forest and regularised regression models for predicting psychosis and its subtypes.

Predictive performance of SCZ3-PRS in combination with clinical traits of BD1

After investigating the predictive power of SCZ3-PRS-CS for phenotypic traits of BD1 in univariate regressions (only the SCZ3-PRS-CS regressed against each outcome), the predictive power of SCZ3-PRS-CS was investigated in combinations with clinical variables (family history of major psychoses in first and second degree relatives, number of depressive and manic episodes, AO-mania, rapid cycling, irritable mania, total number of episodes, AO-depression, general AO) for certain BD1 traits in the RO-UK sample with the random forest method that controls the collinearity between predictor variables (Table 28-30). BD1 cases were randomly allocated to either training, validation or testing sets. To determine the predictive performance, i.e., classification by the cross-validated RF model of the binary outcomes, the ROC (Receiver Operating Characteristic) and its Area under the curve (AUC), sensitivity, specificity, and accuracy were used. Additionally, the Positive Predictive Value (PPV) indicating the risk for false positives is lower with a higher value, and the F1 score, a more accurate metric for prediction accuracy with uneven class distribution, were both reported. Accuracy for the cross-validated RF regression of the continuous outcomes was assessed with R2 and RMSE (Table 28), shows that both the accuracy and AUC-values for binary subphenotypes (psychosis and its subtypes) and R2 and RMSE for continuous subphenotypes indicate a moderate predictive performance of SCZ3-PRS-CS and clinical variables. The best predictions were for psychosis, incongruent psychosis (AUC close to .8) and AO-depression, consistently across methods.

Prediction Models for BD1 Phenotypic Traits

Models using SCZ3-PRS alone do not achieve 100% accuracy in predicting BD1 phenotypic traits [40]. Factors such as family history and age-of-onset may also play a role. To explore this further, prediction models were developed for each BD1 trait, comparing the effectiveness of clinical variables, SCZ3-PRS-CS, and a combination of both. Clinical variables included family history of major psychoses, total number of episodes, number of manic episodes, number of depressive episodes, irritable mania, rapid cycling, general age of onset (AO), AO-depression, and AO-mania. Each variable was excluded when it served as the outcome (Table 29).

Comparison of Predictive Power

For all investigated BD1 traits, the most accurate predictions were obtained from models combining SCZ3-PRS-CS and clinical variables, followed by models using only clinical variables. Models relying solely on SCZ3-PRS-CS showed the weakest prediction indicators (Table 29). Pairwise Bonferroni-corrected one-sample *t*-tests revealed significant differences in metrics between the clinical and clinical plus SCZ3-PRS models, except for congruent psychosis and AO-mania, where only trends were observed. Psychosis, incongruent psychosis, and AO-depression showed the best prediction accuracy.

Variable Importance in Predicting Psychosis

Given that psychosis was the best predicted BD1 subphenotype, the importance of the variables in predicting psychosis and its subtypes was examined using two cross-validation methods: regularized regression (elastic net in "cv.glmnet") and conditional random forest (RF) (cforest) in R (Table 30). In the RF model, higher Mean Depreciation Accuracy (MDA) scores indicate greater importance of a variable for outcome classification. Table 30, illustrates that the importance of variables varies with this method used. The elastic net model highlighted family history, SCZ3-PRS-CS, number of mania episodes, rapid cycling, irritable mania, and general AO as having the highest and equal importance for psychosis prediction. In contrast, the RF model showed diminished importance for these variables, although they remained significant. For mood-incongruent psychosis, family history of major psychoses and irritable mania were the most important predictors in both models, while SCZ3-PRS-CS and general AO of BD1 remained significant but with higher *P*-values.

Table 31 General Age of Onset (AO) in the combined RO/UK sample

RO/UK		General AO				
SCZ3-PRS	Beta/ SE	95%CI	*P*	FDR-*P*	R2	R2 adj.
pT-5 x 10^{-8}	-0.52 (0.24)	-0.98 - -0.06	0.027	0.027	0.003	0.002
pT-1 x 10-7	-0.54 (0.24)	-1.00 - -0.08	0.021	0.025	0.003	0.002
pT-1 x 10-6	-0.75 (0.23)	-1.21 - -0.29	0.0013	0.0018	0.006	0.005
pT-1 x 10-5	-0.94 (0.23)	-1.40 - -0.48	5.46 x 10^{-5}	8.74 x 10^{-5}	0.009	0.008
pT-1 x 10-4	-0.99 (0.23)	-1.44 - -0.54	1.76 x 10^{-5}	3.52 x 10^{-5}	0.01	0.01
pT-0.001	-0.99 (0.23)	-1.44 - -0.54	1.68 x 10^{-5}	4.49 x 10^{-5}	0.01	0.01
pT- 0.01	-1 (0.23)	-1.46 - -0.55	1.63 x 10^{-5}	6.51 x 10^{-5}	0.01	0.01
pT-0.05	-1.51 (0.27)	-2.03 - -0.98	2.00 x 10^{-8}	2.00 x 10^{-7}	0.017	0.017

Results from linear regression models assessing the association between SCZ3-PRS (at various *P*-value thresholds) and the general age of onset for bipolar disorder in the combined RO/UK sample.

Table 32 Age of onset of depression in the combined RO/UK sample

RO/UK	AO depression					
SCZ3-PRS	Beta	SE	95% CI	*P*	FDR-*P*	Adj.R2
pT-5E-08	-1.131	0.365	-0.41 - -1.85	0.0020	0.00232	0.007
pT-1E-07	-1.144	0.366	-0.43 - -1.86	0.0018	0.00232	0.007
pT-1E-06	-1.326	0.361	-0.62 - -2.03	$2.51x10^{-4}$	6.70E-04	0.011
pT-1E-05	-1.472	0.359	-0.77 - -2.18	$4.54x10^{-5}$	1.80E-04	0.014
pT-1E-04	-1.435	0.349	-0.75 - -2.12	$4.36x10^{-5}$	1.80E-04	0.014
pT-0.001	-1.243	0.352	-0.55 - -1.93	$4.29x10^{-4}$	8.60E-04	0.010
pT- 0.01	-1.155	0.355	-0.46 - -1.85	0.0012	0.0019	0.008
pT-0.05	-0.837	0.404	-0.04 - -1.63	0.039	0.039	0.003

Results from linear regression models assessing the association between SCZ3-PRS (at various *P*-value thresholds) and the age of onset for depression in the combined RO/UK sample.

Table 33 Prediction of Psychosis irrespective of subtype in the RO/UK sample

RO/UK			Psychosis			
SCZ3-PRS	OR	SE	95% CI	*P*	FDR-*P*	AdjR2
$pT-5x10^{-8}$	1.06	0.06	0.95 - 1.17	0.30	0.30	<0.00018
$pT-1x10^{-7}$	1.06	0.06	0.96 - 1.18	0.23	0.27	0.001
$pT-1x10^{-6}$	1.18	0.06	1.07 - 1.31	0.0014	0.0019	0.008
$pT-1x10^{-5}$	1.31	0.07	1.18 - 1.45	$4.83x10^{-7}$	$9.65x10^{-7}$	0.020
$pT-1x10^{-4}$	1.37	0.07	1.23 - 1.52	$7.20x10^{-9}$	$1.90x10^{-8}$	0.028
pT-0.001	1.45	0.08	1.30 - 1.62	$1.00x10^{-10}$	$4.00x10^{-10}$	0.036
pT- 0.01	1.41	0.08	1.27 - 1.58	$8.00x10^{-10}$	$3.30x10^{-9}$	0.032
pT-0.05	1.24	0.08	1.09 - 1.41	0.0010	0.0016	0.009

Results from logistic regression models assessing the ability of the SCZ3-PRS (at various *P*-value thresholds) to predict the presence of psychosis in the combined RO/UK sample.

Table 34 Prediction of incongruent Psychosis in combined RO/UK sample

RO/UK			Incongruent			
SCZ3-PRS	OR	S.E.	95% CI	*P*	FDR-*P*	R2 AdjR2
$pT-5x10^{-8}$	1.1	0.05	1.00 - 1.22	0.047	0.047	0.003
$pT-1x10^{-7}$	1.11	0.05	1.01 - 1.22	0.038	0.043	0.003
$pT-1x10^{-6}$	1.21	0.06	1.10 - 1.34	$1.00x10^{-4}$	$1.00x10^{-4}$	0.011
$pT-1x10^{-5}$	1.39	0.07	1.26 - 1.54	$1.00x10^{-10}$	$2.00x10^{-10}$	0.032
$pT-1x10^{-4}$	1.45	0.08	1.31 - 1.61	$8.00x10^{-13}$	$2.00x10^{-12}$	0.041
pT-0.001	1.51	0.08	1.36 - 1.68	$2.00x10^{-14}$	$1.00x10^{-13}$	0.047
pT- 0.01	1.51	0.08	1.36 - 1.68	$2.00x10^{-14}$	$1.00x10^{-13}$	0.047
pT-0.05	1.65	0.11	1.44 - 1.89	$4.00x10^{-13}$	$1.20x10^{-12}$	0.046

Results from logistic regression models assessing the ability of the SCZ3-PRS (at various *P*-value thresholds) to predict the presence of mood-incongruent psychosis in the combined RO/UK sample

Table 35 Prediction of BD1 rapid cycling by SCZ3-PRS (logistic regression)

RO/UK				Rapid Cycling			
SCZ3-PRS	OR	SE	Beta	95% CI	AdjR2	*P*	FDR-*P*
pT-5x10^{-8}	1.041	0.069	0.041	0.91-1.19	0.00	0.556	0.556
pT-1x10^{-7}	1.044	0.069	0.043	0.91-1.19	0.001	0.534	0.610
pT-1x10^{-6}	0.93	0.067	-0.072	0,81-1.06	0.002	0.283	0.377
pT-1x10^{-5}	0.856	0.068	-0.153	0.75-0.98	0.007	0.024	0.0384
pT-1x10^{-4}	0.809	0.068	-0.212	0.71-0.92	0.013	0.002	5.33x10^{-3}
pT-0.001	0.723	0.071	-0.324	0.63-0.83	0.029	4.97x10^{-6}	1.99x10^{-5}
pT- 0.01	0.699	0.072	-0.357	0.61-0.80	0.034	7.354x10^{-7}	5.88x10^{-6}
pT-0.05	0.830	0.080	-0.186	0.70-0.97	0.007	0.020	0.040

Results from logistic regression models assessing the association between SCZ3-PRS (at various *P*-value thresholds) and rapid cycling in individuals with BD1.

Pathway analysis of psychosis in BD1

Twenty-two pathways (Table 36) had a competitive *P*-value of ≤ .05, defined as showing association. All enriched pathways contained at least one gene identified in previous or most recent GWAS of SCZ [13] or BD [41], see Table 40. The highest associated pathways were *ZNF318* (R2 = .951, FDR-*P* = .003), Apoptosis (R2 = .958; FDR-*P* = .003), and Mitochondrion (R2 = .754; FDR-*P* = .037). *ZNF318* (zinc finger protein 318) was identified in the most recent BD GWAS [41]. Other pathways associated with psychosis in the samples were pathways relevant to brain function, including synaptic transmission involving both ion channels and dendrites (regulation_of_dendritic_spine_development; regulation_of_membrane_repolarization; regulation_of_dopamine_receptor_signaling), to the autonomous nervous system (abnormality_of_the_autonomic_nervous_system), to the immune system (regulation_of_immune_system_process).

Table 36 PRSet SCZ3 Individual level pathway analysis in RO-UK sample

Pathway PRS for psychosis	Association with psychosis		Pathway enrichment for psychosis	
	R2[a](%)	FDR P-value[b]	Nr. SNPs	P-value[c]
MITOCHONDRION	0.754	0.037	4899	0.010
ZNF318	0.951	0.003	4333	0.002
REGULATION_OF_IMMUNE_SYSTEM_PROCESS	0.339	0.038	1378	0.043
NCOA2	0.542	0.027	1609	0.011
MIR202_3P	0.54	0.043	1367	0.009
MIR3125	0.7	0.014	1432	0.004
MIR6859_5P	0.634	0.015	1186	0.006
MIR4782_5P	0.665	0.014	1041	0.004
MIR5706	0.665	0.014	1041	0.004
MIR4763_3P	0.659	0.014	1012	0.005
ABNORMALITY_OF_THE_AUTONOMIC_NERVOUS_SYSTEM	0.68	0.014	1022	0.004
MIR10395_3P	0.63	0.015	535	0.005
REGULATION_OF_DENDRITIC_SPINE_DEVELOPMENT	0.607	0.019	458	0.005
MIR197	0.658	0.014	433	0.004
REGULATION_OF_MEMBRANE_REPOLARIZATION	0.548	0.026	336	0.007
APOPTOSIS	0.958	0.003	539	0.000
chr1p21	0.649	0.014	359	0.004
REGULATION_OF_DOPAMINE_RECEPTOR_SIGNALING_PATHWAY	0.317	0.014	136	0.018
MIR625_3P	0.48	0.038	200	0.009
MIR3681_5P	0.636	0.015	223	0.004
MIR6849_5P	0.683	0.014	230	0.003
MIR4669	0.445	0.036	28	0.009

SNP = single nucleotide polymorphism.[a] Pathways presented are the weighted R^2, i.e. R2 devided by the number of SNPs in the pathway.[b] P-values for association after FDR multiple testing correction, significance was set at $p < 0.05$.[c] P-values indicating enrichment were corrected for 10,000 permutations, significance was set at $p < 0.05$.

4.6 Discussion

A strength of this study is the phenotypic homogeneity of strictly diagnosed BD1 cases and the direct investigation of the controls with a psychiatric interview, which is not always the situation in the large-scale GWAS samples, which allows for the highlighting of some new associations. This investigation into the predictive capacity of the latest schizophrenia polygenic risk score (SCZ3-PRS) for bipolar disorder type 1(BD1) phenotypes in a well-characterized European cohort yields several noteworthy findings. It was demonstrated that SCZ3-PRS, while primarily developed for schizophrenia, exhibits a significant, albeit modest, ability to predict various BD1 subphenotypes, including age of onset (for both the disorder generally and for depressive and manic episodes), and the presence of psychosis, particularly the mood-incongruent subtype. This underscores the substantial shared genetic underpinnings between these two major psychiatric disorders, aligning with prior reports of high genetic correlation and a common "psychotic factor". This study is among the first investigating the predictive validity of the SCZ3-PRS for BD1 clinical traits, specifically age of onset and psychosis, in phenotypically homogeneous clinical samples. There is only one published study using SCZ3-PRS for prediction of the clinical course of the disease in psychotic patients (mainly schizophrenia) [42] but not for predicting those clinical traits investigated here. I also evidenced a negative correlation between the number of depressive episodes and psychosis. The results confirm findings of previous studies that used the SCZ-SNP-set 2014 [3] and SCZ3-SNP-set [13] on psychosis in BD1 [6, 43-44] and on AO in BD1 [3, 5, 45]. Moreover, a higher burden of SCZ3-PRS was associated not only with younger general AO of BD1, but also with decreased AO of first depressive episode and of the first manic episode. A relationship between SCZ3-PRS and AO-depression was reported [46] for the AO of unipolar major depression in the UK biobank.

To our knowledge the negative correlation between the number of depressive episodes and psychosis found in the samples both in regularised regressions and RF is a novel finding supported by a meta-analysis of 54 studies of psychotic symptoms in BD [39] showing that psychosis is four times more frequent in manic/mixed episodes than in depressive episodes of BD1. The negative correlation found between the number of depressive episodes and the presence of psychosis, especially incongruent psychosis, warrants further exploration into the complex interplay of mood episodes and psychotic features in BD1. In contrast to previous work [44] who found no effect of SCZ3-PRS on mania in BD, a positive association of SCZ3-PRS on the AO of mania was found here. The difference could originate from ascertainment, the current study contained only BD1 samples, while their sample additionally contained 28.8% BD2 cases.

The predictive power of SCZ3-PRS-CS was further tested in combinations with other nine clinical variables in Random Forest (RF) models that model non-linear relationships and control for the collinearity between predictor variables. According to AUC and accuracy values for dichotomous traits and R2 and RMSE for continuous traits the predictive power of SCZ3-PRS was more modest than in simple linear/logistic regressions, but the best prediction was for incongruent psychosis and AO-depression. Moreover, the RF models that compared the

predictions based on only SCZ3-PRS, on SCZ3-PRS plus clinical variables, and on only clinical variables showed that the worst prediction was provided by the SCZ3-PRS and that the accuracy of the prediction based on only clinical variables was not far from that based on both SCZ3-PRS and clinical variables. This finding is in line with an earlier observation [42] that SCZ3-PRS had minimal value for outcome prediction relative to information from the clinical diagnostic system. Two studies suggested that clinical variables such as psychiatric family history and age of onset improve the predictions based on PRS for clinical purposes [47-48]. This was evident here in the ranking of clinical variables in the prediction of psychosis, as well as in the comparison of the RF model based on only SCZ3-PRS with the model including SCZ3-PRS plus clinical variables.

Therefore, the application of random forest models provided valuable insights into the relative contributions of genetic and clinical factors in predicting BD1 traits. Consistently, the most robust predictions were achieved when SCZ3-PRS was integrated with clinical variables, outperforming models relying solely on clinical information or SCZ3-PRS alone. This highlights the multifactorial nature of BD1 and the necessity of combining genetic predisposition with clinical presentation for improved predictive accuracy. While SCZ3-PRS contributes meaningfully to this predictive power, its clinical utility appears to be maximized within a broader clinical context. Notably, psychosis, particularly its incongruent form, and the age of onset of depression emerged as the most predictable phenotypes in the combined models. A comparison of the variance explained (liability Nagelkerke R2) by the two PRS methods indicated a marked increase in phenotypic variance explained by PRS-CS compared to the pT + clump method. For some investigated BD1 traits (AO-mania, mood-congruent psychosis) the two PRS computation methods gave different results. On the other hand, the pT+clump method showed that pTs stringently associated SCZ3-SNPs (e.g. pT-5 x 10^{-8}; pT-1 x 10^{-7}) offer significant predictions for incongruent psychosis and AO-depression supporting their clinical validity. Intriguingly, I an inverse relationship between SCZ3-PRS loading and the rapid cycling phenotype in BD1 was observed, suggesting a potentially distinct genetic architecture influencing this specific course of illness. Both PRS methods significantly confirmed the trend observed earlier [3] that there is an inverse relationship between SCZ-SNP loading and BD rapid cycling.

The finding that SCZ-PRS most strongly predicts mood-incongruent psychosis has important nosological implications. It suggests that individuals with BD1 who present with this feature may carry a greater burden of the genetic risk typically associated with schizophrenia. This supports a dimensional view where mood-incongruent psychosis in BD represents a point of significant biological overlap on a continuum between affective and non-affective psychoses. Clinically, these individuals may represent a distinct subgroup with a different prognosis or treatment response profile, warranting further investigation into whether this specific genetic signature could be used for patient stratification in the future.

The pathway enrichment analysis identified twenty-two biological pathways associated with schizophrenia and psychosis in BD1, offering potential avenues for future research into the specific molecular mechanisms underlying this critical aspect of the disorder. These findings

may contribute to a more refined understanding of the pathophysiology shared and distinct between schizophrenia and bipolar disorder with psychosis. The enriched pathways are relevant to brain function, including synaptic transmission involving both ion channels and dendrites, and brain development. The pathways that explained the highest variance of psychosis were: *ZNF318*, Apoptosis, Mitochondrion. Neuroimaging studies showed progressive loss of cortical grey matter in first-episode psychosis (FEP) [49], therefore a role for apoptosis mechanisms producing cell or localised synaptic/dendritic loss in psychosis is plausible. Defects in the structure of dendrites of pyramidal neurons may also have direct effects leading to the loss of cortical volume (regulation_of_dendritic_spine_development) [50]. Mitochondrial dysfunction (Mitochondrion) was linked to alterations in dopamine signalling, glutamatergic dysfunction and oxidative stress in schizophrenia [51-52] and in BD [53]. Both the "Mitochondrion" and "*ZNF318*" pathways contain the *CREB3L4*-gene. *CREB3L4* is a subtype of the *CREB1*-gene, expression of which is downregulated in brain tissue of SCZ, BD, MDD patients compared with healthy controls [54]. The chr1p21-pathway with the microRNA encoding gene *MIR137HG* that regulates signalling pathways for neural development is implicated in schizophrenia risk [55] and its early onset [56]. *NCOA2* (R2 = .542, FDR-*P* = .027) was one of 9 genes differentially expressed in the dorsolateral prefrontal cortex (DLPFC) in patients with BD [57]. Regulation of dopamine was implicated in psychosis by the regulation_of_dopamine_ receptor_signaling_pathway. Excessive dopaminergic modulation of striatal function has long been hypothesized to mediate psychosis and antipsychotic drugs target dopaminergic innervation in the striatum [58].

There is also evidence to involve the immune system in the pathogenesis of psychosis (regulation_of_immune_system_process). Increased risk of adulthood psychosis has been linked to high concentrations of proinflammatory cytokines in childhood [59]. In a GWAS of response of BD patients to lithium treatment [11] genes related to the immune system (HLA antigen complex and inflammatory cytokines) were associated with the treatment response and the same genes in the HLA region were also associated with risk for BD [41] and SCZ [13]. In the "negative_regulation_of_immune_system_process" and the "*ZNF318*" pathways appears the *MAD1L1*-gene and in the *NOA2*-pathway appears the *NT5C2*-gene that were associated with BD and SCZ in several GWAS; they were also associated earlier with the AO of BD1 in the RO sample [60]. The immune system PRS pathway, further implicated the gene *FURIN*, recently associated with BD [41] it was linked with decreased neurite outgrowth [61-62]. Altered function of the autonomic nervous system involving heart rate was previously documented in SCZ and psychosis [63-64] and genes present in this pathway were associated with cardiac β-adrenergic signalling and cardiac hypertrophy signalling in BD [57]. Several genes (*CACNA1C*, *GABBR1*, *GABBR2*, *SLC6A9*; *NT5C2*) in pathways linked to psychosis in the BD1 sample are also involved in the epigenomic differential methylation of DNA in SCZ and psychosis [65-66]. DNA methylation was also linked to the AO of SCZ [67] and BD1 [68]. The limited variance explained by PRS alone in predicting BD1 phenotypes also highlights the remaining challenge of "missing heritability" in psychiatric genetics.

4.5 Limitations

There was heterogeneity in BD1 severity. The combined Romanian and UK sample had varying degrees of BD1 severity. The Romanian sample, and partially the UK sample, primarily included hospitalized BD1 cases. Hospitalization is generally an indicator of more severe illness, which could have influenced the findings, especially for phenotypes related to severity, such as the presence of psychotic symptoms [39]. Furthermore, there was incomplete subphenotype information. Not all participants in the UK sample had complete information for all the BD1 subphenotypes being studied. This missing data (Table 37) could have introduced bias or reduced the statistical power for analyses involving those specific subphenotypes within the UK cohort and the combined sample.

4.6 Conclusions

The study is among the first investigating the predictive value of the SCZ3-PRS and shows that these biomarkers have a modest clinical value for predicting some phenotypic traits of BD1 in machine learning models. The findings demonstrate a modest clinical value of SCZ3-PRS. SCZ3-PRS has a limited, or modest, clinical value when used alone for predicting phenotypic traits of bipolar disorder type 1 (BD1). The predictive performance of SCZ3-PRS for BD1 subphenotypes is improved when it is used in combination with clinical variables. The best predictions were achieved by models that integrated both genetic and clinical data. The prediction of certain BD1 traits by an SCZ-derived PRS further supports the idea of shared genetic liability between schizophrenia and bipolar disorder. SCZ3-PRS showed predictive ability for specific BD1 subphenotypes, including psychosis (especially mood-incongruent psychosis), and the age of onset of the disorder and its mood episodes. An inverse relationship was observed between SCZ3-PRS loading and the rapid cycling phenotype, suggesting a potentially different genetic influence on this specific feature of BD1. The study identified several biological pathways associated with psychosis in BD1, offering potential targets for future research into the underlying mechanisms. The findings underscore the complex, multifactorial nature of BD1 and highlight the importance of considering both genetic and clinical information for better prediction and understanding of the disorder. SCZ3-PRS might be used in the clinical counselling for BD1 treatment since previous studies using SCZ-PRS derived from an older SCZ-GWAS [4] showed that a high burden of SCZ-PRS is associated with poor response to antipsychotic and lithium treatment [12, 69-70].

4.7 Supplementary Materials

Table 37 Comparison of clinical traits in BD1 cases across samples

Variable	Overall, N = 1,878	Sample	
		RO = 574	UK = 1304
Sex			
Male	38% (718/1,878)	38% (216/ 574)	38% (502 /1,304
Female	62%(1,160/ 1,878)	62% (358 / 574)	62% (802 /1,304)
Age-at-interview	M =47 y (sd =13)	M = 42y (sd=13)	M = 49y (sd=13)
Age-of-onset BD1	M =25y (sd=10)	M=27y (sd=10)	M = 25y (sd=10)
Psychosis			
No	28% (515/1,878)	16% (92 / 574)	34% (441/ 1,304)
Yes	71% (1,321/1,878)	84% (482 / 574)	66%(849 /1,304)
Missing data	0.75% (14/1864)	0 %	1% (14/1290)
Age-onset-mania	M=29y (sd=11)	M=31y (sd=11)	M=28y (sd=11)
Age onset depression	M=25y (sd=12)	M=25y (sd=13)	M=25y (sd=11)
Rapid cycling			
No	63% (1171/1,878)	90% (519/574)	51% (652/1304)
Yes	16%% (309 /1,878)	10% (55 /574)	19% (254/1304)
Missing data	21% (398/1878)	0 % (0)	30% (398/1304)
Irritable mania			
No	59% (1,110/ 1878)	41% (233 / 574)	67% (877 / 1304)
Yes	19.5% (366 / 1878)	59% (341 / 574)	2% (25 / 1304)
Unknown	21% (402/ 1878)	0	31% (402/1304)
Family history major psychoses			
No	29% (547/ 1878)	40% (231 / 574)	24% (316 / 1304)
Yes	26%(499/ 1878)	60% (343 / 574)	12% (156 /1304)
Unknown	45% (832/1878)	0	64% (832/1304)

A summary of key clinical and demographic traits for Bipolar Disorder I (BD1) cases, stratified by the Romanian (RO) and UK cohorts.

Table 38 Differentiation between BD1 cases and controls in RO sample

	RO					
pT	Beta (ß)	OR	95% CI	R2	*P*	FDR - *P*
pT-5 x 10^{-8}	ß = -.26 s.e. = .062	.77	.68-.87	.022	.000020	2.00 x 10^{-5}
pT-1 x 10^{-7}	ß = -.27 s.e. = .062	.76	.67-.96	.024	.000010	1.14 x 10^{-5}
pT-1 x 10^{-6}	ß = -.31 s.e.=.062	.74	.65-.83	.030	9.19 x 10-^{7}	1.23 x 10^{-6}
pT-1 x 10^{-5}	ß =-.36 s.e. = .063	.70	.62-.79	.040	1.79x 10- ^{10}	3.59 x 10^{-10}
pT-1 x 10^{-4}	ß = .41 s.e.= .064	1.50	1.33-1.70	.051	1.79 x 10-^{10}	2.87 x 10^{-10}
pT-0.001	ß = .45 s.e. =.065	1.57	1.38-1.79	.062	2.46 x 10-^{12}	6.55 x 10^{-12}
pT- 0.01	ß = -.49 s.e.= .065	.61	0.54-0.70	.071	6.48 x 10-^{14}	5.18 x 10^{-13}
pT-0.05	ß = -2.979 s.e.=.41	.051	0.02-0.11	.066	4.91 x 10-^{13}	1.96 x 10^{-12}

Results of logistic regression analyses at various *P*-value thresholds (pT), showing the predictive power of the SCZ3-PRS in distinguishing BD1 cases from controls in the Romanian (RO) sample.

Table 39 Differentiation between BD1 cases and controls in UK sample

	UK						
SCZ3-pT	OR	SE	95% CI	P	R2	Liab.Ng.R2	FDR-P
pT-5E-08	1.770	0.036	1.6 - 1.90	2.79E-58	0.00003	0.000015	3.72E-58
pT-1E-07	1.760	0.036	1.64 - 1.88	9.10E-57	0.002	0.001121	9.10E-57
pT-1E-06	1.760	0.036	1.64 - 1.89	5.67E-57	0.001	0.000307126	6.48E-57
pT-1E-05	1.810	0.036	1.69 - 1.95	3.42E-61	0.037	0.01949031	5.47E-61
pT-1E-04	2.470	0.046	2.26 - 2.71	1.48E-85	0.149	0.08399165	1.18E-84
pT-0.001	2.800	0.056	2.52 - 3.14	1.64E-75	0.163	0.09230233	4.37E-75
pT- 0.01	2.800	0.056	2.52 - 3.14	1.04E-63	0.163	0.092302	2.08E-63
pT-0.05	3.100	0.052	2.74 - 3.56	3.57E-79	0.173	0.098845	1.43E-78

Results of logistic regression analyses at various *P*-value thresholds (pT), showing the predictive power of the SCZ3-PRS in distinguishing BD1 cases from controls in the UK sample.

Prediction of family history for major psychoses in BD1 probands

RO						
pT	Beta	S.E.	Wald	OR	95% CI	*P*-value
pT-5x10^{-8}	0.154	0.088	3.084	1.167	0.982; 1.386	0.079
pT-1x10^{-7}	0.176	0.088	4.001	1.192	1.004; 1.416	0.045
pT-1x10^{-6}	0.175	0.088	3.998	1.192	1.003; 1.415	0.046
pT-1x10^{-5}	0.135	0.086	2.471	1.145	0.967; 1.355	0.116
pT-1x10^{-4}	0.052	0.083	0.393	1.053	0.895; 1.239	0.531
pT-0.001	0.708	0.281	6.377	2.031	1.17; 3.51	0.012
pT- 0.01	0.101	0.087	1.344	1.106	0.933; 1.311	0.246
pT-0.05	0.542	0.544	0.993	1.719	0.592; 4.988	0.319

Results from a pT + clump analysis assessing the association between SCZ3-PRS and a family history of major psychoses in BD1 probands from the Romanian (RO) sample.

Table 40 GWAS genes associated with psychosis included in the enriched pathways

Pathway	Associated GWAS Genes
1. MITOCHONDRION	*BRD8, TRIM31, CKB, SFXN2, CLU, CLIC1, GLYCTK, SLC9B2, LETM2, CREB3L4, METTL15, MARK2, ALAS1, FEN1, FHIT, MLXIP, FOXO3, HARS2, GABBR1, HSPA9, HSPD1, HSPE1, IRF3, YJEFN3, FADS1, MAPT, MSRA, NDUFA2, NRGN, CISD2, PCCB, TMX2, MRPS33, PLEC, POLG, MIEF1, NDFIP2, DARS2, NDUFAF7, AMBRA1, DNAJC11, MAPK3, STARD7, ELAC2, SDHAF1, NDRG4*
2. ZNF318	*CAMKK2, PTK2B, CHRNA2, MATN4, SPECC1, SYNE1, TMTC1, DOCK2, OASL, SLC39A1, GPM6A, TDRD9, CDC25C, NEGR1, TBL1XR1, ZNF664, HARBI1, NMB, CTNND1, TCTN1, MEF2C-AS1, MYO19, MAD1L1, DNAJC11, PLEKHO1, SDCCAG8, SMG6, IGSF9B, WDR76, FOXP1, DARS2, ENOX1, KDM3B, TBC1D5, UBE2D2, RC3H1, SEC11A, RERE, CREB3L4, GATAD2B, DOC2A, MSI2, SPPL3, ZEB2, ATG13, GRIN2A, DLGAP2*
3. REGULATION_OF_IMMUNE_SYSTEM_ PROCESS	*PLK2, RC3H1, DRD2, PLCL2, SCRIB, HSPA9, MDK, FURIN, YTHDF2, MAD1L1, CUL4A, DGKZ, CD47*
4. NCOA2 TARGET GENES	*ITIH1, NT5C2, PACSIN2, OGFOD2, HSPA9, IPO13, ALOX5AP, DYNC1LI2, RBMS3, ABCB9, RBKS, RELA, AGPAT1, SNHG3, MEF2C, MSANTD2*
5. MIR202_3P	*MEF2C, PLEKHO1, PTPRD, BCL7A, ZNF823, SHANK2, MOB4, MSI2, SH3RF3, RD3L, HSPE1-MOB4, TSPAN2*
6. MIR3125	*ELAVL4, ZNF365, RC3H1, CUL4A, ANKRD45, ARL3, SUFU, TRIM8*
7. MIR6859_5P	*ELAVL4, RC3H1, NEBL, ARL3, SUFU, TRIM8*
8. MIR4782_5P	*ADD3, ALAS1, RPS6KA2, TCTN1, CALN1, SUMO2, DNMT3A, DYNC1LI2*
9. MIR5706	*ADD3, ALAS1, RPS6KA2, TCTN1, CALN1, SUMO2, DNMT3A, DYNC1LI2*
10. MIR4763_3P	*SLC6A9, DEF8, ETF1, NGEF, KIF21B, TAF12, SUFU, GATAD2B, KLF6, STAG1, MLXIP, SMARCD1, MEF2C, MARK2*
11. ABNORMALITY_OF_THE_AUTONOMIC _NERVOUS_SYSTEM	*MAPT, CACNA1C, ZEB2, CHRNA3, FGFR1, TUBB3, GIGYF2, ARL3, TCTN1, SNCA, GABBR2, FANCI, FANCA, FANCL, CISD2, SUFU*
12. MIR10395_3P	*ADD3, ATXN7, CHRNA5, KLF6, MOB4, HSPE1-MOB4, CADM2, ACE*
13. REGULATION_OF_DENDRITIC_SPINE _DEVELOPMENT	*NGEF, MEF2C, SHANK3*
14. MIR197	*BCL7A, CTNND1, GATAD2B, SPPL3, ETF1*
15.REGULATION_OF_MEMBRANE_ REPOLARIZATION	*YWHAE, AKAP6*
16. APOPTOSIS	*BNIP3L, CLU, RELA, DPYD, ETF1*
17. CHR1P21	*DPYD, NFU1P2, RPL7P9, RN7SKP270, PTBP2, MIR137HG*
18. REGULATION_OF_DOPAMINE_RECEPTOR _SIGNALING_PATHWAY	*DRD2*
19. MIR625_3P	*WDR76, PTPRD, ALAS1*
20. MIR3681_5P	*CSDE1, FUT10*
21. MIR6849_5P	*CSDE1*
22. MIR4669	*CLIC1, GIGYF2*

Genes associated with psychosis were included in the 22 pathways of SCZ3-PRS identified in the RO/UK sample that overlapped with genes found associated in the most recent GWAS of bipolar disorder [41] and schizophrenia [13].

The focused analysis in Chapter 4 confirmed that while specific transdiagnostic genetic risks, like the SCZ PRS, can predict severe features such as psychosis in BD, it also highlighted that different clinical specifiers (e.g., psychosis versus rapid cycling) have divergent genetic associations. This evidence necessitates a broader approach to systematically map the multifaceted genetic landscape of bipolar disorder, as its considerable clinical heterogeneity suggests various subphenotypes may possess distinct, as well as shared, genetic underpinnings. Therefore, Chapter 5 undertakes a large-scale multi-trait analysis of Genome-Wide Association Studies (MTAG) across eleven clinically defined BD subphenotypes. Leveraging data from multiple cohorts, the objectives are to systematically replicate and assess the consistency of Polygenic Risk Score (PRS) findings for these subphenotypes and to dissect this heterogeneity by identifying specific genomic loci, genes, and biological pathways that contribute to individual clinical presentations, as well as those shared across the broader BD spectrum and with schizophrenia (SCZ) and other closely related cross-disorder traits.

---

5 Bipolar Disorder Subphenotypes

A preprint version of the research in this chapter is available on medRxiv at
doi: https://doi.org/10.1101/2025.06.23.25330155

5.1 Abstract

Background: The clinical heterogeneity of bipolar disorder (BD) is a major obstacle to improving diagnosis, predicting patient outcomes, and developing personalized treatments. A genetic approach is needed to deconstruct the disorder and uncover its fundamental biology. Previous genetic studies focusing on broad diagnostic categories have been limited in their ability to parse this complexity.

Aims: The aim was to test the hypothesis that clinically distinct subphenotypes of BD are associated with different underlying common variant genetic architectures.

Methods: This multicentred study included a primary genome-wide association study (GWAS) of up to 23, 819 bipolar disorder (BD) cases and 163, 839 controls. These results were integrated via multi-trait analysis of GWAS (MTAG) with external summary statistics for BD (59, 287 cases; 781, 022 controls) and schizophrenia (SCZ; 53, 386 cases; 77, 258 controls). Sample overlap was statistically accounted for.

Results: The primary outcomes were the genetic dimensions underlying BD heterogeneity, differentiated by single nucleotide polymorphism (SNP)-heritability (h^2snp), genetic correlations, genomic loci ($P \leq 5\times10^{-8}$), and functional, cell-type, and gene-expression pathway analyses. Four genetically-informed dimensions of BD were identified: Severe Illness, Core Mania, Externalizing/Impulsive Comorbidity, and Internalizing/Affective Comorbidity. The analyses yielded up to 181 subphenotype-associated loci, 53 of which are novel. The Severe Illness Dimension was characterized by a unique neuro-immune signature (a protective association with *HLA-DMA*, $P=2.50\times10^{-273}$) evident only when leveraging SCZ genetic data. The Internalizing/Affective dimension was associated with neurodevelopmental genes (e.g., *DCC*). Notably, the rapid-cycling subphenotype showed a unique signature of strong negative selection, a finding not observed in other traits.

Conclusions: The clinical heterogeneity of bipolar disorder appears to be defined by a complex and multi-layered genetic architecture. The presented findings provide a data-driven, biological framework that may advance psychiatric nosology beyond its current diagnostic boundaries. The delineation of these genetically-informed dimensions offers specific hypotheses functional genomics studies for subsequent therapeutic discovery, laying the foundation for a transition from a uniform treatment model to the paradigm of precision psychiatry. Establishing this framework is an essential step toward refining diagnostic criteria, enabling patient stratification, and developing more effective, and personalized treatments.

5.2 Introduction

Bipolar disorder (BD) is a severe, chronic psychiatric illness affecting around 1% of the population. The disorder has a high heritability of over 80%, and its clinical variability complicates diagnosis, treatment, and research [1-4]. Previous work established distinct genetic overlaps between BD subtypes and other major psychiatric disorders: bipolar disorder I (BD1) shows a high genetic correlation with schizophrenia (SCZ) [3, 5-8], while bipolar disorder II (BD2) links more strongly to major depressive disorder (MDD) and attention-deficit/hyperactivity disorder (ADHD) [2, 6, 9]. This overlap indicates that biological pathways are not constrained by diagnostic manuals, necessitating a data-driven approach to nosology. Given the genetic continuum between BD and SCZ, it was hypothesized deconstructing severe BD requires comparing its genetic architecture with SCZ's to isolate disorder-specific from transdiagnostic risk signals. This heterogeneity impacts treatment, as features including psychosis or comorbidities guide distinct pharmacological strategies, and the iterative process of personalizing an effective regimen may contribute to the illness burden [10]. A summary of these transdiagnostic profiles for key bipolar disorder subphenotypes is provided in Appendix 9.3.

This heterogeneity is evident across multiple clinical domains. Age of onset (AOO) is a critical factor; an earlier AOO typically signifies a greater genetic liability and a more severe disease trajectory [11-12]. An onset before 28 years of age increases the risk for psychotic features, rapid cycling (RC), comorbid anxiety disorders, alcohol or substance use disorders (AlcSUD), and suicide attempts (SA) [13]. RC (defined as ≥4 mood episodes/year) [14], is linked to a family history of mood instability, high psychiatric comorbidity, and a lack of responsiveness to lithium, making it a challenging clinical presentation [15-17]. The long-observed clinical association with other traits, for example, thyroid dysfunction and mood instability in RC may be a key aspect of this profile [18-19]. While preliminary studies suggest benefits from using adjunctive thyroid hormone for RC, a definitive mechanistic link remains unproven [17, 20].

To deconstruct this heterogeneity, eleven subphenotypes were selected for this analysis. These were chosen to represent key domains of the illness based on their established clinical relevance and evidence for familial aggregation, suggesting more genetically homogeneous subgroups of patients which may boost genetic discovery. They encompass core diagnostic subtypes (BD1, BD2, SZA), key course specifiers with significant prognostic value (Psychosis, RC, UM, AOO), and highly prevalent and impactful comorbidities that shape the illness course (AlcSUD, PD, OCD, SA).

5.3 Aims

Genetic research into clinically distinct BD subphenotypes has been hampered by inadequate statistical power. This study tested the hypothesis that the clinical heterogeneity of BD is linked to underlying genetic heterogeneity defined by specific biological pathways. This study employed a two-step MTAG approach, first meta-analysing single subphenotype GWAS with additional BD cases and second by integrating large-scale SCZ GWAS data, to boost power to

identify specific genetic mechanisms. This multivariate approach aimed to reveal genetic factors that confer risk for specific psychopathologies, and those that underlie the observed genetic overlaps with other major psychiatric traits. A robust genetic-clinical framework of four dimensions was first established in the clinical data before GWAS, with subsequent downstream interrogation of the unique and shared biological pathways, spanning neuro-immune, neurodevelopmental, and synaptic systems, that likely define them.

5.4 Methods

To dissect the genetic architecture of bipolar disorder, primary genome-wide association studies (GWAS) were conducted on eleven distinct clinical subphenotypes. To increase statistical power, these results were then integrated with large-scale external summary statistics for BD and schizophrenia using a two-stage Multi-Trait Analysis of GWAS (MTAG). The resulting high-power summary statistics were subjected to extensive downstream analyses, including heritability and genetic correlation estimation, functional genomics, pathway and cell-type enrichment, and transcriptome-wide association studies.

A comprehensive account of the study cohorts, all GWAS and post-GWAS analysis procedures is provided in the General Methods (Chapter 2).

5.5 Results

5.1 Foundational Analyses: Data Quality and Phenotypic Validation

This study included 52% females, with a median age at interview of 22 (interquartile range [IQR], 17-30) years. Clinical characteristics are detailed in Chapter 2, Tables 4-5.

To ensure that phenotype definitions were consistent across international cohorts, an assessment of phenotypic homogeneity was performed. Generalized linear mixed effects (GLMER) models were constructed using pairwise analyses of BD subphenotypes to assess phenotype heterogeneity across geographical sites, termed ‘Region,’ which was used as the random effect (N = 18,800 BD cases). ‘Region’ included symptom-level data from cohorts across Australia, Europe, North America, or Nordic countries. Confidence intervals (95% CI) of predicted probabilities (y-axis) overlapped, indicating homogeneous responses to target phenotypes (x-axis) which met international consensus measures (DSM-IV, DSM-V, ICD-9, or ICD-10).(See Table 41 and Figure 19 below).

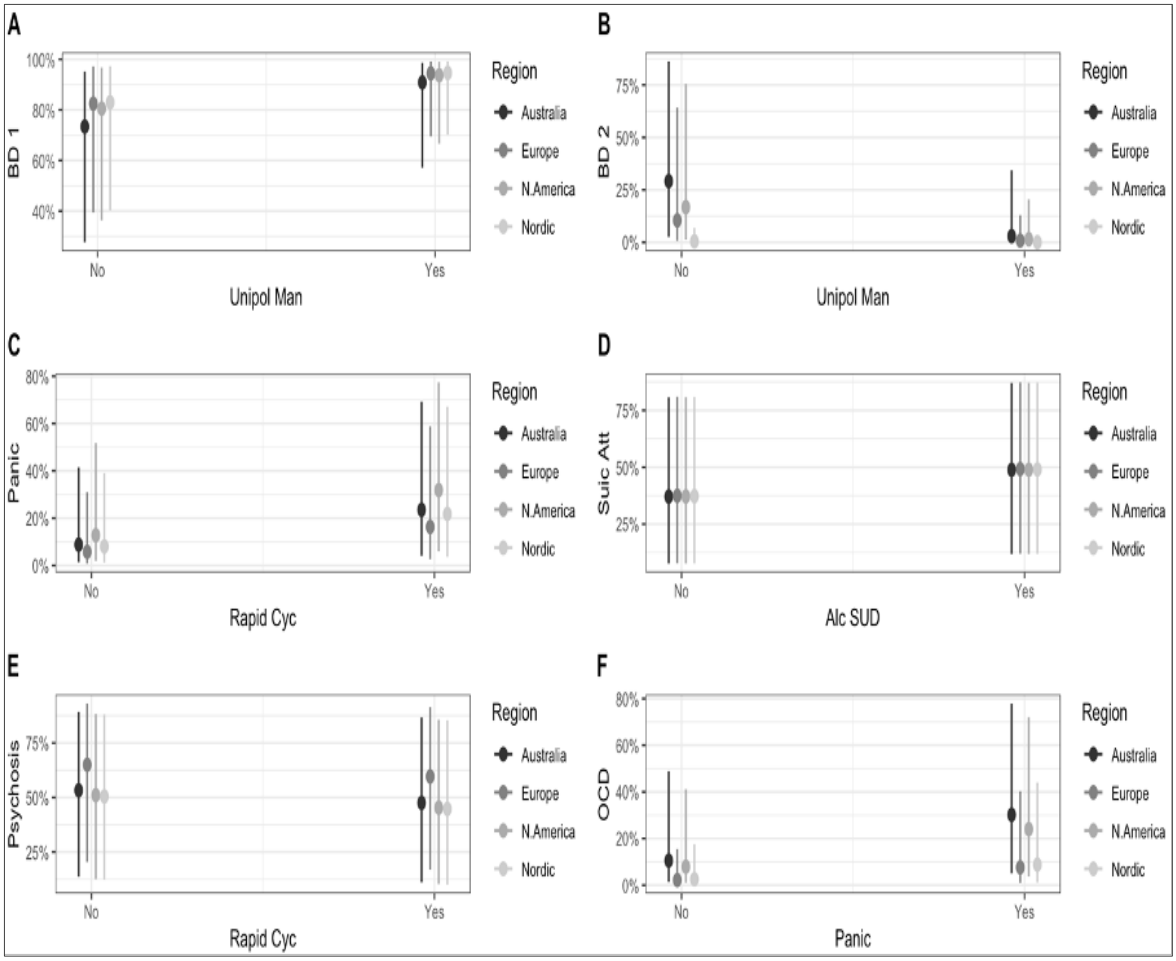


Figure 19 Mixed regression models of homogeneity in phenotype regions.
Generalized linear mixed effects (GLMER) models were constructed using pairwise analyses of BD subphenotypes to assess phenotype heterogeneity across geographical sites, termed 'Region,' which was used as the random effect (N = 18,800 BD cases).

Table 41 Assessment of Phenotypic Homogeneity Across Geographic Regions

BD	BD1	BD2	SZA	PSY	RC	UM	SA	ALC	PD	OCD
PSY	1.09 (.04) ***	-1.95 (.06) ***	2.58 (.20) ***	-	-.25 (.05) ***	.43 (.10) ***	.04 (.05)	.24 (.04) ***	-.12 (.06) *	.02 (.08)
RC	-.57 (.06) ***	.51 (.06) ***	.74 (.17) ***	-.26 (.05) ***	-	-1.92 (.29) ***	.60 (.06) ***	.42 (.06) ***	1.19 (.08) ***	1.09 (.12) ***
UM	1.29 (.18) ***	-2.63 (.41) ***	-.25 (.23)	.41 (.10) ***	-1.92 (.29) ***	-	-1.07 (.11) ***	-.17 (.10)	-.31 (.14) **	-.47 (.22) *
SA	-.01 (.06)	-.11 (.06)	.35 (.12) **	.03 (.05)	.58 (.06) ***	-1.07 (.11) ***	-	.49 (.05) ***	.51 (.07) ***	.36 (.10) ***
ALC	.007 (.05)	-.12 (.06)	.45 (.11) ***	.23 (.04) ***	.41 (.06) ***	-.18 (.10)	.49 (.05) ***	-	.50 (.06) ***	.25 (.08) **
PD	-.22 (.07) **	.15 (.08) *	.72 (.18) ***	-.13 (.06) **	1.19 (.08) ***	-.34 (.14) *	.49 (.07) ***	.50 (.06) ***	-	1.33 (.08) ***
OCD	-.08 (.10)	.12 (.10)	-.21 (.31)	.01 (.08)	1.08 (.12) ***	-.49 (.22) *	.34 (.10) ***	.25 (.08) **	1.33 (.08) ***	-
AOO	-.002 (.00) **	.008 (.00) ***	-.024 (.01) ***	-.017 (.00) ***	-.027 (.00) ***	.023 (.00) ***	-.030 (.00) ***	-.027 (.00) ***	-.024 (.00) ***	-.017 (.00) ***
AO-depr	.009 (.00) **	-.006 (.00) ***	-.029 (.01)	-.013 (.00) ***	-.045 (.00) ***	.023 (.01)	-.027 (.00) ***	-.023 (.00) ***	-.046 (.00) ***	-.027 (.01) ***
AO_man/mix	.021 (.00) ***	-.019 (.00) ***	-.021 (.01)	-.017 (.00) ***	-.034 (.00) ***	-.008 (.01)	-.014 (.00) ***	-.021 (.00) ***	-.022 (.00) ***	-.022 (.01) **

Following these quality control steps, a Confirmatory Factor Analysis (CFA) of the 11 BD subphenotypes empirically derived a robust four-factor clinical model, which indicated acceptable fit indices ($\chi 2$=588.91, P=2.188×10^{-87}; SRMR .084; CFI .936) (Figure 20). Factor analysis was performed using the psych package in R to produce a visualisation of the homogeneous subgroups (subphenotypes) and their interrelatedness. The analysis included clinical data from N = 18,800 BD cases. The factor analysis supported four primary clinical factors for BD heterogeneity, providing an initial framework for understanding BD clinical substructure. The model identified: (1) a Psychosis-Spectrum Factor (schizoaffective disorder, bipolar type [SZA], Psychosis); (2) a Core Bipolar Subtype Factor (BD1, BD2); (3) a Comorbidity and Mood Instability Factor (RC, Panic Disorder [PD], Obsessive Compulsive Disorder [OCD], Alcohol or Substance Use/Dependency [AlcSUD], Suicide Attempt [SA], Unipolar mania [UM] [40-42]); and (4) an Age of Onset [AO/AOO] Factor. This four-factor structure determination was supported by parallel analysis (Figure 24). Support for two dimensions which differentiated risk for Psychosis and Comorbidity was provided further by Principal Component Analysis (PCA) (Figure 21).

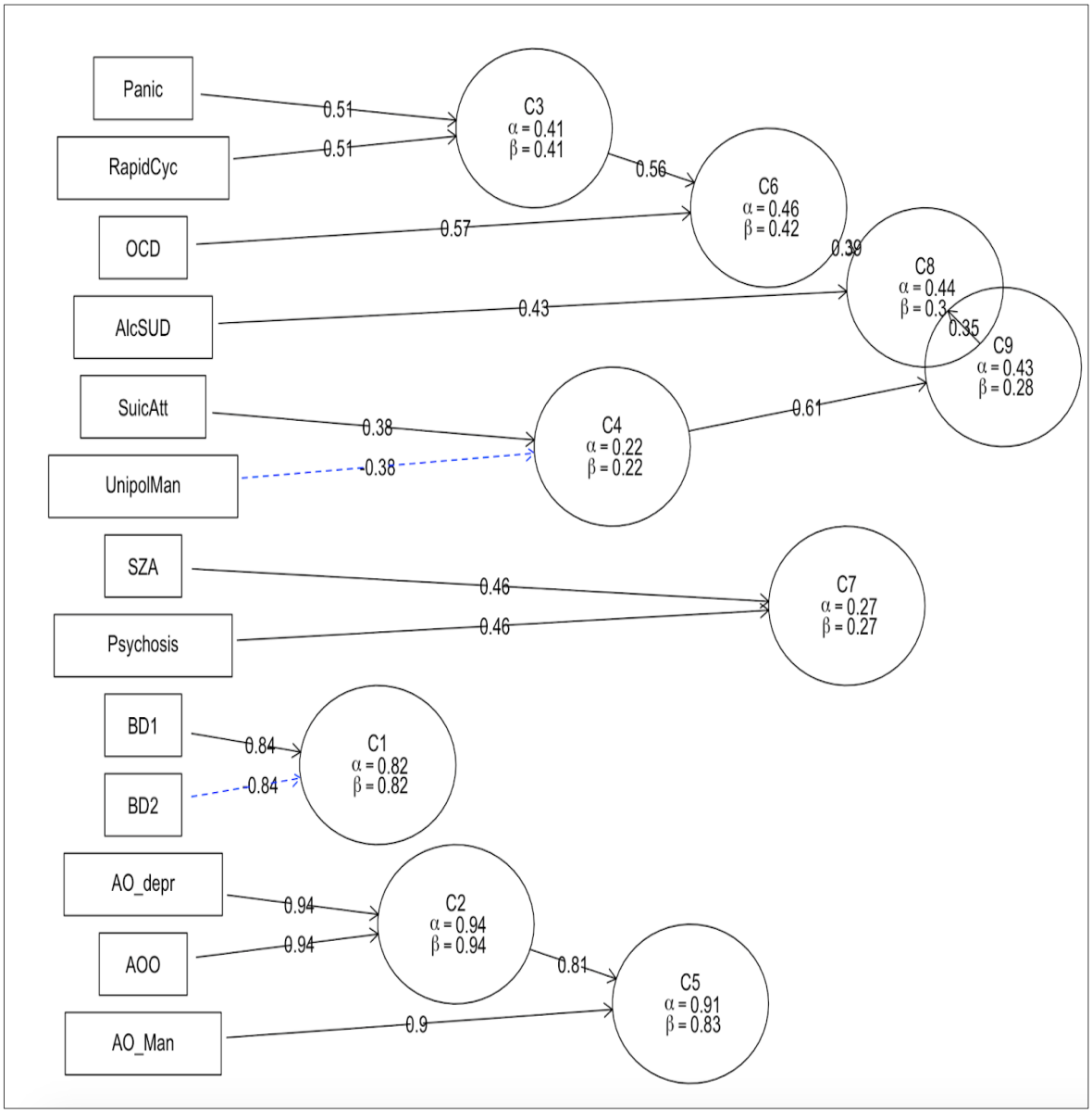


Figure 20 Confirmatory Factor Analysis (CFA) model for BD heterogeneity.
Factor analysis was performed using the psych package in R to produce a visualisation of the homogeneous subgroups (subphenotypes) and their interrelatedness. The analysis included N = 18,800 BD cases. The factor analysis supported four primary clinical factors for BD heterogeneity: (F1) SZA and Psychosis; (F2) BD1 and BD2; (F3) a cluster of RC, PD, OCD, AlcSUD, SA, and UM; and (F4) AOO, AO-depression, and AO-mania/mixed. A Confirmatory Factor Analysis (CFA) of the 11 BD subphenotypes empirically derived this robust four-factor clinical model, which indicated acceptable fit indices (χ2=588.91, P=2.188×10-87; SRMR .084; CFI .936).

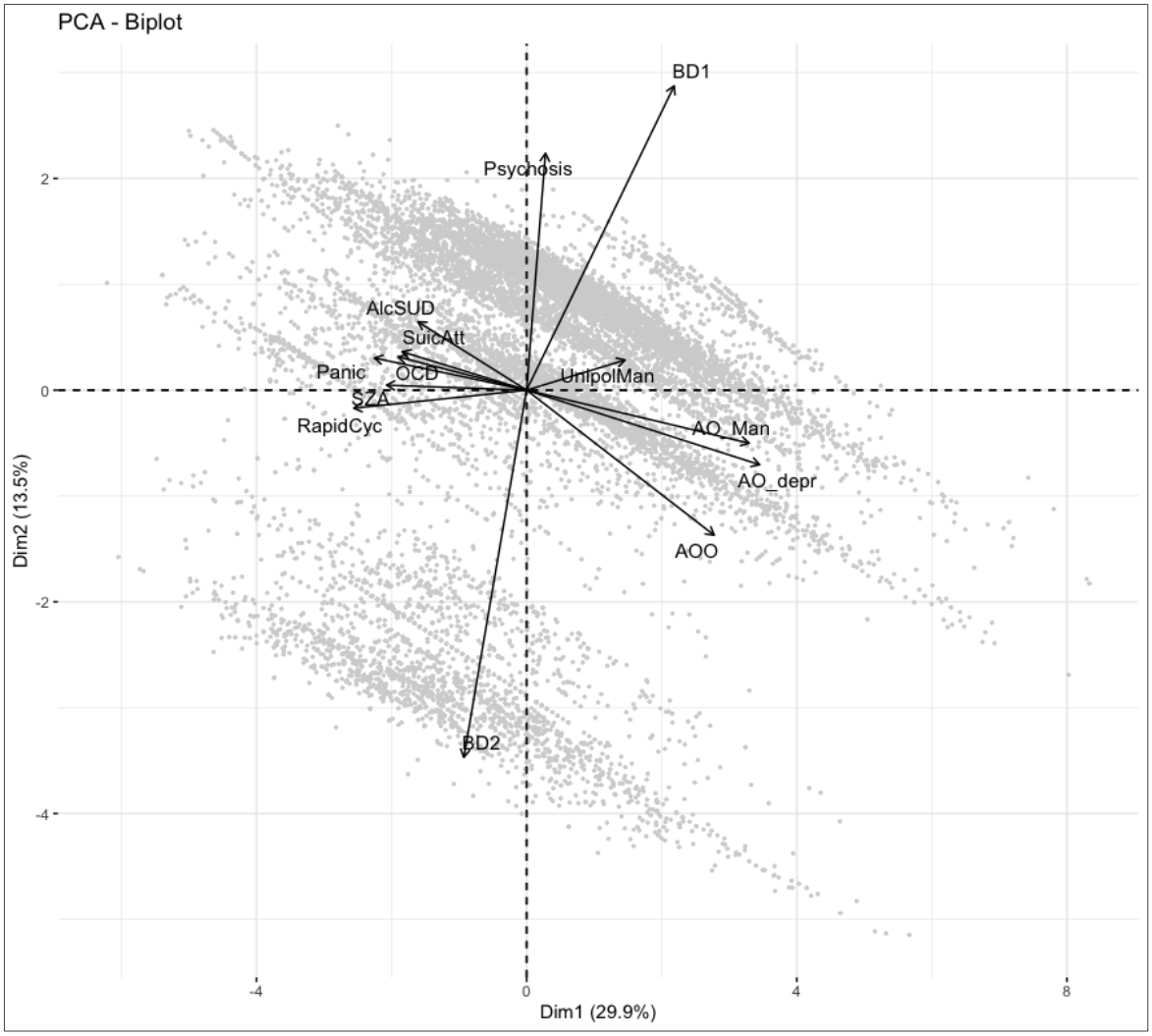
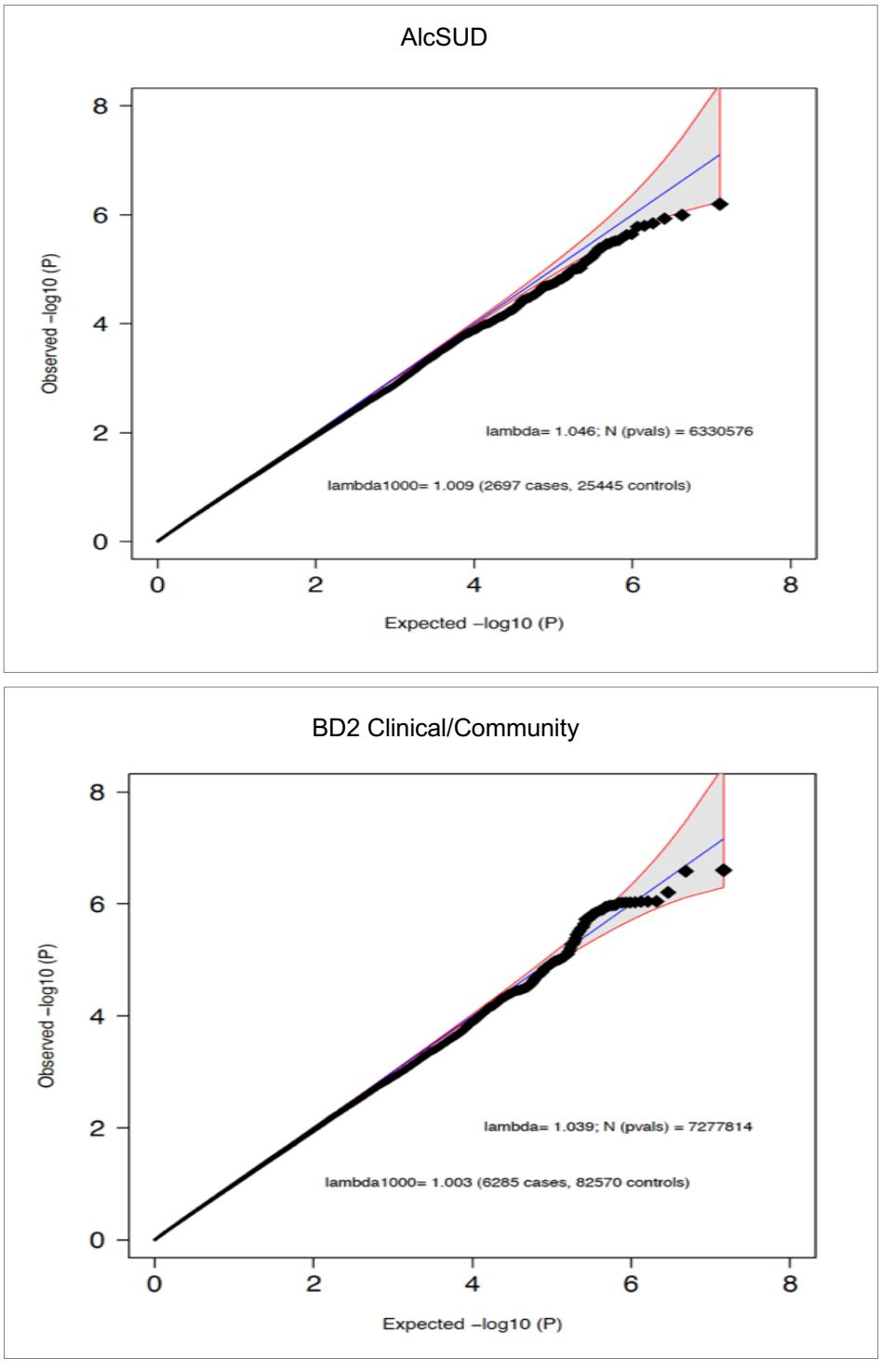


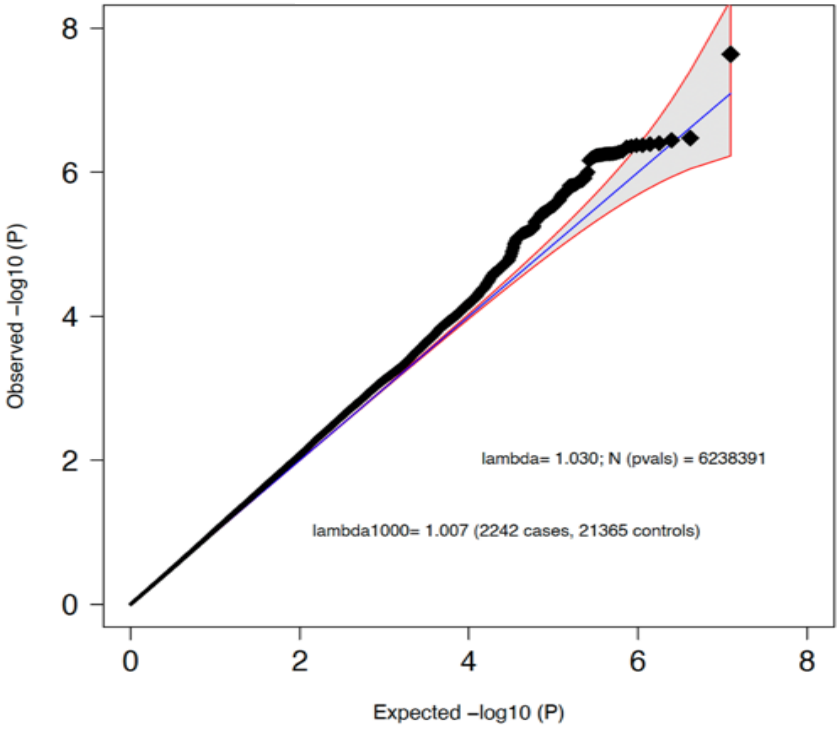
Figure 21 PCA visualization of 11 BD subphenotypes.
PCA visualization of 11 BD subphenotypes, showing clustering. This figure visualizes how components account for variance in the dataset. See Supplementary Table 39 for per-cohort sample sizes for each subphenotype analysis.

5.2 Identification of Four Genetically-Informed Dimensions

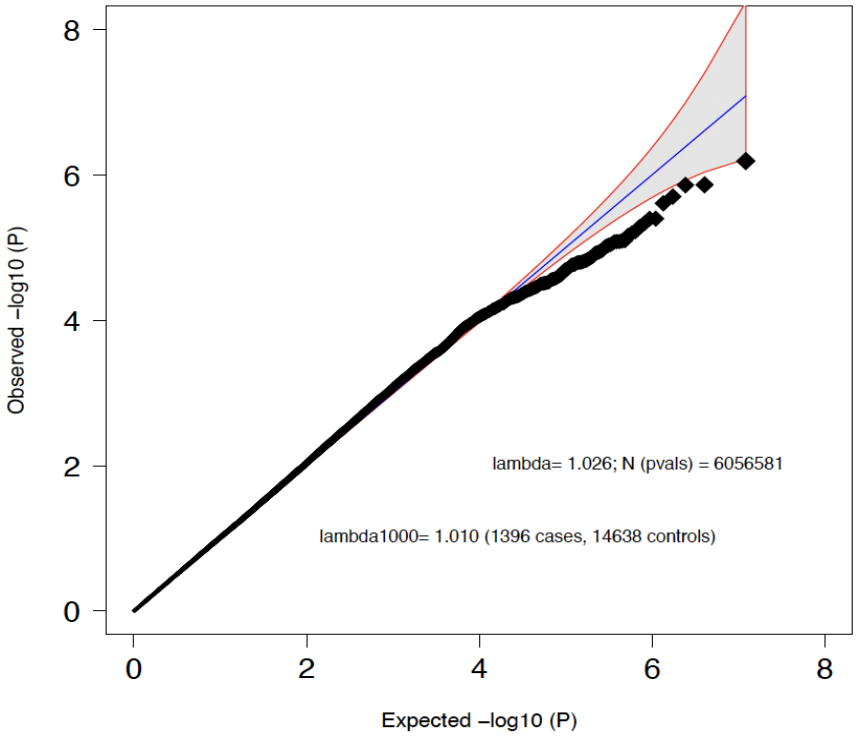
Before dissecting the genetic architecture of bipolar disorder (BD), foundational analyses were conducted to ensure the integrity of the data. The primary genome-wide association studies (GWAS) of eleven clinical subphenotypes showed minimal confounding from uncorrected population stratification or cryptic relatedness, as indicated by Quality Control (QC) and the genomic inflation (GC) factors (λGC) shown in the QQ plots which were close to 1 (Figure 22).



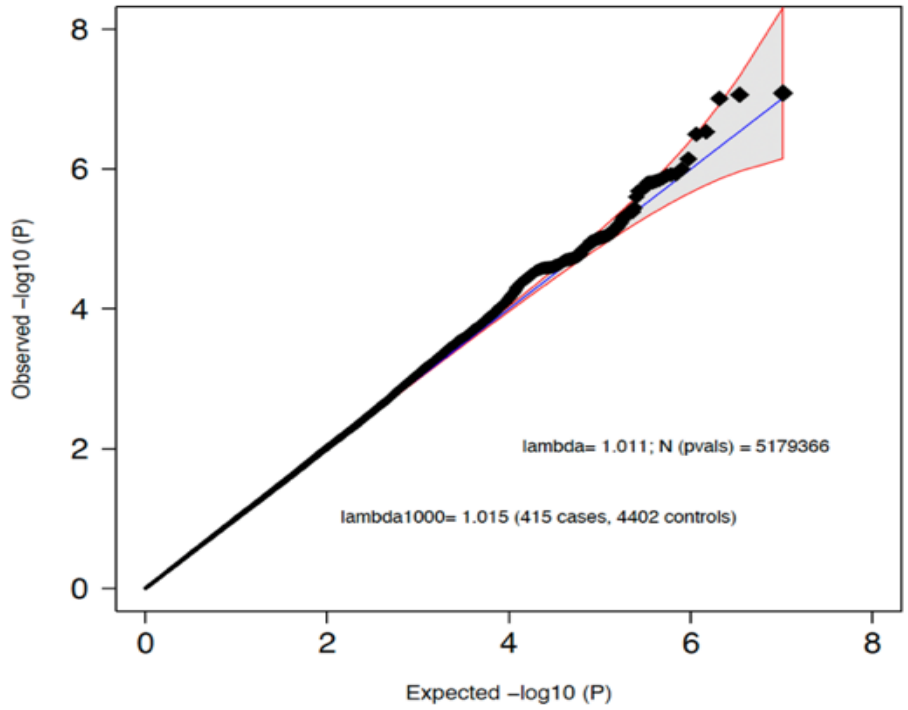
Rapid cycling (RC)



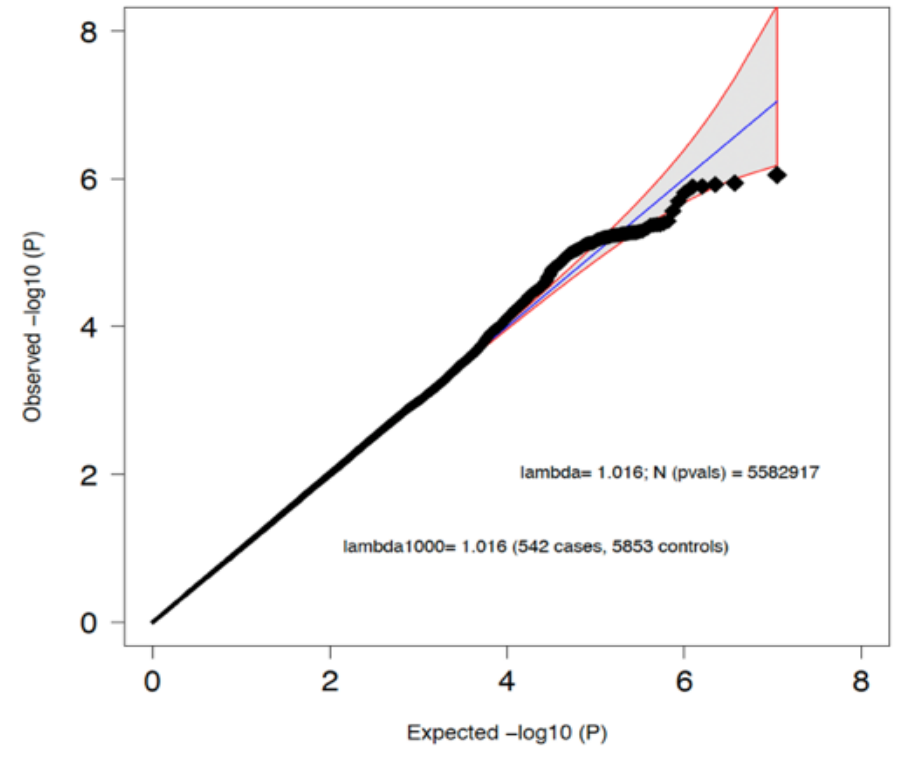
Panic disorder (PD)



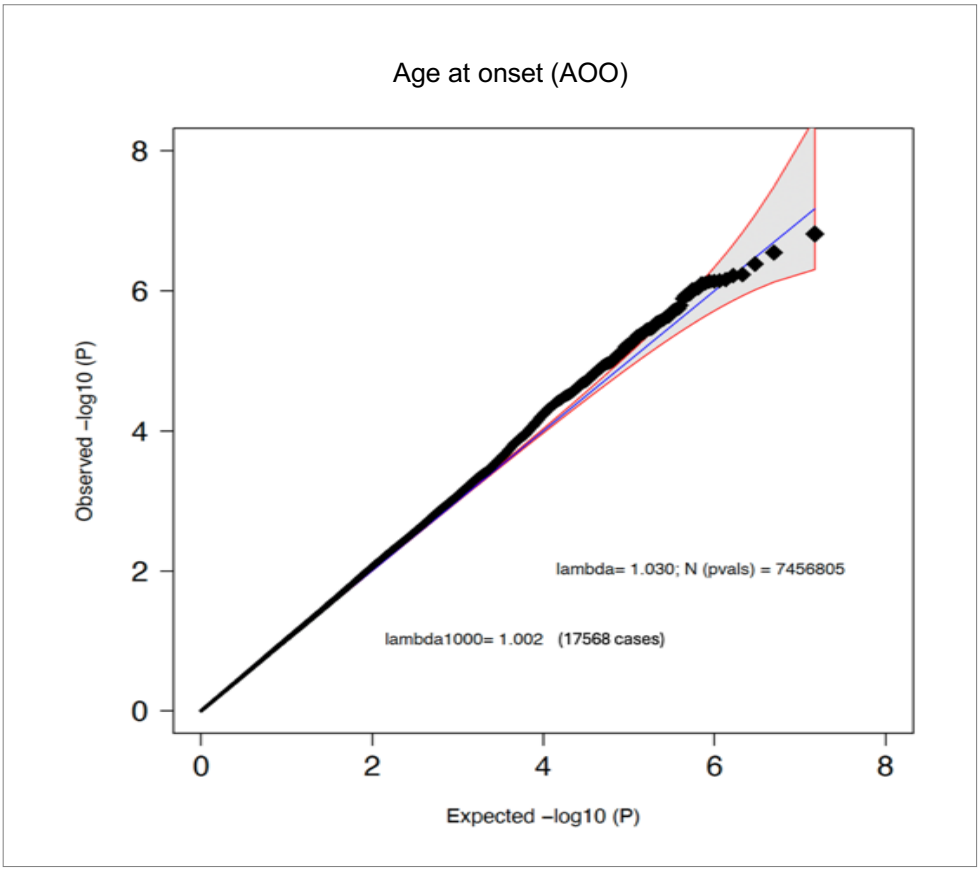
Unipolar mania (UM)



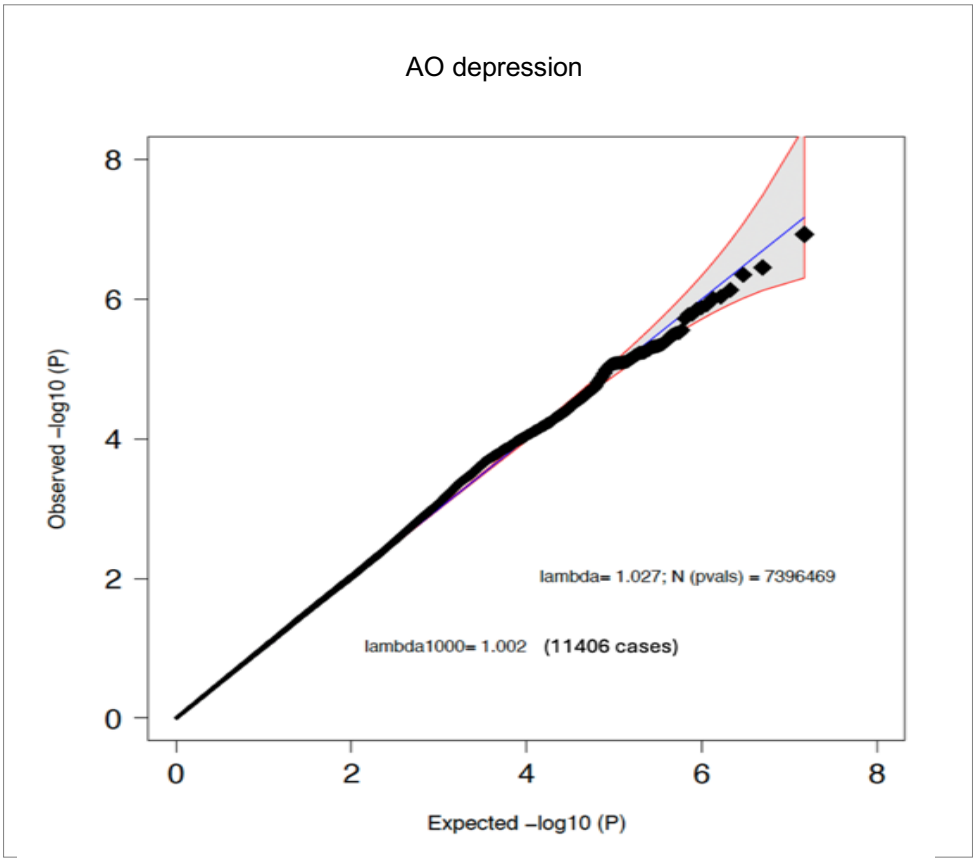
OCD



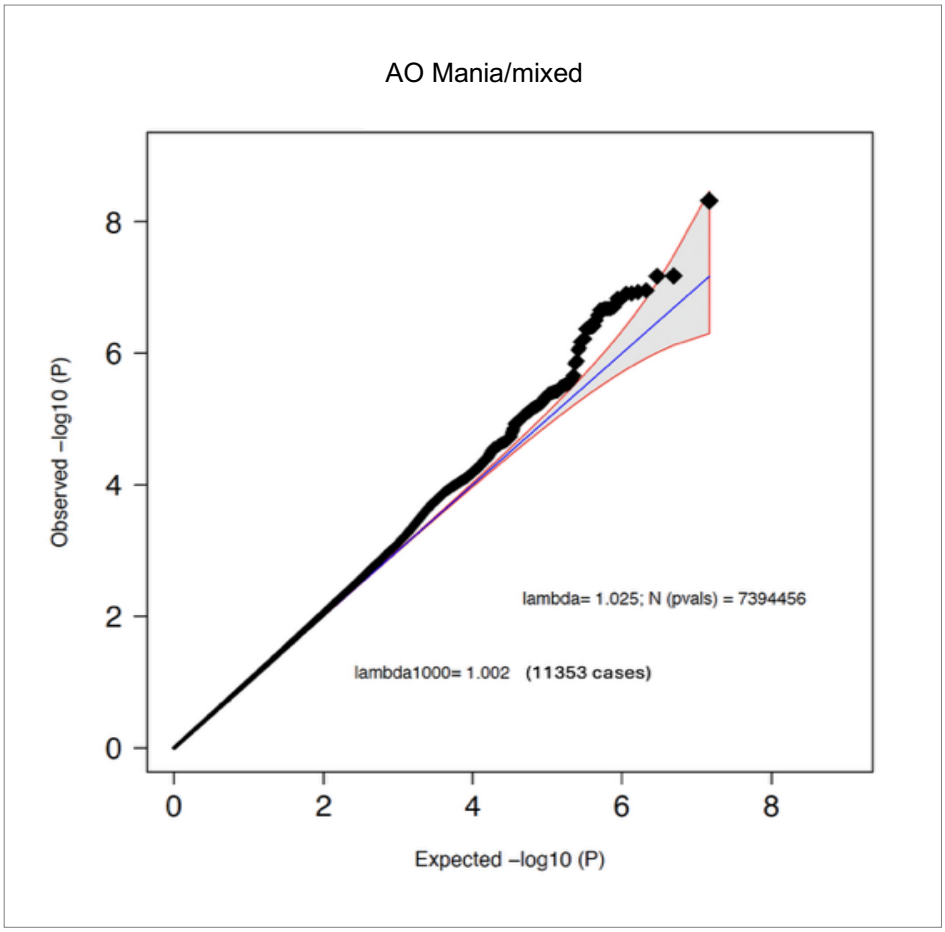
Age at onset (AOO)



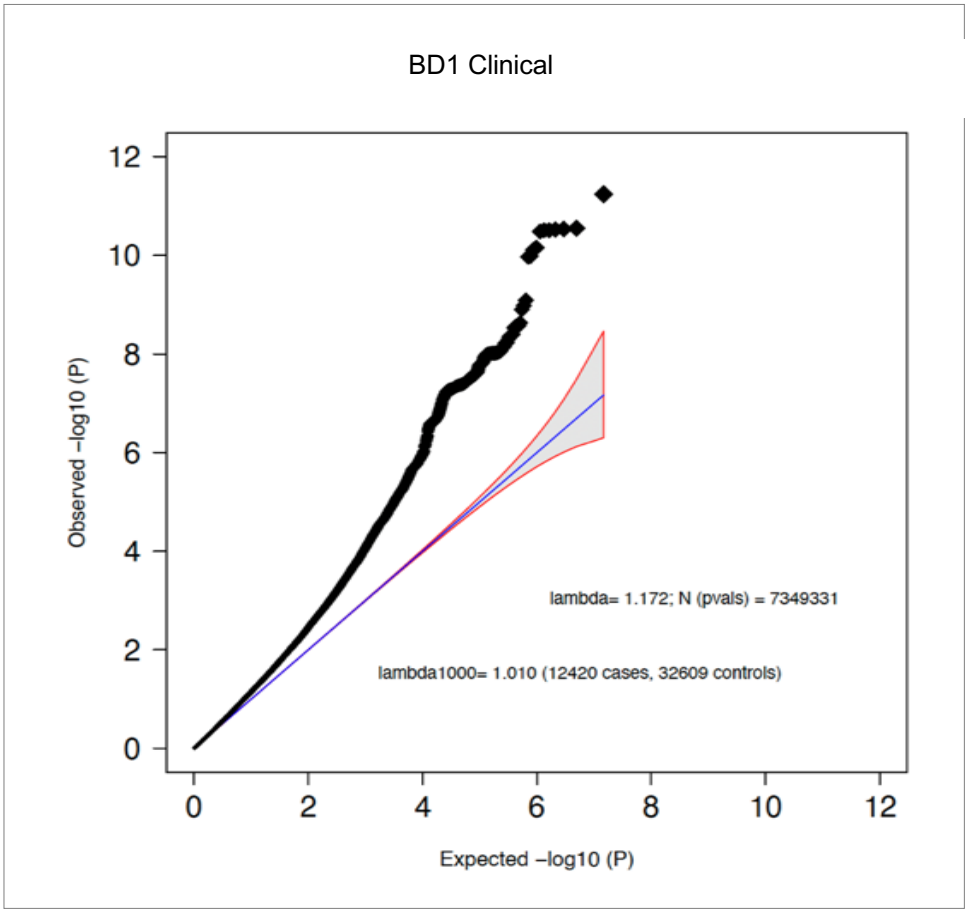
AO depression

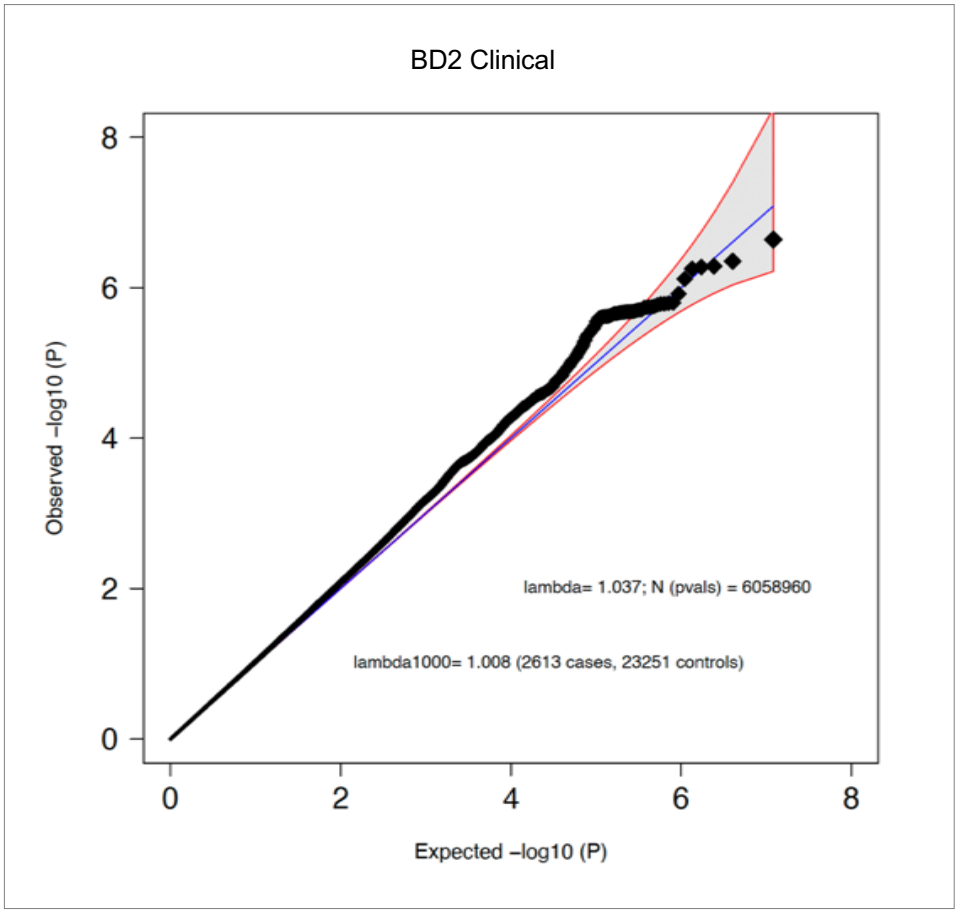


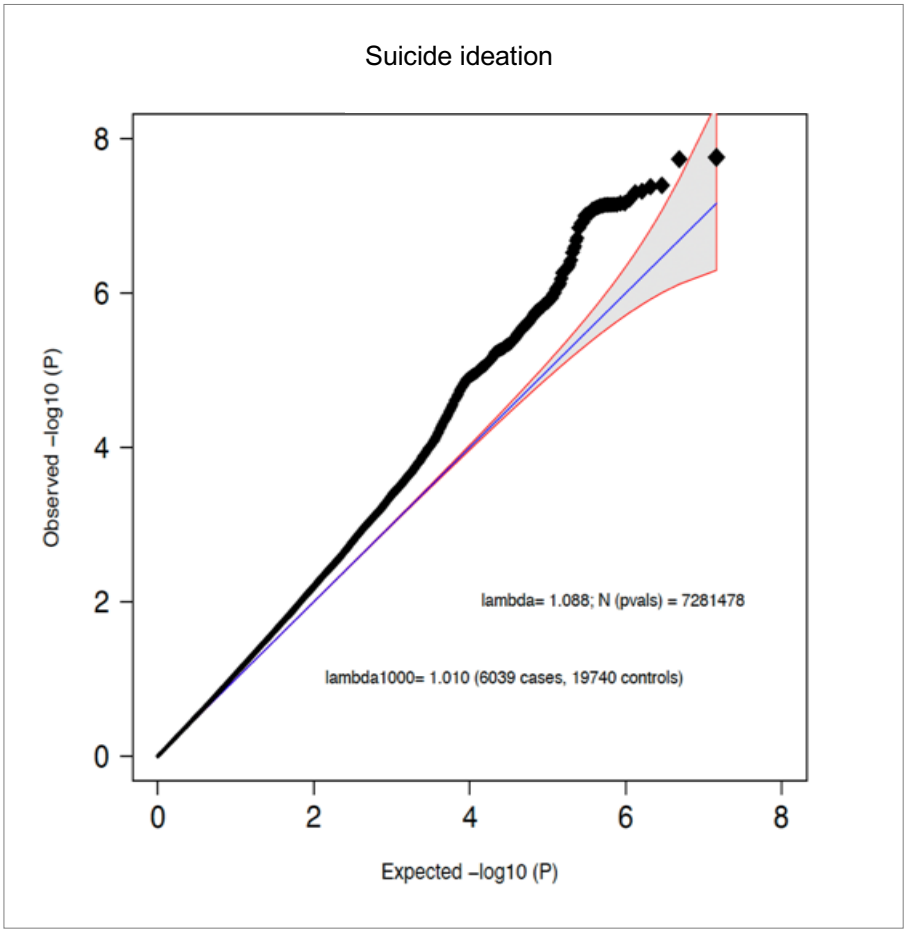
AO Mania/mixed



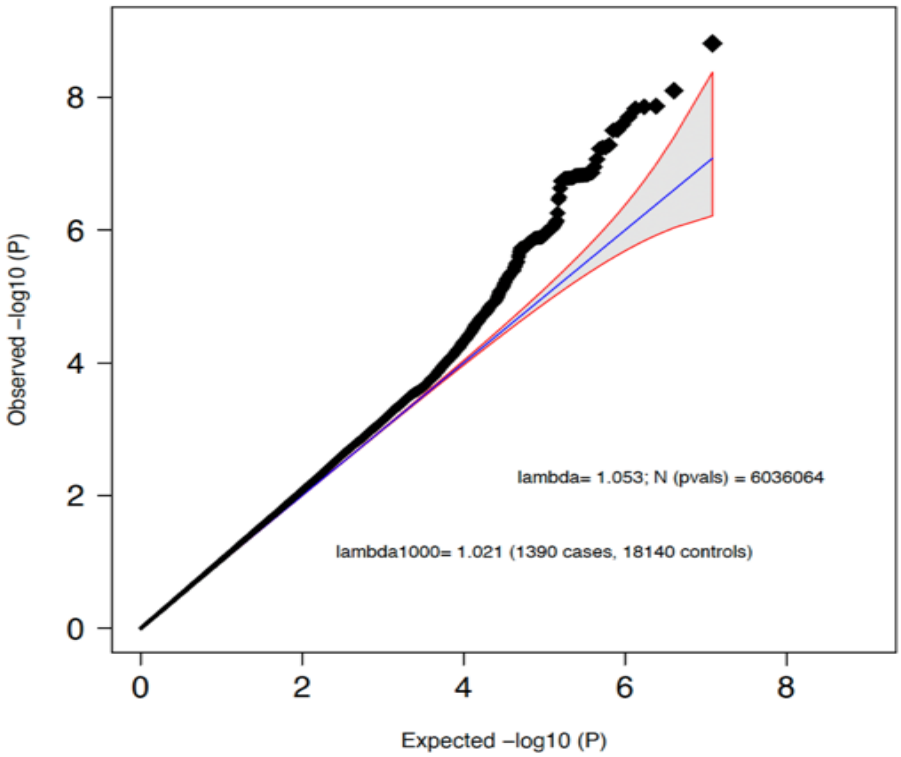
BD1 Clinical



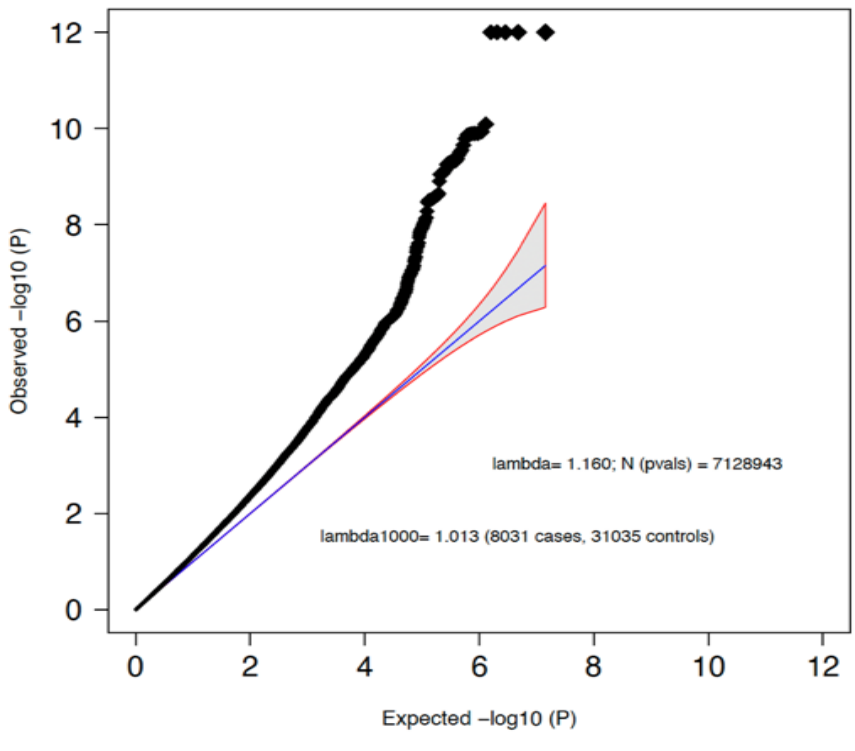

BD2 Clinical
Observed −log10 (P)
Expected −log10 (P)
lambda= 1.037; N (pvals) = 6058960
lambda1000= 1.008 (2613 cases, 23251 controls)
0
2
4
6
8


Suicide ideation
Observed −log10 (P)
Expected −log10 (P)
lambda= 1.088; N (pvals) = 7281478
lambda1000= 1.010 (6039 cases, 19740 controls)
0
2
4
6
8

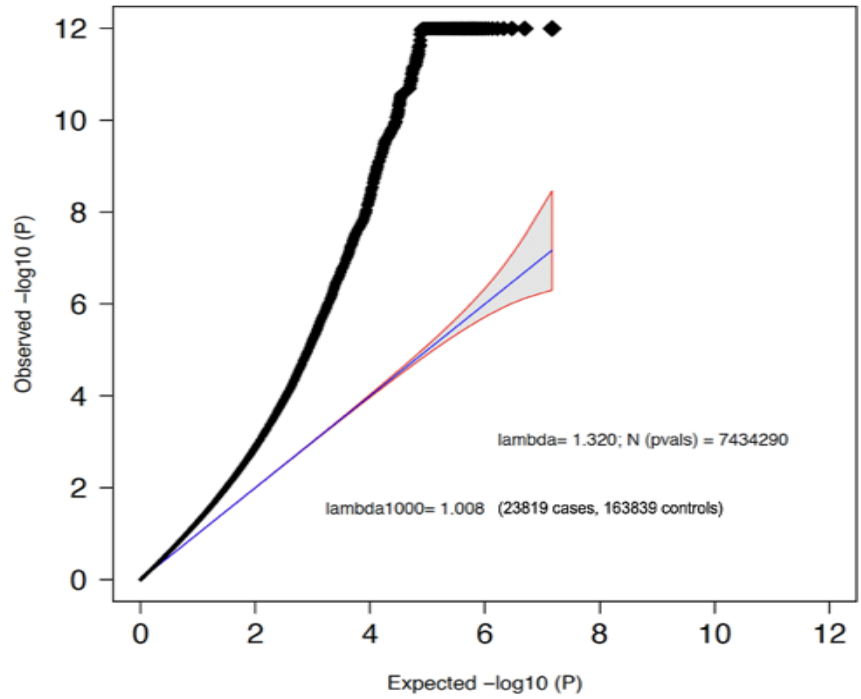
SZA (Schizoaffective)



Psychosis



BD1 Clinical/Community)



Suicide attempt (SA)

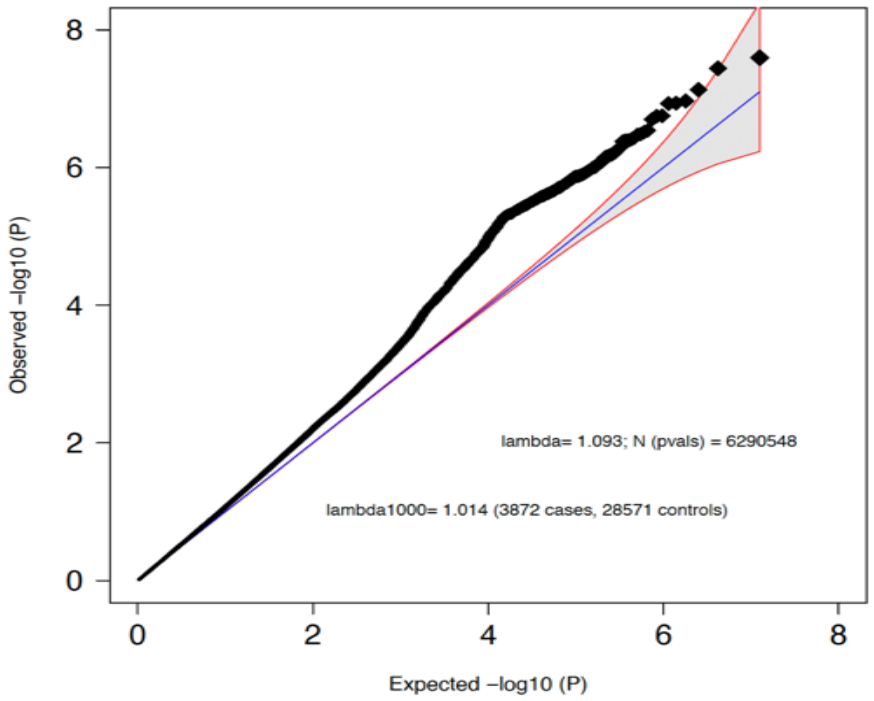
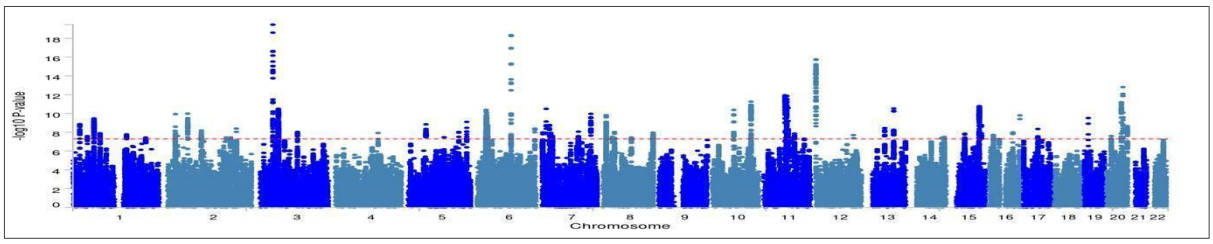


Figure 22 QQ plots for each of the 11 core subphenotype-GWAS

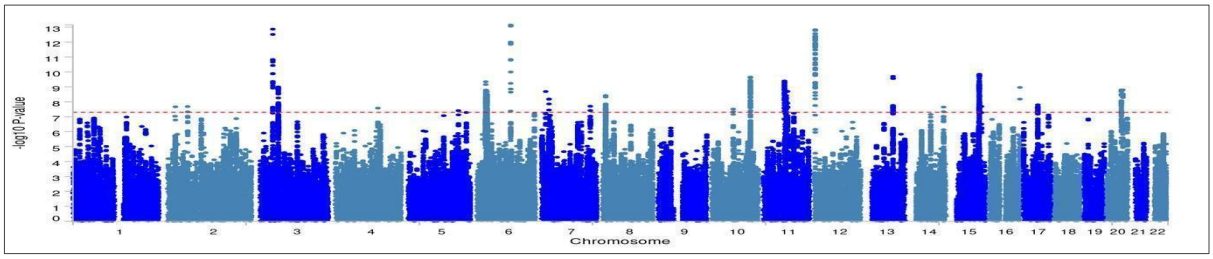
Each plot shows the observed -log10(*P*-values) against the expected -log10(*P*-values) under the null hypothesis of no association. Genomic inflation factors (λGC) are indicated within each plot. These plots indicate minimal confounding from uncorrected population stratification or cryptic relatedness, supporting the validity of the genetic associations. Additional Supplementary Table 53 presents the results from 16 distinct genome-wide association studies (GWAS) conducted on 11 subphenotypes.

With the phenotypic framework established, a two-stage Multi-Trait Analysis of GWAS (MTAG) was employed to boost statistical power and delineate the genetic architecture of the subphenotypes. The initial stage involved integrating primary GWAS results with large-scale external BD summary statistics, followed by a second stage incorporating schizophrenia (SCZ) data, allowing for comparison. The increased statistical power and identified loci from the BD-only MTAGs are visualized in the Manhattan plots (Figure 23). This determination was supported by parallel analysis.

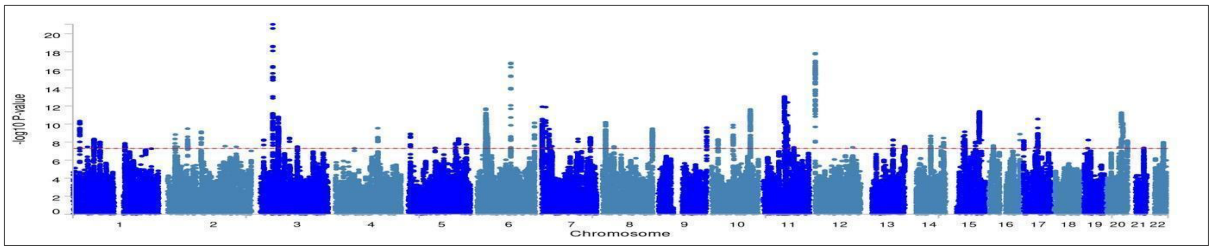
Psychosis



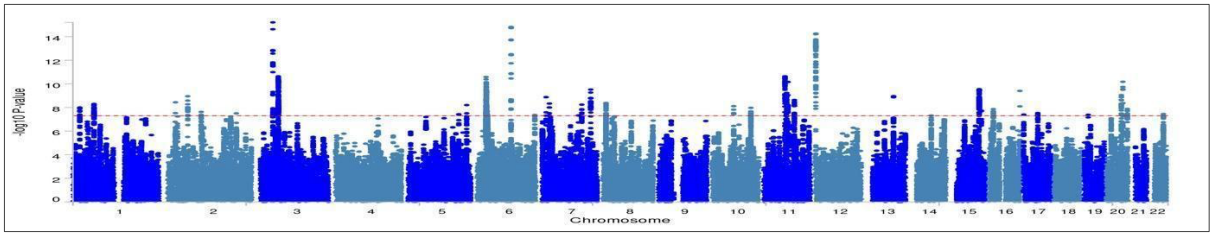
SZA (Schizoaffective)



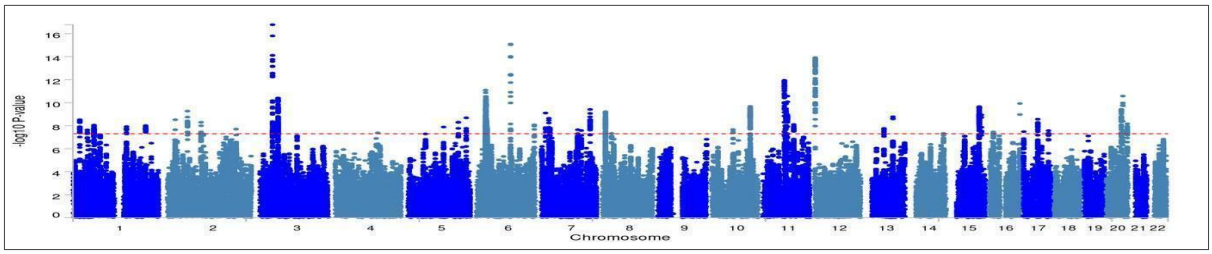
BD1 clinical/community



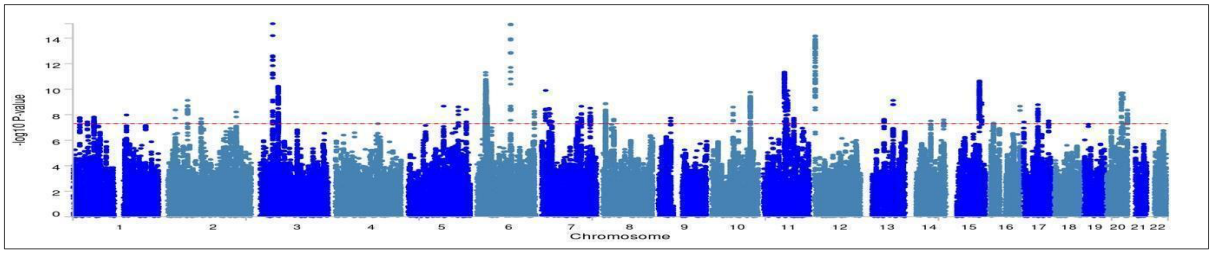
Suicide attempt (SA)



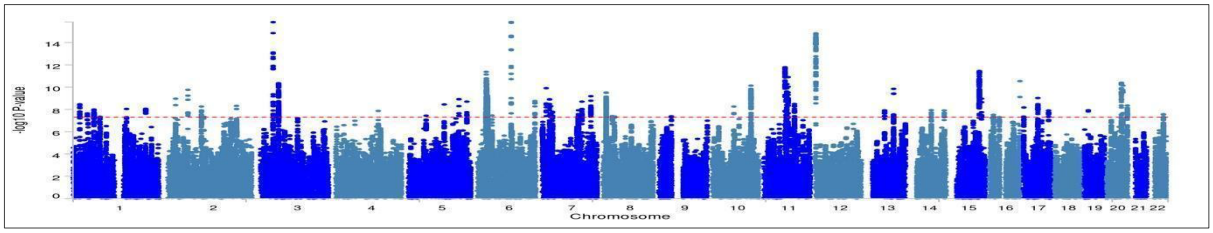
AlcSUD



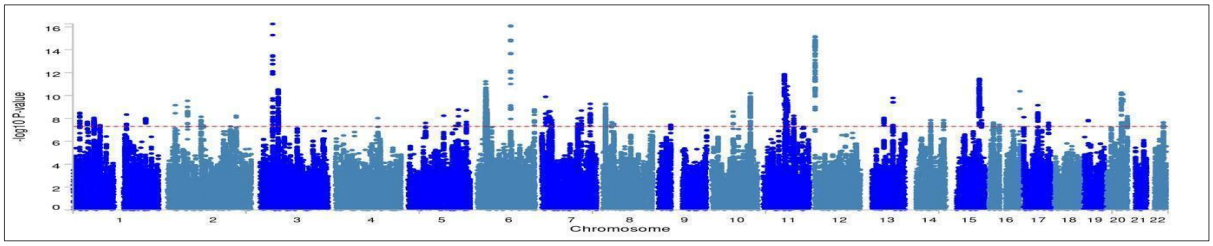
BD2 clinical/community



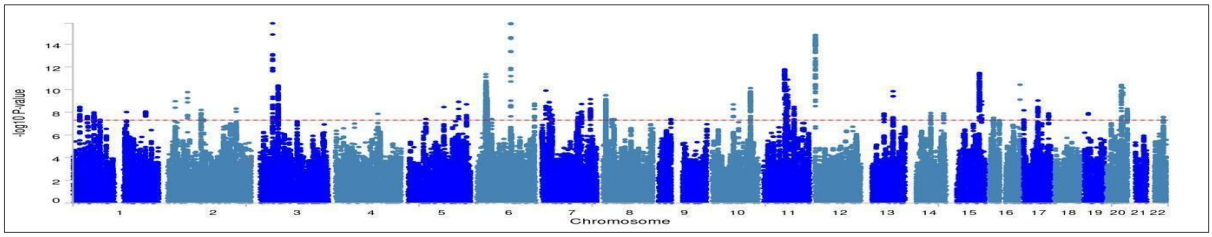
Panic disorder (PD)



Rapid cycling (RC)



OCD



Unipolar mania (UM)

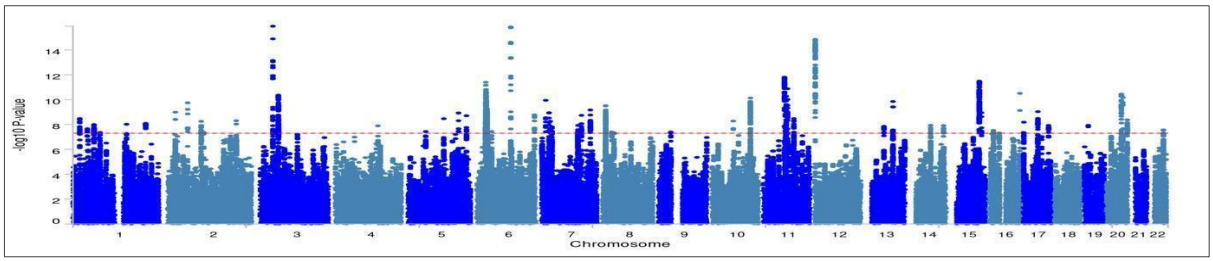


Figure 23 Manhattan plots for each of the 10 subphenotype-BD MTAG analyses.
Each plot displays the -log10(*P*-values) of all SNPs across the genome. The red line indicates the genome-wide significance threshold ($P<5\times10\text{-}^{8}$). These plots visually represent the increased statistical power and identified loci from the BD-only MTAGs. Supplementary Table 48 presents the results from the 10 subphenotype-BD-only and the 10 subphenotype-BD-SCZ MTAGs.

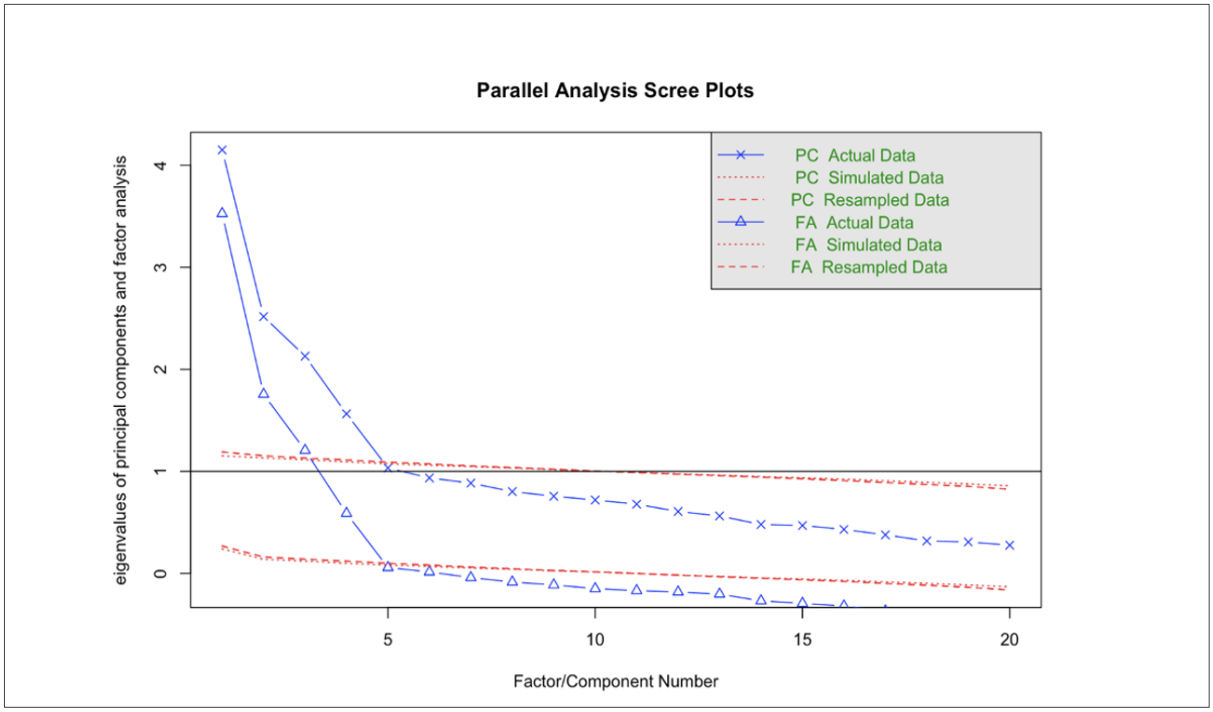


Figure 24 Parallel analysis plot for factor determination.
Parallel analysis determined the number of lower dimensions in the dataset to be four. Eigenvalues for principal components (PC) and factor analysis (FA) converged on four eigenvalues (factors), which are above the PC (upper red line) and FA (lower red line) cutoff, determining four factors were the best fit for the model.

An *a priori* (Figure 21) and subsequent Principal Component Analysis (PCA) (Figure 25) of MTAG loci aligned with these clinical factors, underscoring a genetic basis for the observed clinical distinctions. This genetic PCA explained 81.5% of the variance and revealed four distinct dimensions, or clusters, that may represent points along a biological continuum rather

than discrete entities. The statistical validity of this structure was confirmed by a one-way ANOVA, which revealed a similar robust pattern of difference in LAVA local genetic correlation Rho(ρ) between the clusters, $F(3, 1038) = 203.2$, $P < 2.00\text{x}10^{-16}$.

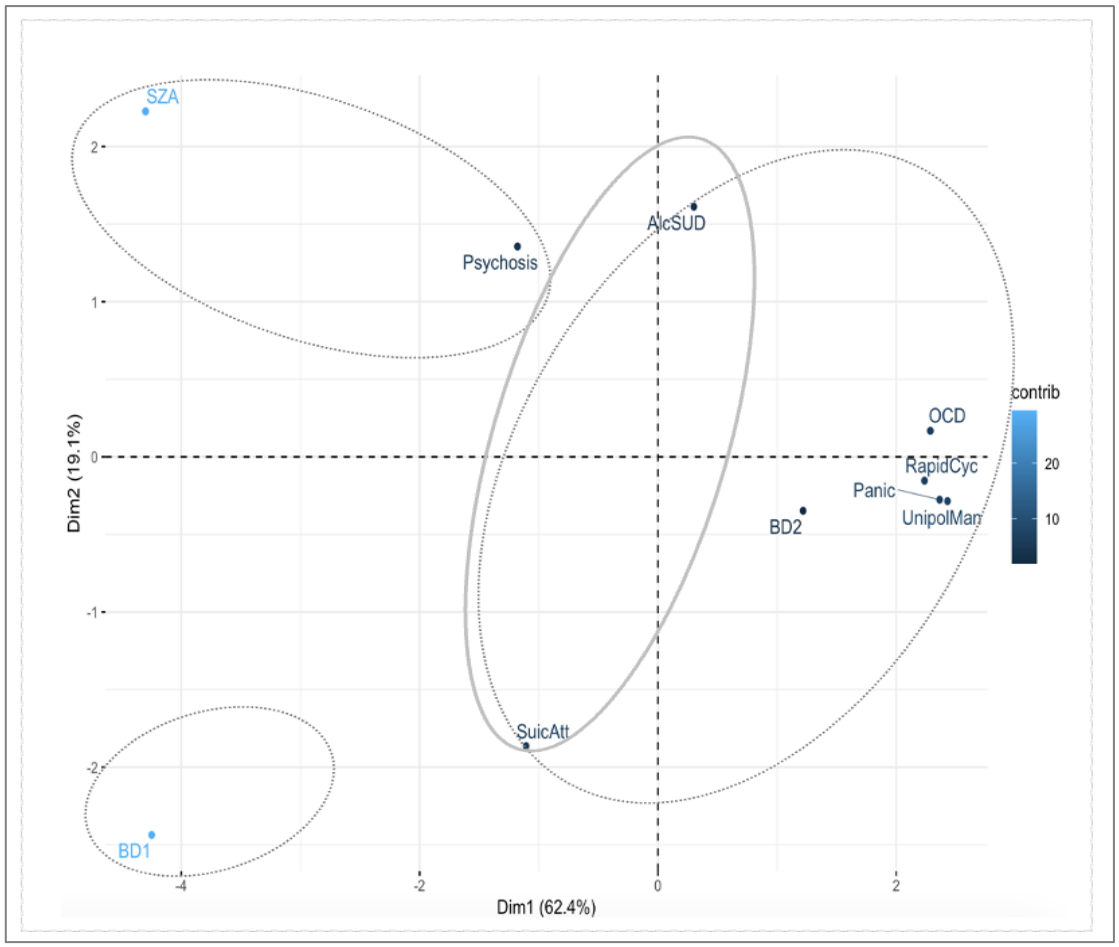


Figure 25 PCA biplot of genomic loci in 10 subphenotype-BD-MTAGs.
Principal component analysis (PCA) biplot of genomic loci from 10 subphenotype-BD MTAG results. The first two dimensions account for 81.5% of the variance. Subphenotypes with similar genetic architectures are closer on the biplot. Lighter colours indicate higher contribution (factor loading) to dimensional variance. A one-way ANOVA revealed a significant difference in LAVA local genetic correlation (ρ) between the PCA.

Table 42 shows direct tests of the incremental power gain from a single subphenotype GWAS to a BD-only MTAG and then to the BD-SCZ MTAG counterpart. This out-of-sample analysis also confirms that MTAG not only increased statistical power but also enhanced the biological coherence of the dimensions, revealing a more valid and meaningful genetic architecture for Bipolar Disorder. See Table 47 for an overview of the external summary statistics.

Table 42 Independent MTAG Validation

'Severe Illness' and 'Core Mania' Dimension (vs. Schizophrenia)

Subphenotype	Univariate *rG* with SCZ	MTAG-BD *rG* with SCZ	MTAG-SCZ-BD *rG* with SCZ
Psychosis	.44	.53	.75
SZA	.50	.57	.79
BD1	.32	.37	.61

'Internalizing' Dimension (vs. Major Depression)

Subphenotype	Univariate *rG* with MDD	MTAG-BD *rG* with MDD	MTAG-SCZ-BD *rG* with MDD
BD2	.55	.60	.65
OCD	.17	.20	.26
PD	.41	.45	.49

'Externalizing' Dimension (vs. ADHD)

Subphenotype	Univariate *rG* with ADHD	MTAG-BD *rG* with ADHD	MTAG-SCZ-BD *rG* with ADHD
AlcSUD	.35	.38	.44
SA	.28	.32	.37

The statistical validity of this transdiagnostic approach was further confirmed by associations of the primary credible gene set (Tables 48-52) with established rare-variant risk genes from the SCHEMA (Chapter 2 [74]) and BipEx (Chapter 2 [75]) consortia using a one-sided Fisher's exact test. Statistical significance was defined as $P < .0125$ (Bonferroni correction for four tests). Our analysis revealed a significant convergence between common- and rare-variant evidence. The enrichment for our primary BD-SCZ credible sets with SCHEMA rare-variant genes was significant (e.g., for BD-SCZ_noMHC set, $P = 4.1 \times 10^{-4}$), driven by overlapping genes *TCF4*, *PBRM1*, and *ZEB2*. The secondary BD-Only sets showed only a nominal enrichment that did not survive correction. While exploratory analyses of the BD-Only sets yielded suggestive trends for *PBRM1* and *TRANK1*, the overall results allow us to begin

genetically dissecting the core components of BD from the broader, transdiagnostic risk factors it shares with SCZ for both common and rare variants. This pattern suggests the convergence is most robust for transdiagnostic factors shared between BD and SCZ.

Based on their genetic and clinical composition, the four dimensions were interpreted as representing:

- A **Severe Illness Dimension** (Psychosis, SZA)
- A **Core Mania Dimension** (BD1)
- An **Externalizing/Impulsive Comorbidity Dimension** (SA, AlcSUD)
- An **Internalizing/Affective Comorbidity Dimension** (BD2, PD, OCD, RC, UM)

5.3 Dimension 1: Severe Illness

This dimension is defined by profound genetic overlap with SCZ (see genetic correlations, Figure 26, Supplementary Table 59.2), a link substantiated by this analyses and consistent with large-scale genomic dissections of the two disorders [43-44].

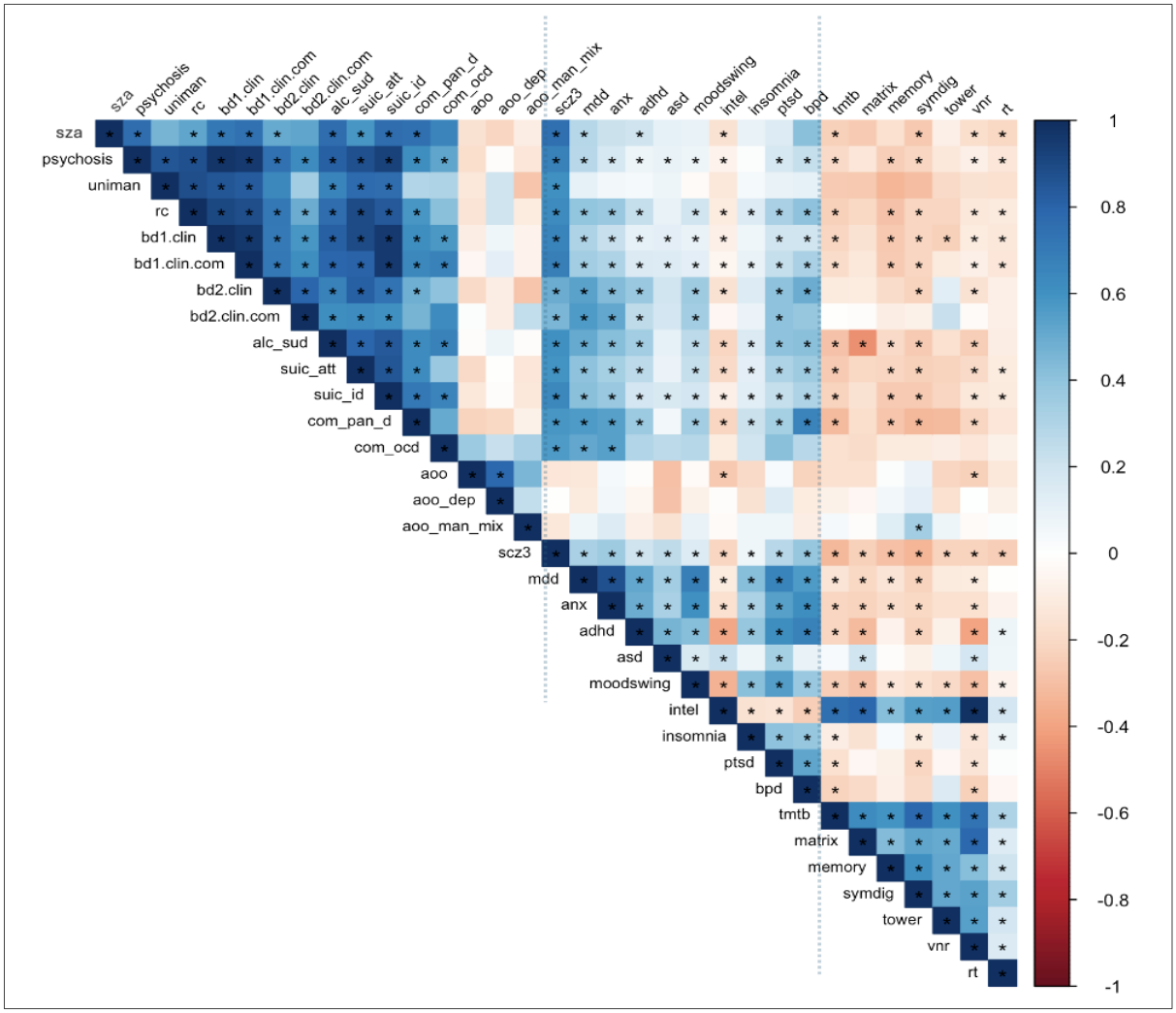
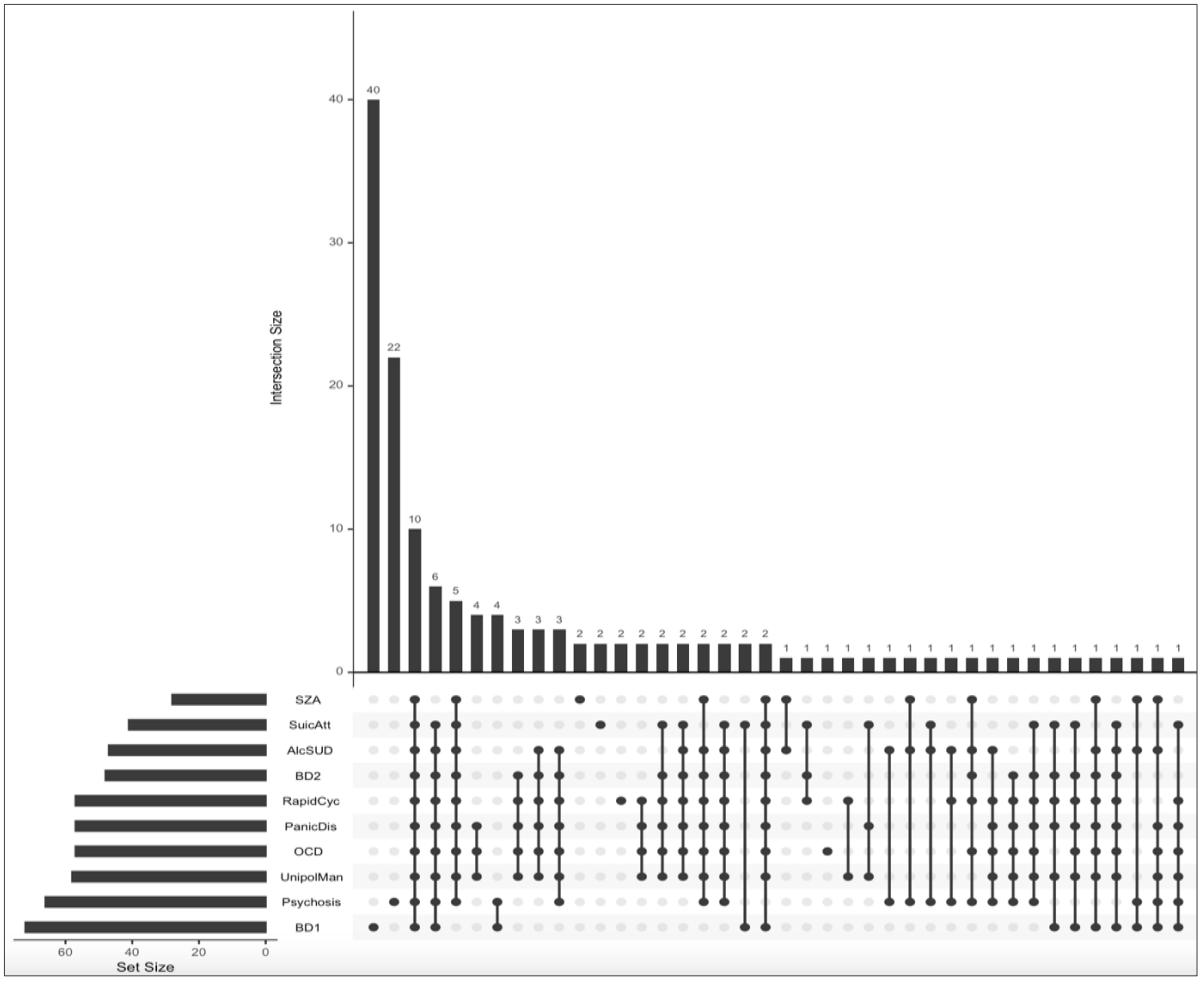


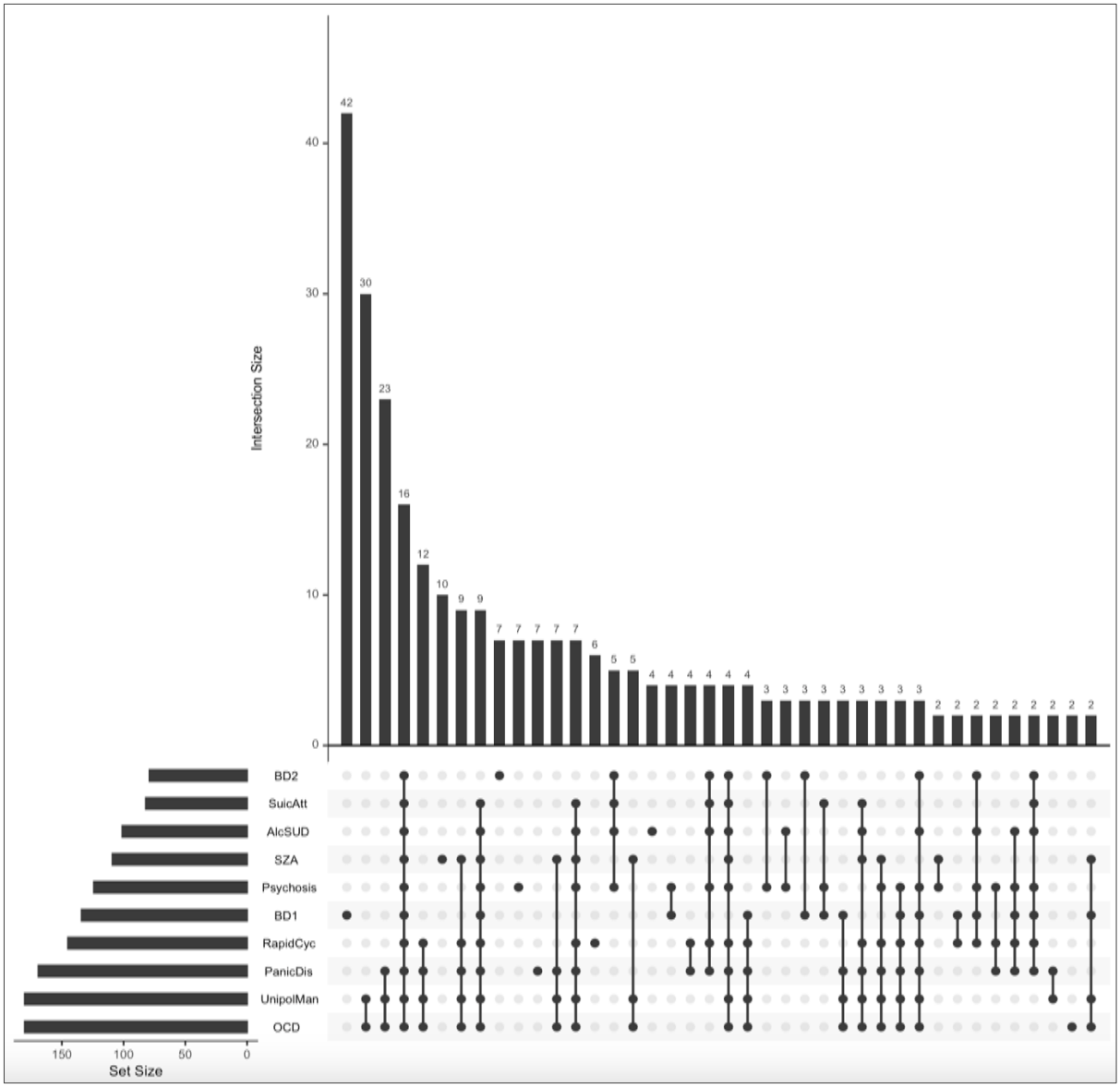
Figure 26 Global Genetic Correlation Heatmap of BD and Cross-Traits.
Bivariate genetic correlations (*rG*) calculated using summary statistics from large-scale GWAS across three trait categories: 13 BD subphenotypes, ten psychiatric disorders, and seven cognitive measures. P-values were Bonferroni-corrected ($P<1.84\times10\text{-}^{4}$) and correlations were standardized in GenomicSEM. See Table 47 for an overview of the external summary statistics.

The inclusion of SCZ variants in the MTAG amplified shared signals; for instance, the number of shared loci between Psychosis and SZA increased by 63% (from 16 to 26) in the BD-SCZ analysis (Supplementary Table 45; Figures 27-28).



	PSY	SZA	BD1	SA	AlcSUD	BD2	PD	RC	OCD	UM
PSY	71	16	62	28	32	26	28	35	30	9
SZA	18.0%	68	52	24	26	25	27	31	29	8
BD1	69.7%	51.5%	70	25	29	24	26	32	27	7
SA	28.3%	23.3%	25.5%	49	25	22	24	28	25	7
AlcSUD	33.0%	26.5%	31.5%	29.8%	38	18	21	26	23	5
BD2	27.1%	25.5%	25.3%	25.3%	21.2%	35	20	25	22	6
PD	27.7%	26.7%	26.5%	26.4%	24.1%	23.0%	43	29	30	8
RC	34.0%	3.1%	32.7%	31.8%	29.5%	28.7%	31.5%	44	29	6
OCD	29.1%	28.2%	27.8%	28.1%	26.1%	25.6%	32.3%	31.5%	33	8
UM	9.7%	8.6%	7.9%	8.8%	6.3%	7.8%	9.9%	7.5%	9.8%	20

Figure 27 UpSet plot of genomic loci overlap (BD-only MTAGs).
Overlap of genomic loci in 10 subphenotype-BD MTAG analyses. The plot and corresponding table visualize the number of shared and unique genomic risk loci across the 10 analyses.



	PSY	SZA	BD1	SA	AlcSUD	BD2	PD	RC	OCD	UM
PSY	123	26	98	37	42	34	43	47	42	21
SZA	16.6%	112	87	31	35	31	40	44	39	20
BD1	63.2%	55.4%	119	33	38	30	39	43	38	19
SA	23.5%	19.9%	21.7%	68	29	24	31	35	31	15
AlcSUD	27.2%	23.3%	25.9%	25.0%	61	21	27	31	27	13
BD2	22.5%	2.7%	2.5%	2.7%	18.1%	58	26	30	26	13
PD	27.2%	25.0%	25.3%	26.5%	22.7%	22.2%	69	36	36	17
RC	28.5%	26.5%	28.1%	29.2%	26.5%	25.9%	29.5%	73	35	17
OCD	25.7%	24.3%	25.0%	26.3%	22.9%	22.8%	3.0%	28.9%	71	17
UM	12.8%	12.2%	12.0%	13.9%	11.0%	11.6%	14.1%	14.2%	14.2%	42

Figure 28 UpSet plot of genomic loci overlap (BD-SCZ MTAGs).

Overlap of genomic loci in 10 subphenotype-BD-SCZ MTAG analyses. The plot and corresponding table visualize the number of shared and unique genomic risk loci across the 10 analyses.

Biologically, this dimension is differentiated by a unique neuro-immune signature. The TWAS analysis revealed that expression of *HLA-DMA* in the cerebellum showed a strong protective association ($P = 2.50 \times 10^{-273}$) only in the BD-SCZ MTAG context; this signal was not robust in the BD-only analysis, indicating this specific immune pathway is a primary feature linking severe BD to SCZ (Supplementary Table 50; Figure 29 below). This synaptic link is mirrored at the cellular level, where the genetic association for GABAergic and cortical neurons became more robust in the BD-SCZ context (*P*-adjusted for Psychosis-BD GABAergic neurons = 3.39×10.0^{-7} vs. 1.96×10^{-11} for Psychosis-BD-SCZ), underscoring a shared cellular vulnerability (see Figure 30). Furthermore, this dimension is characterized by specific synaptic biology. The novel, deleterious variant in the neuronal sodium channel gene *SCN2A* (Combined Annotation Dependent Depletion [CADD] [45] =19.83) was associated specifically with the Psychosis and BD1 subphenotypes, directly implicating fundamental neuronal excitability. This is mirrored in the gene-set analysis, where the significance for pathways including "GOCC_POSTSYNAPTIFIC_SPECIALIZATION", driven by genes involved in scaffolding proteins and glutamatergic receptor subunits, became orders of magnitude stronger for this cluster when SCZ data was added (e.g., for SZA, *P*(Bonferroni) = 1.35×10^{-12}), confirming that the shared biology is concentrated at the synapse (Supplementary Table 49; Figure 31, Figure 34). However, the absence of association could also be interpreted as due to a lack of power to detect an association.

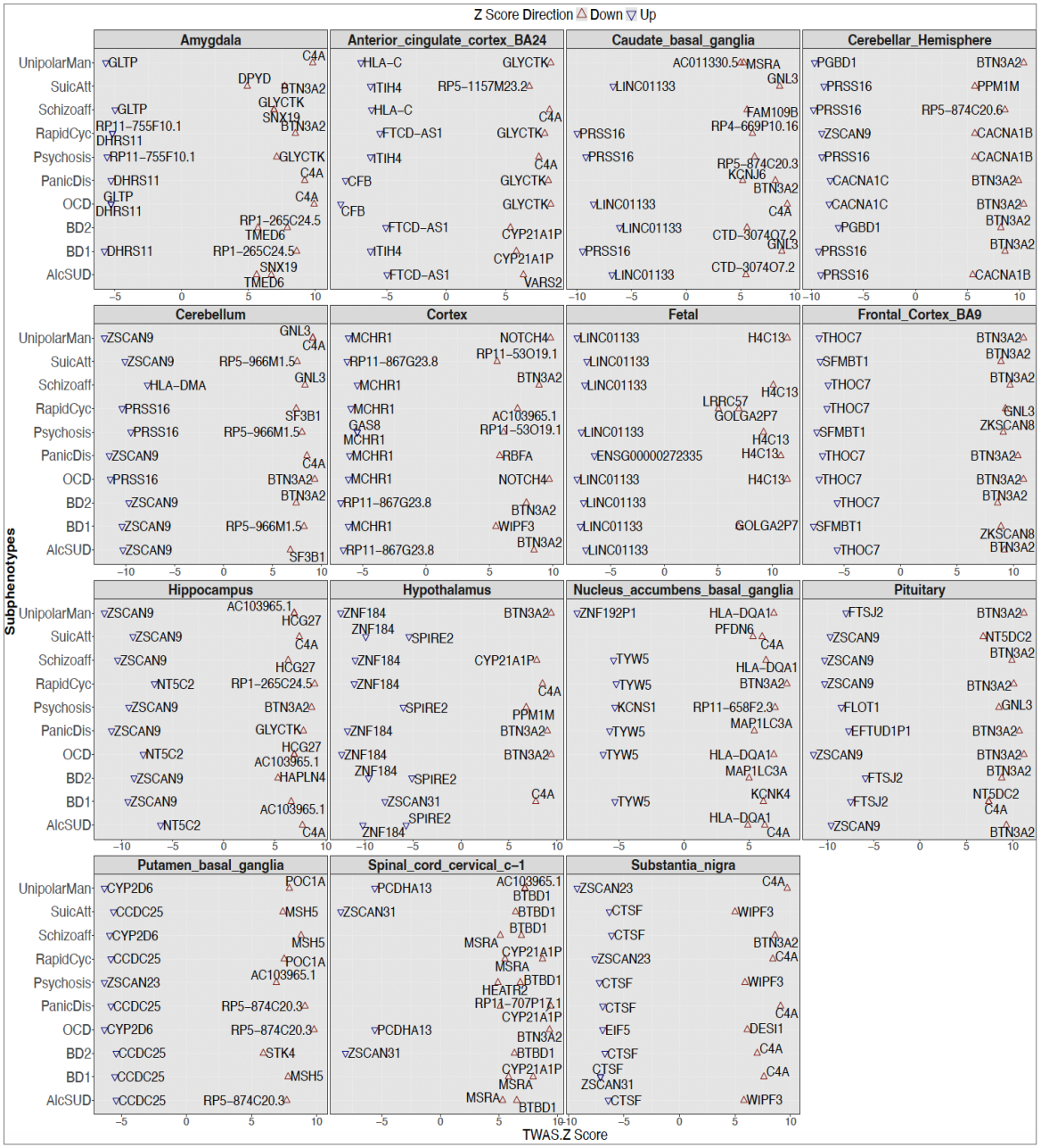


Figure 29 TWAS joint tissue associations in 10 subphenotype-BD-SCZ MTAG.

The plot shows the most robust, conditionally independent gene-tissue associations for each BD subphenotype across 15 brain tissue datasets. The x-axis represents the significance of the association ($-\log_{10}$ *P*-value), corrected for all genes and tissues tested. The direction of effect is indicated by triangles: red for a positive Z-score (increased expression associated with risk) and blue for a negative Z-score (decreased expression associated with risk). See Figure 32 below for the BD-Only TWAS analyses. Subphenotypes included were: Psychosis, Schizoaff, Schizoaffective disorder, BD1, bipolar disorder I, SuicAtt, suicide attempt, AlcSUD, alcohol or substance use disorder, BD2, bipolar disorder II, PanicDis, panic disorder, RapidCyc, rapid cycling, OCD, obsessive compulsive disorder, UnipolarMan, unipolar mania. The Supplementary table 50 provides the full list of gene-tissue associations from the TWAS analysis.

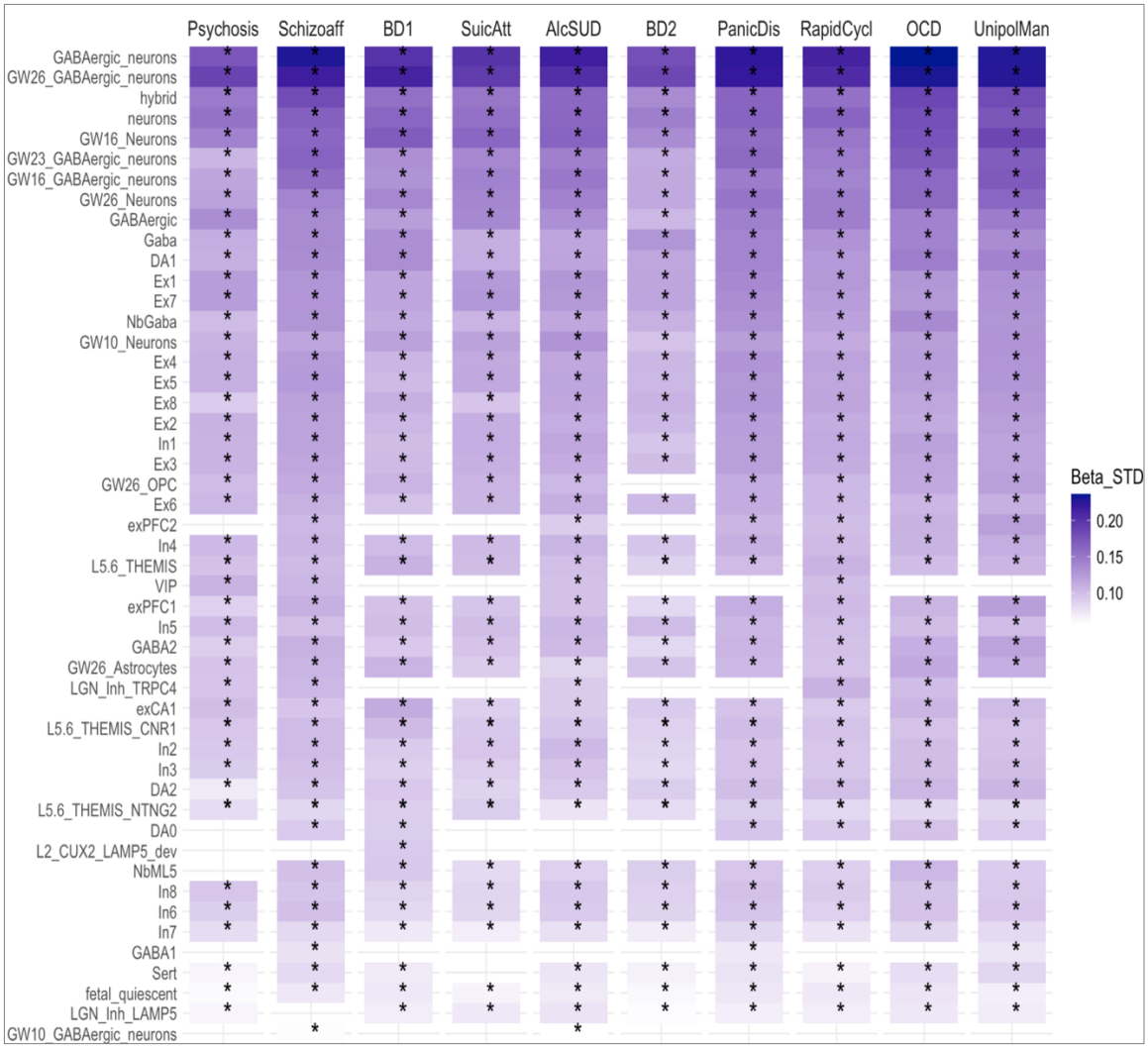


Figure 30 Cell type enrichment analysis in 10 subphenotype-BD-SCZ-MTAGs.

The heatmap displays standardized beta coefficients from cell-type enrichment analysis across 10 BD subphenotypes. Colour intensity corresponds to the strength of the enrichment signal, with subphenotypes ordered by effect size. Absence of colour indicates no association. Asterisks (*) denote associations that remained robust after Bonferroni correction for the number of cell types tested ($P <$ .05). Corresponding results from the BD-Only analysis are shown in Figure 33 below. The Supplementary Table 47 provides the full list of associations from the cell type specificity analysis. The analyses included: Psychosis, Schizoaff, Schizoaffective disorder, BD1, bipolar disorder I, SuicAtt, suicide attempt, AlcSUD, alcohol or substance use disorder, BD2, bipolar disorder II, PanicDis, panic disorder, RapidCycl, rapid cycling, OCD, obsessive compulsive disorder, UnipolMan, unipolar mania.

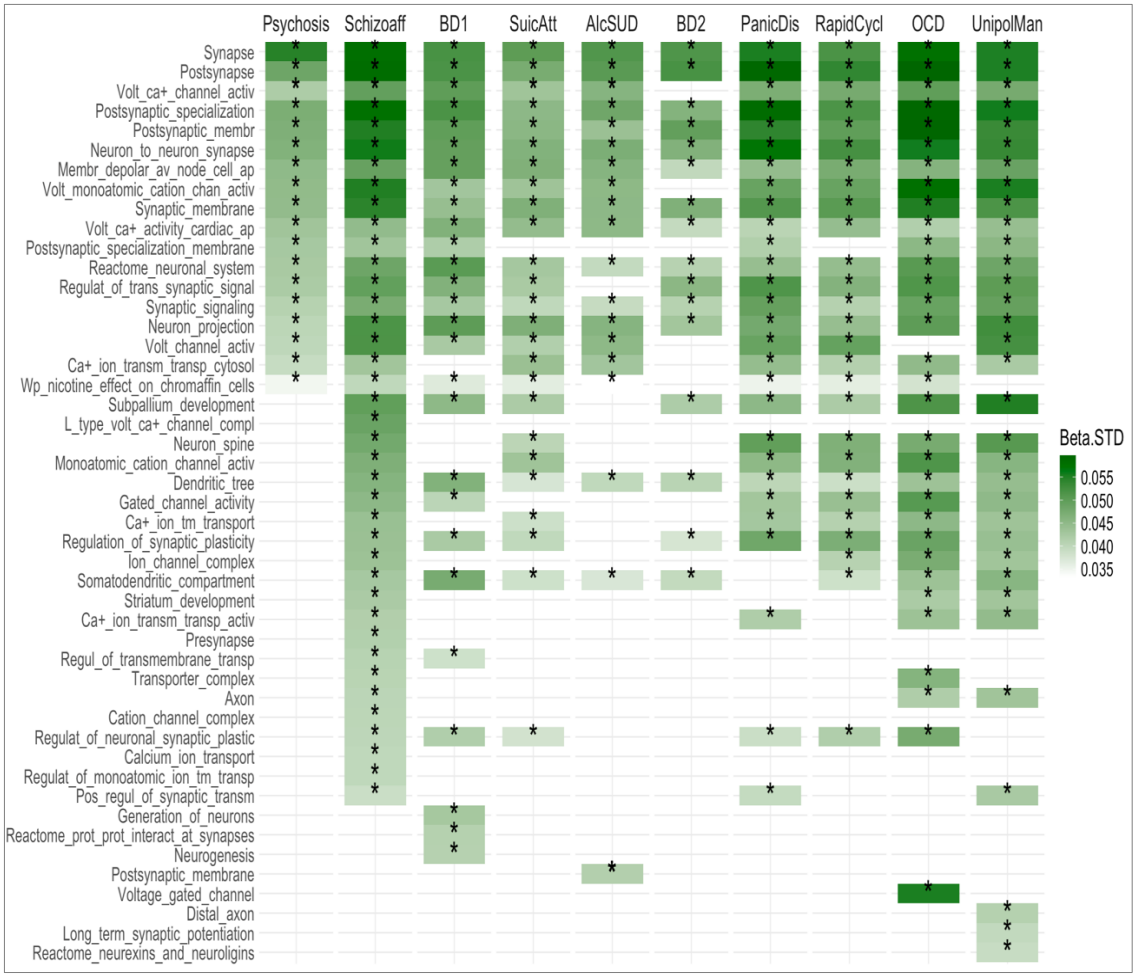


Figure 31 Gene set enrichment analysis in 10 subphenotype-BD-SCZ-MTAGs.

The heatmap displays standardized beta coefficients from MAGMA gene-set enrichment analysis across 10 BD subphenotypes. Colour intensity corresponds to the strength of the enrichment signal, with gene sets ordered by effect size. Absence of colour indicates no association. Asterisks (*) denote associations that remained robust after Bonferroni correction for the number of gene sets tested ($P <$ .05). Corresponding results from the BD-Only analysis are shown in Figure 34 below. The Supplementary Table 49 provides the full list of gene-set associations. The analyses included: Psychosis, Schizoaff, Schizoaffective disorder, BD1, bipolar disorder I, SuicAtt, suicide attempt, AlcSUD, alcohol or substance use disorder, BD2, bipolar disorder II, PanicDis, panic disorder, RapidCycl, rapid cycling, OCD, obsessive compulsive disorder, UnipolMan, unipolar mania.

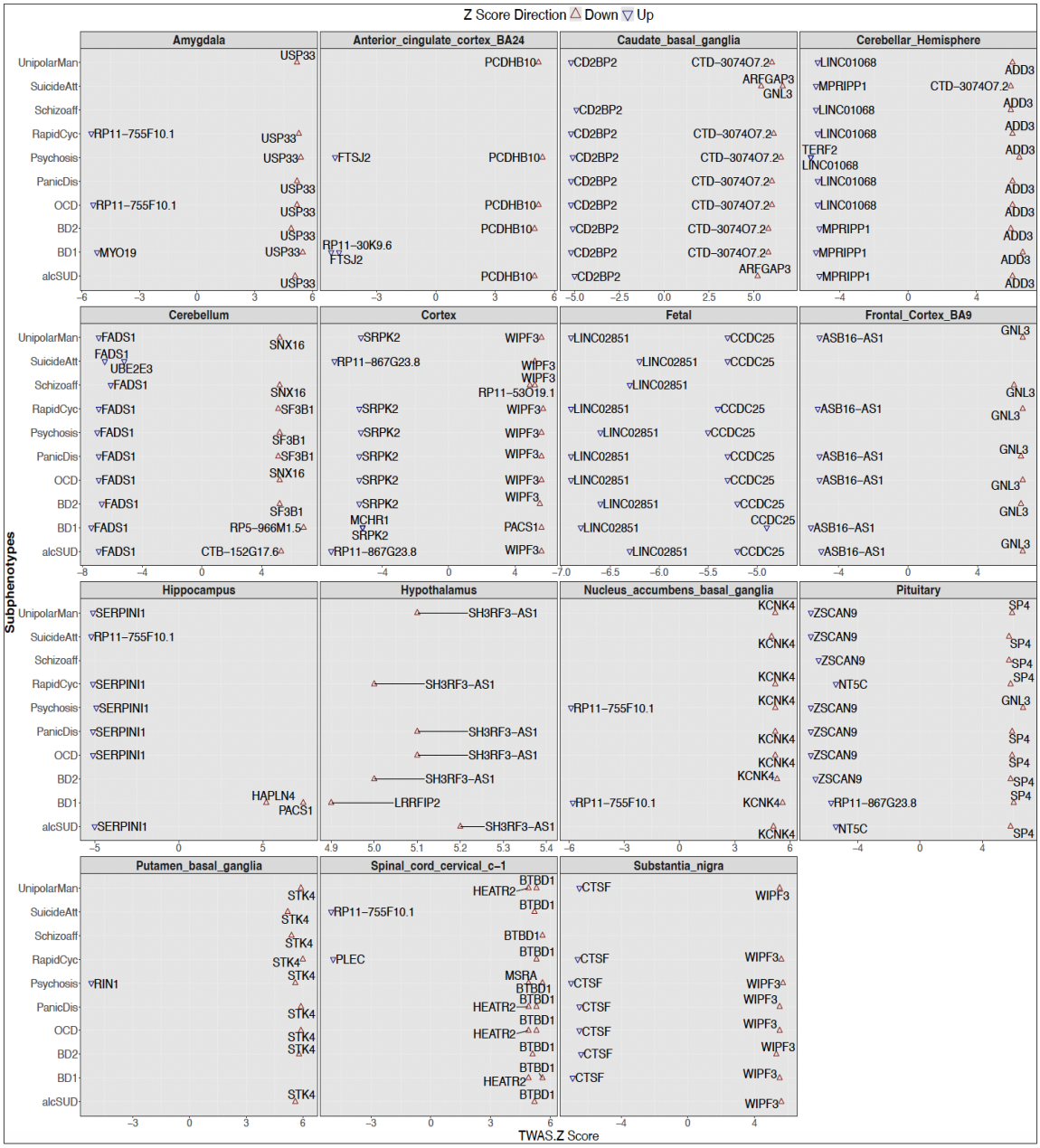


Figure 32 Heatmap of TWAS joint tissue associations (BD-only MTAGs).
Heatmap illustrating TWAS joint tissue associations in 10 subphenotype-BD MTAGs. Effect sizes, categorized by tissue, represent findings across 15 adult and foetal brain tissues. Red (positive) and blue (negative) triangles represent the direction of effect of the TWAS Z-statistic score.

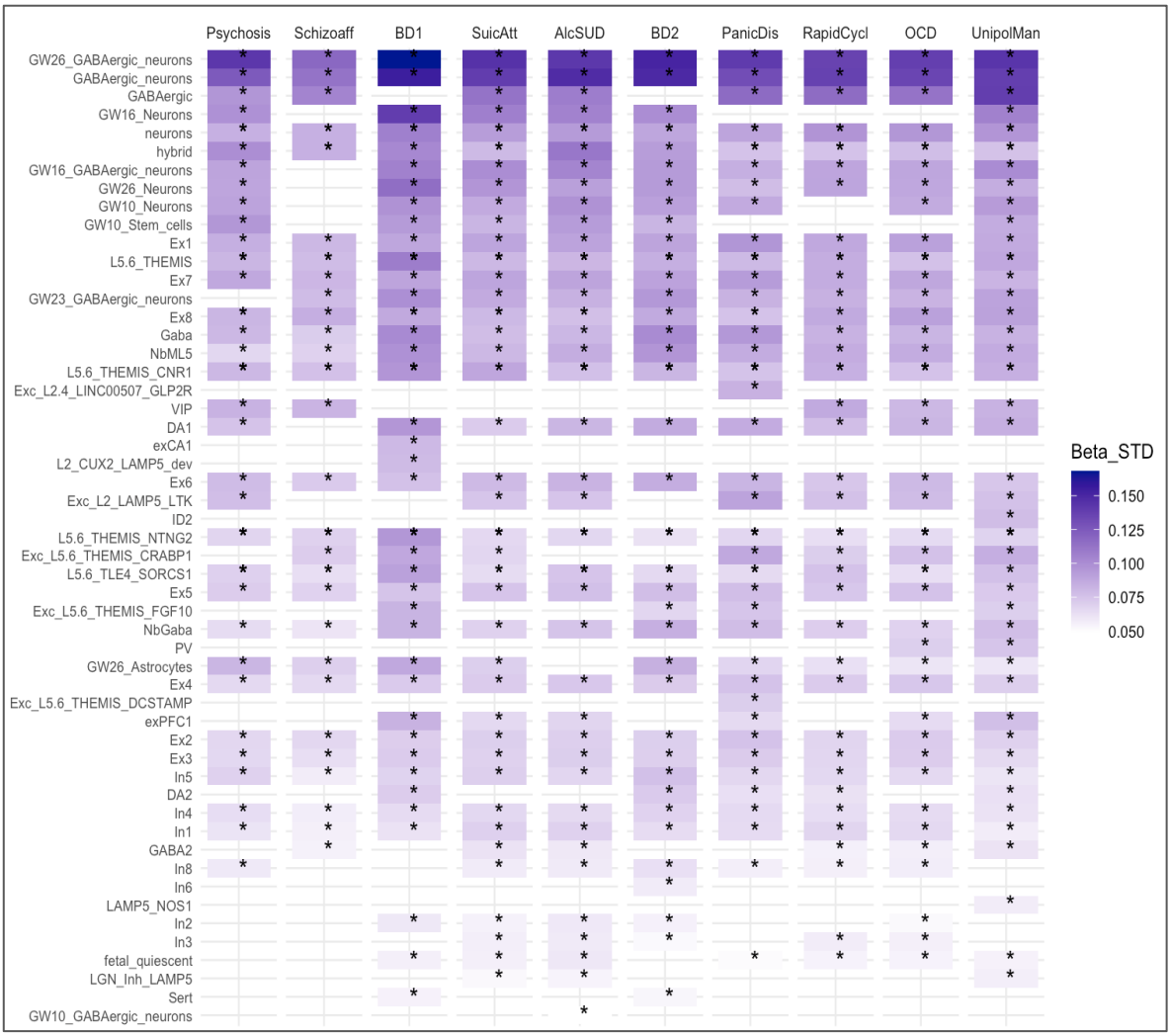


Figure 33 Heatmap illustrating differential cell type enrichment (BD-only MTAGs).
The heatmap illustrates differential cell type enrichment across 10 subphenotype-BD MTAG analyses. Colour intensity corresponds to the strength of the standardized beta. An asterisk (*) marks cell-type associations that survive Bonferroni correction (P<.05).

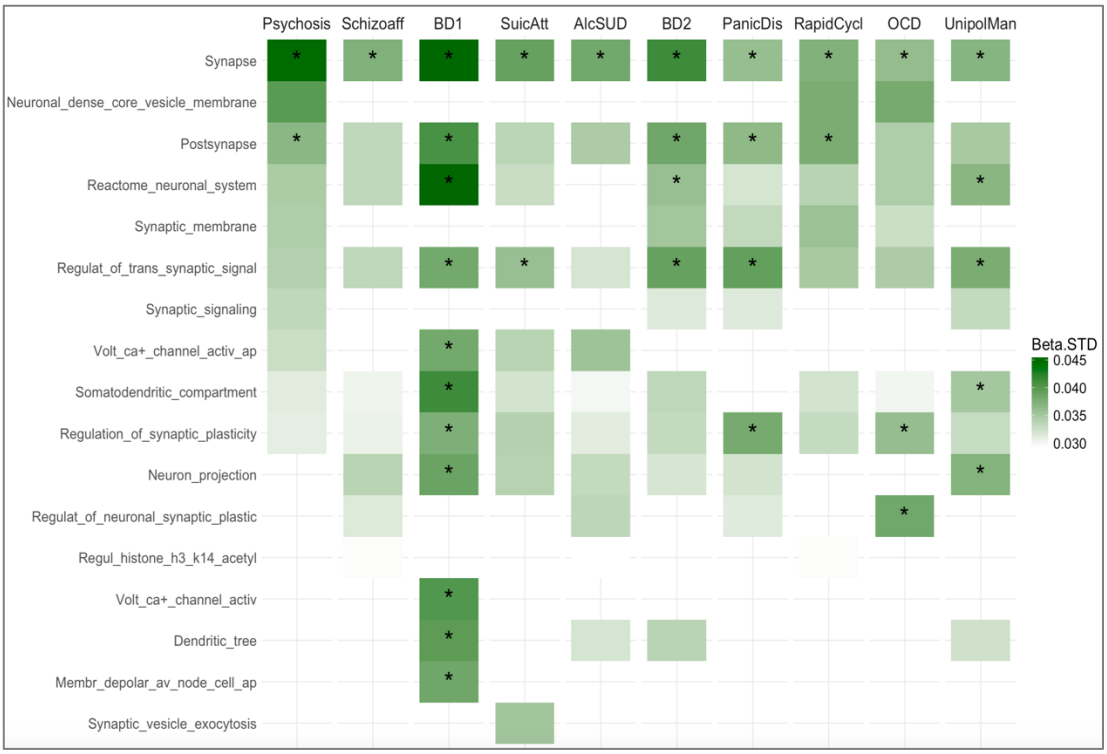
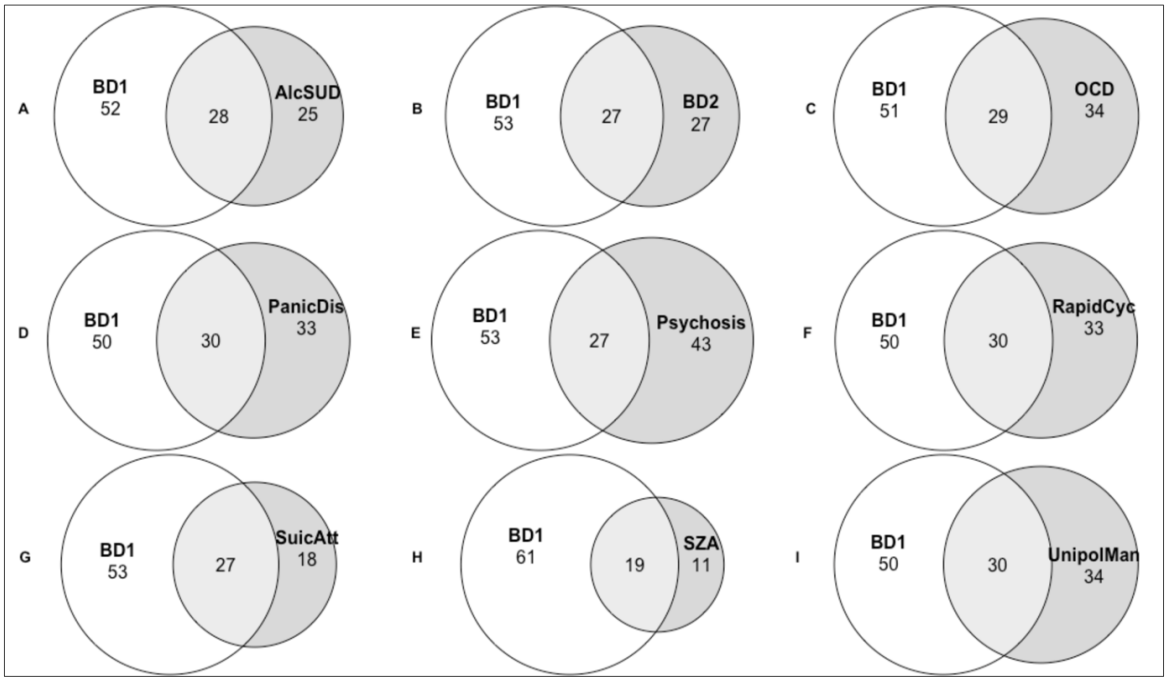


Figure 34 Heatmap illustrating differential gene set enrichment (BD-only MTAGs).
The heatmap illustrates differential gene set enrichment across 10 subphenotype-BD MTAG analyses. Colour intensity corresponds to the strength of the standardized beta. An asterisk (*) marks gene sets that survive Bonferroni correction (P<.05).

5.4 Dimension 2: Core Mania

While genetically related to the Severe Illness Dimension, the BD1 dimension is distinguished by specific loci related to neuronal function and development. The TWAS analysis identified *PACS1*, involved in neuronal protein trafficking, as uniquely associated with BD1 via its expression in the cortex (Supplementary Table 50). The association with *PACS1* suggests altered neurotrophic support may be a specific biological feature of the core manic phenotype. Furthermore, BD1 was specifically associated with a variant in *ADCY2* (rs78308718), a gene previously linked to lithium response [46-49]. This suggests a distinct biological pathway related to treatment response that is characteristic of this core manic phenotype. This was complemented by findings for *CACNA1C*, a well-established risk gene for BD, which showed its strongest association within the Core Mania dimension, reinforcing the importance of calcium channel signalling in mania [3, 5]. This contrast is particularly evident when comparing BD1 and RC; while BD1 shows genetic specificity, RC displays a highly pleiotropic profile, with associated variants overlapping more extensively with other subphenotypes (Figure 35).

A.



B.

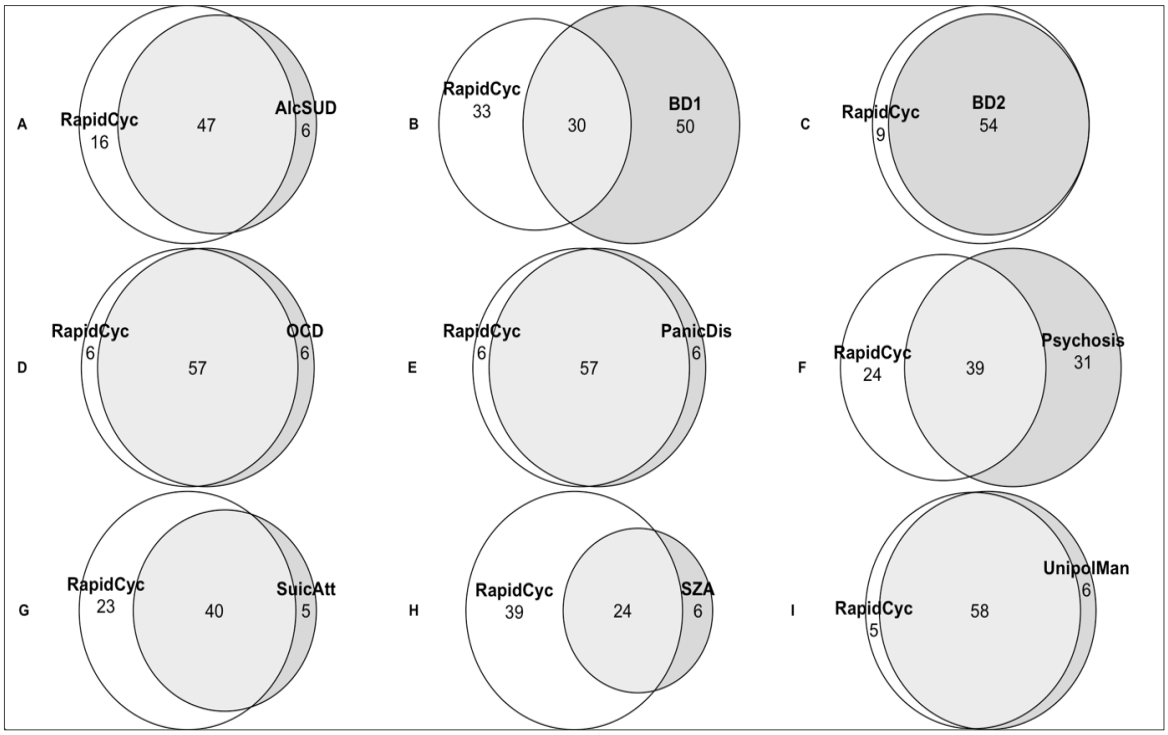


Figure 35 Overlap visualizations of lead SNPs across subphenotypes.
Visualization of shared and unique lead SNPs for bipolar disorder I (BD1) (A) and rapid cycling (RC) (B) from the subphenotype-BD-only MTAG analyses, demonstrating the genetic specificity of BD1 versus the pleiotropic architecture of RC.

5.5 Dimension 3: Externalizing/Impulsive

This dimension is defined by a strong shared liability for impulsive and externalizing behaviours. This was evident in the high global genetic correlation between suicide attempt (SA) and Alcohol/Substance use disorder/dependency (AlcSUD) (*rG* ~ .80, s.e.m.=.056) and was validated by LAVA, which identified three shared local genetic loci between them (see Supplementary Table 51 and 59.2). The three local genetic loci shared between SA and AlcSUD included a region on chromosome 16 containing genes for synaptic vesicle transport, suggesting shared mechanisms of presynaptic function. This dimension shares a common architecture with ADHD; biologically, this dimension is distinguished by a strong enrichment for midbrain dopaminergic neurons, directly implicating reward and motivation pathways in the shared genetic risk for both SA and AlcSUD (Supplementary Table 47). The enrichment for dopaminergic neurons was specific to cells from the ventral tegmental area (VTA), a key hub in the mesolimbic reward circuit, providing a direct anatomical and cellular correlate for the high rates of comorbid substance use in this cluster. The novel association of the gene *MAD1L1* (Table 43, Supplementary Table 49) (critical for neurodevelopment), with the AlcSUD subphenotype in the BD-SCZ MTAG, provides an additional specific biological link for this dimension.

Table 43 Key Genetic and Biological Findings Defining the Dimensions of Bipolar Disorder

Pathway	Key Finding	Primary Evidence	Analysis
Severe Illness Dimension (Psychosis, SZA)			
Neuro-Immune	*HLA-DMA*	Protective; P=2.50×10-^{273}	TWAS (BD-SCZ MTAG)
Synaptic Function	*SCN2A*	Deleterious; CADD = 19.83	Variant Annotation
Synaptic Function	Postsynaptic Specialization	P(Bonferroni) < 1.35×10-^{12}	Gene-Set Enrichment
Cellular	VIP-expressing interneurons	Top enriched cell type	Cell-Type Enrichment
Neurodevelopment	Hippocampal Excitatory Neurons	Enrichment in BD-SCZ analysis	Cell-Type Enrichment
Core Mania Dimension (BD1)			
Synaptic Function	*PACS1*	P=2.00×10-^{19}	TWAS (BD-only)
Externalizing Dimension (SA, AlcSUD)			
Cellular	Midbrain Dopaminergic Neurons	Risk enrichment	Cell-Type Enrichment
Neurodevelopment	*MAD1L1*	Novel Locus; P=1.15×10-^{18}	GWAS (BD-SCZ MTAG)
Internalizing Dimension			
Neurodevelopment	*DCC* (RC, UM, PD, OCD)	Shared Locus; P<1.37×10-^{8}	LAVA
Neuro-Immune	*SMAD3* (RC, PD)	PD/RC Specific Locus; P=3.14×10-^{9}	GWAS (BD-SCZ MTAG)
Cellular (Gut-Brain)	*GLP2R* enrichment (PD)	Specific cell-type enrichment	Cell-Type Enrichment
Cellular	Glutamatergic vs. GABAergic	Weaker glutamatergic assoc.	Cell-Type Enrichment
Evolutionary	Rapid Cycling (RC)	Negative Selection (S) = -1.75	SBayesS
Shared / Foundational (Across Dimensions)			
Foundational	Chromatin Org. & DNA Repair	Top enriched pathway	Gene-Set Enrichment
Systemic (Stress)	Nicotine/Chromaffin Cell Pathway	Enriched in BD-SCZ analysis	Gene-Set Enrichment
Synaptic (Metabolic)	*SLC39A8*, *FADS1*	CADD=23.1; P=2.11×10-^{32}	Variant Annotation, TWAS
Synaptic (Endocannabinoid)	*CNR1* enrichment	Shared vulnerability	Gene-Set Enrichment
Synaptic (Metabolic)	*GLYCTK*	Protective; P=5.20×10-^{110}	TWAS (Amygdala)

Abbreviations: AlcSUD, alcohol/substance use disorder; BD1, bipolar disorder I; BD2, bipolar disorder II; CADD, Combined Annotation Dependent Depletion; GWAS, Genome-Wide Association Study; LAVA, Local Analysis of [co]Variant Annotation; MTAG, Multi-Trait Analysis of GWAS; OCD, obsessive-compulsive disorder; PD, panic disorder; RC, rapid cycling; SA, suicide attempt; SZA, schizoaffective disorder, bipolar type; TWAS, Transcriptome-Wide Association Study; UM, unipolar mania.

5.6 Dimension 4: Internalising/Affective

This broad dimension is underpinned by a complex substructure of shared genetic factors related to mood instability and anxiety. While sharing the core cellular vulnerabilities seen across all dimensions (including GABAergic neurons, astrocytes), its distinction comes from specific gene pathways. The most powerful evidence for this clustering comes from the LAVA analysis, which uncovered a hidden relationship between OCD and PD. Despite a moderate global correlation, these two subphenotypes shared 30 local genetic loci, explaining their clustering and demonstrating a specific, shared genetic architecture for anxiety-compulsive traits that is largely independent of the psychosis axis (Figure 36; Supplementary Table 51). The 30 shared loci between OCD and PD were significantly enriched for genes involved in postsynaptic density scaffolding and calcium signalling, suggesting a shared vulnerability based on the molecular machinery of the synapse in corticostriatal circuits.

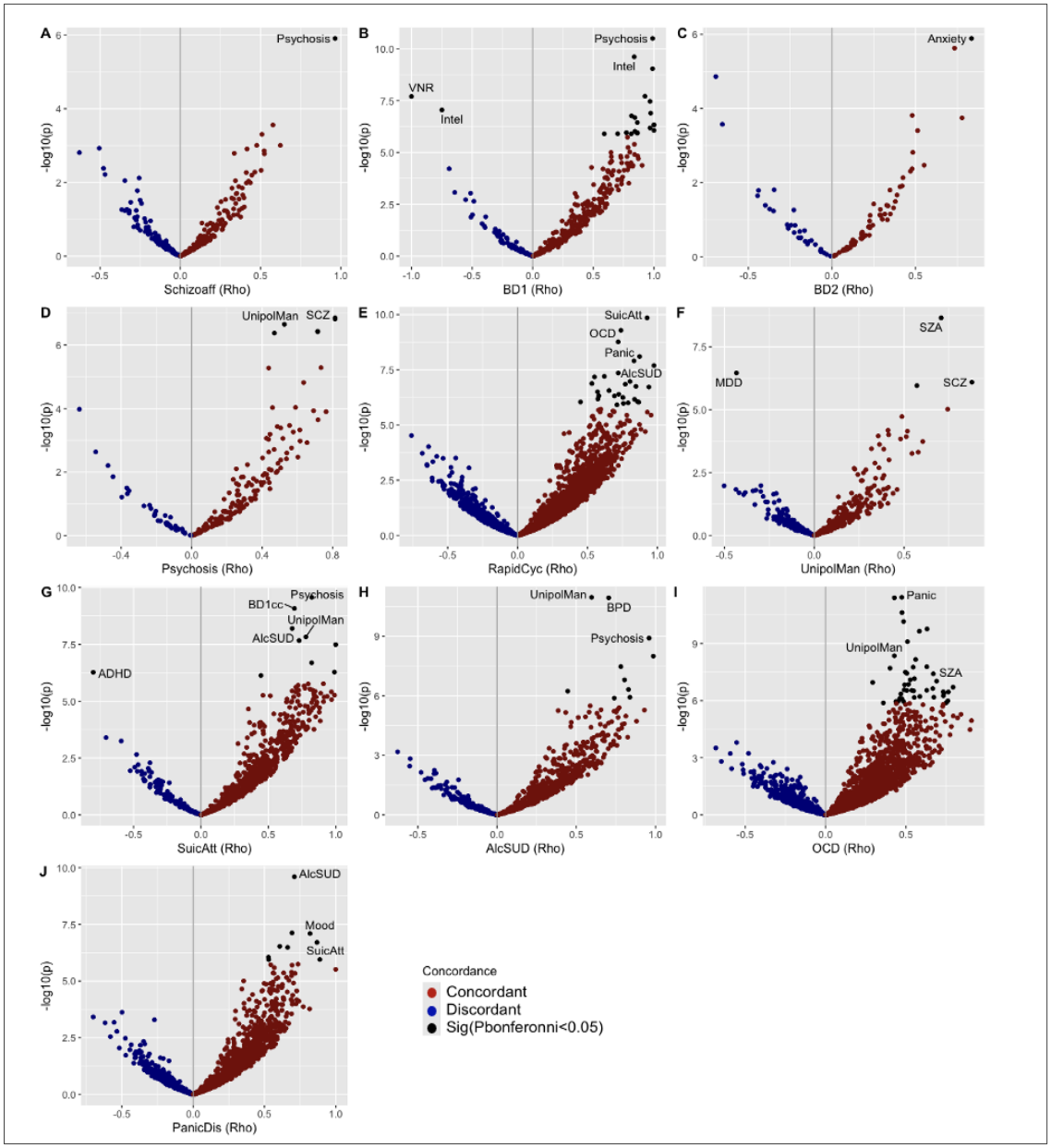


Figure 36 Scatter plots of local genetic correlations.
The Rho(ρ) correlation (x-axis) and log10-p values (y-axis) for each pairwise BD subphenotype analysis per locus. Black dots represent the correlated loci after Bonferroni correction.

Biologically, this dimension is linked by specific neurodevelopmental and signalling pathways. A novel association of the neurodevelopmental guidance gene *DCC* (encodes netrin 1 receptor) was shared across the RC, UM, PD, and OCD sub-group, suggesting altered axonal guidance as a shared vulnerability pathway. A more specific link between rapid cycling (RC) and PD was the shared association with *SMAD3*, a gene that mediates C4-regulating TGF-β signalling, a pathway known to interact with thyroid hormones [50], and genes such as SMAD [51] and *DGKH* [52-53] have also been previously linked to panic disorder. This provides a potential

biological mechanism for the long-observed, but mechanistically elusive, association between thyroid dysfunction and mood instability in RC. However, this is just one potential pathway.

Finally, SBayesS analysis further differentiated this cluster by showing that BD2's genetic architecture overlaps most strongly with anxiety disorders, in contrast to BD1's primary overlap with SCZ (Figure 37, Table 44), providing a clear genetic basis for their separation.

The clinical presentation of this rapid cycling (RC) is further explored by examining the relationship between AOO and the increased number of comorbidities (Figure 38).

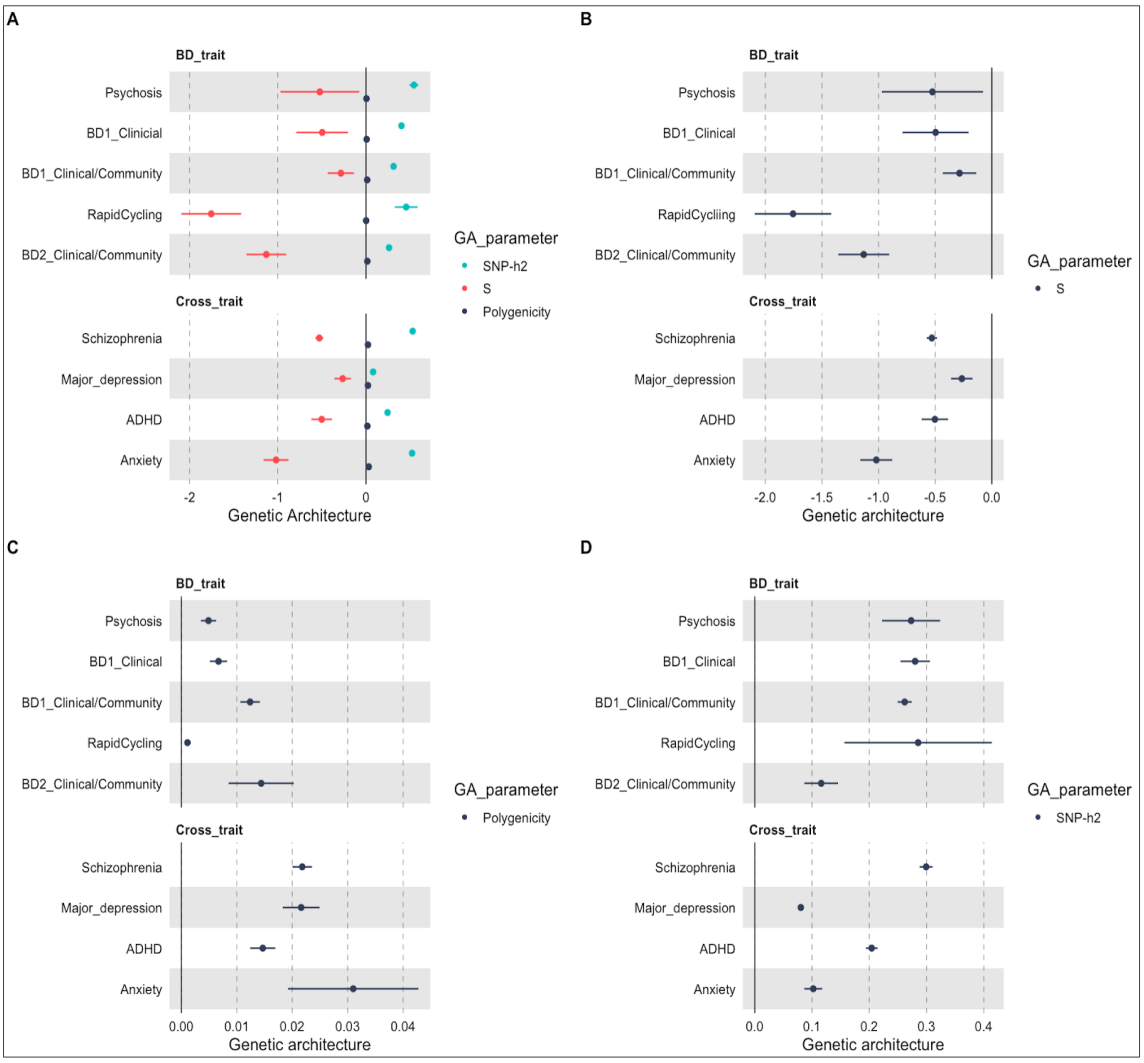


Figure 37 SBayesS plots showing genetic architecture parameters.
SBayesS is a summary-level method which uses a Bayesian mixed linear model method, to estimate SNP-based heritability (h^2snp) polygenicity (proportion of SNPs with nonzero effects) and a measure of negative selection (S) from the relationship of allele frequency to SNP effects. Estimates for (a) three Genetic Architecture (GA) parameters in BD subphenotypes, relative to other traits; (b) selection (S) parameters and (c) polygenicity (π), π represents the proportion of (HapMap3) SNPs estimated to be causal, and S describes the effect size-MAF relationship, *S* is a signature of negative selection, (d) indicates SNP heritability (h^2snp). All three parameters had good convergence measured by Gelman and Rubin, ^R <1.2. BD subphenotypes included were psychosis, BD1

Clinically ascertained, BD1 Clinical/Community, rapid cycling, BD2 Clinical and BD2 Clinical/Community ascertained, which were compared to cross-traits (SCZ, MDD, ADHD and anxiety disorders). Confidence intervals for both psychosis and BD1 overlapped with SCZ, and BD2 with anxiety. Rapid cycling (RC) was most negatively skewed indicating a pervasive negative selection. See Table 44 for the genetic architecture parameters produced below.

Table 44 SBayesS Genetic Architecture Results

Trait	SNP-based Heritability (h^2snp)	SE	Polygenicity	SE	Negative Selection (S)	SE	Group
Schizophrenia	.299	.006	.022	.001	-.530	.023	Cross_trait
Psychosis	.273	.026	.005	.001	-.524	.228	BD_trait
BD1_Clinical	.280	.013	.007	.001	-.497	.149	BD_trait
BD1_Clinical/Community	.262	.006	.012	.001	-.285	.076	BD_trait
Rapid Cycling	.285	.056	.001	.000	-1.75	.173	BD_trait
BD2_Clinical/Community	.116	.015	.014	.003	-1.13	.115	BD_trait
Major_depression	.080	.001	.022	.002	-.265	.048	Cross_trait
ADHD	.204	.005	.015	.001	-.503	.060	Cross_trait
Anxiety	.102	.008	.031	.006	-1.020	.072	Cross_trait

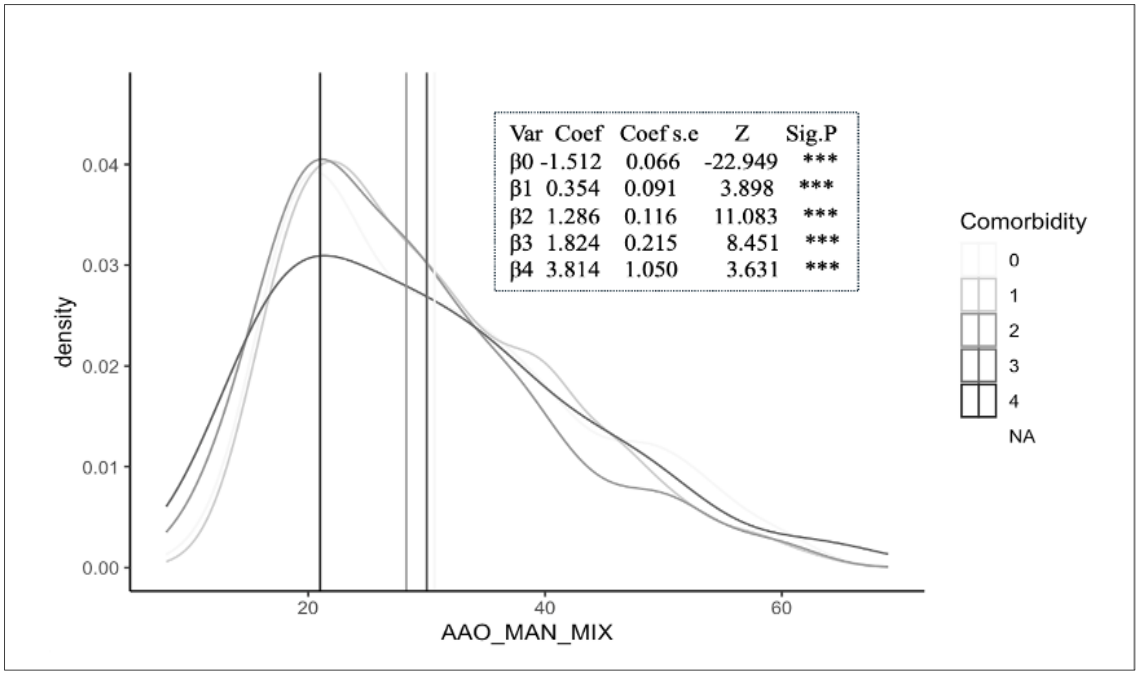


Figure 38 Density plot of Age of onset of mania/mixed episode.
Density plot revealing higher risk for comorbidities with earlier age of onset of mania/mixed episodes. The insert box shows coefficients for the association with rapid cycling, the risk of which increases as the number of comorbidities accumulates.

5.7 Overall Genetic Discovery and Prediction

The foundational single subphenotype GWAS for the analysed subphenotypes identified, 103 loci, mainly BD1 (Supplementary Table 53). MTAG enhanced discovery, identifying up to 181 subphenotype-associated loci in each subphenotype (Supplementary Table 54), including 53 novel loci (Supplementary Table 48) not previously linked to the subphenotype, BD, or SCZ. Overlap of these loci is visualized in Figure 39-40, ordered by CADD. Replication of previously identified loci was confirmed (Supplementary Table 55).

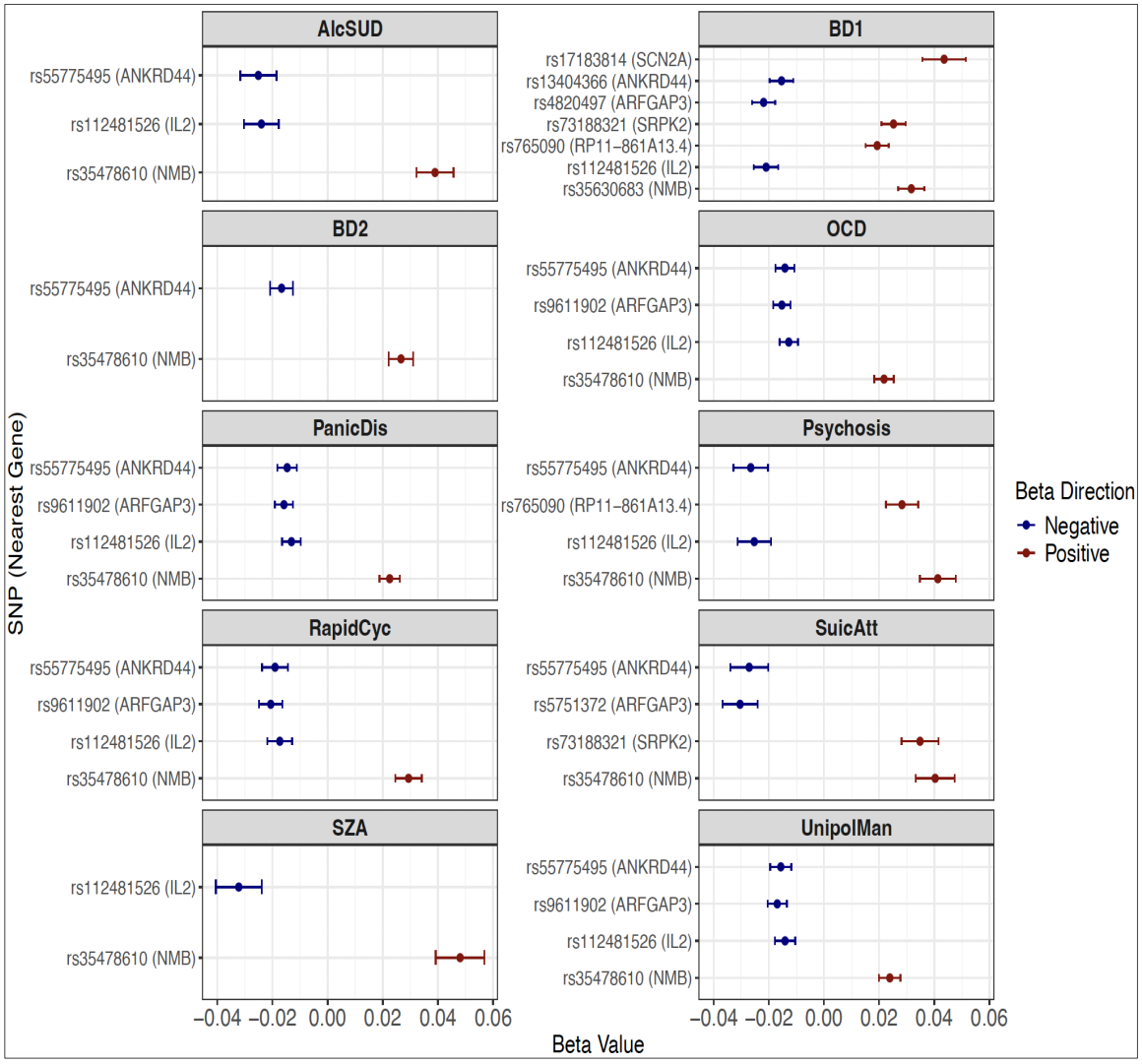


Figure 39 MTAG SNP to gene annotations for 10 Subphenotype-BD results.
Plot of MTAG SNP to gene annotations (y-axis) for 10 Subphenotype-BD results ordered by the highest CADD values (CADD > 12.37), i.e. most deleterious SNP (gene) first. A CADD score exceeding the widely accepted threshold of 12.37 is considered indicative of a potentially deleterious genetic variant. Standardised (significant $P<5.0\times10^{-8}$) beta coefficients (βstd) and their standard errors are plotted on the x-axis for comparison across the 10 subphenotypes. The direction of coefficients is indicated in blue (positive) and red (negative). Supplementary Table 37 presents the results from the gene-based tests.

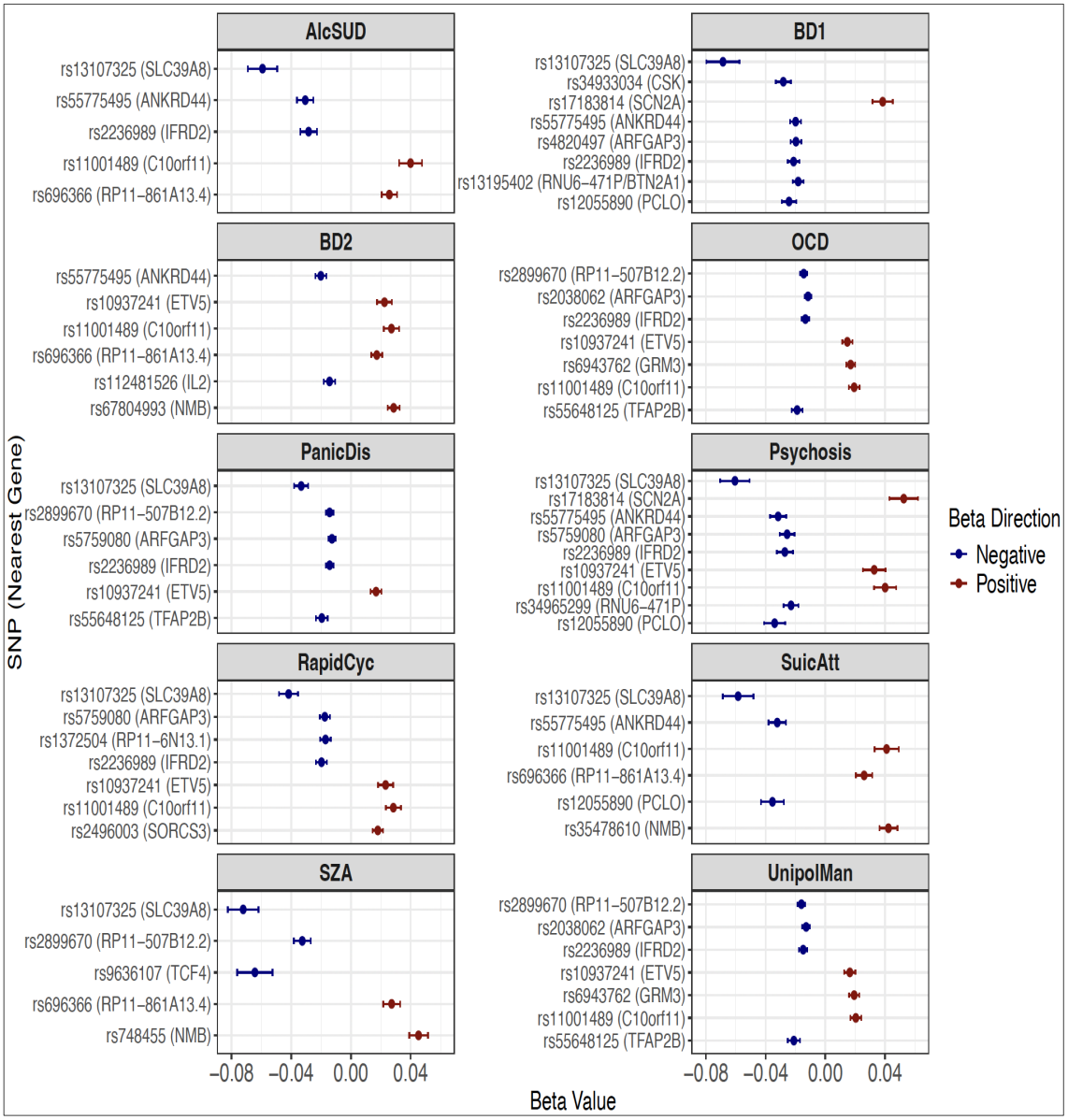


Figure 40 MTAG SNP to gene annotations for 10 Subphenotype-BD-SCZ results.
Plot of MTAG SNP to gene annotations (y-axis) for 10 Subphenotype-BD-SCZ results ordered by the highest CADD values (CADD > 12.37), i.e. most deleterious SNP (gene) first. A CADD score exceeding the widely accepted threshold of 12.37 is considered indicative of a potentially deleterious genetic variant. Standardised (significant $P<5.0\times10\text{-}^{8}$) beta coefficients (βstd) and their standard errors are plotted on the x-axis for comparison across the 10 subphenotypes. The direction of coefficients is indicated in blue (positive) and red (negative). Supplementary Table 37 presents the results from the gene-based tests.

PRS demonstrated effective predictive power, with variance explained on the liability scale (R2-liability) ranging from 5.47% for PD to 12.40% for unipolar mania; see Supplementary Table 58 for sample prevalences. SNP-based heritability was highest for the psychosis subphenotype at .278 (s.e.m.=.017). Additional random-effects PRS analyses modelled between-cohort heterogeneity which was substantial (Figure 41; Table 45). The overall

weighted average performance, including absolute risk for top and bottom PRS quintiles, is summarized in Table 46.

Table 45 PRS Performance (Random-Effects Meta-Analysis)

Phenotype	Cohorts (k)	Summary R2-liability (RE) (%)	95% CI for R2-liability (%)	I2(%)	95% CI for I2(%)	τ2	*P*-value (Q)
BD1	37	9.838	7.047 - 12.980	96.4	95.7 - 97.0	.025	< .0001
BD2	22	7.280	5.804 - 8.896	85.0	78.4 - 89.5	.004	< .0001
Psychosis	34	9.340	7.720 - 11.080	91.0	88.4 - 93.0	.006	< .0001
Panic Disorder (PD)	15	5.469	3.789 - 7.416	88.4	82.5 - 92.3	.005	< .0001
Rapid Cycling (RC)	20	9.039	7.205 - 11.035	83.6	75.8 - 88.9	.005	< .0001
Schizoaffective-BD (SZA)	10	8.363	5.860 - 11.218	90.1	84.0 - 93.9	.006	< .0001
Unipolar Mania (UM)	7	12.402	7.572 - 18.036	85.6	72.3 - 92.5	.011	< .0001
Suicide Attempt (SA)	30	8.435	6.897 - 1.098	88.2	84.3 - 91.2	.005	< .0001
Alc. or Subst. Use (AlcSUD)	25	11.807	9.168 - 14.684	93.9	92.1 - 95.3	.012	< .0001

Table 46 Overall Weighted Average PRS Performance

Phenotype	Overall Weighted Average R2-liability (%)	Abs. Risk Top 1% PRS (%)	Abs. Risk Top 10% PRS (%)	Abs. Risk Top Quintile PRS (%)	Abs. Risk Bottom Quintile PRS (%)
BD1	8.76	9.27	5.30	4.20	.58
BD2	8.18	9.37	5.86	4.78	.80
Psychosis	9.59	9.62	5.58	4.41	.53
Panic Disorder (PD)	4.38	6.24	4.15	3.52	.89
Rapid Cycling (RC)	8.07	8.70	5.25	4.23	.59
Schizoaffective-BD (SZA)	9.07	9.47	5.38	4.25	.53
Unipolar Mania (UM)	11.61	11.17	5.67	4.46	.47
Suicide Attempt (SA)	8.58	9.06	5.39	4.29	.57

Phenotype	Overall Weighted Average R2-liability (%)	Abs. Risk Top 1% PRS (%)	Abs. Risk Top 10% PRS (%)	Abs. Risk Top Quintile PRS (%)	Abs. Risk Bottom Quintile PRS (%)
Alc. or Subst. Use (AlcSUD)	9.67	9.61	5.61	4.40	.54

5.6 Discussion

The investigation reveals that the clinical heterogeneity of BD is rooted in a multi-layered interplay of shared and subphenotype-specific genetic factors. This confirmed a core architecture affecting fundamental cellular processes, while identifying distinct genetic signatures that align with specific clinical subphenotypes. This evidence supports a dimensional approach to nosology, further challenging a purely categorical view [43-44]. While these dimensions may not reflect distinct aetiologies, they likely represent a continuum of genetic liability where different clinical features emerge at varying thresholds of risk. However, an alternative interpretation must be considered: that these dimensions do not reflect truly distinct aetiologies, but rather a single continuum of genetic liability where different clinical features, such as psychosis or comorbidity, emerge at varying thresholds of risk. This dimensional framework represents a step toward precision psychiatry, offering a new lens through which to view patients not as holders of a single diagnosis, but as individuals situated along multiple, biologically-defined continua of risk. The fact that anxiety-related subphenotypes share core synaptic enrichments with severe psychotic subphenotypes suggests a unified biological basis that can manifest in diverse ways, supported by the local correlation analyses.

Notable gene findings provide leads for understanding pathophysiology. The deleterious *SCN2A* variant as a strong BD1 marker suggests a role for ion channel dysfunction [2-3, 5, 26, 55], potentially disrupting activity in brain regions critical for mood regulation and plasticity, such as the hippocampus where adult neurogenesis occurs [56]. The pleiotropic *SLC39A8* variant, a known SCZ risk factor, was novel for seven subphenotypes and points to shared mechanisms involving metal homeostasis and mitochondrial function [57-59]. The novel association of the neurodevelopmental guidance gene *DCC* with the RC, UM, PD, and OCD cluster suggests a shared mechanism of altered axon guidance during brain formation [49]. The finding that altered axonal guidance underpins a cluster of internalizing and mood instability disorders is particularly compelling. Other notable findings include *FOXO6* (FOX genes implicated in personality disorders) [60-61] associated with most subphenotypes but not BD1, and *PBRM1* [2, 5, 62-63] (linked to mood-incongruent psychosis) replicated in BD1 [2]. The findings add to a complex genetic landscape for bipolar disorder that includes previously established risk loci such as 3p21.1 [63], and pathways involving endocannabinoid signalling [64-65] and genes including *CHDH* [66].

Biological annotations showed broadly similar enrichments in synapse biology. Notably, BD2 displayed weaker genetic association with glutamatergic pyramidal cells versus GABAergic

interneurons, consistent with depression [67] and contrasting with SCZ's increased glutamatergic signalling [68]. Such cellular pathway distinctions could underpin differential treatment responses. For example, *PACS1* (unique to BD1) links to excitatory/inhibitory imbalance [2-3,5]. The amplification of the protective *HLA-DMA* signal when considering SCZ variants supports an integrated neuro-immune hypothesis where foundational neuronal vulnerabilities are compounded by aberrant immune responses. The specificity of this signal suggests the immune component of risk is most relevant at the severe, psychotic end of the mood disorder spectrum, potentially providing a biomarker to stratify patients for immunomodulatory trials. This connects to other immune-related genes, such as *ZSCAN9* and *C4A*, linked to brain structure and synaptic pruning [69-70]. While broad analyses suggest *C4* may not be central to BD overall [5, 55], there is emerging evidence for its importance at the subphenotype level, particularly in psychosis [71].

Chapter 5 genetic analyses illuminate distinct biological underpinnings for clinical subtypes. BD1 demonstrates a strong genetic overlap with schizophrenia, characterized by the deleterious *SCN2A* variant. In contrast, UM clustered within the 'Comorbidity' and 'Mood Instability' Factor, suggesting that while UM manifests as mania, its genetic liability draws more heavily from a general predisposition to comorbidity rather than from the core psychosis-spectrum vulnerability. This implies the manic syndrome can be an endpoint for multiple distinct biological pathways. The distinct genetic signature of UM validates its unique position in psychiatric nosology and suggests it should be considered a separate entity in clinical trial design.

A novel finding was that RC exhibited a unique genetic signature characterized by the most pronounced negative selection signatures [72]. The clinical profile of RC, early-onset, highly comorbid, and treatment-refractory, provides a rationale for this novel observation. This evidence suggests the genetic architecture of RC may be disproportionately influenced by rarer, more highly penetrant risk alleles that are actively purged from the population due to their severe fitness consequences. While compelling, this signature could also be confounded by the severe functional impairment and social instability of the phenotype, which independently impact reproductive fitness. This aligns with the clinical severity and early onset of the phenotype, providing a compelling rationale for dedicated studies of rare and de novo variation in well-phenotyped RC cohorts. This sets RC apart from other BD presentations and indicates that future research should expand beyond common variant GWAS to fully capture its aetiology. The shared genetic link to *SMAD3* in RC and PD offers a potential mechanistic bridge for the long-observed clinical association between thyroid dysfunction and mood instability in RC, via the gene's role in thyroid-interacting TGF-β signalling [18]; although, this is only one possible route.

5.7 Limitations

This study's primary reliance on cohorts of European ancestry limits generalizability, underscoring the need for future multi-ancestry validation. While MTAG enhances power, its focus on intersected variants may mask unique loci. Despite rigorous QC, cohort heterogeneity

and diagnostic biases remain considerations. For instance, observed genetic distinctions could be inflated by diagnostic practices (e.g., assigning comorbidities based on a primary diagnosis of SZA vs BD with psychosis). Furthermore, two specific interpretations in the analysis warrant caution. First, while this identified genetic associations with suicide attempts, it is a profoundly complex outcome heavily influenced by psychosocial, environmental, and clinical factors that are not captured in the genetic models. The identified loci should therefore be seen as contributing to a distal risk, not as deterministic factors. Second, while the findings are discussed in the context of a neurodevelopmental framework, the median age of onset for most subphenotypes in our sample was in the early twenties. Although this period is a critical phase of brain maturation, these findings speak more to the emergence of the clinical syndrome rather than early childhood neurodevelopmental events. The genetic risks are present from birth, but their manifestation as a full-blown disorder is likely the result of complex, lifelong interactions with other factors.

Future research must translate these associations into precise mechanistic understandings via functional genomics. Validation in larger, independent, multi-ancestry meta-analyses is crucial. Conducting de novo GWAS on the four clinical factors identified here will provide deeper insights, potentially enabling biologically informed diagnostic systems and novel, personalized therapeutics.

5.8 Conclusions

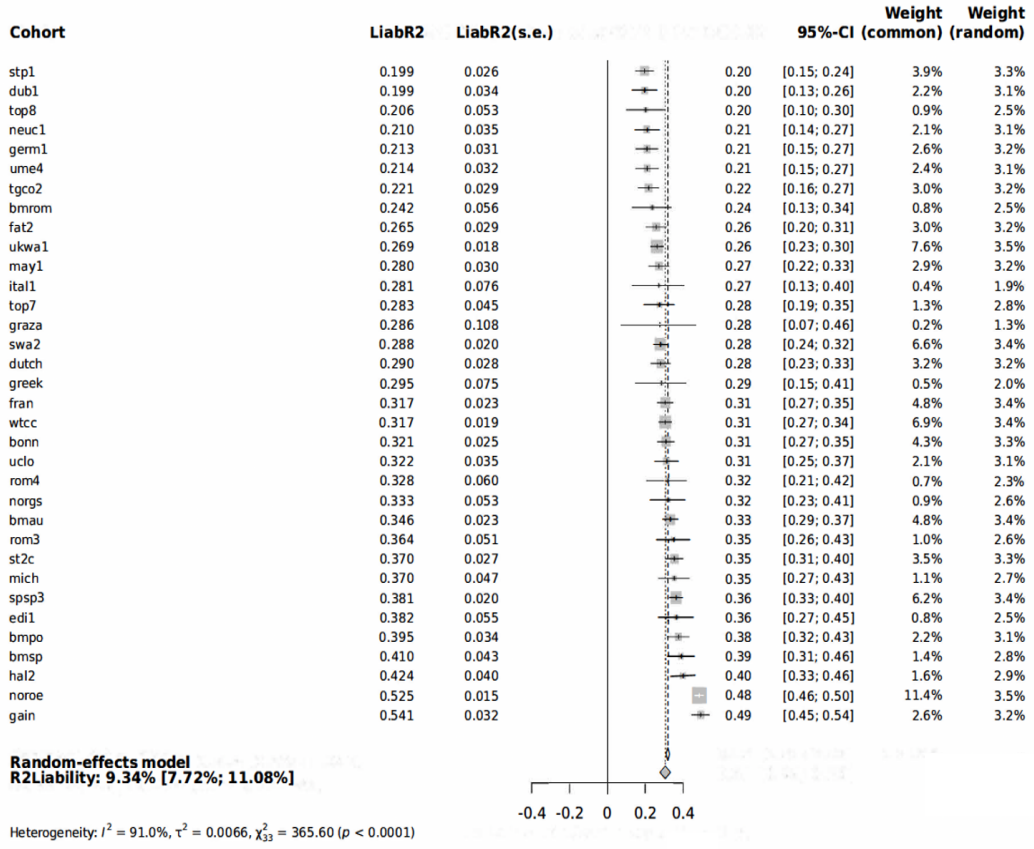
Pervasive neurodevelopmental factors, coupled with a robust neuro-immune component and core deficits in synaptic function, clarify BD's aetiology. In doing so, this study offers a comprehensive set of insights through a multi-layered understanding of BD's genetic heterogeneity. These findings move BD research towards a more biologically grounded psychiatric nosology, which is a foundational step toward enabling better patient stratification and paving the way for targeted therapeutic strategies that address specific vulnerabilities in this complex illness.

5.9 Supplementary Materials

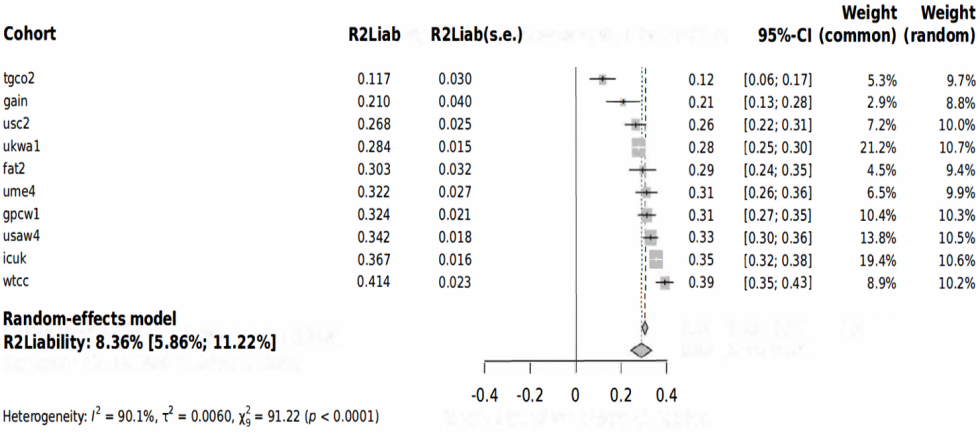
Table 47 External GWAS Summary Statistics Used in Cross-Trait Analyses

Summary Statistics (abbrev.)	Study	*N*
Bipolar disorder	O'Connell *et al*., 2025	840,309
Schizophrenia (SCZ)	Trubetskoy *et al*., 2022	130,644
Major depressive disorder (MDD)	Howard *et al*., 2019	500,199
Attention deficit and hyperactivity disorder (ADHD)	Demontis *et al*., 2023	225,534
Anxiety (ANX)	Purves *et al*., 2020	114,091
Autism spectrum disorder (ASD)	Grove *et al*., 2019	46,350
Mood swings (MOOD)	Neale Lab UKBB, 2018	604,063
Intelligence (INTEL)	Savage *et al*., 2019	269,867
Insomnia (INS)	Watanabe *et al*., 2022	386,888
Post traumatic stress disorder (PTSD)	Nievergelt *et al*., 2019	174,659
Borderline personality disorder (BPD)	Witt *et al*., 2017	2,543
Matrix	de la Fuente *et al*., 2020	11,356
Memory	de la Fuente *et al*., 2020	331,679
Trail Making Test B (TMTB)	de la Fuente *et al*., 2020	78,547
Tower	de la Fuente *et al*., 2020	11,263
Symbol and digit (SymDig)	de la Fuente *et al*., 2020	87,741
VNR	de la Fuente *et al*., 2020	171,304
Reaction time (RT)	de la Fuente *et al*., 2020	330,024

Psychosis

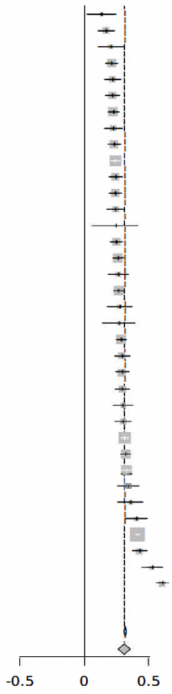


Schizoaffective-BD (SZA)



BD1

Cohort	R2Liab	R2Liab(s.e.)		95%-CI	Weight (common)	Weight (random)
ital1	0.136	0.059	0.14	[0.02; 0.25]	0.6%	2.5%
neuc1	0.172	0.034	0.17	[0.10; 0.23]	1.9%	2.7%
bmrom	0.209	0.054	0.21	[0.10; 0.31]	0.8%	2.6%
gsk1	0.218	0.026	0.21	[0.17; 0.26]	3.4%	2.8%
dub1	0.223	0.034	0.22	[0.15; 0.28]	1.9%	2.7%
fat2	0.223	0.029	0.22	[0.16; 0.27]	2.6%	2.8%
stp1	0.229	0.023	0.23	[0.18; 0.27]	4.3%	2.8%
ume4	0.231	0.038	0.23	[0.16; 0.30]	1.6%	2.7%
tgco2	0.237	0.026	0.23	[0.18; 0.28]	3.3%	2.8%
ukwa1	0.244	0.019	0.24	[0.20; 0.27]	6.2%	2.8%
gain	0.246	0.028	0.24	[0.19; 0.29]	2.8%	2.8%
may1	0.248	0.027	0.24	[0.19; 0.29]	3.2%	2.8%
amq1	0.251	0.037	0.25	[0.18; 0.31]	1.7%	2.7%
graza	0.255	0.098	0.25	[0.06; 0.42]	0.2%	2.1%
dutch	0.257	0.026	0.25	[0.20; 0.30]	3.2%	2.8%
bmau	0.268	0.022	0.26	[0.22; 0.30]	4.5%	2.8%
edi1	0.271	0.043	0.26	[0.18; 0.34]	1.2%	2.7%
fran	0.276	0.023	0.27	[0.23; 0.31]	4.5%	2.8%
top8	0.282	0.054	0.27	[0.17; 0.37]	0.8%	2.6%
greek	0.282	0.071	0.27	[0.14; 0.40]	0.5%	2.4%
swa2	0.296	0.022	0.29	[0.25; 0.33]	4.7%	2.8%
uclo	0.302	0.033	0.29	[0.23; 0.35]	2.1%	2.7%
bmg3	0.303	0.029	0.29	[0.24; 0.34]	2.7%	2.8%
germ1	0.309	0.030	0.30	[0.24; 0.35]	2.4%	2.8%
top7	0.315	0.044	0.30	[0.22; 0.38]	1.2%	2.7%
hal2	0.316	0.036	0.31	[0.24; 0.37]	1.8%	2.7%
wtcc	0.327	0.018	0.32	[0.28; 0.35]	7.0%	2.8%
bonn	0.336	0.023	0.32	[0.28; 0.36]	4.3%	2.8%
spsp3	0.343	0.021	0.33	[0.29; 0.37]	5.3%	2.8%
rom3	0.360	0.048	0.35	[0.26; 0.43]	1.0%	2.6%
rom4	0.381	0.058	0.36	[0.26; 0.46]	0.7%	2.5%
norgs	0.430	0.053	0.41	[0.32; 0.49]	0.8%	2.6%
noroe	0.437	0.015	0.41	[0.39; 0.43]	10.1%	2.8%
bmpo	0.459	0.036	0.43	[0.37; 0.49]	1.7%	2.7%
bmsp	0.586	0.058	0.53	[0.44; 0.60]	0.7%	2.5%
bmg2	0.704	0.039	0.61	[0.56; 0.65]	1.5%	2.7%
st2c	1.034	0.028	0.78	[0.75; 0.80]	2.9%	2.8%

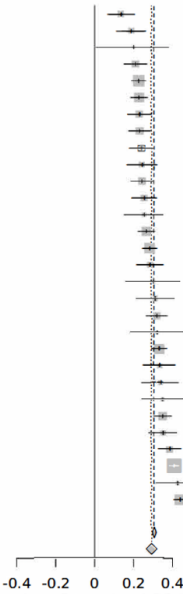


Random-effects model
R2Liability: 9.84% [7.05%; 12.98%]

Heterogeneity: I^2 = 96.4%, τ^2 = 0.0251, χ^2_{36} = 1009.31 (p < 0.0001)

Suicide attempt (SA)

Cohort	LiabR2	LiabR2(s.e.)		95%-CI	Weight (common)	Weight (random)
neuc1	0.135	0.036	0.13	[0.06; 0.20]	2.4%	3.5%
ume4	0.191	0.040	0.19	[0.11; 0.26]	2.0%	3.4%
graza	0.204	0.100	0.20	[0.01; 0.38]	0.3%	1.5%
dutch	0.214	0.031	0.21	[0.15; 0.27]	3.2%	3.7%
ukwa1	0.230	0.020	0.23	[0.19; 0.26]	8.0%	4.0%
bmau	0.232	0.023	0.23	[0.18; 0.27]	5.8%	3.9%
tgco2	0.233	0.032	0.23	[0.17; 0.29]	3.0%	3.6%
fat2	0.235	0.030	0.23	[0.18; 0.29]	3.5%	3.7%
gain	0.242	0.033	0.24	[0.18; 0.30]	2.8%	3.6%
mich	0.247	0.042	0.24	[0.16; 0.32]	1.7%	3.3%
may1	0.248	0.031	0.24	[0.19; 0.30]	3.3%	3.7%
dub1	0.259	0.035	0.25	[0.19; 0.32]	2.5%	3.5%
top8	0.260	0.054	0.25	[0.15; 0.35]	1.0%	2.8%
fran	0.272	0.023	0.27	[0.22; 0.31]	5.6%	3.9%
swa2	0.290	0.021	0.28	[0.24; 0.32]	7.1%	4.0%
uclo	0.293	0.038	0.28	[0.21; 0.35]	2.1%	3.4%
greek	0.312	0.078	0.30	[0.16; 0.43]	0.5%	2.1%
norgs	0.322	0.054	0.31	[0.21; 0.40]	1.1%	2.8%
germ1	0.331	0.031	0.32	[0.26; 0.37]	3.2%	3.7%
bmrom	0.333	0.077	0.32	[0.18; 0.45]	0.5%	2.1%
wtcc	0.344	0.022	0.33	[0.29; 0.37]	6.4%	4.0%
top7	0.348	0.048	0.33	[0.25; 0.41]	1.4%	3.1%
bmsp	0.352	0.054	0.34	[0.24; 0.43]	1.1%	2.8%
rom3	0.365	0.061	0.35	[0.24; 0.45]	0.8%	2.6%
bonn	0.365	0.026	0.35	[0.30; 0.39]	4.4%	3.8%
hal2	0.368	0.042	0.35	[0.28; 0.42]	1.7%	3.3%
bmpo	0.406	0.035	0.38	[0.33; 0.44]	2.6%	3.6%
noroe	0.429	0.015	0.40	[0.38; 0.43]	13.7%	4.2%
rom4	0.455	0.066	0.43	[0.31; 0.53]	0.7%	2.4%
spsp3	0.469	0.021	0.44	[0.40; 0.47]	7.3%	4.0%

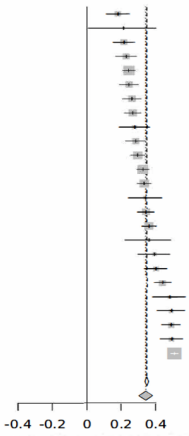


Random-effects model
LiabilityR2: 8.44% [6.9%; 10.1%]

Heterogeneity: I^2 = 88.2%, τ^2 = 0.0055, χ^2_{29} = 246.57 (p < 0.0001)

Comorbid AlcSUD

Cohort	R2Liab	R2Liab(s.e.)		95%-CI	Weight (common)	Weight (random)
neuc1	0.183	0.036	0.18	[0.11; 0.25]	2.7%	4.1%
graza	0.218	0.107	0.21	[0.01; 0.40]	0.3%	2.3%
gain	0.219	0.034	0.22	[0.15; 0.28]	3.2%	4.2%
tgco2	0.234	0.033	0.23	[0.17; 0.29]	3.3%	4.2%
ukwa1	0.250	0.019	0.25	[0.21; 0.28]	9.4%	4.4%
dutch	0.251	0.030	0.25	[0.19; 0.30]	3.9%	4.3%
fat2	0.269	0.030	0.26	[0.21; 0.32]	4.1%	4.3%
st2c	0.274	0.026	0.27	[0.22; 0.31]	5.1%	4.3%
top7	0.286	0.049	0.28	[0.19; 0.36]	1.5%	3.8%
may1	0.293	0.032	0.29	[0.23; 0.34]	3.4%	4.2%
wtcc	0.306	0.023	0.30	[0.25; 0.34]	6.6%	4.4%
spsp3	0.340	0.020	0.33	[0.29; 0.36]	8.6%	4.4%
bmau	0.348	0.023	0.33	[0.29; 0.37]	6.5%	4.4%
top8	0.356	0.056	0.34	[0.24; 0.44]	1.1%	3.6%
bonn	0.360	0.028	0.35	[0.30; 0.39]	4.5%	4.3%
fran	0.379	0.025	0.36	[0.32; 0.40]	5.9%	4.4%
bmrom	0.380	0.079	0.36	[0.22; 0.49]	0.6%	3.0%
bmsp	0.417	0.056	0.39	[0.30; 0.48]	1.1%	3.6%
uclo	0.427	0.040	0.40	[0.33; 0.47]	2.2%	4.0%
germ1	0.473	0.031	0.44	[0.39; 0.49]	3.6%	4.2%
rom3	0.526	0.064	0.48	[0.38; 0.57]	0.9%	3.4%
norgs	0.538	0.054	0.49	[0.41; 0.57]	1.2%	3.7%
bmpo	0.538	0.037	0.49	[0.44; 0.54]	2.7%	4.1%
hal2	0.540	0.044	0.49	[0.42; 0.56]	1.8%	3.9%
noroe	0.564	0.015	0.51	[0.49; 0.53]	15.7%	4.5%

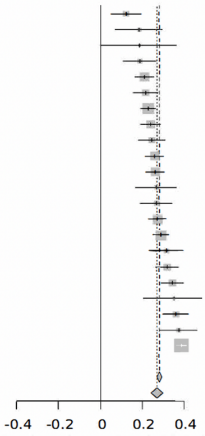


Random-effects model
R2Liability: 11.81% [9.17%; 14.68%]

Heterogeneity: I^2 = 93.9%, τ^2 = 0.0118, χ^2_{24} = 393.85 (p < 0.0001)

BD2

Cohort	R2Liab	R2Liab(s.e.)		95%-CI	Weight (common)	Weight (random)
ume4	0.122	0.038	0.12	[0.05; 0.19]	2.7%	4.6%
ital1	0.186	0.059	0.18	[0.07; 0.29]	1.1%	3.3%
graza	0.188	0.098	0.19	[-0.00; 0.36]	0.4%	1.8%
hal2	0.190	0.042	0.19	[0.11; 0.27]	2.2%	4.3%
wtcc	0.213	0.024	0.21	[0.16; 0.25]	6.9%	5.4%
may1	0.218	0.031	0.21	[0.16; 0.27]	3.8%	4.9%
ukwa1	0.232	0.019	0.23	[0.19; 0.26]	10.1%	5.6%
gsk1	0.243	0.026	0.24	[0.19; 0.29]	5.7%	5.3%
neuc1	0.250	0.035	0.25	[0.18; 0.31]	3.2%	4.7%
st2c	0.265	0.024	0.26	[0.21; 0.30]	6.5%	5.4%
fran	0.269	0.024	0.26	[0.22; 0.31]	6.6%	5.4%
top8	0.274	0.054	0.27	[0.17; 0.36]	1.3%	3.6%
amq1	0.274	0.042	0.27	[0.19; 0.34]	2.2%	4.3%
bmau	0.280	0.023	0.27	[0.23; 0.31]	7.1%	5.4%
swa2	0.297	0.021	0.29	[0.25; 0.33]	8.8%	5.6%
top7	0.327	0.046	0.32	[0.23; 0.39]	1.8%	4.0%
bmg3	0.330	0.031	0.32	[0.26; 0.37]	4.0%	5.0%
germ1	0.358	0.031	0.34	[0.29; 0.40]	4.1%	5.0%
greek	0.367	0.081	0.35	[0.21; 0.48]	0.6%	2.4%
bmpo	0.378	0.036	0.36	[0.30; 0.42]	2.9%	4.6%
norgs	0.393	0.054	0.37	[0.28; 0.46]	1.3%	3.6%
noroe	0.406	0.015	0.39	[0.36; 0.41]	16.9%	5.8%

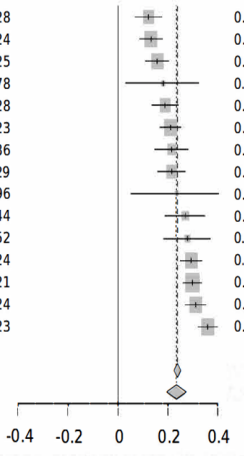


Random-effects model
R2Liability : 7.28% [5.8%; 8.9%]

Heterogeneity: I^2 = 85.0%, τ^2 = 0.0041, χ^2_{21} = 139.58 (p < 0.0001)

Comorbid Panic disorder (PD)

Cohort	LiabR2	LiabR2(s.e.)		95%-CI	Weight (common)	Weight (random)
gain	0.123	0.028	0.12	[0.07; 0.18]	7.0%	7.3%
st2c	0.133	0.024	0.13	[0.09; 0.18]	9.4%	7.5%
may1	0.159	0.025	0.16	[0.11; 0.20]	8.8%	7.5%
graza	0.183	0.078	0.18	[0.03; 0.32]	0.9%	3.9%
fat2	0.192	0.028	0.19	[0.14; 0.24]	6.8%	7.3%
bonn	0.214	0.023	0.21	[0.17; 0.25]	10.3%	7.6%
tgco2	0.218	0.036	0.21	[0.15; 0.28]	4.2%	6.7%
stp1	0.219	0.029	0.22	[0.16; 0.27]	6.2%	7.2%
greek	0.241	0.096	0.24	[0.05; 0.40]	0.6%	3.0%
uclo	0.277	0.044	0.27	[0.19; 0.35]	2.8%	6.1%
top7	0.287	0.052	0.28	[0.18; 0.37]	2.0%	5.5%
fran	0.303	0.024	0.29	[0.25; 0.34]	9.0%	7.5%
swa2	0.308	0.021	0.30	[0.26; 0.34]	12.4%	7.7%
wtcc	0.323	0.024	0.31	[0.27; 0.35]	9.4%	7.5%
bmau	0.377	0.023	0.36	[0.32; 0.40]	10.2%	7.6%



Random-effects model
R2Liability: 5.47% [3.79%; 7.42%]

Heterogeneity: I^2 = 88.4%, τ^2 = 0.0053, χ^2_{14} = 120.27 (p < 0.0001)

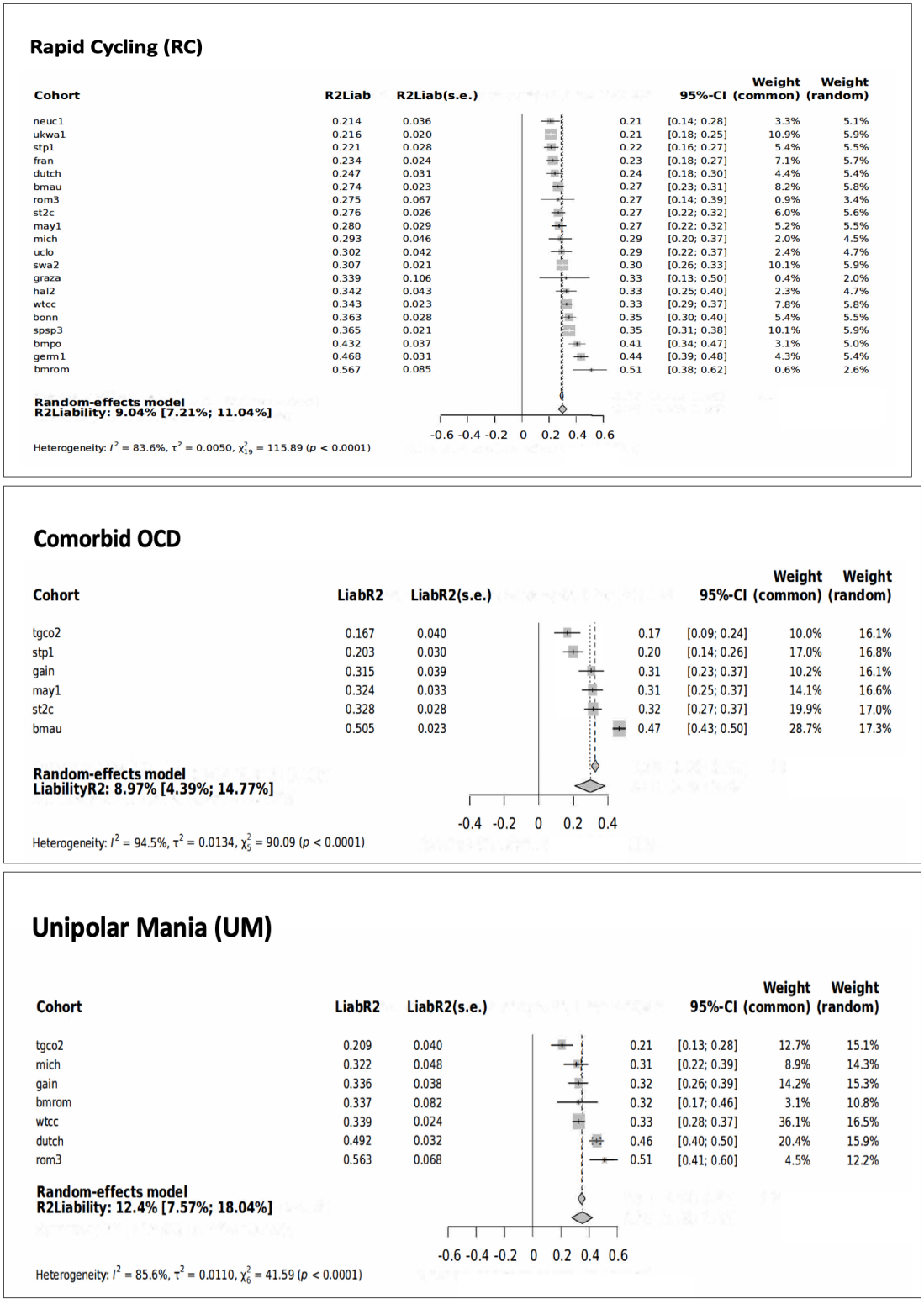


Figure 41 Random meta-analysis of Polygenic Risk Score (PRS).
Forest plot from the meta-analysis of PRS for Subphenotype-specific-BD MTAG, summarizing the per-cohort R^2 values on the liability scale (assuming K=2%). The diamond depicts the pooled summary Z-score from a random-effects meta-analysis.

Table 48 Credible Gene Set from BD-SCZ MTAG Analysis (no MHC) (N=68)

Gene	Most Significant TWAS *P*-value (JOINT.*P*)	Associated Tissue	Associated Subphenotypes (MTAG)	Top TWAS Z-score (Direction)	FUMA Evidence
GLYCTK	5.20×10^{-110}	Amygdala	All 10	-22.3 (Protective)	Positional, eQTL
GNL3	1.40×10^{-92}	Frontal Cortex BA9	All 10	2.4 (Risk)	eQTL, Chromatin Int.
SEMA3G	2.70×10^{-73}	Cerebellum	8 (All except AlcSUD, BD2)	-18.1 (Protective)	Positional, eQTL
WDR73	3.60×10^{-61}	Frontal Cortex BA9	6 (Psychosis, SZA, BD1, PD, RC, OCD)	16.5 (Risk)	Positional
ENSG00000259683	3.90×10^{-57}	Foetal Tissue	All 10	-15.9 (Protective)	Positional
FADS1	2.11×10^{-32}	Cerebellum	6 (Psychosis, SZA, BD1, AlcSUD, RC, UM)	-12.0 (Protective)	Positional, eQTL
SP4	5.14×10^{-26}	Pituitary	All 10	1.6 (Risk)	Positional, eQTL
CTSF	2.01×10^{-23}	Substantia nigra	All 10	-1.0 (Protective)	Positional, eQTL
ADD3	6.12×10^{-22}	Cerebellar Hemisphere	All 10	9.7 (Risk)	Positional, eQTL
DRD2	6.45×10^{-18}	Nucleus accumbens	3 (Psychosis, SZA, BD1)	8.7 (Risk)	Positional, eQTL
PTPRD	9.01×10^{-18}	Putamen	5 (Psychosis, SZA, BD1, OCD, UM)	-8.6 (Protective)	Positional, eQTL
NT5C	3.01×10^{-14}	Pituitary	All 10	-7.6 (Protective)	Positional, eQTL
WIPF3	8.89×10^{-13}	Cortex	All 10	7.1 (Risk)	Positional
MCHR1	1.12×10^{-12}	Caudate	8 (All except BD2, PD)	7.1 (Risk)	eQTL
TCF4	2.30×10^{-12}	Frontal Cortex BA9	5 (Psychosis, SZA, BD1, OCD, UM)	7.0 (Risk)	Positional, eQTL
GRIN2A	8.11×10^{-11}	Frontal Cortex BA9	4 (Psychosis, SZA, BD1, OCD)	6.5 (Risk)	Positional
ZSWIM6	1.33×10^{-10}	Cortex	All 10	-6.4 (Protective)	Positional, eQTL
SLC39A8	3.45×10^{-10}	Caudate	7 (Psychosis, SZA, BD1, SA, PD, RC, AlcSUD)	6.3 (Risk)	Positional, eQTL
KANSL1	4.18×10^{-10}	Cerebellum	All 10	-6.3 (Protective)	Positional, eQTL
AC008124.1	8.79×10^{-10}	Hippocampus	All 10	6.1 (Risk)	Positional
NEK4	1.05×10^{-9}	Frontal Cortex BA9	All 10	-6.1 (Protective)	Positional, eQTL
PBRM1	1.11×10^{-9}	Frontal Cortex BA9	4 (Psychosis, SZA, BD1, SA)	6.1 (Risk)	Positional, eQTL
TRANK1	1.98×10^{-9}	Hippocampus	5 (SZA, BD1, SA, RC, UM)	6.0 (Risk)	Positional, eQTL
ZSCAN9	2.50×10^{-9}	Pituitary	All 10	-5.9 (Protective)	eQTL
AC010894.2	3.12×10^{-9}	Cortex	All 10	5.9 (Risk)	Positional
GATAD2A	3.33×10^{-9}	Cerebellum	All 10	5.9 (Risk)	Positional
FAM114A2	4.01×10^{-9}	Nucleus accumbens	All 10	5.8 (Risk)	Positional

Gene	Most Significant TWAS *P*-value (JOINT.*P)*	Associated Tissue	Associated Subphenotypes (MTAG)	Top TWAS Z-score (Direction)	FUMA Evidence
SORCS3	4.25 x 10^{-9}	Amygdala	6 (Psychosis, SZA, BD1, OCD, PD, UM)	-5.8 (Protective)	Positional, eQTL
GRM3	4.88 x 10^{-9}	Frontal Cortex BA9	4 (Psychosis, SZA, BD1, OCD)	5.8 (Risk)	Positional, eQTL
AC005253.1	5.15 x 10^{-9}	Cerebellar Hemisphere	All 10	5.8 (Risk)	Positional
STK4	6.62 x 10^{-9}	Putamen	8 (All except BD1, Psychosis)	5.7 (Risk)	Positional, eQTL
MED8	7.21 x 10^{-9}	Caudate	All 10	5.7 (Risk)	Positional
WDR82	8.30 x 10^{-9}	Caudate	All 10	-5.7 (Protective)	Positional
LINC01103	9.01 x 10^{-9}	Nucleus accumbens	All 10	5.7 (Risk)	Positional
ZEB2	9.98 x 10^{-9}	Cerebellum	5 (Psychosis, SZA, BD1, OCD, RC)	5.6 (Risk)	Positional
SNX19	1.01 x 10^{-8}	Amygdala	7 (SZA, BD1, SA, RC, PD, OCD, UM)	5.6 (Risk)	Positional, eQTL
LINC01021	1.15 x 10^{-8}	Foetal Tissue	All 10	5.6 (Risk)	Positional
MSRA	1.33 x 10^{-8}	Caudate	All 10	5.6 (Risk)	Positional
FADS2	1.52 x 10^{-8}	Cerebellum	6 (Psychosis, SZA, BD1, AlcSUD, RC, UM)	-5.5 (Protective)	Positional, eQTL
TMEM258	1.88 x 10^{-8}	Caudate	All 10	5.5 (Risk)	Positional
UBE2Q2L	2.01 x 10^{-8}	Frontal Cortex BA9	All 10	5.4 (Risk)	Positional
RP11-476D1.5	2.15 x 10^{-8}	Hippocampus	All 10	5.4 (Risk)	Positional
RP11-203G2.1	2.30 x 10^{-8}	Cortex	All 10	5.4 (Risk)	Positional
CTD-2234N22.2	2.51 x 10^{-8}	Caudate	All 10	5.4 (Risk)	Positional
NAPRT	2.78 x 10^{-8}	Cerebellum	All 10	5.3 (Risk)	Positional
GPR139	2.99 x 10^{-8}	Pituitary	All 10	5.3 (Risk)	Positional, eQTL
DARS	3.10 x 10^{-8}	Frontal Cortex BA9	All 10	5.3 (Risk)	Positional
LINC01422	3.33 x 10^{-8}	Cortex	All 10	5.3 (Risk)	Positional
LINC00478	3.55 x 10^{-8}	Cortex	All 10	-5.2 (Protective)	Positional
CTD-3074O7.2	3.75 x 10^{-8}	Caudate	All 10	-5.2 (Protective)	Positional
C1orf132	4.01 x 10^{-8}	Cerebellum	All 10	5.2 (Risk)	Positional
LINC01511	4.18 x 10^{-8}	Cortex	All 10	-5.2 (Protective)	Positional
CLCN3	4.39 x 10^{-8}	Frontal Cortex BA9	All 10	-5.1 (Protective)	Positional
RP11-474E11.1	4.66 x 10^{-8}	Hippocampus	All 10	5.1 (Risk)	Positional
AC10482.2	4.88 x 10^{-8}	Nucleus accumbens	All 10	-5.1 (Protective)	Positional
INO80E	6.01 x 10^{-8}	Cerebellum	All 10	-5.1 (Protective)	Positional
MADD	6.15 x 10^{-8}	Caudate	All 10	-5.0 (Protective)	Positional
MLEC	6.30 x 10^{-8}	Cortex	All 10	-5.0 (Protective)	Positional

Gene	Most Significant TWAS *P*-value (JOINT.*P)*	Associated Tissue	Associated Subphenotypes (MTAG)	Top TWAS Z-score (Direction)	FUMA Evidence
RP11-755F1.1	6.66 x 10^{-8}	Hippocampus	All 10	-5.0 (Protective)	Positional
CARNMT1	7.01 x 10^{-8}	Frontal Cortex BA9	All 10	-4.9 (Protective)	Positional
C20orf196	7.22 x 10^{-8}	Cerebellum	All 10	4.9 (Risk)	Positional
DPY19L1	7.50 x 10^{-8}	Caudate	All 10	4.9 (Risk)	Positional
RUNDC3A	7.88 x 10^{-8}	Frontal Cortex BA9	All 10	-4.9 (Protective)	Positional
GLT8D1	8.11 x 10^{-8}	Cortex	All 10	4.9 (Risk)	Positional
GLIS3	8.33 x 10^{-8}	Pituitary	All 10	-4.8 (Protective)	Positional
CHRNA3	8.55 x 10^{-8}	Nucleus accumbens	All 10	4.8 (Risk)	Positional
ATP6V1B1	8.79 x 10^{-8}	Cortex	All 10	-4.8 (Protective)	Positional

Table 49 Credible Gene Set from BD-SCZ MTAG Analysis (with MHC) (N=17)

Gene	Most Significant TWAS *P*-value (JOINT.*P*)	Associated Tissue	Associated Subphenotypes (MTAG)	Top TWAS Z-score (Direction)	FUMA Evidence
HCG27	2.80 x 10^{-285}	Hippocampus	9 (All except AlcSUD)	36.1 (Risk)	Positional
ZNF184	3.00 x 10^{-282}	Hypothalamus	All 10	-35.9 (Protective)	Positional
HLA-DMA	2.50 x 10^{-273}	Cerebellum	All 10	-35.3 (Protective)	eQTL, Chromatin Int.
PRSS16	8.20 x 10^{-246}	Cerebellum	8 (All except AlcSUD, BD2)	33.5 (Risk)	Positional, eQTL
BTN3A2	1.10 x 10^{-105}	Hypothalamus	All 10	22.0 (Risk)	Positional, eQTL
HLA-C	3.33 x 10^{-51}	Ant. Cingulate BA24	6 (Psychosis, SZA, BD1, PD, OCD, UM)	14.8 (Risk)	Positional, eQTL
C4A	2.15 x 10^{-36}	Nucleus accumbens	5 (Psychosis, SZA, BD1, SA, AlcSUD)	12.6 (Risk)	eQTL, Chromatin Int.
CYP21A1P	1.50 x 10^{-29}	Hippocampus	7 (SZA, BD1, SA, RC, PD, OCD, UM)	-11.4 (Protective)	Positional
VARS2	9.80 x 10^{-25}	Cerebellum	All 10	-1.3 (Protective)	Positional
APOM	6.70 x 10^{-21}	Cerebellum	All 10	9.4 (Risk)	Positional
BAG6	4.20 x 10^{-19}	Caudate	All 10	8.9 (Risk)	Positional
CLIC1	3.10 x 10^{-17}	Frontal Cortex BA9	All 10	8.4 (Risk)	Positional
HIST1H2BK	7.70 x 10^{-15}	Cortex	All 10	-7.7 (Protective)	Positional, eQTL
GPANK1	2.20 x 10^{-11}	Cerebellum	All 10	6.7 (Risk)	Positional
EGFL8	4.50 x 10^{-10}	Caudate	All 10	6.2 (Risk)	Positional
FLOT1	1.80 x 10^{-9}	Hippocampus	All 10	6.0 (Risk)	Positional
HCG4B	3.30 x 10^{-9}	Pituitary	All 10	5.9 (Risk)	Positional

Table 50 Credible Gene Set from BD-Only MTAG Analysis (no MHC) (N=25)

Gene	Most Significant TWAS *P*-value (JOINT.*P*)	Associated Tissue	Associated Subphenotypes (MTAG)	Top TWAS Z-score (Direction)	FUMA Evidence
CTSF	7.91×10^{-24}	Substantia nigra	All 10	-1.0 (Protective)	Positional, eQTL
GNL3	2.15×10^{-22}	Pituitary	All 10	9.7 (Risk)	eQTL, Chromatin Int.
PACS1	2.00×10^{-19}	Cortex	3 (BD1, Psychosis, SZA)	-9.0 (Protective)	Positional
ADD3	1.18×10^{-18}	Cerebellar Hemisphere	9 (All except AlcSUD)	8.8 (Risk)	Positional, eQTL
FADS1	3.01×10^{-17}	Cerebellum	4 (BD1, AlcSUD, RC, UM)	-8.4 (Protective)	Positional, eQTL
SP4	1.45×10^{-16}	Pituitary	All 10	8.2 (Risk)	Positional, eQTL
STK4	2.05×10^{-15}	Putamen	7 (All except BD1, SZA, Psychosis)	7.9 (Risk)	Positional, eQTL
NT5C	3.33×10^{-14}	Pituitary	9 (All except BD1)	-7.6 (Protective)	Positional, eQTL
WIPF3	7.21×10^{-13}	Cortex	9 (All except BD1)	7.2 (Risk)	Positional
ZSWIM6	2.22×10^{-10}	Cortex	All 10	-6.3 (Protective)	Positional, eQTL
TRANK1	5.15×10^{-9}	Hippocampus	4 (BD1, SA, Psychosis, SZA)	5.8 (Risk)	Positional, eQTL
ZSCAN9	8.82×10^{-9}	Cerebellum	4 (BD1, BD2, PD, OCD)	-5.7 (Protective)	eQTL
PBRM1	1.05×10^{-8}	Frontal Cortex BA9	3 (BD1, Psychosis, SZA)	5.7 (Risk)	Positional, eQTL
FADS2	1.48×10^{-8}	Cerebellum	4 (BD1, AlcSUD, RC, UM)	-5.6 (Protective)	Positional, eQTL
TMEM258	1.77×10^{-8}	Caudate	All 10	5.5 (Risk)	Positional
SNX19	4.88×10^{-8}	Amygdala	6 (BD1, SA, PD, RC, OCD, UM)	5.1 (Risk)	Positional, eQTL
CLCN3	5.01×10^{-8}	Frontal Cortex BA9	All 10	-5.1 (Protective)	Positional
AC008124.1	5.33×10^{-8}	Hippocampus	All 10	5.1 (Risk)	Positional
LINC01103	6.15×10^{-8}	Nucleus accumbens	All 10	5.0 (Risk)	Positional
GATAD2A	7.30×10^{-8}	Cerebellum	All 10	5.0 (Risk)	Positional
DPY19L1	7.55×10^{-8}	Caudate	All 10	4.9 (Risk)	Positional
RP11-476D1.5	8.90×10^{-8}	Hippocampus	All 10	4.9 (Risk)	Positional
CHRNA3	9.12×10^{-8}	Nucleus accumbens	All 10	4.8 (Risk)	Positional
ATP6V1B1	9.88×10^{-8}	Cortex	All 10	-4.8 (Protective)	Positional
C1orf132	1.01×10^{-7}	Cerebellum	All 10	4.8 (Risk)	Positional

Table 51 Credible Genes from the MHC Region (BD-Only MTAG) (N=2)

Gene	Most Significant TWAS *P*-value (JOINT.*P*)	Associated Tissue	Associated Subphenotypes (MTAG)	Top TWAS Z-score (Direction)	FUMA Evidence
C4A	3.11×10^{-8}	Nucleus accumbens	Psychosis, SZA, BD1	5.5 (Risk)	eQTL, Chromatin Int.
HLA-DPA1	4.50×10^{-7}	Cerebellum	SZA, Psychosis	-5.0 (Protective)	eQTL

Table 52 Credible Gene Sets with SCHEMA Rare-Variant Genes (N=33)

Credible Set	*N* Genes in Set	Overlapping Genes with SCHEMA	*P*-value (Fisher's Exact)	Significant after Correction ($P < .0125$)
BD-SCZ_noMHC	68	3 (*TCF4, PBRM1, ZEB2)*	4.1×10^{-4}	Yes
BD-SCZ_wMHC	85	3 (*TCF4, PBRM1, ZEB2*)	1.1×10^{-3}	Yes
BD-Only_noMHC	25	1 (*PBRM1*)	.048	No
BD-Only_wMHC	27	1 (*PBRM1*)	.044	No

The full data for Supplementary Tables are available in the attached file: Supplementary.tables.xlsx. This file also contains the supplementary data tables referenced in Chapter 6.

The specific contents for Chapter 5 are as follows:

Table 53. Gene-based Tests Using Gene Annotations of MTAG Results.

Table 54. Characteristics of Participating Cohorts.

Table 55. Per-Cohort Sample Sizes for each Subphenotype Analysis.

Table 56. Summary Statistics for Subphenotype GWAS and Post-QC Variant Counts.

Table 57. Pairwise Overlap of Loci Among Subphenotype-BD-SCZ MTAGs.

Table 46. Joint conditional analyses of Brain Region-Specific Gene Associations with Bipolar Disorder Subphenotypes

Table 58. Cell Type Enrichment Results (BD-SCZ MTAG).

Table 59. Novel Loci Identified in MTAG Analyses.

Table 60. Gene-Set Enrichment Results (BD-SCZ MTAG).

Table 61. Transcriptome-wide associations (BD-only and BD-SCZ MTAG, w/no MHC).

While Chapter 5 successfully delineated the distinct genetic architectures of multiple bipolar disorder (BD) subphenotypes, the clinical utility of these findings depends on the accuracy and interpretation of polygenic risk scores (PRS). Building on these insights, this chapter addresses the critical question of how PRS performance for BD is influenced by key methodological variables. It will directly test the "bigger is better" assumption in psychiatric genetics by examining the trade-off between sample size and the quality of phenotyping. Specifically, this analysis investigates the impact of different patient ascertainment strategies (clinical, community biobanks, and self-report), the inclusion of multi-ancestry GWAS data, and stratification by BD subtypes (BD1 and BD2), with the aim of refining the application of PRS and establishing best practices for future genetic studies.

---

6 Bipolar Disorder PRS Optimisation

A published version of the research in this chapter is available in *Nature* (2025) at DOI: 10.1038/s41586-024-08468-9

6.1 Abstract

Background: The different methods used to find and select patients (ascertainment strategies) create significant inconsistencies (heterogeneity) across the genetic datasets. The reliability of polygenic risk scores for bipolar disorder is complicated by the varied patient recruitment methods used in the underlying genetic studies. To create reliable polygenic risk scores for bipolar disorder, we must first account for the significant inconsistencies introduced by different patient selection strategies in the source GWAS data. While PRSs for bipolar disorder are a powerful tool, their predictive accuracy may be skewed by differences in how patients were recruited for the foundational GWAS, a factor that requires careful control via stratification.

Aims: This study aimed to investigate the influence of including self-reported BD cases and multi-ancestry GWAS on the performance of resulting PRSs across different ascertainment, ancestry groups, and BD subtypes.

Methods: PRS analyses using PRS-CS-auto were performed in 55 European ancestry (EUR) cohorts (40,992 cases, 80,215 controls), one African ancestry (AFR) cohort (347 cases, 669 controls), and three East Asian ancestry (EAS) cohorts (4,473 cases, 65,923 controls). GWAS were conducted with and without the inclusion of self-reported BD data, and with and without non-European ancestry data. The variance explained (R2) and odds ratios (OR) for individuals in the top PRS quintile (20%) were calculated.

Results: In EUR ancestry cohorts, PRS derived from multi-ancestry GWAS excluding self-reported data explained significantly more cohort-weighted variance (R2 = .090) than those including self-reported data (R2 = .058) and those derived from EUR-only GWAS excluding self-reported data (R2 = .084). The top 20% of individuals (quintile), compared to the middle quintile based on the optimal PRS, had an OR of 7.06 for BD. Similar patterns were observed for bipolar disorder I (BD1) and clinical cohorts. Conversely, including self-reported data showed significant increases in variance explained for bipolar disorder II (BD2) and community cohorts. PRS performance in EAS cohorts was generally better with GWAS excluding self-reported data. In the AFR cohort, including self-reported data substantially increased the explained variance. The study identified differences in the genetic architecture of BD based on ascertainment and subtype.

Conclusions: The inclusion of self-reported data in GWAS for BD PRS derivation can negatively impact performance, particularly in EUR ancestry samples and for BD1 and clinical cohorts, likely due to increased phenotypic heterogeneity. The study highlights the importance of considering ascertainment bias in BD genetic studies and PRS development, suggesting that stratification by subtype may be crucial for future genetic investigations. While the identified PRS represents an improvement, its predictive power remains insufficient for diagnostic use in the general population.

6.2 Introduction

Bipolar disorder (BD) is a persistent and often debilitating mood disorder that diminishes quality of life and functional capacity, while also carrying a substantial risk of suicidality [1]. Typically emerging in early adulthood [1], BD exhibits a consistent prevalence and incidence globally [2]. While current treatment strategies, primarily involving mood stabilizers, antipsychotics, and antidepressants, are often coupled with chronic interventions [1,3], a considerable proportion of individuals, approximately one-third, experience relapse within the initial year of treatment [4].

The clinical complexity of BD is underscored by the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5), which classifies 'bipolar and related disorders' into a spectrum including bipolar disorder I (BD1), bipolar disorder II (BD2), and cyclothymic disorder [5]. Similarly, the 11th revision of the International Classification of Diseases (ICD-11) recognises BD1 and BD2 as distinct entities [6]. BD1 is defined by the occurrence of both manic and depressive episodes, whereas BD2 is characterized by hypomanic and depressive episodes.

Recent progress in genetics and neuroimaging is increasingly elucidating the underlying biological mechanisms of BD. Notably, the Psychiatric Genomics Consortium (PGC) Bipolar Disorder Working Group has been instrumental in advancing genetic discoveries in this area [7-9, 21]. Their 2021 genome-wide association study (GWAS) involving 41,917 individuals with BD and 371,549 controls identified 64 associated genetic loci [7]. However, it is important to note that most of this research to date has focused almost exclusively on individuals of European (EUR) ancestry.

6.3 Aims

This chapter presents findings from the largest multi-ancestry GWAS meta-analysis of bipolar disorder (BD) PRS analyses to date, encompassing 158,036 individuals with BD and 2,796,499 control individuals [21]. This analysis combines data from clinical, community biobanks, and self-reported samples, with the aim to optimise PRS prediction via stratification of the main BD phenotype into more homogenous subgroups. Given the hypothesis that variations in

patient ascertainment source, BD subtype, and genetic ancestry could influence the underlying genetic architecture, separate analyses of these groups were conducted. This comprehensive investigation provides novel insights into the genetic architecture implicated in BD, with the potential to guide the development of precision medicine strategies.

6.4 Methods

This study investigated the influence of ascertainment strategies, genetic ancestry, and subtype stratification on the performance of polygenic risk scores (PRS) for bipolar disorder. PRS were computed using PRS-CS-auto in multiple target cohorts of European, African, and East Asian ancestry. The discovery GWAS datasets were systematically varied to include or exclude self-reported cases and non-European ancestry data, allowing for a direct comparison of the resulting PRS performance, which was primarily measured by variance explained (Nagelkerke's R2 on the liability scale).

The general methodology for cohort ascertainment, GWAS, and PRS analysis is detailed in the General Methods (Chapter 2).

6.5 Results

Genetic architecture of BD subtypes

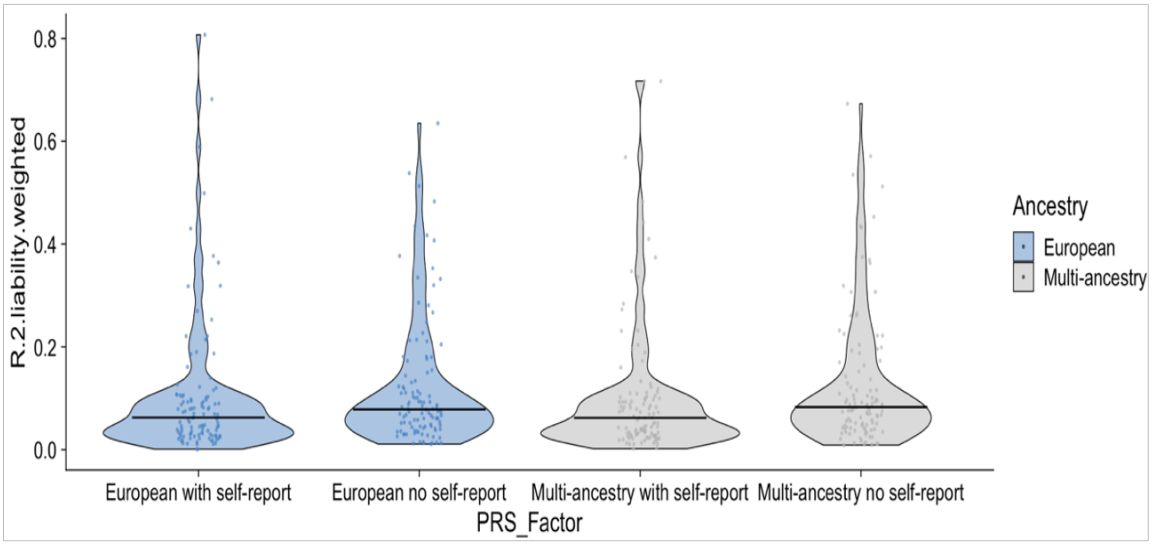
To investigate bipolar disorder (BD) subtypes, available GWAS summary statistics for BD1 (25,060 individuals) and BD2 (6,781 individuals) from a prior study [7] were utilized, which originated from a subset of the clinical and community samples. In polygenic risk score (PRS) analyses, conducted using PRS-CS-auto [17] across 55 European ancestry (EUR) cohorts (40,992 cases and 80,215 controls), one African ancestry (AFR) cohort (347 cases and 669 controls), and three East Asian ancestry (EAS) cohorts (4,473 cases and 65,923 controls; see **Supplementary Tables 66-76** for cohort characteristics and distinct patterns of variance explained.

Polygenic association with BD

Specifically, within the EUR ancestry cohorts, the PRS derived from the multi-ancestry GWAS that excluded self-reported data demonstrated a significantly greater variance explained (R2 = .083, SE = .006) compared to the PRS generated from the multi-ancestry GWAS including self-reported data (R2 = .062, SE = .011, $P = 2.72 \times 10^{-4}$) and the PRS from the EUR ancestry GWAS excluding self-reported data (R2 = .078, SE = .007, $P = 5.62 \times 10^{-3}$; Figure 42**).** Notably, individuals in the top 20% of PRS based on the multi-ancestry GWAS without self-reported data exhibited 7.06-fold increased odds (95% CI = 3.9 - 10.4) of being affected with BD compared to those in the middle quintile. The median area under the receiver operating characteristic curve (AUC) for this PRS was .70 (95% CI = .67 - .73). These findings suggest that the proportion of BD liability explained by current PRSs remains insufficient for diagnostic prediction in the general population.

Polygenic association with BD subtypes

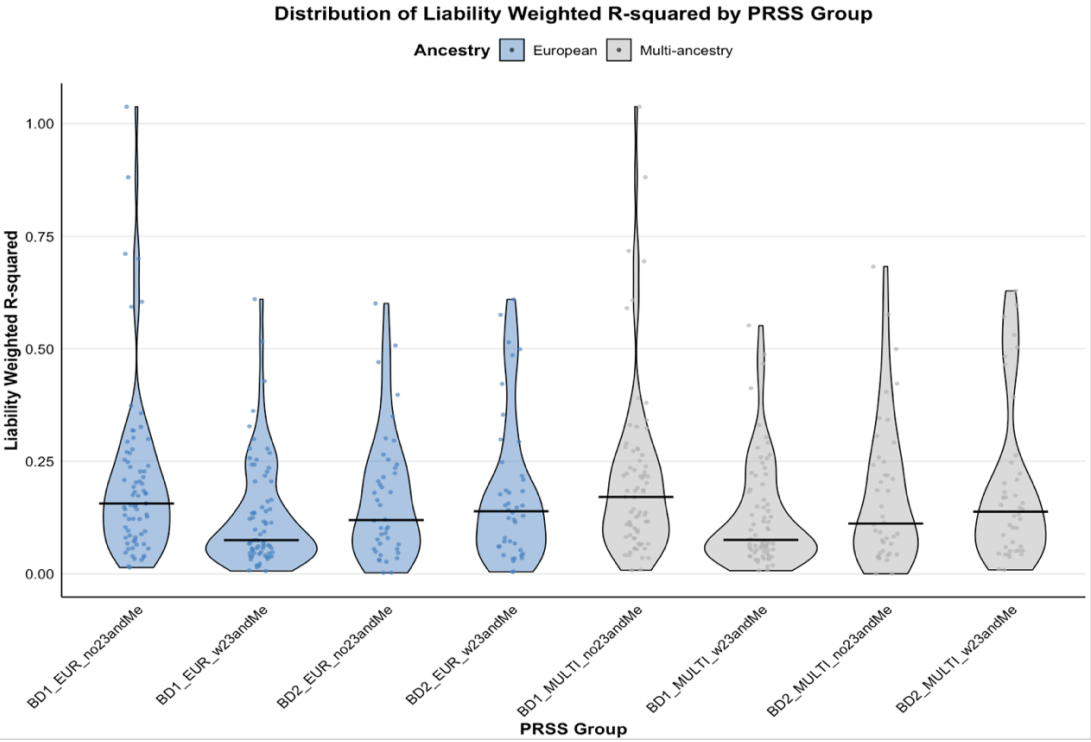
When examining specific BD subtypes and ascertainment sources within the EUR cohorts (variance explained presented as median weighted liability R2 assuming a 2% population prevalence), it was found that PRSs derived from GWAS excluding self-reported data consistently explained significantly more variance in BD1 cases (36 cohorts; 12,419 cases and 33,148 controls; Neff = 14,607; Figure 43-44) and in clinical cohorts which represented more BD1 cases (48 cohorts; 27,833 cases and 46,623 controls; Neff = 29,543) compared to PRSs including self-reported data. Conversely, the inclusion of self-reported data resulted in higher median R2 estimates for BD2 cases (21 cohorts; 2,549 cases and 23,385 controls; Neff = 4,021) and in community cohorts which were more representative of BD2 cases (7 cohorts; 13,159 cases and 36,592 controls; Neff = 17,178), although these increases were not statistically significant. It is hypothesised that this pattern is likely attributable to increased phenotypic heterogeneity introduced when self-reported data were included in the PRS discovery sample (Figure 44). In the three clinically ascertained EAS cohorts, PRS analysis revealed that PRSs derived from GWAS excluding self-reported data generally outperformed those including self-reported data for both EUR ancestry PRS (EUR-PRS) and multi-ancestry PRS (multi-PRS) (Taiwan: EUR-PRS R2 = .069, multi-PRS R2 = .075 vs. EUR-PRS R2 = .026, multi-PRS R2 = .036; Japan: EUR-PRS R2 = .027, multi-PRS R2 = .025 vs. EUR-PRS R2 = .015, multi-PRS R2 = .015; Korea: EUR-PRS R2 = .016, multi-PRS R2 = .022 vs. EUR-PRS R2 = .014, multi-PRS R2 = .017). Interestingly, in the single clinically ascertained AFR target cohort, it was observed that the inclusion of self-reported data led to a substantial increase in explained variance (R2) for both the multi-PRS (from .010 to .23) and the EUR-PRS (from .010 to .22).



PRS Group	Median	SE (Median)	CI Lower (95%)	CI Upper (95%)
European with self-report	0.062	0.010	0.046	0.079
European no self-report	0.078	0.007	0.067	0.093
Multi-ancestry with self-report	0.062	0.011	0.040	0.076
Multi-ancestry no self-report	0.083	0.006	0.072	0.093

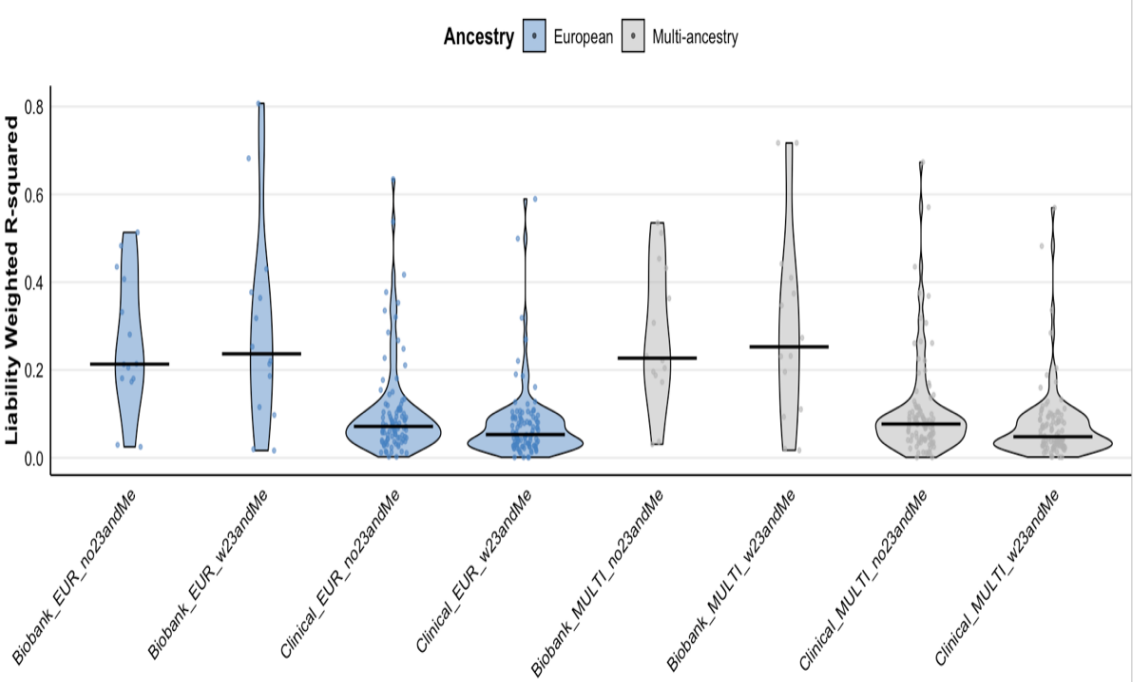
Figure 42 Liability R-squared by PRS across ancestry

The figure displays the distribution of R-squared values (liability weighted) across different PRS (Polygenic Risk Score) groups, stratified by reported ancestry (European and Multi-ancestry). Violin plots illustrate the density of R-squared within each PRS group, while individual data points are overlaid to show the spread of the data. A horizontal black line within each violin represents the median R-squared for that specific PRS group. The accompanying table provides a numerical summary of these results, presenting the median R-squared, the standard error of the median (estimated via bootstrapping, N=10,000 bootstraps), and the 95% bias-corrected and accelerated bootstrap confidence intervals for the median for each PRS group. This representation of the visual comparison of the central tendency and spread of R-squared values across the four different PRS and GWAS-ancestry (blue and grey) categories, is presented in the statistical estimates and their uncertainty in the table.



PRSS Group	Median	SE (Median)	CI Lower (95%)	CI Upper (95%)
BD1_EUR_no23andMe	0.156	0.016	0.128	0.189
BD1_EUR_w23andMe	0.075	0.017	0.062	0.116
BD2_EUR_no23andMe	0.119	0.034	0.079	0.195
BD2_EUR_w23andMe	0.139	0.021	0.072	0.177
BD1_MULTI_no23andMe	0.170	0.026	0.121	0.214
BD1_MULTI_w23andMe	0.075	0.015	0.063	0.116
BD2_MULTI_no23andMe	0.112	0.038	0.077	0.207
BD2_MULTI_w23andMe	0.138	0.022	0.086	0.170

Figure 43 Liability R-squared by PRS across subtypes.
The violin plot visualizes the distribution of liability-weighted R-squared values for each PRSS group (subtype group), with the black lines indicating the bootstrapped median for each group. The table complements this by providing a numerical summary of these distributions. For each PRSS group, the table presents the bootstrapped median, its standard error (a measure of the variability of the median estimate), and the 95% confidence interval. This confidence interval gives a range where the true median value is likely to fall, based on the bootstrap resampling. By comparing these medians and their confidence intervals across the PRSS groups, one can infer the magnitude and statistical significance of differences in predictive power, as measured by R-squared on the liability scale, between the PRSS groups by ascertainment (BD1 or BD2) and GWAS-ancestry (colour blue and grey).



PRSA Group	Median	SE (Median)	CI Lower (95%)	CI Upper (95%)
Biobank_EUR_no23andMe	0.213	0.054	0.177	0.358
Biobank_EUR_w23andMe	0.237	0.066	0.116	0.371
Clinical_EUR_no23andMe	0.072	0.006	0.061	0.084
Clinical_EUR_w23andMe	0.053	0.008	0.040	0.070
Biobank_MULTI_no23andMe	0.227	0.062	0.180	0.395
Biobank_MULTI_w23andMe	0.253	0.067	0.111	0.378
Clinical_MULTI_no23andMe	0.077	0.006	0.064	0.088
Clinical_MULTI_w23andMe	0.048	0.009	0.038	0.069

Figure 44 Liability R-squared by PRS across ascertainment.
The violin plot visualises the distribution of liability-weighted R-squared values for each PRSA group (ascertainment group), with the black lines indicating the bootstrapped median for each group. The table complements this by providing a numerical summary of these distributions. For each PRSA group, the table presents the bootstrapped median, its standard error (a measure of the variability of the median estimate), and the 95% confidence interval. This confidence interval gives a range where the true median value is likely to fall, based on the bootstrap resampling. By comparing these medians and their confidence intervals across the PRSA groups, one can infer the magnitude and statistical significance of differences in predictive power, as measured by R-squared on the liability scale, between the PRSA groups by ascertainment (clinical or biobank/community) and GWAS-ancestry (colour blue and grey)

6.6 Discussion

This study represents the largest PRS analyses of GWAS of BD to date, encompassing a diverse range of ancestries (EUR, EAS, AFR, and LAT). The results corroborate the initial hypothesis that variations in ascertainment and BD subtype are associated with differences in genetic architecture. Subsequent post-GWAS analyses in O'Connell *et al*., provided novel insights into the biological underpinnings and genetic architecture of BD, highlighting further distinctions based on participant ascertainment and BD subtype. Furthermore, it was demonstrated that the inclusion of multi-ancestry data enhanced the polygenic prediction accuracy.

The genetic correlation findings from the latest large-scale BD GWAS (O'Connell, Koromina and van der Veen *et al*., 2025) coupled with these PRS analyses, underscore that the genetic architecture of BD varies across ascertainment methods and subtypes, a phenomenon seemingly driven by the relative representation of each subtype within the sample. In O'Connell *et al*., an analysis of BD subtypes revealed a strong, albeit imperfect (rG = .88, SE = .05), genetic correlation between BD1 and BD2 [21]. Notably, this study observed high genetic correlations between both BD1 (rG = .85, SE = .03) and BD2 (rG = .95, SE = .06) with community-ascertained samples. In contrast, the genetic correlation between BD1 and self-reported BD (rG = .42, SE = .02) was significantly lower ($P = 7.1 \times 10^{-13}$) than that between BD2 and self-reported BD (rG = .76, SE = .05). Furthermore, assuming a 1% population prevalence [22], heritability estimates indicated a higher SNP-based heritability (h^2snp) for BD1 (h^2snp = .21, s.e. = .01) compared to BD2 (h^2snp = .11, SE = .01). Considering the differing proportions of BD1 and BD2 individuals in clinical and community cohorts, the study also examined the genetic correlation between BD in these settings and self-reported BD, conditioning on the genetic risk for BD1 and BD2. Following this adjustment, the genetic correlation between self-reported BD and BD in community cohorts (rG = .92, s.e. = .09) was not significantly different (P = .10) from that observed in clinical cohorts (rG = .71, SE = .13). As expected, schizophrenia was more strongly genetically correlated with the main BD phenotype meta-analysis excluding self-reported data and with BD1 and BD in clinical samples.

This highlights a critical trade-off in psychiatric genetics between sample characteristics and PRS performance. Clinically ascertained, hospital-based samples are often enriched for more severe illness (e.g., more BD1, higher rates of psychosis) and may have a higher underlying genetic burden. While a GWAS of such a sample can yield larger effect sizes, the resulting PRS may have limited generalizability to the wider community. Conversely, biobank and self-report samples offer massive sample sizes but may capture a broader, more heterogeneous, and potentially less severe spectrum of the disorder. Our finding that excluding self-report data improved prediction for BD1 and clinical cohorts, while including it was neutral-to-positive for BD2 and community cohorts, empirically demonstrates this ascertainment-specific genetic architecture. Future PRS development must grapple with this trade-off, perhaps by developing ascertainment-specific PRS or by using methods that can model and account for this heterogeneity.

The observed differences in the genetic architecture of BD subtypes appear to be related to the method of ascertainment. Specifically, BD in clinical and community samples exhibited a strong but imperfect genetic correlation, with varying degrees of correlation with self-reported BD. The lower genetic correlation and limited genetic overlap between clinically ascertained cases and self-reported cases are likely driven by a higher proportion of BD1 within the clinical and community samples (O'Connell *et al*.). Consistent with this, the PRS derived from meta-analyses excluding self-reported data performed better in clinical and BD1 target samples, whereas the inclusion of self-reported data improved PRS performance in community and BD2 target samples. Moreover, the pattern of genetic correlations between BD and other psychiatric disorders shifted with the inclusion of self-reported data, with schizophrenia showing the strongest correlation in the absence of self-reported data, and major depressive disorder (MDD) exhibiting the strongest correlation when self-reported data were included (O'Connell *et al*. [21]).

These findings suggest that self-reported samples may be enriched for individuals with BD2, aligning with recent reports of increasing depression and ADHD PRS and decreasing BD PRS over time in BD2 diagnoses [23]. However, O'Connell *et al*. recognise the potential for overdiagnosis of BD. This is especially a concern in outpatient settings, among individuals with conditions such as, chronic depression or borderline personality disorder, characterised by higher comorbidity rates [24-25] which warrants consideration.

The multi-ancestry PRS yielded the most substantial improvement over the EUR-PRS in two of the three EAS ancestry target cohorts (Korea and Taiwan), with more modest gains observed in EUR target cohorts. The limited improvement in the AFR target cohort may be attributable to the genetic heterogeneity within this population [26]. These results underscore the value of incorporating multi-ancestry representation in PRS training data, consistent with findings in other complex diseases [27]. While the predictive power of the BD PRS developed in this study represents a notable advancement compared to previous efforts [7], it still falls short of the threshold for clinical utility [28].

6.7 Limitations

The study lacked in-sample linkage disequilibrium estimates for all cohorts and relied on a EUR reference panel for multi-ancestry analyses. A EUR LD reference panel was used in the PRS analyses. This approach may not fully capture the LD patterns and interindividual heterogeneity present within the diverse ancestry groups included in the meta-analyses. Also, it was noted that some out-of-sample PRS predictions exceeded the meta-analysed SNP-based heritability statistic, a phenomenon that could potentially indicate inflated or spurious results and warrants careful consideration. The inclusion of samples with minimal phenotyping, while increasing sample size, may have introduced noise and reduced specificity, along with the influence of the effects of between-cohort heterogeneity.

6.7 Conclusions

In conclusion, this large-scale multi-ancestry GWAS of BD has identified differences in its genetic architecture based on both ascertainment and subtype. This suggests that future genetic studies of BD will benefit from stratification by subtype. However, it is crucial to consider the potential impact of between-cohort heterogeneity in PRS analysis. For instance, this heterogeneity can lead to PRS predictions that exceed the meta-analysed SNP-based heritability statistic. Chapter 5 investigations offer a potential explanation for this observation, based on existing research [29]. Analysis supported underlying genetic architecture differences across cohorts, which could be mitigated by using MTAG. This approach acknowledges and accounts for the heterogeneity that can confound standard meta-analysis results, and post-GWAS PRS analyses, as does the random-effects PRS models used in Chapter 5 of this thesis.

6.9 Supplementary Materials

The full data for Supplementary Tables are available in the attached file: Supplementary.tables.xlsx. This file also contains the supplementary data tables referenced in Chapter 5.

The specific contents for Chapter 6 are as follows:

- Supplementary Table 66: Summary of 79 cohorts included in the PGC4 bipolar disorder meta-analyses.
- Supplementary Table 67: Sample size (cases/controls), assessment/ascertainment type, and discovery from ancestry-specific and multi-ancestry meta-analyses.
- Supplementary Table 68: Liability scale SNP-heritability estimates in EUR meta-analyses using LD score regression.
- Supplementary Table 69: Genetic correlation of bipolar disorder with other psychiatric disorders.
- Supplementary Table 70: Multi-Ancestry PRS excluding self-report data in European target samples.
- Supplementary Table 71: Multi-Ancestry PRS excluding self-report data in European BD1 target samples.
- Supplementary Table 72: Multi-Ancestry PRS excluding self-report data in European BD2 target samples.
- Supplementary Table 73: Multi-Ancestry PRS excluding self-report data in European Clinical target samples.
- Supplementary Table 4: Multi-Ancestry PRS excluding self-report data in European Community target samples.
- Supplementary Table 75: Comparison of variance explained by different PRS in European target samples.
- Supplementary Table 76: Multi-ancestry and European PRS in non-European target samples.

The preceding chapters have demonstrated that deconstructing the heterogeneity of bipolar disorder (BD) requires both novel dimensional frameworks (Chapter 2) and large-scale genetic dissection (Chapters 3 & 4), and that the utility of these findings is shaped by critical methodological factors like ascertainment and ancestry (Chapter 5). The collective results converge on a key point: a more robust and nuanced understanding of BD is achievable, but requires moving beyond broad diagnostic categories. The final chapter will now synthesize these findings, critically discuss their broader implications and limitations, and outline future directions for research and clinical practice.

---

7 General Discussion

Findings, Limitations, and Future Directions

The rigorous, multi-stage research strategy employed throughout this thesis has demonstrated its fundamental importance in advancing our understanding of bipolar disorder (BD). By progressively incorporating independent cohorts for validation, thereby substantially increasing sample sizes, and by consistently focusing on elucidating the genetic architectures of more homogeneous BD subphenotypes, this body of work has made a tangible impact on how we approach the study of this complex condition. Specifically, this thesis has enhanced the reliability and generalisability of identified genetic associations, boosted the statistical power crucial for detecting subtle etiological effects, and enabled a more precise dissection of the pathways contributing to BD's diverse clinical presentations. This methodological commitment, central to the research presented, represents a consequential step beyond broad categorisations towards a more granular, biologically informed understanding of the disorder.

Building upon these specific contributions and the insights they have generated, this concluding chapter now aims not only to synthesise the findings of this thesis through a critical appraisal but also to offer insights into how key challenges, particularly those illuminated by the investigations herein, could guide and be addressed by future research.

7.1 Foundational Challenges In Psychiatric Genomics

Genome-wide association studies (GWAS) have limitations, primarily focusing on common variants, which excludes rarer variants. GWAS only account for a portion of the heritability of complex traits, leaving a substantial part unexplained. Additionally, the stringent multiple testing burden associated with including many genetic variants necessitates larger sample sizes to identify smaller effects.

If case and control groups or cohorts are not well-matched, Wang *et al*. (2023) suggest this can lead to different sub-populations being represented and reduce statistical power for association [1]. This recent study proposed that between-cohort heterogeneity could be addressed by using Multi-Trait Analysis of GWAS (MTAG) analysis [2]. There was potential evidence to support this in Chapter 5 which utilized MTAG to derive SNP effect size estimates. This choice was motivated by a key limitation in standard fixed-effect meta-analysis: its assumption of homogeneity across all contributing cohorts. When this assumption is violated, a common occurrence in large-scale genetic studies, standard methods can produce biased SNP effect sizes, which are often underestimated. The analyses in this thesis were particularly susceptible to this issue, as demonstrated by the high heterogeneity statistics observed across all subphenotypes (e.g., $I^2 = 96.4\%$ for BD1).

To address this, MTAG provides a more robust estimation by treating each cohort as a distinct but genetically correlated trait. This model leverages the shared genetic signals across cohorts to improve the precision of the SNP effect estimates, thereby correcting for the bias introduced by heterogeneity. The practical benefit of this improved methodology is evident in the enhanced predictive power of the resulting Polygenic Risk Scores (PRS). The PRS for BD1, for example, achieved an R2-liability of 9.8%, and the score for the unipolar mania subphenotype reached 12.4%. These results exceed the ~4.6% reported for BD by the large-scale [3] study, which used a standard fixed-effects approach, and are highly competitive with the 8.4%-9.0% R2-liability reported by [4], and even exceeding this estimate for certain specific subphenotypes. This may demonstrate that MTAG's superior handling of heterogeneity, as argued by Wang *et al*. (2023) [1], leads to more powerful and predictive genetic scores, alongside the increased power of a multi-trait phenotypic approach.

In contrast, the standard fixed-effects meta-analysis in Chapter 6 does not directly address heterogeneity. MTAG meta-analysis could take clinical cohorts as the focal trait, and community cohorts as genetically correlated traits, which may produce a different pattern of PRS variance. MTAG-PRS could show a more balanced predictive power across the two ascertainment groups. Under the standard model, PRS built solely on a clinical cohort (with potentially inflated effect sizes due to enrichment) might over-predict risk in a community sample. Conversely, a PRS built primarily on a community sample might under-predict risk in a clinical sample where the genetic burden is likely higher. Using MTAG, integrating the clinical cohorts as a focal trait and community data as a closely related trait, could potentially moderate these extreme biases, leading to a more reliable PRS performance. Subsequently, stratification was applied in the PRS-CS analyses in Chapter 6 to address this ascertainment bias.

Chapter 6 sensitivity analyses highlighted the importance of modelling cohort-specific characteristics and potential heterogeneity. The MTAG method was therefore employed in the analysis of the BD subphenotypes in Chapter 5. Treating BD phenotypes as highly correlated traits within an MTAG framework may offer several advantages, including increased power to detect genetic associations, and more precise and potentially less biased SNP estimates which led to better-performing PRS. Furthermore, a sensitivity analysis was also performed that applied a random-effects model in out-of-sample PRS-CS subphenotype predictions, which improved prediction.

Regardless, most BD subphenotype studies remain largely Eurocentric, that inherently limits discovery, prediction and the generalisability of findings. This is particularly relevant given recent evidence that depression risk can differ among people of different ancestry [5-6]. Studying ancestrally diverse populations in GWAS is therefore essential [7]. Tools have been developed to improve disease risk prediction across diverse populations. For example, the PRS-CS method used in this thesis, has more recently been extended to improve polygenic prediction in ancestrally diverse populations (PRS-CSx) [8].

Foundational Limitations of Genome-Wide Association Studies (GWAS)

GWAS have been instrumental in identifying genetic variants associated with bipolar disorder (BD), however they have additional limitations to fully elucidating the disorder's complex genetic basis:

Missing Heritability: GWAS primarily focus on common genetic variants (Single Nucleotide Polymorphisms or SNPs) with a minor allele frequency of at least 1-5% in the population. This approach overlooks the potential contribution of rare variants, which, although individually infrequent, could collectively account for a portion of the "missing heritability", the gap between the estimated heritability of BD from twin studies (60-85%) and the variance explained by identified common variants (around 25%), see Chapter 1. Also, GWAS typically assess the independent effect of each SNP, often overlooking complex interactions between multiple genes (epistasis) or between genes and environmental factors (gene-environment interactions), which are likely crucial in the development of BD.

Small Effect Sizes of Individual Variants: Most genetic variants identified by GWAS for BD have small effect sizes, meaning each variant only contributes a small increase in the risk of developing the disorder. This makes it challenging to translate these findings into clinically useful predictions or diagnostic tools at the individual level. Translating GWAS findings into new therapeutics for psychiatry is an ongoing effort. Detecting these small effects requires very large sample sizes (tens of thousands of cases and controls) to achieve sufficient statistical power. Misclassified cases/controls in GWAS, a problem for all disorders, is especially impactful for BD due to its spectrum nature. For instance, studies by Zimmerman *et al*. 2008 [9] and Zimmerman *et al*. 2010 [10] have explored whether bipolar disorder is over diagnosed or if previous overdiagnoses impact psychiatric classifications. This misclassification, due to diagnostic challenges, subtypes, and comorbidities, reduces statistical power, biases effect sizes, and that hinders replication. Likewise, non-random mating (mating is influenced by existing traits), could further bias heritability estimates. Non-random mating can violate the core assumptions also of Mendelian Randomization (MR) introducing bias into the results.

Applying Mendelian Randomization to BD

Mendelian Randomization (MR) is a valuable epidemiological tool that employs genetic variants as instrumental variables for exposures [11]. By leveraging the random assignment of genetic variants at conception, MR aims to explore causal relationships with outcomes, such as bipolar disorder (BD), while mitigating the confounding and reverse causation that often affect traditional observational studies. However, applying MR to complex diseases such as BD using GWAS data presents several challenges that can test MR's core assumptions.
Large-scale GWAS often necessitate multi-cohort designs, which can introduce sample heterogeneity. This lack of homogeneity may undermine key MR assumptions and prevent the consistent definition of BD phenotypes across studies. The commonly used two-sample MR approach, which utilises separate GWAS datasets for the exposure and outcome, assumes relative homogeneity between these samples. This assumption is a particular concern in BD

research, where outcome data frequently come from clinically ascertained participants, while risk factor (exposure) data may originate from broader population-based cohorts. Whilst the one-sample MR method, using a single cohort, can relax this homogeneity requirement, it carries a higher risk of data overfitting and weak instrument bias [12-13].

MR is therefore also not immune to potential biases. Although the fixed nature of genotypes helps protect against some reverse causality between the phenotypic exposure and outcome, violations of MR assumptions, such as horizontal pleiotropy (where genetic variants affect the outcome through pathways independent of the specific exposure being investigated), can still lead to false associations. The replicability of MR findings in BD research has also been a notable concern. For instance, initial MR associations reported between BD and cardiovascular diseases (CVD) and their subtypes did not persist after meta-analysis [14]. Similarly, MR studies examining the relationship between circulating metabolites and BD risk have shown limited replicability, often due to discrepancies between discovery and replication datasets [15-16].

Evidence of MR Limitations in the Thesis BD Research

The application of MR and sensitivity analyses to the initial GWAS results for eleven BD subphenotype GWAS in the current thesis work (Chapter 5), revealed heterogeneity evidenced by Cochran's Q (QEgger) and horizontal pleiotropy, indicated by the MR-Egger Intercept [18]. While multivariable Mendelian Randomization (MVMR) could potentially address some horizontal pleiotropy, it would not eliminate it entirely [19-20]. Causal inference for heritable phenotypic risk factors using heterogeneous genetic instruments requires careful consideration of these limitations. Horizontal pleiotropy, where the genetic instrument affects the outcome through pathways independent of the exposure, poses a challenge in BD MR studies [21]. A genetic variant might influence specific BD subphenotypes, such as the development or severity of a rapid-cycling course, through mechanisms beyond its impact on the overt circadian disruption. For example, an SNP in a clock gene could affect both an individual's intrinsic circadian rhythm (the intended exposure) and, independently, modulate critical intracellular calcium signalling pathways or glutamatergic neurotransmission, both of which are known to be involved in mood episode recurrence and the underlying pathophysiology of BD. In such cases, the SNP's effect on a rapid-cycling BD subphenotype would not be solely mediated by the 'propensity towards circadian disruption,' leading to a "horizontal" pathway and potentially spurious causal associations.

Environmental factors, such as inconsistent daily routines, major psychosocial stressors, or even seasonal changes in light exposure, also play a substantial role in the course of bipolar disorder and could interact with these pleiotropic pathways, further complicating the interpretation of MR results. Even if a genuine causal effect of circadian disruption on rapid cycling in BD exists, this type of horizontal pleiotropy can bias the estimated magnitude of this effect, either inflating or deflating it.

Summary: In summary, MVMR might not be a complete solution for horizontal pleiotropy in BD research due to the disorder's complexity. BD has a multifaceted aetiology involving numerous interacting genetic and environmental factors. Identifying and accurately measuring all potential "other factors" influenced by a genetic variant associated with a risk factor is incredibly difficult, if not impossible, as all relevant biological pathways are currently unknown. MVMR can only mitigate some biases from weak instruments and horizontal pleiotropy for known confounders. BD likely has many unknown confounders, therefore MVMR may still be biased by residual confounding. Genetic variants identified for many BD risk factors might be weakly associated with those risk factors (weak instruments), which can amplify bias from even small amounts of horizontal pleiotropy, rendering MR results unreliable [22]. Given these challenges, developing methods beyond MVMR is crucial for more reliable causal inference in BD research using MR. Future research should focus on developing better instrument selection strategies to minimize the use of variants with widespread pleiotropic effects on pathways unrelated to BD. This ongoing effort to disentangle the complex web of genetic influences and causal pathways involved in the development and manifestation of bipolar disorder highlights the need for continued methodological advancements in the application of MR to this challenging condition.

7.2 Addressing Confounding And Latent Dimensions In BD

The factor analysis-based Multiple Indicators and Multiple Causes (MIMIC) model (Chapter 3) is more advanced because it's designed to study complex, underlying traits (e.g., anxiety, social or cognitive deficits) that cannot be measure with a single number. It does this by combining multiple factors at once and also accounts for the fact that our measurements are never perfectly accurate, due to unobservable (latent) constructs. Single regression models offer simplicity and ease of interpretation for examining direct relationships between observed variables, but they lack the ability to model latent constructs and account for measurement error in the same way as MIMIC models.

Inverse Probability Weighting (IPW) Limitations

Inverse Probability Weighting (IPW) (Chapter 3), while a powerful tool for addressing potential bias, is inherently limited by unmeasured confounding and potential model mis-specification. IPW using propensity scores is a statistical method employed to mitigate confounding bias in observational studies, including case-control studies investigating BD. This technique aims to create a pseudo-population where the distribution of measured covariates is balanced between the exposure groups (e.g., cases and controls) by weighting individuals based on the inverse of their probability of belonging to their observed group, given their measured characteristics. The importance of considering confounding when assessing bias in observational research cannot be overstated.

However, the effectiveness of IPW with propensity scores in fully adjusting for bias, particularly in the complex landscape of BD research, is subject to crucial assumptions and limitations. A core constraint is IPW's complete dependence on observed covariates. Although

the propensity scores constructed in Chapter 3 were based on factors that could be identified and measured within the statistical models, the fundamental challenge persists: any confounders that are unmeasured, poorly measured, or entirely unknown will not be accounted for by this method. Because the full spectrum of factors that could potentially confound an observed association is unknown (and may never be perfectly captured by measured variables alone), IPW despite its utility in balancing observed covariates, cannot entirely eliminate the risk of bias stemming from these unobserved influences. Therefore, a degree of caution must always be applied when interpreting results adjusted using IPW, as the potential for residual confounding from unknown or unmeasured factors remains.

7.3 PRS For BD: Strengths, Caveats And Heterogeneity

The SCZ3-PRS study in Chapter 4, among the first to investigate this specific Polygenic Risk Score (PRS) for bipolar disorder I (BD1), highlighted several practical limitations and findings. Heterogeneity in BD1 severity within the combined Romanian and UK study sample (e.g., due to varying proportions of hospitalized, more severe cases) and incomplete phenotype information for some participants potentially influenced the results and complicated the interpretation of PRS effects. This is a known challenge, particularly when a PRS is derived from a disorder such as schizophrenia (SCZ) which may have distinct severity profiles from BD [22]. While the SCZ3-PRS demonstrated modest clinical value for some BD1 phenotypic traits, providing an incremental predictive improvement when combined with clinical variables in machine learning models, its utility as a standalone predictor was limited. This underscores that while transdiagnostic PRS can be informative, their predictive power for specific subphenotypes of another disorder may be constrained by partially distinct genetic architectures. This point is relevant to findings of both shared and distinct genetic factors across major psychiatric disorders (Chapter 5).

Generally, BD exhibits a strong polygenic component, and PRS serve to quantify an individual's aggregate genetic liability from numerous small-effect variants identified through GWAS. PRS are valuable for research stratification (as employed in Chapter 3 to 6) by integrating with traditional risk factors for a more comprehensive risk assessment. However, it is crucial to recognise that PRS are correlational, do not imply causation, and are not absolute predictors as they omit many other developmental and environmental influences. Nevertheless, PRS can help mitigate some GWAS limitations by aggregating many small effect sizes, including from variants that do not meet stringent GWAS significance thresholds, thereby contributing to addressing some of the 'missing heritability.' They also offer broader applications in risk prediction, such as for offspring of affected individuals (though current standalone predictive power for BD is modest), and in exploring potential cross-disorder genetic contributions. However, caution is warranted when interpreting PRS derived from GWAS with limited statistical power or those based on weak genetic instruments.

Utilizing PRS within a Mendelian Randomization (MR) framework for BD subgroups, for instance, to assess features such as psychosis or age of onset in BD1, also presents distinct challenges. PRS for subgroups often derive from smaller effective sample sizes and lower

heritability estimates than PRS for the primary BD phenotype. As PRS typically explain only a fraction of an exposure's variance, their use as instruments can lead to weak instrument bias in MR, potentially flawing causal effect assessments, especially given the existing 'missing heritability.' For example, a sensitivity analysis in Chapter 5 using the SlopeHunter method to adjust for potential collider bias highlighted potential concerns for using MR, as results showed evidence of over-correction. This was likely attributable to weak instruments, specifically a limited number of 'index-specific' SNPs for the BD subphenotype, possibly reflecting high genetic correlations between the main BD phenotype and its subphenotypes, alongside a lack of clear temporal separation from the overarching BD phenotype.

Ultimately, the development of more powerful and reliable PRS for BD relies on large-scale GWAS to obtain robust effect size estimates and capture a greater spectrum of genetic variants. While pooling data from multiple cohorts can substantially boost statistical power in GWAS, it also introduces the considerable challenge of between-cohort heterogeneity, which may arise from differences in diagnostic criteria, sample ascertainment, and environmental factors. Such heterogeneity can introduce noise, potentially obscuring true genetic associations and leading to less generalizable PRS. Therefore, researchers face a balancing act: maximizing statistical power through large sample sizes while meticulously addressing and mitigating the impact of between-cohort heterogeneity. This thesis has attempted to address such heterogeneity in Chapters 3 through 6 using methods including a subphenotypic approach, mixed regression modelling, random-effects meta-analyses of MTAG PRS results, and stratification by BD ascertainment and subtypes.

7.4 Subphenotyping BD: Limits Of Genetic Stratification

This section explores the advantages and disadvantages of focusing research on specific subphenotypes of bipolar disorder (BD) compared to studying the broader BD phenotype.

Despite limitations, the thesis work benefited from a subphenotypic approach which identified novel genetic insights beyond those found when analysing the main BD phenotype. This revealed interrelated and subphenotype-specific mechanisms within BD clinical characteristics, as well as shared genetic architecture with other psychiatric and somatic disorders, evidenced by concordant pleiotropic effects. For instance, multivariate GWAS have begun to reveal underlying dimensional genetic liabilities across psychiatric disorders [29]. Similar subphenotype-specific genetic signatures aligned in Chapter 5 with this in the pre-defined four-factor model, potentially informing the future development of BD nosology and treatments. BD is a highly heterogeneous condition. Studying more homogeneous subgroups (subphenotypes) reduced noise. Focusing on more specific subgroups allowed for the detection of risk factors with greater statistical power compared to analysing the entire heterogeneous BD group.

Different subphenotypes exhibited distinct underlying biological mechanisms, allowing for a more precise understanding of the BD's aetiology. Identifying biomarkers associated with specific subphenotypes could lead to more accurate predictions of illness course, treatment

response, and comorbidity risk within those subgroups. Studying BD subphenotypes helped illuminate some genetic and clinical overlap between BD and other psychiatric conditions. This could help refine classification of bipolar disorders in the future.

The Chapter 5 analysis focused exclusively on individuals of European ancestry due to the lack of phenotyping in non-European ancestries. This limits the applicability of these findings to other populations. Heterogeneity between cohorts likely also limited discovery. While multivariate analyses increased the effective sample size, they reduced the analysis to intersected variants, potentially masking unique genetic loci associated with individual traits. Furthermore, clinical misdiagnoses and cross-trait assortative mating could have introduced biases requiring further investigation. Future research should incorporate formal fine-mapping beyond the TWAS conditional analyses to pinpoint causal variants in larger multi-ancestry analyses.

Limitations of a Subphenotypic Approach: There are also general limitations of studying BD subphenotypes. There is a lack of a universally agreed-upon and biologically validated system for defining BD subphenotypes remains a pivotal challenge. This initiative encourages a more dimensional, mechanistic, and integrative approach to mental health research, with the long-term goal of improving diagnosis and treatment. Dividing the overall BD sample into smaller subphenotype groups reduces statistical power, potentially hindering the detection of genetic associations, especially in genetic studies. An attempt was made to mitigate this by increasing power to subphenotype-specific SNPs in the MTAG analyses. However, dividing BD into too many subgroups based on superficial differences might not reflect underlying biological distinctions and could lead to non-replicable findings. Individuals may exhibit characteristics of multiple subphenotypes, making clear categorisation difficult. Therefore, future efforts will require more detailed, accurate subphenotype classification which requires comprehensive and standardised clinical assessments, which can be resource intensive.

In summary, studying bipolar disorder subphenotypes offers promise for dissecting the disorder's heterogeneity and advancing the understanding of its specific biological and clinical features. However, as highlighted by the limitations encountered in in the current research (European ancestry bias, cohort heterogeneity, focus on on intersected variants), and the general challenges associated with subphenotype research (definition, sample size, statistical power), careful consideration of both the benefits and limitations will be crucial for designing and interpreting research in this complex area.

7.5 Advancements And Path Forward In BD Genomics

The polygenic risk score (PRS) for bipolar disorder (BD) developed in this study represents a notable advancement compared to previous efforts [3], yet its predictive power still falls short of the threshold required for clinical utility.

Additional methodological short-comings are address below:

Multi-ancestry BD Polygenic Risk Score (PRS) prediction: A critical challenge in applying BD PRS is achieving accurate and generalisable predictions across diverse ancestries. Current PRS development is often hampered by the predominant reliance on European (EUR) linkage disequilibrium (LD) reference panels. This approach may not adequately capture the distinct LD patterns and interindividual genetic heterogeneity within various ancestral groups included in large-scale meta-analyses, thereby limiting the portability and utility of PRS in non-EUR populations. Recognising these disparities, initiatives including the PRIMED Consortium are actively working to reduce them in polygenic risk assessment [30].

Increased GWAS Sample Sizes: While increasing the sample size of GWAS for BD is crucial for improving the power to detect more associated genetic variants, it is unlikely that this factor alone will be sufficient to bring PRS into routine clinical use. However, larger GWAS are expected to identify more genetic variants associated with BD, including those with smaller effect sizes. Incorporating these into PRS should lead to a modest but important increase in the variance explained and thus, potentially better risk stratification at the population level. Larger samples will likely yield more accurate estimates of the effect sizes of individual variants, which are used as weights in PRS calculation. This could improve the reliability and predictive power of the scores. With very large sample sizes, GWAS might be able to reliably identify common variants with lower minor allele frequencies that still contribute to BD risk. Including these could further enhance PRS accuracy.

Future Challenges: PRS based solely on common variants are unlikely to capture the portion of heritability attributed to rare variants, copy number variations (CNVs), and complex gene-gene and gene-environment interactions. These factors require different research approaches (e.g., WES/WGS). Even with more variants identified, the individual variant effect sizes for BD are likely to remain small. This inherent polygenic architecture of the disorder means that even a PRS incorporating thousands or millions of variants might only explain a limited amount of the overall risk. The improvement in predictive accuracy with increasing sample size may eventually plateau.

The current predictive accuracy of BD PRS is far below the threshold generally considered necessary for routine clinical decision-making (e.g., for diagnosis or guiding treatment in individuals). While the largest to date BD GWAS (Chapter 6) will improve prediction, it remains uncertain whether further increases in sample size will be sufficient to reach this threshold. If a PRS for bipolar disorder were to achieve AUCs (Area Under the Curve) of .90-.95 in well-powered, across independent validation studies, it could have clinical utility, particularly in already identified high-risk populations [31].

PRS however will only account for genetic predisposition and do not incorporate the role of environmental factors in BD development and course. Clinical risk prediction will likely require integrating PRS with environmental and clinical risk factors for meaningful individual-level assessment. This will need to be integrated further with familial risk, clinical features,

and information from other types of genetic variation (rare variants, CNVs), focusing on specific, well-defined subphenotypes of BD, for establishing clinical utility and actionable strategies based on PRS results. Only through such a multifactorial approach coupled with larger and more diverse WGS/WES/GWAS, advancements in methodology and translational research, might there be the potential of genetic risk scores to be realised in the clinical management of BD. Given that only a fraction of CNV carriers develop psychiatric disorders [31], it will be vital to determine how CNVs and PRS jointly contribute to risk. While SCZ studies indicate an additive effect for total risk [32], interactive effects between specific CNVs and PRS require further research for combining these relative risk weights together for prediction.

7.6 Research Challenges And Promising PRS Developments

Addressing and Leveraging Bipolar Disorder Heterogeneity

A crucial future direction, consistently highlighted throughout this thesis, lies in the continued exploration of inter- and intra-individual heterogeneity in BD. Moving away from analysing broad, heterogeneous patient groups towards dissecting more homogeneous subgroups is essential for clarifying the complex genetic and clinical landscape of the disorder. The approach taken in this thesis aimed to identify more specific subgroups, which in turn revealed shared and differential underlying biological mechanisms potentially informative for treatment response. This will require deepening phenotypic characterisation for subgroup refinement. To power future dissections of this heterogeneity, even deeper phenotyping will be required. This includes the comprehensive and detailed assessment of family history, longitudinal clinical characteristics, treatment history, comorbidities, cognitive function, personality traits, and environmental exposures, which could allow for the identification of patterns of co-occurring features that define previously unrecognised BD subgroups.

Employing Data-Driven Strategies and Navigating Methodological Considerations

Alongside deeper phenotyping, advanced data-driven approaches will be vital. Statistical techniques like cluster analysis, applied to rich phenotypic data, can help identify natural groupings of individuals with similar clinical profiles, thereby defining more homogeneous phenotypic subgroups. Simultaneously, methods that group individuals and genetic variants based on shared patterns of association can identify genetically homogeneous subgroups without relying on pre-defined phenotypic categories. While misdiagnosis can introduce noise into case and control groups, the study's approach of using genetic data to help define subgroups, rather than relying solely on potentially heterogeneous phenotypic diagnoses, can mitigate some of this impact. Furthermore, phenomena including assortative mating, where individuals with similar traits preferentially partner, could indirectly affect the genetics of BD by concentrating risk genes within families, potentially increasing the genetic heterogeneity observed among BD patients.

Recent methodological advancements offer promise for refining BD PRS. For instance, a novel approach using genetic data to identify genetically homogeneous subgroups (biclusters) within BD subphenotypes, without relying on pre-defined clinical categories, demonstrated improved polygenic risk prediction for BD1 using only a small subset of bicluster-specific SNPs. This suggests that focusing on such genetically defined subgroups might enhance the replication of associated SNPs in future studies [33]. The gene-set enrichment analysis of this identified genetic subgroup also revealed an over-representation of pathways related to neuronal development and maintenance, aligning with the subtype and subphenotype differential gene-set enrichment analyses conducted in this thesis (Chapter 5) and the individual-level pathway (PRSet-PRS) discrimination of psychosis in BD1 (Chapter 4). These methods underscore a shift towards prioritising SNPs more likely to have a functional impact, rather than treating all SNPs equally.

Building on the importance of biological context, annotated genes and gene sets are crucial for understanding the pathways and functions potentially disrupted in BD, which may manifest as distinct endophenotypes. By studying these, researchers can pinpoint specific biological pathways for investigation. Gene-set specific PRS could then be constructed to explore associations with these particular endophenotypes. If a disease's genetic architecture is enriched within certain biological pathways, a targeted gene-set specific PRS might capture a stronger, more relevant signal compared to a genome-wide PRS that includes many variants unrelated to that specific pathway. For complex diseases with diverse genetic underpinnings such as BD, different gene-set specific PRS could be particularly relevant for the distinct subgroups of individuals identified in Chapter 5.

Computational and Genomic Advancements in PRS

Further enhancing PRS utility, methods leveraging predicted epigenomic features from whole-genome sequencing data and advanced computational techniques including machine learning and deep learning are emerging. As identified in Chapter 4, machine learning can capture complex non-linear relationships, indicating potential for improved PRS prediction through sophisticated feature selection and weighting by deep learning methods, such as Deep Convolutional Neural Networks (DCNN) [34] and Deep Ensemble Encoder Networks (DEEN) [35].

Beyond common variants, efforts are underway to incorporate a broader spectrum of genetic variation into PRS. While increasing GWAS sample size is crucial for better-powered common variant PRS, including rare SNVs and structural variations (including CNVs), often captured by WGS rather than SNP arrays holds promise for increasing predictability. These variants, though individually rare, can have larger effect sizes and contribute substantially to BD risk. Consequently, developing methods to effectively weight and combine rare variants with common variants in a PRS is an active research area. Models such as RICE [36], which integrates common and rare variants using ensemble learning and burden scores, and EPRS [37], focusing on prioritizing gene clusters with specific rare variants for risk stratification,

exemplify this trend. Additionally, gene-based burden scores (GBS) can identify rare variant associations for incorporation into PRS models [38].

7.7 Relating Subphenotypes And Endophenotypes

Bridging Subphenotypes with Endophenotypes through Genetic Insights and Advanced Methodologies

From Clinical Subphenotypes to Biological Endophenotypes: The Rationale: A primary goal of the Subphenotyping efforts in this thesis is to bridge clinically defined subgroups of BD with more biologically grounded endophenotypes, thereby advancing clinical translation. While the subphenotypes explored herein (e.g., characterized by predominant polarity, illness course, specific symptom dimensions, or age of onset) refine the clinical picture, endophenotypes represent heritable, measurable traits (including cognitive deficits, affective temperaments, or neurobiological markers) considered closer to BD's underlying genetic vulnerabilities. Identifying such endophenotypes helps delineate more genetically homogeneous groups sharing specific biological susceptibilities.

Polygenic Risk Scores: A Bridge to Endophenotype Discovery: PRS, particularly gene-set specific PRS, offer a crucial link between genetic risk and these endophenotypes. For example, genetic variants influencing early brain development, assessed via a gene-set PRS, have been associated with neurophysiological endophenotypes such as reduced P300 amplitude in psychosis [39]. The annotated genes, gene sets, and PRS analyses from this thesis (Chapters 4 and 5) provide a critical foundation for future work aimed at identifying and validating such endophenotypes in BD, including specific biological components, such as *HLA-DMA* and the complement component 4A [40].

Integrating Multi-Omics Data for Deeper Biological Insights: Recent methodological developments further empower this transition from clinical subphenotypes to biologically anchored endophenotypes. Multi-omics integration, using tools such as Weighted Gene Co-expression Network Analysis (WGCNA;[41]) to find dysregulated gene networks from transcriptomic data, and Multi-Omics Factor Analysis (MOFA; [42]) to uncover comprehensive molecular signatures from diverse data types (e.g., genomics, transcriptomics, epigenomics), could substantially aid in patient stratification and understanding BD's heterogeneity by identifying how different biological systems interact.

Tissue-Specific Approaches and "Biotypes": Refining Biological Subgroups: Further refining this quest for biologically meaningful subgroups, methods such as CASTom-iGEx (Context-Aware Stratification based on Tissue-specific imputed Gene Expression; [43]) leverage tissue-specific gene expression data to define "biotypes." These biotypes, representing a convergence of genetic risk onto specific biological pathways, can form more homogeneous groups at a molecular level than traditional PRS groupings and may correlate strongly with

distinct clinical features and endophenotype profiles, as shown in other complex disorders [43]. Although the application of these sophisticated multi-omics and biotyping methods in BD is currently limited by the need for well-characterized cohorts with comprehensive data and advanced analytical expertise, they may offer powerful pathways forward.

Synthesizing Approaches for a Biologically-Informed Future in BD: By integrating the refined Subphenotyping approaches developed in this thesis with PRS strategies focused on biological pathways and these advanced multi-omics and biotyping techniques, future research may more effectively connect observable clinical heterogeneity in BD to robust, underlying endophenotypes. This convergence is essential for developing a biologically informed classification system and ultimately, more targeted and effective treatments for bipolar disorder.

7.8 Advances Towards Personalised Bipolar Disorders Treatment

The comprehensive investigations undertaken in this thesis, from delineating novel dimensional models (Chapter 3) and subphenotype-specific genetic architectures (Chapter 5) to refining polygenic risk prediction (Chapters 3 and 6), could lay critical groundwork for advancing personalized approaches in BD. This section outlines how these specific contributions may eventually be leveraged to improve clinical translation.

1. Leveraging Thesis-Defined Dimensions and Subphenotypes for Enhanced Clinical Assessment: A core advance of this thesis is the identification of more homogeneous patient subgroups based on distinct clinical and genetic profiles, such as the novel 'chronic functioning' dimension (Chapter 3) and the four genetically-informed dimensions (Chapter 5).

Future work could focus on translating these refined classifications into clinically applicable tools:

- **Holistic Patient Profiling**: The imperative, as highlighted by the thesis, is to move beyond current diagnostic systems that do not fully capture individual heterogeneity. The dimensional and subphenotype frameworks developed herein can inform more comprehensive assessments that integrate mood polarity with cognitive function, personality facets, and functional impairments. This holistic view is essential for understanding the individual's experience and predicting their illness course, particularly given findings such as the association of the *FOXO6* gene (Chapter 5) with BD trajectory and its potential longitudinal role in hippocampal function and memory [44].
- **Early Identification within At-Risk Populations**: The transdiagnostic genetic links explored (Chapters 3 to 6), particularly the shared genetic liabilities between BD, SCZ, MDD, and ADHD, can refine early identification strategies. For instance, research has

elucidated genetic overlap between BD, MDD, and SCZ, even extending to borderline personality disorder [45]. Broader efforts are charting the landscape of genetic overlap across mental disorders and related traits such as MOOD [46], and identifying specific loci that highlight shared risk with both mental and somatic disorders [47]. The thesis's insights into how ADHD or anxiety genetic risk contributes to specific BD presentations (Chapter 3) could guide vigilance for youth exhibiting such early comorbid symptoms, potentially signalling higher BD risk (especially for a chronic, more complex BD trajectory).

2. Advancing Predictive Tools for Personalised Risk Stratification and Treatment Planning: This thesis has explored both the utility and limitations of PRS (Chapters 3 to 6). The path to personalized treatment involves building on these insights:

- **Refined PRS for Specific Subgroups**: The finding that PRS performance is influenced by ascertainment, ancestry, and subphenotype definition (Chapter 6) underscores the need to develop and validate PRS tailored to more homogeneous patient groups, such as those identified in Chapter 4. Future efforts should aim to integrate common and rare variants with a subphenotypic approach, potentially using advanced methods such as RICE [36] or EPRS [37]), to enhance predictive power for the four genetically-informed specific thesis-derived dimensions (Chapter 5).
- **Integrating PRS with Clinical and Biological Data**: As demonstrated by the improved prediction when SCZ3-PRS was combined with clinical variables (Chapter 3), future predictive models should aim to integrate genetic markers such as the thesis-derived ones, with deep phenotypic data (longitudinal course, comorbidities, cognitive profiles, environmental exposures) and other biomarkers. This multi-modal approach, potentially leveraging machine learning (Chapter 4), is key to moving PRS towards clinical utility for predicting modifiable outcomes such as treatment response and suicidality.

3. Targeting Novel Biological Pathways for Tailored Therapeutic Interventions: The gene and gene set pathway analyses conducted (Chapters 4 and 5), which implicated specific biological mechanisms, genes such as *FOXO6* in thesis-defined subphenotypes, offer avenues for novel therapeutic development.

- **Subphenotype-Specific Drug Discovery**: Identifying distinct biological pathways for different subphenotypes (Chapter 4-5) could lead to interventions targeted at the specific underlying biology of a patient subgroup, rather than a one-size-fits-all approach. For example, disruptions in pathways involving *FOXO6* (regulated via the PI3K/PKB pathway; [44]) or other hub genes identified in specific BD subphenotypes could become future novel targets for functional studies.
- **Biologically-Informed Repurposing**: Shared gene set pathways between BD and comorbid conditions (e.g., ADHD, anxiety) or related disorders (SCZ, MDD), as explored throughout the thesis, might also allow for the informed repurposing of

existing treatments for specific BD subgroups characterized by these shared genetic signatures.

4. Overcoming Translational Hurdles: The Path Forward: Translating these research advances into routine clinical practice requires continued effort:

- **Developing Robust, Generalisable Models**: This necessitates new, deeply phenotyped, multi-ancestry cohorts to validate and extend the subphenotype and PRS findings of this thesis. Advanced data-driven approaches, such as the multi-omics integration techniques (WGCNA [41]), MOFA [42])) and biotyping methods (CASTom-iGEx [43]) discussed previously, will be vital for identifying robust, biologically-grounded patient segments.
- **Bridging the Gap to Clinical Practice**: The development of practice guidelines for utilizing complex, multi-modal data (including genomics) in patient assessment and management will be essential, alongside improved genetics training for clinicians.

By building upon the specific dimensional and Subphenotyping frameworks established in this thesis and by rigorously addressing methodological challenges, future research could pave the way for more precise diagnostics, targeted interventions, and ultimately, more personalized and effective care for individuals with bipolar disorder.

THESIS CONCLUSION

Genomics Insights into Bipolar Disorder Architectures - Pleiotropic Genes and Polygenic Burdens: A Step Towards Personalized Treatment

In conclusion, this thesis has systematically dissected the intricate genetic landscape of bipolar disorder, making an impactful contribution by demonstrating how a detailed understanding of pleiotropic genes and polygenic burdens can reframe our approach to this complex condition. By moving beyond traditional diagnostic categories to explore its dimensional nature (Chapter 3) and, critically, by dissecting its heterogeneity across clinically defined subphenotypes (Chapter 5), this work has not only underscored the limitations of a monolithic view of bipolar disorder but has also yielded tangible resources for the wider research community. Notably, based on the subphenotype-specific genetic architectures elucidated in Chapter 5, summary statistics for each of the subphenotypes will be released (including those incorporating SCZ3-SNPs effects), offering a novel and valuable foundation for future investigations into these more homogeneous BD groups. The examination of transdiagnostic polygenic risk scores (Chapter 4) further emphasized the pervasive pleiotropy influencing BD by highlighting shared genetic underpinnings with other psychiatric conditions. Moreover, the comprehensive multi-ancestry analysis of Polygenic Risk Score (PRS) performance (Chapter 6) has illuminated how methodological rigor in accounting for population structure and ascertainment is crucial when navigating the complexities of polygenic traits.

Collectively, these in-depth explorations into polygenicity and pleiotropy have provided a more nuanced understanding of BD's aetiology, directly enhancing our capacity to appreciate distinct genetic contributions to its varied clinical presentations. The identification of specific genetic loci and pathways associated with different subphenotypes (Chapter 5), informed by their unique polygenic and pleiotropic profiles, offers clearer and more promising avenues for the development of targeted biomarkers and interventions. Furthermore, the insights gained into the impact of ascertainment bias and ancestral diversity on PRS accuracy (Chapter 6) reinforce the impact of this thesis in advocating for robust, personalized approaches to risk assessment that respect the complex interplay of an individual's genetic background and clinical manifestation.

While the journey towards personalized treatment for bipolar disorder is ongoing, this thesis represents a pivotal step forward, driven by its detailed characterization of how pleiotropy and polygenicity manifest across diverse BD presentations and methodological contexts. By elucidating these complex genetic interplays, providing actionable data through resources such as the forthcoming summary statistics, and addressing key methodological challenges, this work lays a more robust foundation for future research aimed at translating genomic insights into clinically meaningful tools. Ultimately, the deeper, more stratified understanding of bipolar disorder's genetic architecture achieved and promoted in this thesis holds substantial promise for enabling more precise diagnostic strategies and fostering the development of

personalized treatment approaches that may genuinely improve outcomes for individuals living with this challenging condition.

8 References

8.1 CHAPTER 1

1. McGuffin P, Rijsdijk F, Andrew M, Sham P, Katz R, Cardno A. The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Arch Gen Psychiatry*. 2003;60(5):497-502.
2. Serafini G, Pardini M, Monacelli F, Orso B, Girtler N, Brugnolo A, et al. Neuroprogression as an Illness Trajectory in Bipolar Disorder: A Selective Review of the Current Literature. *Brain Sciences*. 2021;11(2):276.
3. Hibar DP, Westlye LT, Doan NT, Jahanshad N, Cheung JW, Ching CRK, et al. Cortical abnormalities in bipolar disorder: an MRI analysis of 6503 individuals from the ENIGMA Bipolar Disorder Working Group. *Molecular Psychiatry*. 2018;23(4):932-42.
4. Rødevand L, Bahrami S, Frei O, Chu Y, Shadrin A, O'Connell KS, et al. Extensive bidirectional genetic overlap between bipolar disorder and cardiovascular disease phenotypes. *Translational Psychiatry*. 2021;11(1):563.
5. Smeland OB, Bahrami S, Frei O, Shadrin A, O'Connell K, Savage J, et al. Genome-wide analysis reveals extensive genetic overlap between schizophrenia, bipolar disorder, and intelligence. *Molecular Psychiatry*. 2020;25(4):844-53.
6. Léda-Rêgo G, Studart-Bottó P, Abbade P, Rabelo-Da-Ponte FD, Casqueiro JS, Sarmento S, et al. Lifetime prevalence of psychiatric comorbidities in patients with bipolar disorder: A systematic review and meta-analysis. *Psychiatry research*. 2024;337:115949.
7. Oliva V, Fico G, De Prisco M, Gonda X, Rosa AR, Vieta E. Bipolar disorders: an update on critical aspects. *The Lancet Regional Health - Europe*. 2025;48:101037.
8. Spoorthy MS, Chakrabarti S, Grover S. Comorbidity of bipolar and anxiety disorders: An overview of trends in research. *World Journal of Psychiatry*. 2019;9(1):7-29.
9. Lu W, Dimick MK, Fiksenbaum LM, Timmins V, Mitchell RHB, Zou Y, et al. Clinical correlates of socioeconomic status in adolescent bipolar disorder. *Comprehensive Psychiatry*. 2020;101:152186.
10. Vieta E, Salagre E, Grande I, Carvalho AF, Fernandes BS, Berk M, et al. Early Intervention in Bipolar Disorder. *The American journal of psychiatry*. 2018;175(5):411-26.
11. World Health Organization. *International classification of diseases and related health problems, 10th revision*. Geneva; 1992.
12. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders, (DSM-5®)*. 5th ed. Washington, DC; 2013.
13. Angst J, Marneros A. Bipolarity from ancient to modern times: conception, birth and rebirth. *Journal of affective disorders*. 2001;67(1-3):3-19.
14. Kraepelin E. *Clinical Psychiatry: A Text-Book for Students and Physicians*. Translated and edited by A. R. Diefendorf. Macmillan; 1902. (6th edn).
15. van der Markt A, Klumpers UMH, Draisma S, Dols A, Nolen WA, Post RM, et al. Testing a clinical staging model for bipolar disorder using longitudinal life chart data. *Bipolar Disorders*. 2019;21(3):228-34.
16. Lake CR, Hurwitz N. Schizoaffective disorders are psychotic mood disorders; thereare no schizoaffective disorders. *Psychiatry Research*. 2006;143(2-3):255-87.
17. Zivanovic O, Nedic A. Kraepelin's concept of manic-depressive insanity: One hundred years later. Journal of Affective Disorders. 2012;137(1-3):15-24.
18. Mendel E. *Die Manie*. Urban und Schwarzenberg; 1881.

19. Dunner DL, Fleiss JL, Fieve RR. The course of development of mania in patients with recurrent depression. *American Journal of Psychiatry*. 1976;133(8):905-8.
20. Griesinger W. *Die Pathologie und Therapie der psychischen Krankheiten fur Arzte und Studierende*. Krabbe; 1845.
21. Dunner DL, Fieve RR. Clinical Factors in Lithium Carbonate Prophylaxis Failure. *Archives of General Psychiatry*. 1974;30(2):229-33.
22. Kramlinger KG, Post RM. Ultra-rapid and ultradian cycling in bipolar affective illness. *The British journal of psychiatry*. 1996;168(3):314-23.
23. Geller B, Williams M, Zimerman B, Frazier J, Beringer L, Warner KL. Prepubertal and early adolescent bipolarity differentiate from ADHD by manic symptoms, grandiose delusions, ultra-rapid or ultradian cycling. *Journal of affective disorders*. 1998;51(2):81-91.
24. Chapman J, Jamil RT, Fleisher C, Torrico TJ. Borderline Personality Disorder. In: *Cultural Sociology of Mental Illness: An A-to-Z Guide*. SAGE Publications; 2024.
25. Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Medicine*. 2013;11(1):126.
26. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (5th ed., text rev.)*. 2022.
27. World Health Organization. *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*. Geneva; 1992.
28. Reed GM, First MB, Kogan CS, Hyman SE, Gureje O, Gaebel W, et al. Innovations and changes in the ICD-11 classification of mental, behavioural and neurodevelopmental disorders. *World Psychiatry*. 2019;18(1):3-19.
29. Chakrabarti S. Bipolar disorder in the International Classification of Diseases-Eleventh version: A review of the changes, their basis, and usefulness. *World Journal of Psychiatry*. 2022;12(12):1335-47.
30. Craddock N, Sklar P. Genetics of bipolar disorder. *Lancet*. 2013;381(9878):1654-62.
31. Schöttle D, Schimmelmann BG, Conus P, Cotton SM, Michel C, McGorry PD, et al. Differentiating schizoaffective and bipolar I disorder in first-episode psychotic mania. *Schizophrenia Research*. 2012;140(1-3):31-6.
32. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC; 2000.
33. Geoffroy PA, Etain B, Scott J, Henry C, Jamain S, Leboyer M, et al. Reconsideration of bipolar disorder as a developmental disorder: Importance of the time of onset. *Journal of Physiology-Paris*. 2013;107(4):278-85.
34. Grande I, Berk M, Birmaher B, Vieta E. Bipolar disorder. *Lancet*. 2016;387(10027):1561-72.
35. McIntyre RS, Berk M, Brietzke E, Goldstein BI, López-Jaramillo C, Kessing LV, et al. Bipolar disorders. *Lancet*. 2020;396(10265):1841-56.
36. Lin JC, Chan A, Zaman H, Sundram F. Bipolar Disorder. In: *Encyclopedia of Pharmacy Practice and Clinical Pharmacy*. Elsevier; 2023. p. 655-71.
37. Lee S, McAfee JC, Lee J, Gomez A, Ledford AT, Clarke D, et al. Massively parallel reporter assay investigates shared genetic variants of eight psychiatric disorders. *Cell*. 2025;0(0).
38. Hirschfeld RMA, Lewis L, Vornik LA. Perceptions and impact of bipolar disorder: how far have we really come? Results of the national depressive and manic-depressive association 2000 survey of individuals with bipolar disorder. *The Journal of Clinical Psychiatry*. 2003;64(2):161-74.
39. Singh T, Rajput M. Misdiagnosis of bipolar disorder. *Psychiatry (Edgmont)*. 2006;3(10):57-63.

40. Lish JD, Dime-Meenan S, Whybrow PC, Price RA, Hirschfeld RMA. The National Depressive and Manic-depressive Association (DMDA) survey of bipolar members. *Journal of Affective Disorders*. 1994;31(4):281-94.
41. Perugi G, Micheli C, Akiskal HS, Madaro D, Socci C, Quilici C, et al. Polarity of the first episode, clinical characteristics, and course of manic depressive illness: a systematic retrospective investigation of 320 bipolar I patients. *Comprehensive psychiatry*. 2000;41(1):13-8.
42. Dunner DL, Tay LK. Diagnostic reliability of the history of hypomania in bipolar II patients and patients with major depression. *Comprehensive psychiatry*. 1993;34(5):303-7.
43. Kassem L, Lopez V, Hedeker D, Steele J, Zandi P, McMahon FJ. Familiality of polarity at illness onset in bipolar affective disorder. *American Journal of Psychiatry*. 2006;163(10):1754-9.
44. Kessing LV, Hansen HV, Christensen EM, Dam H, Gluud C, Wetterslev J. Do young adults with bipolar disorder benefit from early intervention? *Journal of Affective Disorders*. 2014;152-154:403-8.
45. Kalman JL, Loohuis LMO, Vreeker A, McQuillin A, Stahl EA, Ruderfer D, et al. Characterisation of age and polarity at onset in bipolar disorder. *The British Journal of Psychiatry*. 2021;219(6):659-69.
46. Mann-Wrobel MC, Carreno JT, Dickinson D. Meta-analysis of neuropsychological functioning in euthymic bipolar disorder: an update and investigation of moderator variables. *Bipolar Disorders*. 2011;13(4):334-42.
47. Carvalho AF, Quevedo J, McIntyre RS, Soeiro-De-Souza MG, Fountoulakis KN, Berk M, et al. Treatment implications of predominant polarity and the polarity index: a comprehensive review. *The international journal of neuropsychopharmacology*. 2014;18(2):pyu015.
48. Culpepper L. The Diagnosis and Treatment of Bipolar Disorder: Decision-Making in Primary Care. *The Primary Care Companion for CNS Disorders*. 2014;16(3):PCC.13r01609.
49. Perroud N, Cordera P, Zimmermann J, Michalopoulos G, Bancila V, Prada P, et al. Comorbidity between attention deficit hyperactivity disorder (ADHD) and bipolar disorder in a specialized mood disorders outpatient clinic. *Journal of affective disorders*. 2014;168:161-6.
50. Gergel T, Adiukwu F, McInnis M. Suicide and bipolar disorder: opportunities to change the agenda. *The Lancet Psychiatry*. 2024;11(10):781-4.
51. Goldberg JF, Truman CJ. Antidepressant-induced mania: an overview of current controversies. *Bipolar disorders*. 2003;5(6):407-20.
52. Chiang KS, Miklowitz DJ. Psychotherapy in Bipolar Depression: Effective Yet Underused. *Psychiatric annals*. 2023;53(2):58-62.
53. Merikangas KR, Jin R, He J-P, Kessler RC, Lee S, Sampson NA, et al. Lifetime and 12-Month Prevalence of Bipolar Spectrum Disorder in the National Comorbidity Survey Replication. *Archives of general psychiatry*. 2007;64(5):543-52.
54. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Psychiatry*. 2022;9(2):137-50.
55. O'Connell KS, Koromina M, van der Veen T, Boltz T, David FS, Yang JMK, et al. Genomics yields biological and phenotypic insights into bipolar disorder. *Nature*. 2025:1-12.
56. Yang J, O'Connell K, Escott-Price V, Florio AD. 43. EXPLORING SEX DIFFERENCES IN THE GENETIC ARCHITECTURE OF BIPOLAR DISORDER. *European Neuropsychopharmacology*. 2023;75:S78-S9.

57. Dell'Osso B, Cafaro R, Ketter TA. Has Bipolar Disorder become a predominantly female gender related condition? Analysis of recently published large sample studies. *International Journal of Bipolar Disorders*. 2021;9(1):3.
58. Gordovez FJA, McMahon FJ. The genetics of bipolar disorder. *Molecular Psychiatry*. 2020;25(3):544-59.
59. Malaspina D, Owen MJ, Heckers S, Tandon R, Bustillo J, Schultz S, et al. Schizoaffective Disorder in the DSM-5. *Schizophrenia Research*. 2013;150(1):21-5.
60. Keramatian K, Pinto JV, Tsang VWL, Chakrabarty T, Yatham LN. Duration of untreated or undiagnosed bipolar disorder and clinical characteristics and outcomes: systematic review and meta-analysis. *Br J Psychiatry*. 2025;227(3):622-632.
61. Goldberg JF, Harrow M, Whiteside JE. Risk for bipolar illness in patients initially hospitalized for unipolar depression. *American Journal of Psychiatry*. 2001;158(8):1265-70.
62. Grunze H, Born C. The Impact of Subsyndromal Bipolar Symptoms on Patient's Functionality and Quality of Life. *Frontiers in Psychiatry*. 2020;11:510.
63. Kalman JL, Papiol S, Forstner AJ, Heilbronner U, Degenhardt F, Strohmaier J, et al. Investigating polygenic burden in age at disease onset in bipolar disorder: Findings from an international multicentric study. *Bipolar Disord*. 2019;21(1):68-75.
64. Perlis RH, Miyahara S, Marangell LB, Wisniewski SR, Ostacher M, DelBello MP, et al. Long-term implications of early onset in bipolar disorder: Data from the first 1000 participants in the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD). *Biological Psychiatry*. 2004;55(9):875-81.
65. Baldessarini RJ, Tondo L, Vázquez GH, Undurraga J, Bolzani L, Yildiz A, et al. Age at onset versus family history and clinical outcomes in 1,665 international bipolar-I disorder patients. *World psychiatry*. 2012;11(1):40-6.
66. Wiström ED, O'Connell KS, Karadag N, Bahrami S, Hindley GFL, Lin A, et al. Genome-wide analysis reveals genetic overlap between alcohol use behaviours, schizophrenia and bipolar disorder and identifies novel shared risk loci. *Addiction*. 2022;117(3):600-10.
67. Bolton S, Warner J, Harriss E, Geddes J, Saunders KEA. Bipolar disorder: Trimodal age-at-onset distribution. *Bipolar Disorders*. 2021;23(4):341-56.
68. Miola A, Fountoulakis KN, Baldessarini RJ, Veldic M, Solmi M, Rasgon N, et al. Prevalence and outcomes of rapid cycling bipolar disorder: Mixed method systematic meta-review. *Journal of Psychiatric Research*. 2023;164:404-15.
69. Bourla A, Ferreri F, Baudry T, Panizzi V, Adrien V, Mouchabac S. Rapid cycling bipolar disorder: Literature review on pharmacological treatment illustrated by a case report on ketamine. *Brain and Behavior*. 2022;12(2):e2483.
70. Coryell W, Solomon D, Turvey C, Keller M, Leon AC, Endicott J, et al. The Long-term Course of Rapid-Cycling Bipolar Disorder. *Archives of General Psychiatry*. 2003;60(9):914-20.
71. Andreasen NC, Rice J, Endicott J, Coryell W, Grove WM, Reich T. Familial rates of affective disorder. A report from the National Institute of Mental Health Collaborative Study. *Archives of general psychiatry*. 1987;44(5):461-9.
72. Akiskal HS, Pinto O. The evolving bipolar spectrum. Prototypes I, II, III, and IV. *Psychiatr Clin North Am*. 1999;22(3):517-34, vii.
73. Jauhar S, McCutcheon R, Borgan F, et al. The relationship between cortical glutamate and striatal dopamine in first-episode psychosis: a cross-sectional multimodal PET and magnetic resonance spectroscopy imaging study. *Lancet Psychiatry*. 2018;5(10):816-823.
74. Jauhar S, McCutcheon RA, Nour MM, et al. Dopamine and Mood in Psychotic Disorders: An 18F-DOPA PET Study. *JAMA Psychiatry*. Published online August 13, 2025.

75. O'Connell KS, Coombes BJ. Genetic contributions to bipolar disorder: current status and future directions. *Psychological Medicine*. 2021;51(13):2156-67.
76. Stapp EK, Fullerton JM, Musci RJ, Zandi PP, McInnis MG, Mitchell PB, et al. Family environment and polygenic risk in the bipolar high-risk context. *JCPP Advances*. 2023;3(2):e12143.
77. Birmaher B, Merranko J, Hafeman D, Goldstein BI, Diler R, Levenson JC, et al. A Longitudinal Study of Psychiatric Disorders in Offspring of Parents With Bipolar Disorder From Preschool to Adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2021;60(11):1419-29.
78. Swann AC, Geller B, Post RM, Altshuler L, Chang KD, Delbello MP, et al. Practical Clues to Early Recognition of Bipolar Disorder: A Primary Care Approach. *Primary Care Companion to The Journal of Clinical Psychiatry*. 2005;7(1):15-21.
79. Arango C, Fraguas D, Parellada M. Differential Neurodevelopmental Trajectories in Patients With Early-Onset Bipolar and Schizophrenia Disorders. *Schizophrenia Bulletin*. 2014;40(Suppl 2):S138-46.
80. Kloiber S, Rosenblat JD, Husain MI, Ortiz A, Berk M, Quevedo J, et al. Neurodevelopmental pathways in bipolar disorder. *Neuroscience & Biobehavioral Reviews*. 2020;112:213-26.
81. Passos IC, Mwangi B, Vieta E, Berk M, Kapczinski F. Areas of controversy in neuroprogression in bipolar disorder. *Acta Psychiatrica Scandinavica*. 2016;134(2):91-103.
82. Owen MJ, O'Donovan MC. Schizophrenia and the neurodevelopmental continuum:evidence from genomics. *World Psychiatry*. 2017;16(3):227-35.
83. Birmaher B, Axelson D, Strober M, Gill MK, Yang M, Ryan N, et al. Comparison of Manic and Depressive Symptoms between Children and Adolescents with Bipolar Spectrum Disorders. *Bipolar disorders*. 2009;11(1):52-64.
84. Luby JL, Navsaria N. Pediatric bipolar disorder: evidence for prodromal states and early markers. *Journal of child psychology and psychiatry*. 2010;51(4):459-67.
85. Spear LP. Adolescent Neurodevelopment. *The Journal of adolescent health*. 2013;52(2 Suppl 2):S7-13.
86. Najt P, Wang F, Spencer L, Johnston JAY, Cox Lippard ET, Pittman BP, et al. Anterior Cortical Development During Adolescence in Bipolar Disorder. *Biological Psychiatry*. 2016;79(4):303-10.
87. Wegbreit E, Cushman GK, Puzia ME, Weissman AB, Kim KL, Laird AR, et al. Developmental Meta-Analyses of the Functional Neural Correlates of Bipolar Disorder. *JAMA psychiatry*. 2014;71(8):926-35.
88. Cannon M, Caspi A, Moffitt TE, Harrington H, Taylor A, Murray RM, Poulton R. Evidence for early-childhood, pan-developmental impairment specific to schizophreniform disorder: results from a longitudinal birth cohort. *Arch Gen Psychiatry*. 2002;59(5):449-56.
89. Valli I, Fabbri C, Young AH. Uncovering neurodevelopmental features in bipolar affective disorder. *The British Journal of Psychiatry*. 2019;215(1):383-385.
90. Kremen WS, Prom-Wormley E, Panizzon MS, Eyler LT, Fischl B, Neale MC, et al. Genetic and environmental influences on the size of specific brain regions in midlife: the VETSA MRI study. *NeuroImage*. 2010;49(2):1213-23.
91. Padmanabhan JL, Tandon N, Haller CS, Mathew IT, Eack SM, Clementz BA, et al. Correlations Between Brain Structure and Symptom Dimensions of Psychosis in Schizophrenia, Schizoaffective, and Psychotic Bipolar I Disorders. *Schizophrenia Bulletin*. 2015;41(1):154-62.

92. Panizzon MS, Fennema-Notestine C, Eyler LT, Jernigan TL, Prom-Wormley E, Neale M, et al. Distinct genetic influences on cortical surface area and cortical thickness. *Cerebral cortex*. 2009;19(11):2728-35.
93. Winkler AM, Kochunov P, Blangero J, Almasy L, Zilles K, Fox PT, et al. Cortical Thickness or Grey Matter Volume? The Importance of Selecting the Phenotype for Imaging Genetics Studies. *NeuroImage*. 2010;53(3):1135-46.
94. Abé C, Liberg B, Klahn AL, Petrovic P, Landén M. Mania-related effects on structural brain changes in bipolar disorder - a narrative review of the evidence. *Molecular Psychiatry*. 2023;28(7):2674-82.
95. Hanford LC, Nazarov A, Hall GB, Sassi RB. Cortical thickness in bipolar disorder: a systematic review. *Bipolar Disorders*. 2016;18(1):4-18.
96. Phillips ML, Swartz HA. A critical appraisal of neuroimaging studies of bipolar disorder: toward a new conceptualization of underlying neural circuitry and a road map for future research. *The American journal of psychiatry*. 2014;171(8):829-43.
97. Arnone D, Cavanagh J, Gerber D, Lawrie SM, Ebmeier KP, McIntosh AM. Magnetic resonance imaging studies in bipolar disorder and schizophrenia: meta-analysis. *The British journal of psychiatry*. 2009;195(3):194-201.
98. Keramatian K, Dhanoa T, McGirr A, Lang DJ, Honer WG, Lam RW, et al. Structural brain changes in first episode mania with and without psychosis: Data from the Systematic Treatment Optimization Program for Early Mania (STOP-EM). *The World Journal of Biological Psychiatry*. 2018;19(sup2):S30-S40.
99. Berkovitch L, Charles L, Del Cul A, Hamdani N, Delavest M, Sarrazin S, et al. Disruption of Conscious Access in Psychosis Is Associated with Altered Structural Brain Connectivity. *The Journal of Neuroscience*. 2021;41(3):513-22.
100. Lim CS, Baldessarini RJ, Vieta E, Yucel M, Bora E, Sim K. Longitudinal neuroimaging and neuropsychological changes in bipolar disorder patients: review of the evidence. *Neuroscience and biobehavioral reviews*. 2013;37(3):418-35.
101. Abé C, Ekman CJ, Sellgren C, Petrovic P, Ingvar M, Landén M. Manic episodes are related to changes in frontal cortex: a longitudinal neuroimaging study of bipolar disorder 1. *Brain*. 2015;138(Pt 11):3440-8.
102. Jakobsson J, Stridsberg M, Zetterberg H, Blennow K, Ekman CJ, Johansson AGM, et al. Decreased cerebrospinal fluid secretogranin II concentrations in severe forms of bipolar disorder. *Journal of Psychiatry & Neuroscience*. 2013;38(4):E21-E23.
103. Beyer JL, Young R, Kuchibhatla M, Krishnan KRR. Hyperintense MRI lesions in bipolar disorder: A meta-analysis and review. *International review of psychiatry*. 2009;21(4):394-409.
104. Bourne C, Aydemir Ö, Balanzá-Martínez V, Bora E, Brissos S, Cavanagh JTO, et al. Neuropsychological testing of cognitive impairment in euthymic bipolar disorder: an individual patient data meta-analysis. *Acta Psychiatrica Scandinavica*. 2013;128(3):149-62.
105. Sassi RB, Brambilla P, Hatch JP, Nicoletti MA, Mallinger AG, Frank E, et al. Reduced left anterior cingulate volumes in untreated bipolar patients. *Biological Psychiatry*. 2004;56(7):467-75.
106. Heinrichs RW, Zakzanis KK. Neurocognitive deficit in schizophrenia: A quantitative review of the evidence. *Neuropsychology*. 1998;12(3):426-45.
107. Bora E, Yücel M, Pantelis C. Neurocognitive markers of psychosis in bipolar disorder: A meta-analytic study. *Journal of Affective Disorders*. 2010;127(1-3):1-9.
108. Cotter D, Hudson L, Landau S. Evidence for orbitofrontal pathology in bipolar disorder and major depression, but not in schizophrenia. *Bipolar Disorders*. 2005;7(4):358-69.

109. Jackowski AP, de Araújo Filho GM, de Almeida AG, de Araújo CM, Reis M, Nery F, et al. The involvement of the orbitofrontal cortex in psychiatric disorders: an update of neuroimaging findings. *Revista brasileira de psiquiatria*. 2012;34(2):207-12.
110. Cotrena C, Damiani Branco L, Ponsoni A, Samamé C, Milman Shansis F, Paz Fonseca R. Executive functions and memory in bipolar disorders I and II: new insights from meta-analytic results. *Acta psychiatrica Scandinavica*. 2020;141(2):110-30.
111. Sánchez-Morla EM, López-Villarreal A, Jiménez-López E, Aparicio AI, Martínez-Vizcaíno V, Rodriguez-Jimenez R, et al. Impact of number of episodes on neurocognitive trajectory in bipolar disorder patients: a 5-year follow-up study. *Psychological medicine*. 2019;49(8):1299-307.
112. Perugi G, Pallucchini A, Rizzato S, Pinzone V, De Rossi P. Current and emerging pharmacotherapy for the treatment of adult attention deficit hyperactivity disorder (ADHD). *Expert Opinion on Pharmacotherapy*. 2019;20(12):1457-70.
113. Askeland RB, Hannigan LJ, O'Connell KS, Corfield EC, Frei O, Thapar A, et al. Developmental manifestations of polygenic risk for bipolar disorder from infancy to middle childhood. *Translational psychiatry*. 2023;13(1):29.
114. Ellard KK, Zimmerman JP, Kaur N, Van Dijk KRA, Roffman JL, Nierenberg AA, et al. Functional connectivity between anterior insula and key nodes of frontoparietal executive control and salience networks distinguish bipolar depression from unipolar depression and healthy controls. *Biological psychiatry Cognitive neuroscience and neuroimaging*. 2018;3(5):473-83.
115. Phillips ML, Ladouceur CD, Drevets WC. A neural model of voluntary and automatic emotion regulation: implications for understanding the pathophysiology and neurodevelopment of bipolar disorder. *Molecular Psychiatry*. 2008;13(9):833-57.
116. Nelson SB, Valakh V. Excitatory/Inhibitory Balance and Circuit Homeostasis in Autism Spectrum Disorders. *Neuron*. 2015;87(4):684-98.
117. Zoghbi HY, Bear MF. Synaptic dysfunction in neurodevelopmental disorders associated with autism and intellectual disabilities. *Cold Spring Harbor perspectives in biology*. 2012;4(3):a009886.
118. Lee E, Lee J, Kim E. Excitation/Inhibition Imbalance in Animal Models of Autism Spectrum Disorders. *Biological psychiatry*. 2017;81(10):838-47.
119. Kehrer C, Maziashvili N, Dugladze T, Gloveli T. Altered Excitatory-Inhibitory Balance in the NMDA-Hypofunction Model of Schizophrenia. *Frontiers in molecular neuroscience*. 2008;1:6.
120. Khayachi A, Ase A, Liao C, Kamesh A, Kuhlmann N, Schorova L, et al. Chronic lithium treatment alters the excitatory/ inhibitory balance of synaptic networks and reduces mGluR5-PKC signalling in mouse cortical neurons. *Journal of psychiatry & neuroscience*. 2021;46(3):E402-E14.
121. Rao JS, Harry GJ, Rapoport SI, Kim HW. Increased excitotoxicity and neuroinflammatory markers in postmortem frontal cortex from bipolar disorder patients. *Molecular psychiatry*. 2010;15(4):384-92.
122. Eastwood SL, Harrison PJ. Markers of glutamate synaptic transmission and plasticity are increased in the anterior cingulate cortex in bipolar disorder. *Biological psychiatry*. 2010;67(11):1010-6.
123. Jun C, Choi Y, Lim SM, Bae S, Hong YS, Kim JE, et al. Disturbance of the Glutamatergic System in Mood Disorders. *Experimental Neurobiology*. 2014;23(1):28-35.

124. Levinson AJ, Young LT, Fitzgerald PB, Daskalakis ZJ. Cortical inhibitory dysfunction in bipolar disorder: a study using transcranial magnetic stimulation. *Journal of clinical psychopharmacology*. 2007;27(5):493-7.
125. Fayed N, Andrés E, Viguera L, Modrego PJ, Garcia-Campayo J. Higher glutamate+glutamine and reduction of N-acetylaspartate in posterior cingulate according to age range in patients with cognitive impairment and/or pain. *Academic radiology*. 2014;21(9):1211-7.
126. Soares JC, Young AH, editors. *Bipolar disorders: basic mechanisms and therapeutic implications*. CRC Press; 2007.
127. Brady RO, McCarthy JM, Prescot AP, Jensen JE, Cooper AJ, Cohen BM, et al. Brain gamma-aminobutyric acid (GABA) abnormalities in bipolar disorder. *Bipolar disorders*. 2013;15(4):434-9.
128. Bearden CE, Woogen M, Glahn DC. Neurocognitive and neuroimaging predictors of clinical outcome in bipolar disorder. *Current psychiatry reports*. 2010;12(6):499-504.
129. Wang AY, Lohmann KM, Yang CK, Zimmerman EI, Pantazopoulos H, Herring N, et al. Bipolar disorder type 1 and schizophrenia are accompanied by decreased density of parvalbumin- and somatostatin-positive interneurons in the parahippocampal region. *Acta neuropathologica*. 2011;122(5):615-26.
130. Goddard AW, Mason GF, Almai A, Rothman DL, Behar KL, Petroff OAC, et al. Reductions in Occipital Cortex GABA Levels in Panic Disorder Detected With 1H-Magnetic Resonance Spectroscopy. *Archives of General Psychiatry*. 2001;58(6):556-61.
131. Arora I, Mal P, Arora P, Paul A, Kumar M. GABAergic implications in anxiety and related disorders. *Biochemical and Biophysical Research Communications*. 2024;724:150218.
132. Dharavath RN, Pina-Leblanc C, Tang VM, Sloan ME, Nikolova YS, Pangarov P, et al. GABAergic signaling in alcohol use disorder and withdrawal: pathological involvement and therapeutic potential. *Frontiers in Neural Circuits*. 2023;17:1218737.
133. Biria M, Banca P, Healy MP, Keser E, Sawiak SJ, Rodgers CT, et al. Cortical glutamate and GABA are related to compulsive behaviour in individuals with obsessive compulsive disorder and healthy controls. *Nature Communications*. 2023;14(1):2527.
134. Goodwin FK, Jamison KR. *Manic-depressive illness: bipolar disorders and recurrent depression*. 2nd ed. Oxford University Press; 2007.
135. Janowsky DS, Davis JM, El-Yousef MK, Sekerke HJ. A cholinergic-adrenergic hypothesis of mania and depression. *Lancet*. 1972;2(7778):632-5.
136. Van Enkhuizen J, Janowsky DS, Olivier B, Minassian A, Perry W, Young JW, et al. The catecholaminergic-cholinergic balance hypothesis of bipolar disorder revisited. *European journal of pharmacology*. 2015;753:114-26.
137. Lykouras L, Markianos M, Hatzimanolis J, Malliaras D, Stefanis C. Association of biogenic amine metabolites with symptomatology in delusional (psychotic) and nondelusional depressed patients. *Progress in neuro-psychopharmacology & biological psychiatry*. 1995;19(5):877-87.
138. Emilien G, Maloteaux JM, Geurts M, Hoogenberg K, Cragg S. Dopamine receptors—physiological understanding to therapeutic intervention potential. *Pharmacology & Therapeutics*. 1999;84(2):133-56.
139. Ashok AH, Marques TR, Jauhar S, Nour MM, Goodwin GM, Young AH, Howes OD. The dopamine hypothesis of bipolar affective disorder: the state of the art and implications for treatment. *Molecular psychiatry*. 2017;22(5):666-79.
140. Tye KM, Mirzabekov JJ, Warden MR, Ferenczi EA, Tsai HC, Finkelstein J, et al. Dopamine neurons modulate neural encoding and expression of depression-related behaviour. *Nature*. 2013;493(7433):537-41.

141. Kaalund SS, Newburn EN, Ye T, Tao R, Li C, Deep-Soboslay A, et al. Contrasting changes in DRD1 and DRD2 splice variant expression in schizophrenia and affective disorders, and associations with SNPs in postmortem brain. *Molecular psychiatry*. 2014;19(12):1258-66.
142. Jauhar S, McCutcheon RA, Nour MM, et al. Dopamine and Mood in Psychotic Disorders: An 18F-DOPA PET Study. *JAMA Psychiatry*. Published online August 13, 2025.
143. Potter WZ, Manji HK. Catecholamines in depression: an update. *Clinical chemistry*. 1994;40(2):279-87.
144. Nutt DJ. Altered central alpha 2-adrenoceptor sensitivity in panic disorder. *Archives of general psychiatry*. 1989;46(2):165-9.
145. Muneer A. Kynurenine Pathway of Tryptophan Metabolism in Neuropsychiatric Disorders: Pathophysiologic and Therapeutic Considerations. *Clinical psychopharmacology and neuroscience*. 2020;18(4):507-26.
146. Lee JG, Woo YS, Park SW, Seog DH, Seo MK, Bahk WM. Neuromolecular Etiology of Bipolar Disorder: Possible Therapeutic Targets of Mood Stabilizers. *Clinical Psychopharmacology and Neuroscience*. 2022;20(2):228-41.
147. Goodwin FK. Suicide, aggression, and depression. A theoretical framework for future research. *Annals of the New York Academy of Sciences*. 1986;487:351-5.
148. Åsberg M, Nordström P, Träskman-Bendz L. Cerebrospinal Fluid Studies in Suicidal Behavior. *Annals of the New York Academy of Sciences*. 1986;487:243-55.
149. Hughes JH, Dunne F, Young AH. Effects of acute tryptophan depletion on mood and suicidal ideation in bipolar patients symptomatically stable on lithium. *The British journal of psychiatry*. 2000;177:447-51.
150. Shelton RC. Intracellular mechanisms of antidepressant drug action. *Harvard Review of Psychiatry*. 2000;8(4):161-74.
151. Friedman E, Wang HY. Receptor-mediated activation of G proteins is increased in postmortem brains of bipolar affective disorder subjects. *Journal of neurochemistry*. 1996;67(3):1145-52.
152. Soares JC, Chen G, Dippold CS, Wells KF, Frank E, Kupfer DJ, et al. Concurrent measures of protein kinase C and phosphoinositides in lithium-treated bipolar patients and healthy individuals: a preliminary study. *Psychiatry research*. 2000;95(2):109-18.
153. Förstner U, Bohus M, Gebicke-Härter PJ, Bäumer B, Berger M, van Calker D. Decreased agonist-stimulated Ca2+ response in neutrophils from patients under chronic lithium therapy. *European archives of psychiatry and clinical neuroscience*. 1994;243(5):240-3.
154. Van Calker D, Förstner U, Bohus M, Gebicke-Härter P, Hecht H, Wark HJ, et al. Increased sensitivity to agonist stimulation of the Ca2+ response in neutrophils of manic-depressive patients: effect of lithium therapy. *Neuropsychobiology*. 1993;27(3):180-3.
155. Dubovsky SL, Thomas M, Hijazi A, Murphy J. Intracellular calcium signalling in peripheral cells of patients with bipolar affective disorder. *European archives of psychiatry and clinical neuroscience*. 1994;243(5):229-34.
156. Muresanu DF. Neuroprotection and neuroplasticity - a holistic approach and future perspectives. *Journal of the neurological sciences*. 2007;257(1-2):38-43.
157. Pap M, Cooper GM. Role of glycogen synthase kinase-3 in the phosphatidylinositol 3-Kinase/Akt cell survival pathway. *The Journal of biological chemistry*. 1998;273(32):19929-32.
158. de Oliveira GS, Ceresér KM, Fernandes BS, Kauer-Sant'Anna M, Fries GR, Stertz L, et al. Decreased brain-derived neurotrophic factor in medicated and drug-free bipolar patients. *Journal of Psychiatric Research*. 2009;43(14):1171-4.

159. Beurel E, Jope RS. The paradoxical pro- and anti-apoptotic actions of GSK3 in the intrinsic and extrinsic apoptosis signaling pathways. *Progress in neurobiology*. 2006;79(4):173-89.
160. Muneer A. Wnt and GSK3 Signaling Pathways in Bipolar Disorder: Clinical and Therapeutic Implications. *Clinical psychopharmacology and neuroscience*. 2017;15(2):100-14.
161. Palmer DS, Howrigan DP, Chapman SB, Adolfsson R, Bass N, Blackwood D, et al. Exome sequencing in bipolar disorder identifies AKAP11 as a risk gene shared with schizophrenia. *Nature genetics*. 2022;54(5):541-7.
162. Budd SL, Nicholls DG. Mitochondria in the life and death of neurons. *Essays in biochemistry*. 1998;33:43-52.
163. Mattson MP, Gleichmann M, Cheng A. Mitochondria in Neuroplasticity and Neurological Disorders. *Neuron*. 2008;60(5):748-66.
164. Silver IA, Deas J, Erecińska M. Ion homeostasis in brain cells: Differences in intracellular ion responses to energy limitation between cultured neurons and glial cells. *Neuroscience*. 1997;78(2):589-601.
165. el-Mallakh RS, Barrett JL, Wyatt RJ. The Na,K-ATPase Hypothesis for Bipolar Disorder: Implications of Normal Development. *Journal of child and adolescent psychopharmacology*. 1993;3(1):37-52.
166. Massaad CA, Klann E. Reactive Oxygen Species in the Regulation of Synaptic Plasticity and Memory. *Antioxidants & Redox Signaling*. 2011;14(10):2013-54.
167. Osete JR, Akkouh IA, de Assis DR, et al. Lithium increases mitochondrial respiration in iPSC-derived neural precursor cells from lithium responders. *Mol Psychiatry*. 2021;26(11):6789-6805.
168. Birner A, Platzer M, Bengesser SA, Dalkner N, Fellendorf FT, Queissner R, et al. Increased breakdown of kynurenine towards its neurotoxic branch in bipolar disorder. *PLoS ONE*. 2017;12(2):e0172699.
169. Savitz J, Dantzer R, Wurfel BE, Victor TA, Ford BN, Bodurka J, et al. Neuroprotective kynurenine metabolite indices are abnormally reduced and positively associated with hippocampal and amygdalar volume in bipolar disorder. *Psychoneuroendocrinology*. 2015;52:200-11.
170. Husain MI, Chaudhry IB, Khoso AB, et al. Minocycline and celecoxib as adjunctive treatments for bipolar depression: a multicentre, factorial design randomised controlled trial. *Lancet Psychiatry*. 2020;7(6):515-527.
171. Deshauer D, Duffy A, Alda M, Grof E, Albuquerque J, Grof P. The cortisol awakening response in bipolar illness: a pilot study. *Canadian journal of psychiatry*. 2003;48(7):462-6.
172. Bartels M, van den Berg M, Sluyter F, Boomsma DI, de Geus EJC. Heritability of cortisol levels: Review and simultaneous analysis of twin studies. *Psychoneuroendocrinology*. 2003;28(2):121-37.
173. Lewis KJS, Di Florio A, Forty L, Gordon-Smith K, Perry A, Craddock N, et al. Mania triggered by sleep loss and risk of postpartum psychosis in women with bipolar disorder. *Journal of Affective Disorders*. 2018;225:624-9.
174. Ramirez-Mahaluf JP, Rozas-Serri E, Ivanovic-Zuvic F, Risco L, Vöhringer PA, BaHammam AS, et al. Effectiveness of Sleep Deprivation in Treating Acute Bipolar Depression as Augmentation Strategy: A Systematic Review and Meta-Analysis. *Frontiers in Psychiatry*. 2020;11:70.
175. Smeraldi E, Benedetti F, Barbini B, Campori E, Colombo C. Sustained Antidepressant Effect of Sleep Deprivation Combined with Pindolol in Bipolar Depression: A Placebo-Controlled Trial. *Neuropsychopharmacology*. 1999;20(4):380-5.

176. Kripke DF, Nievergelt CM, Joo EJ, Shekhtman T, Kelsoe JR. Circadian polymorphisms associated with affective disorders. *Journal of Circadian Rhythms*. 2009;7:2.
177. Rodriguez de los Santos M, Kopell BH, Buxbaum Grice A, Ganesh G, Yang A, Amini P, et al. Divergent landscapes of A-to-I editing in postmortem and living human brain. *Nature Communications*. 2024;15(1):1283.
178. Manji HK, Henter ID, Zarate Jr CA. Bipolar disorder: a neurobiological synthesis. Curr Top Behav Neurosci. 2011;5:331-40.
179. Belvederi Murri M, Prestia D, Mondelli V, Pariante C, Patti S, Olivieri B, et al. The HPA axis in bipolar disorder: Systematic review and meta-analysis. Psychoneuroendocrinology. 2016;63:327-42.
180. Daban C, Vieta E, Mackin P, Young AH. Hypothalamic-pituitary-adrenal axis and bipolar disorder. *The Psychiatric clinics of North America*. 2005;28(2):469-80.
181. Fries GR, Vasconcelos-Moreno MP, Gubert C, dos Santos BT, Sartori J, Eisele B, et al. Hypothalamic-Pituitary-Adrenal Axis Dysfunction and Illness Progression in Bipolar Disorder. *International Journal of Neuropsychopharmacology*. 2015;18(1):pyu043.
182. Manji HK, Quiroz JA, Payne JL, Singh J, Lopes BP, Viegas JS, et al. The underlying neurobiology of bipolar disorder. World Psychiatry. 2003;2(3):136-46.
183. D'Addario C, Dell'Osso B, Palazzo MC, Benatti B, Lietti L, Cattaneo E, et al. Selective DNA Methylation of BDNF Promoter in Bipolar Disorder: Differences Among Patients with BDI and BDII. Neuropsychopharmacology. 2012;37(7):1647-55.
184. Ludwig B, Dwivedi Y. Dissecting bipolar disorder complexity through epigenomic approach. Molecular Psychiatry 2016 21:11. 2016;21(11):1490-8.
185. Duffy A, Goodday SM, Keown-Stoneman C, Scotti M, Maitra M, Nagy C, et al. Epigenetic markers in inflammation-related genes associated with mood disorder: a cross-sectional and longitudinal study in high-risk offspring of bipolar parents. International journal of bipolar disorders. 2019;7(1).
186. Fries GR, Quevedo J, Zeni CP, Kazimi IF, Zunta-Soares G, Spiker DE, et al. Integrated transcriptome and methylome analysis in youth at high risk for bipolar disorder: a preliminary analysis. Translational Psychiatry 2017 7:3. 2017;7(3):e1059.
187. Roth TL, Lubin FD, Funk AJ, Sweatt JD. LASTING EPIGENETIC INFLUENCE OF EARLY-LIFE ADVERSITY ON THE BDNF GENE. Biological psychiatry. 2009;65(9):760-9.
188. Mitchell C, Schneper LM, Notterman DA. DNA methylation, early life environment, and health outcomes. Pediatric research. 2016;79(1-2):212-9.
189. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009;460(7256):748-52.
190. Ruderfer DM, Ripke S, McQuillin A, Boocock J, Stahl EA, Pavlides JMW, et al. Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. Cell. 2018;173(7):1705-15.e16.
191. Murray RM, Sham P, Van Os J, Zanelli J, Cannon M, McDonald C. A developmental model for similarities and dissimilarities between schizophrenia and bipolar disorder. Schizophrenia Research. 2004;71(2-3):405-16.
192. Clifton NE, Hannon E, Harwood JC, Di Florio A, Thomas KL, Holmans PA, et al. Dynamic expression of genes associated with schizophrenia and bipolar disorder across development. Translational Psychiatry 2019 9:1. 2019;9(1):1-9.

193. Marangoni C, Hernandez M, Faedda GL. The role of environmental exposures as risk factors for bipolar disorder: A systematic review of longitudinal studies. Journal of Affective Disorders. 2016;193:165-74.
194. Shintani AO, Rabelo-da-Ponte FD, Marchionatti LE, Watts D, Ferreira de Souza F, Machado CdS, et al. Prenatal and perinatal risk factors for bipolar disorder: A systematic review and meta-analysis. Neuroscience & Biobehavioral Reviews. 2023;144:104960.
195. Haukvik UK, McNeil T, Lange EH, Melle I, Dale AM, Andreassen OA, et al. Pre- and perinatal hypoxia associated with hippocampus/amygdala volume in bipolar disorder. Psychological Medicine. 2013;44(5):975-85.
196. Chudal R, Brown AS, Gissler M, Suominen A, Sourander A. Is maternal smoking during pregnancy associated with bipolar disorder in offspring? Journal of Affective Disorders. 2015;171:132-6.
197. Gonzalez-Calvo I, Ronald A, Shakoor S, Taylor MJ, Eley TC, Hosang GM. Perinatal risk factors and subclinical hypomania: A prospective community study. Journal of Affective Disorders. 2024;362:885-92.
198. Pacchiarotti I, Bond DJ, Baldessarini RJ, Nolen WA, Grunze H, Licht RW, et al. The International Society for Bipolar Disorders (ISBD) task force report on antidepressant use in bipolar disorders. The American journal of psychiatry. 2013;170(11):1249-62.
199. Saiardi A, Mudge AW. Lithium and fluoxetine regulate the rate of phosphoinositide synthesis in neurons: a new view of their mechanisms of action in bipolar disorder. Transl Psychiatry. 2018;8(1):175.
200. Manchia M, Adli M, Akula N, Ardau R, Aubry JM, Backlund L, et al. Assessment of Response to Lithium Maintenance Treatment in Bipolar Disorder: A
201. Nunes A, Trappenberg T, Alda M. Asymmetrical reliability of the Alda score favours a dichotomous representation of lithium responsiveness. PloS one. 2020;15(1).
202. Song J, Bergen SE, Di Florio A, Karlsson R, Charney A, Ruderfer DM, et al. Genome-wide association study identifies SESTD1 as a novel risk gene for lithium-responsive bipolar disorder. Molecular Psychiatry 2016 21:9. 2015;21(9):1290-7.
203. Hou L, Heilbronner U, Degenhardt F, Adli M, Akiyama K, Akula N, et al. Genetic variants associated with response to lithium treatment in bipolar disorder: a genome-wide association study. Lancet (London, England). 2016;387(10023):1085.
204. Perlis RH, Smoller JW, Ferreira MA, McQuillin A, Bass N, Lawrence J, et al. A genomewide association study of response to lithium for prevention of recurrence in bipolar disorder. Am J Psychiatry. 2009;166(6):718-25.
205. Chen CH. Variant GADL1 and response to lithium therapy in bipolar I disorder. N Engl J Med. 2014;370(2):119-28.
206. Cruceanu C, Alda M, Dion PA, Turecki G, Rouleau GA. No evidence for GADL1 variation as a bipolar disorder susceptibility factor in a Caucasian lithium-responsive cohort. Am J Psychiatry. 2015;172(1):94-5.
207. Birnbaum R, Shin JH, Weinberger D. Variant GADL1 and response to lithium in bipolar I disorder. N Engl J Med. 370. United States2014. p. 1855-6
208. Kotambail A, Mathur A, Bhat SM, Rai PS, Sharma PS, Satyamoorthy K. GADL1 gene polymorphisms and lithium response in bipolar I disorder: lack of association from an Indian population. Psychiatr Genet. 2015;25(1):39-40.
209. Song J, Bergen SE, Di Florio A, Karlsson R, Charney A, Ruderfer DM, et al. Erratum: Genome-wide association study identifies SESTD1 as a novel risk gene for lithium-responsive bipolar disorder. Molecular Psychiatry 2017 22:8. 2017;22(8):1223.

210. McQuillin A, Yao K, Nadeem A, Veen TVD, Thygesen J, Jonsson L, et al. Implication of the ADCY1 Gene in Lithium Response in Bipolar Disorder by Genome-wide Association Meta-analysis. 2024.
211. Ho AMC, Coombes BJ, Nguyen TTL, Liu D, McElroy SL, Singh B, et al. Mood-Stabilizing Antiepileptic Treatment Response in Bipolar Disorder: A Genome-Wide Association Study. Clinical pharmacology and therapeutics. 2020;108(6):1233-42.
212. Cipriani A, Saunders K, Attenburrow MJ, Stefaniak J, Panchal P, Stockton S, et al. A systematic review of calcium channel antagonists in bipolar disorder and some considerations for their future development. Molecular psychiatry. 2016;21(10):1324-32.
213. Colbourne L, Harrison PJ. Brain-penetrant calcium channel blockers are associated with a reduced incidence of neuropsychiatric disorders. Molecular psychiatry. 2022;27(9):3904-12.
214. Mosayebi-Samani M, Melo L, Agboada D, Nitsche MA, Kuo MF. Ca2+ channel dynamics explain the nonlinear neuroplasticity induction by cathodal transcranial direct current stimulation over the primary motor cortex. European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology. 2020;38:63-72.
215. Alda M. LITHIUM IN THE TREATMENT OF BIPOLAR DISORDER: PHARMACOLOGY AND PHARMACOGENETICS. Molecular psychiatry. 2015;20(6):661-70.
216. O'Brien WT, Klein PS. Validating GSK3 as an in vivo target of lithium action. Biochemical Society transactions. 2009;37(Pt 5):1133-8.
217. Moore GJ, Bebchuk JM, Parrish JK, Faulk MW, Arfken CL, Strahl-Bevacqua J, et al. Temporal dissociation between lithium-induced changes in frontal lobe myo-inositol and clinical response in manic-depressive illness. The American journal of psychiatry. 1999;156(12):1902-8.
218. Wang J, Michelhaugh SK, Bannon MJ. Valproate robustly increases Sp transcription factor-mediated expression of the dopamine transporter gene within dopamine cells. European Journal of Neuroscience. 2007;25(7):1982-6.
219. Yatham LN, Liddle PF, Gonzalez M, et al. A Positron Emission Tomography Study of Dopamine Transporter Density in Patients With Bipolar Disorder With Current Mania and Those With Recently Remitted Mania. *JAMA Psychiatry*. 2022;79(12):1217-1224.
220. Gitlin M. Treatment-resistant bipolar disorder. Molecular psychiatry. 2006;11(3):227-40.
221. Yildiz A, Vieta E, Leucht S, Baldessarini RJ. Efficacy of Antimanic Treatments: Meta-analysis of Randomized, Controlled Trials. Neuropsychopharmacology 2011 36:2. 2010;36(2):375-89.
222. Jawad MY, Qasim S, Ni M, Guo Z, Di Vincenzo JD, d'Andrea G, et al. The Role of Ketamine in the Treatment of Bipolar Depression: A Scoping Review. Brain Sciences. 2023;13(6):909-.
223. Lee M-C, Yang Y-C, Chen Y-C, Chang B-S, Li Y-C, Huang S-C. Treatment of acute bipolar depression. Tzu-Chi Medical Journal. 2018;30(3):141.
224. Swann AC, Bowden CL, Calabrese JR, Dilsaver SC, Morris DD. Differential effect of number of previous episodes of affective disorder on response to lithium or divalproex in acute mania. The American journal of psychiatry. 1999;156(8):1264-6.
225. Scott J, Paykel E, Morriss R, Bentall R, Kinderman P, Johnson T, et al. Cognitive-behavioural therapy for severe and recurrent bipolar disorders: randomised controlled trial. The British journal of psychiatry : the journal of mental science. 2006;188(APR.):313-20.
226. Bienvenu OJ, Davydow DS, Kendler KS. Psychiatric 'diseases' versus behavioral disorders and degree of genetic influence. Psychological medicine. 2011;41(1):33-40.

227. Johansson V, Kuja-Halkola R, Cannon TD, Hultman CM, Hedman AM. A population-based heritability estimate of bipolar disorder - In a Swedish twin sample. Psychiatry Research. 2019;278:180-7.
228. Barnett JH, Smoller JW. The Genetics of Bipolar Disorder. Neuroscience. 2009;164(1):331-.
229. Rice J, Cloninger CR, Reich T. Multifactorial inheritance with cultural transmission and assortative mating. I. Description and basic properties of the unitary models. American Journal of Human Genetics. 1978;30(6):618.
230. Tsuang MT, Faraone SV. The genetics of mood disorders1990.
231. Kessler RC, Rubinow DR, Holmes C, Abelson JM, Zhao S. The epidemiology of DSM-III-R bipolar I disorder in a general population survey. Psychological medicine. 1997;27(5):1079-89.
232. Gershon ES, Hamovit J, Guroff JJ, Dibble E, Leckman JF, Sceery W, et al. A family study of schizoaffective, bipolar I, bipolar II, unipolar, and normal control probands. Archives of general psychiatry. 1982;39(10):1157-67.
233. Andreasen NC, Rice J, Endicott J, Coryell W, Grove WM, Reich T. Familial rates of affective disorder. A report from the National Institute of Mental Health Collaborative Study. Archives of general psychiatry. 1987;44(5):461-9.
234. Heun R, Maier W. The distinction of bipolar II disorder from bipolar I and recurrent unipolar depression: results of a controlled family study. Acta psychiatrica Scandinavica. 1993;87(4):279-84.
235. Smoller JW, Finn CT. Family, twin, and adoption studies of bipolar disorder. American journal of medical genetics Part C, Seminars in medical genetics. 2003;123C(1):48-58.
236. Song J, Bergen SE, Kuja-Halkola R, Larsson H, Landén M, Lichtenstein P. Bipolar disorder and its relation to major psychiatric disorders: a family-based study in the Swedish population. Bipolar disorders. 2015;17(2):184-93.
237. Laursen TM, Labouriau R, Licht RW, Bertelsen A, Munk-Olsen T, Mortensen PB. Family history of psychiatric illness as a risk factor for schizoaffective disorder: a Danish register-based cohort study. Archives of general psychiatry. 2005;62(8):841-8.
238. Andlauer TFM, Guzman-Parra J, Streit F, Strohmaier J, Gonzalez MJ, Gil Flores S, et al. Bipolar multiplex families have an increased burden of common risk variants for psychiatric disorders. Mol Psychiatry. 2019.
239. Kendler KS, Ohlsson H, Sundquist J, Sundquist K. An extended swedish national adoption study of bipolar disorder illness and cross-generational familial association with schizophrenia and major depression. JAMA Psychiatry. 2020;77(8):814-22.
240. Faraone SV, Biederman J, Wozniak J. Examining the comorbidity between attention deficit hyperactivity disorder and bipolar I disorder: a meta-analysis of family genetic studies. The American journal of psychiatry. 2012;169(12):1256-66.
241. Patel RS, Manikkara G, Chopra A. Bipolar Disorder and Comorbid Borderline Personality Disorder: Patient Characteristics and Outcomes in US Hospitals. Medicina. 2019;55(1):13-.
242. Lopes LM, Matias M, Marques M, Lopes IM, Reis J. Bipolar Disorder and Borderline Personality Disorder: A Diagnostic Challenge. European Psychiatry. 2023;66(Suppl 1):S701-S.
243. Kuja-Halkola R, Lind Juto K, Skoglund C, Rück C, Mataix-Cols D, Pérez-Vigil A, et al. Do borderline personality disorder and attention-deficit/hyperactivity disorder co-aggregate in families? A population-based study of 2 million Swedes. Molecular psychiatry. 2021;26(1):341-9.
244. Badner JA, Gershon ES. Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. Molecular Psychiatry 2002 7:4. 2002;7(4):405-11.

245. Segurado R, Detera-Wadleigh SD, Levinson DF, Lewis CM, Gill M, Nurnberger JI, et al. Genome Scan Meta-Analysis of Schizophrenia and Bipolar Disorder, Part III: Bipolar Disorder. The American Journal of Human Genetics. 2003;73(1):49-62.
246. McQueen MB, Devlin B, Faraone SV, Nimgaonkar VL, Sklar P, Smoller JW, et al. Combined Analysis from Eleven Linkage Studies of Bipolar Disorder Provides Strong Evidence of Susceptibility Loci on Chromosomes 6q and 8q. The American Journal of Human Genetics. 2005;77(4):582-95.
247. Hamshere ML, Bennett P, Williams N, Segurado R, Cardno A, Norton N, et al. Genomewide Linkage Scan in Schizoaffective Disorder Significant Evidence for Linkage at 1q42 Close to DISC1, and Suggestive Evidence at 22q11 and 19p13. 2005.
248. Park N, Juo SH, Cheng R, Liu J, Loth JE, Lilliston B, et al. Linkage analysis of psychosis in bipolar pedigrees suggests novel putative loci for bipolar disorder and shared susceptibility with schizophrenia. Molecular Psychiatry 2004 9:12. 2004;9(12):1091-9.
249. Maziade M, Roy MA, Chagnon YC, Cliche D, Fournier JP, Montgrain N, et al. Shared and specific susceptibility loci for schizophrenia and bipolar disorder: a dense genome scan in Eastern Quebec families. Molecular Psychiatry 2005 10:5. 2004;10(5):486-99.
250. Kerem BS, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, et al. Identification of the cystic fibrosis gene: genetic analysis. Science (New York, NY). 1989;245(4922):1073-80.
251. Craddock N, Forty L. Genetics of affective (mood) disorders. European journal of human genetics : EJHG. 2006;14(6):660-8.
252. Ferreira MAR, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L, et al. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. Nature Genetics 2008 40:9. 2008;40(9):1056-8.
253. Craddock N, Sklar P. Genetics of bipolar disorder: successful start to a long journey. Trends in Genetics. 2009;25(2):99-105.
254. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. Nature Genetics 2019 51:5. 2019;51(5):793-803.
255. Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, Qiao Z, et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nature Genetics 2021 53:6. 2021;53(6):817-29.
256. Sul JH, Service SK, Huang AY, Ramensky V, Hwang S-G, Teshiba TM, et al. Contribution of common and rare variants to bipolar disorder susceptibility in extended pedigrees from population isolates. 2020.
257. Toma C, Shaw AD, Allcock RJN, Heath A, Pierce KD, Mitchell PB, et al. An examination of multiple classes of rare variants in extended families with bipolar disorder. Translational psychiatry. 2018;8(1).
258. Goes FS, Pirooznia M, Parla JS, Kramer M, Ghiban E, Mavruk S, et al. Exome Sequencing of Familial Bipolar Disorder. JAMA psychiatry. 2016;73(6):590-7.
259. Maaser A, Forstner AJ, Strohmaier J, Hecker J, Ludwig KU, Sivalingam S, et al. Exome sequencing in large, multiplex bipolar disorder families from Cuba. PLoS One. 2018;13(10):e0205895.
260. Forstner AJ, Fischer SB, Schenk LM, Strohmaier J, Maaser-Hecker A, Reinbold CS, et al. Whole-exome sequencing of 81 individuals from 27 multiply affected bipolar disorder families. Translational Psychiatry. 2020;10(1):57.

261. McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S, et al. Microduplications of 16p11.2 are associated with schizophrenia. Nature genetics. 2009;41(11):1223-7.
262. Malhotra D, McCarthy S, Michaelson JJ, Vacic V, Burdick KE, Yoon S, et al. High frequencies of de novo CNVs in bipolar disorder and schizophrenia. Neuron. 2011;72(6):951-63.
263. Green EK, Rees E, Walters JTR, Smith KG, Forty L, Grozeva D, et al. Copy number variation in bipolar disorder. Molecular Psychiatry. 2016;21(1):89-93.
264. Kirov G. CNVs in neuropsychiatric disorders. Human molecular genetics. 2015;24(R1):R45-R9.
265. Charney AW, Stahl EA, Green EK, Chen CY, Moran JL, Chambert K, et al. Contribution of Rare Copy Number Variants to Bipolar Disorder Risk Is Limited to Schizoaffective Cases. Biol Psychiatry. 2019;86(2):110-9.
266. Tesli M, Espeseth T, Bettella F, Mattingsdal M, Aas M, Melle I, et al. Polygenic risk score and the psychosis continuum model. Acta psychiatrica Scandinavica. 2014;130(4):311-7.
267. Kataoka M, Matoba N, Sawada T, Kazuno AA, Ishiwata M, Fujii K, et al. Exome sequencing for bipolar disorder points to roles of de novo loss-of-function and protein-altering mutations. Molecular Psychiatry. 2016;21(7):885-93.
268. Georgieva L, Rees E, Moran JL, Chambert KD, Milanova V, Craddock N, et al. De novo CNVs in bipolar affective disorder and schizophrenia. Human molecular genetics. 2014;23(24):6677-83.
269. Goes FS, Pirooznia M, Tehan M, Zandi PP, McGrath J, Wolyniec P, et al. De Novo Variation in Bipolar Disorder. Molecular psychiatry. 2021 Aug;26(8):4127-36
270. Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, et al. Rate of de novo mutations and the importance of father's age to disease risk. Nature. 2012;488(7412):471-5.
271. Frans EM, Sandin S, Reichenberg A, Lichtenstein P, Långström N, Hultman CM. Advancing paternal age and bipolar disorder. Archives of general psychiatry. 2008;65(9):1034-40.
272. Chudal R, Gissler M, Sucksdorff D, Lehti V, Suominen A, Hinkka-Yli-Salomäki S, et al. Parental age and the risk of bipolar disorders. Bipolar disorders. 2014;16(6):624-32.
273. Malaspina D, Harlap S, Fennig S, et al. Advancing paternal age and the risk of schizophrenia. *Arch Gen Psychiatry*. 2001;58(4):361-367.
274. Ament SA, Szelinger S, Glusman G, Ashworth J, Hou L, Akula N, et al. Rare variants in neuronal excitability genes influence risk for bipolar disorder. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(11):3576-81.
275. Gandal MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S, et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science (New York, NY). 2018;362(6420).
276. Vornholt E, Liharska LE, Cheng E, Hashemi A, Park YJ, Ziafat K, et al. Characterizing cell type specific transcriptional differences between the living and postmortem human brain. medRxiv : the preprint server for health sciences. 2024.
277. Aas M, Haukvik UK, Djurovic S, Tesli M, Athanasiu L, Bjella T, et al. Interplay between childhood trauma and BDNF val66met variants on blood BDNF mRNA levels and on hippocampus subfields volumes in schizophrenia spectrum and bipolar disorders. Journal of psychiatric research. 2014;59:14-21.
278. Hosang GM, Fisher HL, Cohen-Woods S, McGuffin P, Farmer AE. Stressful life events and catechol-O-methyl-transferase (COMT) gene in bipolar disorder. Depression and Anxiety. 2017;34(5):419-26.

279. Tosato S, Bonetto C, De Santi K, Lasalvia A, Gennarelli M, Cristofalo D, et al. COMT but Not 5HTTLPR Gene Is Associated with Depression in First-Episode Psychosis: The Role of Stressful Life Events. Genes (Basel). 2023;14(2).
280. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007;449(7164):851-61.
281. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al. A global reference for human genetic variation. Nature 2015 526:7571. 2015;526(7571):68-74.
282. Baum AE, Akula N, Cabanero M, Cardona I, Corona W, Klemens B, et al. A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. Molecular psychiatry. 2008;13(2):197-207.
283. Zeng Z, Wang T, Li T, Li Y, Chen P, Zhao Q, et al. Common SNPs and haplotypes in DGKH are associated with bipolar disorder and schizophrenia in the Chinese Han population. Molecular psychiatry. 2011;16(5):473-5.
284. Takata A, Kawasaki H, Iwayama Y, Yamada K, Gotoh L, Mitsuyasu H, et al. Nominal association between a polymorphism in DGKH and bipolar disorder detected in a meta-analysis of East Asian case-control samples. Psychiatry and clinical neurosciences. 2011;65(3):280-5.
285. Cichon S, Mühleisen TW, Degenhardt FA, Mattheisen M, Miró X, Strohmaier J, et al. Genome-wide Association Study Identifies Genetic Variation in Neurocan as a Susceptibility Factor for Bipolar Disorder. American Journal of Human Genetics. 2011;88(3):372-81.
286. Dannlowski U, Kugel H, Grotegerd D, Redlich R, Suchy J, Opel N, et al. NCAN Cross-Disorder Risk Variant Is Associated With Limbic Gray Matter Deficits in Healthy Subjects and Major Depression. Neuropsychopharmacology 2015 40:11. 2015;40(11):2510-6.
287. Einarsdottir E, Peyrard-Janvid M, Darki F, Tuulari JJ, Merisaari H, Karlsson L, et al. Identification of NCAN as a candidate gene for developmental dyslexia. Scientific Reports 2017 7:1. 2017;7(1):1-11.
288. Miró X, Meier S, Dreisow ML, Frank J, Strohmaier J, Breuer R, et al. Studies in humans and mice implicate neurocan in the etiology of mania. The American journal of psychiatry. 2012;169(9):982-90.
289. Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, et al. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007 447:7145. 2007;447(7145):661-78.
290. Ekholm JM, Kieseppä T, Heikkalinna T, Partonen T, Paunio T, Perola M, et al. Evidence of susceptibility loci on 4q32 and 16p12 for bipolar disorder. Human Molecular Genetics. 2003;12(15):1907-15.
291. Cheng R, Juo SH, Loth JE, Nee J, Iossifov I, Blumenthal R, et al. Genome-wide linkage scan in a large bipolar disorder sample from the National Institute of Mental Health genetics initiative suggests putative loci for bipolar disorder, psychosis, suicide, and panic disorder. Molecular psychiatry. 2006;11(3):252-60.
292. Krystal JH, Sanacora G, Blumberg H, Anand A, Charney DS, Marek G, et al. Glutamate and GABA systems as targets for novel antidepressant and mood-stabilizing treatments. Molecular Psychiatry 2002 7:1. 2002;7(1):S71-S80.
293. Craddock N, Jones L, Jones IR, Kirov G, Green EK, Grozeva D, et al. Strong genetic evidence for a selective influence of GABAA receptors on a component of the bipolar disorder phenotype. Molecular psychiatry. 2010;15(2):146-53.
294. Sklar P, Smoller JW, Fan J, Ferreira MAR, Perlis RH, Chambert K, et al. Whole-genome Association Study of Bipolar Disorder. Molecular psychiatry. 2008;13(6):558-.

295. Alliey-Rodriguez N, Zhang D, Badner JA, Lahey BB, Zhang X, Dinwiddie S, et al. Genome-wide association study of personality traits in bipolar patients. Psychiatric genetics. 2011;21(4):190-.
296. Kandaswamy R, McQuillin A, Curtis D, Gurling H. Allelic Association, DNA Resequencing and Copy Number Variation at the Metabotropic Glutamate Receptor GRM7 Gene Locus in Bipolar Disorder. American Journal of Medical Genetics. 2014;165(4):365-72.
297. Sklar P, Ripke S, Scott LJ, Andreassen OA, Cichon S, Craddock N, et al. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nature genetics. 2011;43(10):977-85.
298. Mühleisen TW, Leber M, Schulze TG, Strohmaier J, Degenhardt F, Treutlein J, et al. Genome-wide association study reveals two new risk loci for bipolar disorder. Nature Communications. 2014;5(1):3339.
299. Schrode N, Ho SM, Yamamuro K, Dobbyn A, Huckins L, Matos MR, et al. Synergistic effects of common schizophrenia risk variants. Nature genetics. 2019;51(10):1475-85.
300. Traylor M, Markus H, Lewis CM. Homogeneous case subgroups increase power in genetic association studies. European Journal of Human Genetics. 2014;23(6):863-.
301. Berridge MJ, Downes CP, Hanley MR. Neural and developmental actions of lithium: a unifying hypothesis. Cell. 1989;59(3):411-9.
302. Berridge MJ. Neuronal calcium signaling. Neuron. 1998;21(1):13-26.
303. Schlecker C, Boehmerle W, Jeromin A, DeGray B, Varshney A, Sharma Y, et al. Neuronal calcium sensor-1 enhancement of InsP3 receptor activity is inhibited by therapeutic levels of lithium. The Journal of clinical investigation. 2006;116(6):1668-74.
304. Kabir ZD, Martínez-Rivera A, Rajadhyaksha AM. From Gene to Behavior: L-Type Calcium Channel Mechanisms Underlying Neuropsychiatric Symptoms. Neurotherapeutics. 2017;14(3):588-.
305. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. PLOS Computational Biology. 2015;11(4):e1004219.
306. Desfossés J, Stip E, Bentaleb LA, Potvin S. Endocannabinoids and Schizophrenia. Pharmaceuticals. 2010;3(10):3101-26.
307. Zamberletti E, Rubino T, Parolaro D. The endocannabinoid system and schizophrenia: integration of evidence. Current pharmaceutical design. 2012;18(32):4980-90.
308. Serra G, Fratta W. A possible role for the endocannabinoid system in the neurobiology of depression. Clinical practice and epidemiology in mental health : CP & EMH. 2007;3.
309. Singh T, Poterba T, Curtis D, Akil H, Al Eissa M, Barchas JD, et al. Rare coding variants in ten genes confer substantial risk for schizophrenia. Nature. 2022;604(7906):509-16.
310. Palmer DS, Howrigan DP, Chapman SB, Adolfsson R, Bass N, Blackwood D, Boks MP, Chen CY, Churchhouse C, Corvin AP, Craddock N. Exome sequencing in bipolar disorder reveals shared risk gene AKAP11 with schizophrenia. medRxiv. 2021 Mar 12:2021-03.
311. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. Nature genetics. 2015;47(9):1091-8.
312. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nature Genetics 2016 48:3. 2016;48(3):245-52.
313. Kim B, Kim D, Schulmann A, Patel Y, Caban-Rivera C, Kim P, et al. Cellular Diversity in Human Subgenual Anterior Cingulate and Dorsolateral Prefrontal Cortex by Single-Nucleus RNA-Sequencing. J Neurosci. 2023;43(19):3582-97

314. Goff KM, Goldberg EM. A Role for VIP Interneurons in Neurodevelopmental Disorders. Developmental neuroscience. 2021;43(3-4):168-80.
315. Akula N, Barb J, Jiang X, Wendland JR, Choi KH, Sen SK, et al. RNA-sequencing of the brain transcriptome implicates dysregulation of neuroplasticity, circadian rhythms and GTPase binding in bipolar disorder. Molecular psychiatry. 2014;19(11):1179-85.
316. Gigase FAJ, Snijders GJLJ, Boks MP, de Witte LD. Neurons and glial cells in bipolar disorder: A systematic review of postmortem brain studies of cell number and size. Neuroscience and biobehavioral reviews. 2019;103:150-62.
317. Holmgren A, Akkouh I, O'Connell KS, Osete JR, Bjørnstad PM, Djurovic S, et al. Bipolar patients display stoichiometric imbalance of gene expression in post-mortem brain samples. Molecular Psychiatry 2024 29:4. 2024;29(4):1128-38.
318. Zandi PP, Jaffe AE, Goes FS, Burke EE, Collado-Torres L, Huuki-Myers L, et al. Amygdala and anterior cingulate transcriptomes from individuals with bipolar disorder reveal down-regulated neuroimmune and synaptic pathways. Nature neuroscience. 2022;25(3):381-9.
319. Jaffe AE, Straub RE, Shin JH, Tao R, Gao Y, Collado-Torres L, et al. Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. Nature neuroscience. 2018;21(8):1117-25.
320. Wartchow KM, Scaini G, Quevedo J. Glial-Neuronal Interaction in Synapses: A Possible Mechanism of the Pathophysiology of Bipolar Disorder. Advances in experimental medicine and biology. 2023;1411:191-208.
321. Lawrence JM, Breunig S, Foote IF, Tallis CB, Grotzinger AD. Genomic SEM applied to explore etiological divergences in bipolar subtypes. Psychological Medicine. 2024;54(6):1152-9.
322. Zimmerman M, Ruggero CJ, Chelminski I, Young D. Is bipolar disorder overdiagnosed? The Journal of clinical psychiatry. 2008;69(6):935-40.
323. Zimmerman M, Ruggero CJ, Chelminski I, Young D. Psychiatric diagnoses in patients previously overdiagnosed with bipolar disorder. The Journal of clinical psychiatry. 2010;71(1):26-31.
324. Guzman-Parra J, Streit F, Forstner AJ, Strohmaier J, Gonzalez MJ, Gil Flores S, et al. Clinical and genetic differences between bipolar disorder type 1 and 2 in multiplex families. Transl Psychiatry. 2021;11(1):31.
325. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. Bioinformatics. 2015;31(9):1466-8.
326. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nature Communications 2019 10:1. 2019;10(1):1-10.
327. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nature genetics. 2010;42(7):565-9.
328. Pain O, Glanville KP, Hagenaars SP, Selzam S, Furtjes AE, Gaspar HA, et al. Evaluation of polygenic prediction methodology within a reference-standardized framework. PLOS Genetics. 2021;17(5):e1009021.
329. Hiser J, Koenigs M. The multifaceted role of ventromedial prefrontal cortex in emotion, decision-making, social cognition, and psychopathology. Biological psychiatry. 2017;83(8):638-.
330. Birmaher B, Hafeman D, Merranko J, Zwicker A, Goldstein B, Goldstein T, et al. Role of Polygenic Risk Score in the Familial Transmission of Bipolar Disorder in Youth. JAMA Psychiatry. 2022;79(2):160-8.

331. Hafeman DM, Uher R, Merranko J, Zwicker A, Goldstein B, Goldstein TR, et al. Person-level contributions of bipolar polygenic risk score to the prediction of new-onset bipolar disorder in at-risk offspring. Journal of affective disorders. 2024;368:359-65.
332. Yao K, van der Veen T, Thygesen J, Bass N, McQuillin A. Multiple psychiatric polygenic risk scores predict associations between childhood adversity and bipolar disorder. Journal of affective disorders. 2023;341:137-46.
333. Kimbrel NA, Ashley-Koch AE, Qin XJ, Lindquist JH, Garrett ME, Dennis MF, et al. A Genome-Wide Association Study of Suicide Attempts in the Million Veterans Program Identifies Evidence of Pan-Ancestry and Ancestry-Specific Risk Loci. Molecular psychiatry. 2022;27(4):2264-72.
334. Zai CC, Dimick MK, Young LT, Kennedy JL, Goldstein BI. Polygenic risk scores in relation to suicidality among youth with or at risk for bipolar disorder. Journal of affective disorders. 2025;375:44-8.
335. Mullins N, Bigdeli TB, Børglum AD, Coleman JRI, Demontis D, Mehta D, et al. GWAS of suicide attempt in psychiatric disorders and association with major depression polygenic risk scores. American Journal of Psychiatry. 2019;176(8):651-60.
336. Wilcox HC, Fullerton JM, Glowinski AL, Benke K, Kamali M, Hulvershorn LA, et al. Traumatic Stress Interacts With Bipolar Disorder Genetic Risk to Increase Risk for Suicide Attempts HHS Public Access. J Am Acad Child Adolesc Psychiatry. 2017;56(12):1073-80.
337. Lee PH, Doyle AE, Li X, Silberstein M, Jung JY, Gollub RL, et al. Genetic Association of Attention-Deficit/Hyperactivity Disorder and Major Depression With Suicidal Ideation and Attempts in Children: The Adolescent Brain Cognitive Development Study. Biological psychiatry. 2021;92(3):236-45.
338. Thiel K, Lemke H, Winter A, Flinkenflügel K, Waltemate L, Bonnekoh L, et al. White and gray matter alterations in bipolar I and bipolar II disorder subtypes compared with healthy controls - exploring associations with disease course and polygenic risk. Neuropsychopharmacology 2024 49:5. 2024;49(5):814-23.
339. Lei D, Li W, Tallman MJ, Patino LR, McNamara RK, Strawn JR, et al. Changes in the brain structural connectome after a prospective randomized clinical trial of lithium and quetiapine treatment in youth with bipolar disorder. Neuropsychopharmacology 2021 46:7. 2021;46(7):1315-23.
340. Calafato MS, Thgesen JH, Ranlund S, Zartaloudi E, Cahn W, Crespo-Facorro B, et al. Use of schizophrenia and bipolar disorder polygenic risk scores to identify psychotic disorders. The British Journal of Psychiatry. 2018;213(3):535-.
341. Mistry S, Escott-Price V, Florio AD, Smith DJ, Zammit S. Genetic risk for bipolar disorder and psychopathology from childhood to early adulthood. Journal of Affective Disorders. 2019;246:633-9.
342. Grigoroiu-Serbanescu M, Giaroli G, Thygesen JH, Shenyan O, Bigdeli TB, Bass NJ, et al. Predictive power of the ADHD GWAS 2019 polygenic risk scores in independent samples of bipolar patients with childhood ADHD. Journal of Affective Disorders. 2020;265:651-9.
343. Nunez NA, Coombes BJ, Romo-Nava F, Bond DJ, Vande Voort J, Croarkin PE, et al. Clinical and Genetic Correlates of Bipolar Disorder With Childhood-Onset Attention Deficit Disorder. Frontiers in Psychiatry. 2022;13:884217-45.
344. Reginsson GW, Ingason A, Euesden J, Bjornsdottir G, Olafsson S, Sigurdsson E, et al. Polygenic risk scores for schizophrenia and bipolar disorder associate with addiction. Addiction Biology. 2017;23(1):485-92.
345. Du Rietz E, Coleman J, Glanville K, Choi SW, O'Reilly PF, Kuntsi J. Association of Polygenic Risk for Attention-Deficit/Hyperactivity Disorder With Co-occurring Traits and

Disorders. Biological psychiatry Cognitive neuroscience and neuroimaging. 2018;3(7):635-43.
346. Demontis D, Walters GB, Athanasiadis G, Walters R, Therrien K, Nielsen TT, et al. Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. Nature Genetics 2023 55:2. 2023;55(2):198-208.
347. Patel KHS, Walters GB, Stefánsson H, Stefánsson K, Degenhardt F, Nothen M, et al. Predicting ADHD in alcohol dependence using polygenic risk scores for ADHD. American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics. 2024;195(4).
348. Blumberg HP, Kaufman J, Martin A, Charney DS, Krystal JH, Peterson BS. Significance of adolescent neurodevelopment for the neural circuitry of bipolar disorder. Annals of the New York Academy of Sciences. 2004;1021:376-83.
349. Vonk R, Van Der Schot AC, Van Baal GCM, Van Oel CJ, Nolen WA, Kahn RS. Premorbid school performance in twins concordant and discordant for bipolar disorder. Journal of affective disorders. 2012;136(3):294-303.
350. Coombes BJ, Markota M, Mann JJ, Colby C, Stahl E, Talati A, et al. Dissecting clinical heterogeneity of bipolar disorder using multiple polygenic risk scores. Translational Psychiatry 2020 10:1. 2020;10(1):1-8.
351. Schubert KO, Thalamuthu A, Amare AT, Frank J, Streit F, Adl M, et al. Combining schizophrenia and depression polygenic risk scores improves the genetic prediction of lithium response in bipolar disorder patients. Translational Psychiatry 2021 11:1. 2021;11(1):1-8.
352. Amare AT, Schubert KO, Hou L, Clark SR, Papiol S, Heilbronner U, et al. Association of Polygenic Score for Schizophrenia and HLA Antigen and Inflammation Genes With Response to Lithium in Bipolar Affective Disorder: A Genome-Wide Association Study. JAMA psychiatry. 2018;75(1):65-74.
353. Amare AT, Schubert KO, Hou L, Clark SR, Papiol S, Cearns M, et al. Association of polygenic score for major depression with response to lithium in patients with bipolar disorder. Molecular Psychiatry. 2020.
354. Allardyce J, Cardno AG, Gordon-Smith K, Jones L, Di Florio A, Walters JTR, et al. Specificity of polygenic signatures across symptom dimensions in bipolar disorder: an analysis of UK Bipolar Disorder Research Network data. The Lancet Psychiatry. 2023;10(8):623-31.
355. David FS, Stein F, Andlauer TFM, Streit F, Witt SH, Herms S, et al. Genetic contributions to transdiagnostic symptom dimensions in patients with major depressive disorder, bipolar disorder, and schizophrenia spectrum disorders. Schizophrenia Research. 2023;252:161-71.
356. Sugaya L, Hasin DS, Olfson M, Lin KH, Grant BF, Blanco C. Child Physical Abuse and Adult Mental Health: A National Study. Journal of traumatic stress. 2012;25(4):384-92.
357. Aas M, Henry C, Andreassen OA, Bellivier F, Melle I, Etain B. The role of childhood trauma in bipolar disorders. International Journal of Bipolar Disorders. 2016;4(1):1-10.
358. Baldwin JR, Sallis HM, Schoeler T, Taylor MJ, Kwong ASF, Tielbeek JJ, et al. A genetically informed Registered Report on adverse childhood experiences and mental health. Nature Human Behaviour 2022 7:2. 2022;7(2):269-90.
359. Dell'Osso B, Cafaro R, Ketter TA. Has Bipolar Disorder become a predominantly female gender related condition? Analysis of recently published large sample studies. Int J Bipolar Disord. 2021;9(1):3.
360. Arnold LM. Gender differences in bipolar disorder. The Psychiatric clinics of North America. 2003;26(3):595-620.

361. Ercis M, Sanchez-Ruiz JA, Webb LM, Solares-Bravo M, Betcher HK, Moore KM, et al. Sex differences in effectiveness and adverse effects of mood stabilizers and antipsychotics: A systematic review. Journal of Affective Disorders. 2024;352:171-92.
362. Schwaba T, Mallard TT, Maihofer AX, Rhemtulla M, Lee PH, Smoller JW, et al. Comparison of the multivariate genetic architecture of eight major psychiatric disorders across sex. Nature Genetics 2025 57:3. 2025;57(3):583-90.
363. Williams J, Chen T, Hua X, Wong W, Yu K, Kraft P, et al. Integrating Common and Rare Variants Improves Polygenic Risk Prediction Across Diverse Populations. medRxiv. 2024:2024.11.05.24316779-2024.11.05.
364. Krapohl E, Patel H, Newhouse S, Curtis CJ, Von Stumm S, Dale PS, et al. Multi-polygenic score approach to trait prediction. Molecular psychiatry. 2018;23(5):1368-74.
365. Craig JE, Han X, Qassim A, Hassall M, Cooke Bailey JN, Kinzy TG, et al. Multitrait analysis of glaucoma identifies new risk loci and enables polygenic prediction of disease susceptibility and progression. Nature genetics. 2020;52(2):160-6.
366. Hansen L, Bernstorff M, Enevoldsen K, Kolding S, Damgaard JG, Perfalk E, et al. Predicting Diagnostic Progression to Schizophrenia or Bipolar Disorder via Machine Learning. JAMA Psychiatry. 2025.

8.2 CHAPTER 2

1. Nurnberger JI, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, et al. Diagnostic interview for genetic studies: rationale, unique features, and training. Archives of General Psychiatry. 1994;51(11):849-59.
2. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.) Washington, DC: Author.
3. Endicott J, Spitzer RL. A diagnostic interview: the schedule for affective disorders andschizophrenia. Archives of General Psychiatry. 1978;35(7):837-44.
4. McGuffin P, Farmer A, Harvey I. A polydiagnostic application of operational criteria in studies of psychotic illness: development and reliability of the OPCRIT system. Archives of General Psychiatry. 1991;48(8):764-70.
5. American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.
6. American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed. Text revision). Washington, DC: Author.
7. American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
8. World Health Organization. (1978). *International classification of diseases* (9th ed.). Geneva: Author.
9. World Health Organization. (1992). *International classification of diseases* (10th ed.). Geneva: Author.
10. First, M.B., Spitzer, R.L., Gibbon, M., & Williams, J.B.W. (1997). *Structured Clinical Interview for DSM-IV Axis I disorders (SCID I)*. New York: Biometric Research Department.
11. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*. 2016;48(10):1279-83.

12. Loh PR, Danecek P, Palamara PF, Fuchsberger C, Reshef YA, Finucane HK, et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nature Genetics. 2016;48(11):1443-8.
13. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nature Genetics*. 2016;48(10):1284-7.
14. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.
15. Merikangas KR, Akiskal HS, Angst J, Greenberg PE, Hirschfeld RMA, Petukhova M, et al. Lifetime and 12-Month Prevalence of Bipolar Spectrum Disorder in the National Comorbidity Survey Replication. Archives of general psychiatry. 2007;64(5):543-52.
16. Perälä J, Suvisaari J, Saarni SI, Kuoppasalmi K, Isometsä E, Pirkola S, Partonen T, Tuulio-Henriksson A, Hintikka J, Kieseppä T, Härkänen T, Koskinen S, Lönnqvist J. Lifetime prevalence of psychotic and bipolar I disorders in a general population. Arch Gen Psychiatry. 2007 Jan;64(1):19-28.
17. Tondo L, Baldessarini RJ, Hennen J, Floris G, Silvetti F, Tohen M. Lithium treatment and risk of suicidal behavior in bipolar disorder patients. J Clin Psychiatry. 1998 Aug;59(8):405-14.
18. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry. 2005 Jun;62(6):593-602. doi: 10.1001/archpsyc.62.6.593. Erratum in: Arch Gen Psychiatry. 2005 Jul;62(7):768.
19. Ruscio AM, Hallion LS, Lim CCW, et al. Cross-sectional Comparison of the Epidemiology of DSM-5 Generalized Anxiety Disorder Across the Globe. *JAMA Psychiatry*. 2017;74(5):465-475.
20. Grant BF, Goldstein RB, Saha TD, et al. Epidemiology of DSM-5 Alcohol Use Disorder: Results From the National Epidemiologic Survey on Alcohol and Related Conditions III. *JAMA Psychiatry*. 2015;72(8):757-766.
21. Nock MK, Borges G, Bromet EJ, Cha CB, Kessler RC, Lee S. Suicide and suicidal behavior. *Epidemiol Rev*. 2008;30(1):133-154.
22. Angst J, Marneros A. Bipolarity from ancient to modern times: conception, birth and rebirth. *J Affect Disord*. 2001;67(1-3):3-19.
23. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291-295.
24. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015;47(11):1236-1241.
25. Turley P, Walters RK, Maghzian O, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet*. 2018;50(2):229-237.
26. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol*. 2015;11(4):e1004219. Published 2015 Apr 17.
27. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8(1):1826. Published 2017 Nov 28
28. Gusev A, Ko A, Shi H, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*. 2016;48(3):245-252.
29. Werme J, van der Sluis S, Posthuma D, de Leeuw CA. An integrated framework for local genetic correlation analysis. *Nat Genet*. 2022;54(3):274-282.
30. Zeng J, Xue A, Jiang L, et al. Widespread signatures of natural selection across human complex traits and functional genomic categories. *Nat Commun*. 2021;12(1):1164.

31. Lee SH, Goddard ME, Wray NR, Visscher PM. A better coefficient of determination for genetic profile analysis. *Genet Epidemiol*. 2012;36(3):214-224.
32. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575.
33. Chen W, Wu Y, Zheng Z, et al. Improved analyses of GWAS summary statistics by reducing data heterogeneity and errors. *Nat Commun*. 2021;12(1):7117.
34. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190-2191.
35. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics*. 2015;31(9):1466-1468.
36. Ge T, Chen CY, Ni Y, Feng YA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun*. 2019;10(1):1776.
37. Palla L, Dudbridge F. A Fast Method that Uses Polygenic Scores to Estimate the Variance Explained by Genome-wide Marker Panels and the Proportion of Variants Affecting a Trait. *Am J Hum Genet*. 2015;97(2):250-259.
38. R Core Team. R: A language and environment for statistical computing [Computer software]. Version 4.4.2. Vienna, Austria: R Foundation for Statistical Computing; 2024.
39. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 2007;39(2):175-191.
40. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk of complex disease. *Curr Opin Genet Dev*. 2008;18(3):257-263.
41. Nakazawa, M. (2023). fmsb: Functions for Medical Statistics Book with some Demographic Data. R package version 0.7.6.
42. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.
43. Mahmoud O, Dudbridge F, Davey Smith G, Munafo M, Tilling K. A robust method for collider bias correction in conditional genome-wide association studies. Nat Commun. 2022 Feb 2;13(1):619.
44. O'Connell KS, Koromina M, van der Veen T, et al. Genomics yields biological and phenotypic insights into bipolar disorder. *Nature*. 2025;639(8056):968-975.
45. Trubetskoy V, Pardiñas AF, Qi T, et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*. 2022;604(7906):502-508.
46. Howard DM, Adams MJ, Clarke TK, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci*. 2019;22(3):343-352.
47. Demontis D, Walters GB, Athanasiadis G, Walters R, Therrien K, Nielsen TT, et al. Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. Nature Genetics. 2023;55(2):198-208.
48. Purves KL, Coleman JRI, Meier SM, et al. A major role for common genetic variation in anxiety disorders. *Mol Psychiatry*. 2020;25(12):3292-3303.
49. Grove J, Ripke S, Als TD, et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet*. 2019;51(3):431-444.
50. Neale Lab UK Biobank. (2018). GWAS Round 2. http://www.nealelab.is/uk-biobank.
51. Nievergelt CM, Maihofer AX, Klengel T, et al. International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. *Nat Commun*. 2019;10(1):4558.

52. Witt SH, Streit F, Jungkunz M, et al. Genome-wide association study of borderline personality disorder reveals genetic overlap with bipolar disorder, major depression and schizophrenia. *Transl Psychiatry*. 2017;7(6):e1155.
53. Watanabe K, Jansen PR, Savage JE, et al. Genome-wide meta-analysis of insomnia prioritizes genes associated with metabolic and psychiatric pathways. *Nat Genet*. 2022;54(8):1125-1132.
54. Savage JE, Jansen PR, Stringer S, et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet*. 2018;50(7):912-919.
55. de la Fuente J, Davies G, Grotzinger AD, Tucker-Drob EM, Deary IJ. A general dimension of genetic sharing across diverse cognitive traits inferred from molecular data. *Nat Hum Behav*. 2021;5(1):49-58.
56. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11(5):550-560.
57. Mitchell RE, Hartley AE, Walker VM, Gkatzionis A, Yarmolinsky J, Bell JA, et al. Strategies to investigate and mitigate collider bias in genetic and Mendelian randomisation studies of disease progression. *PLoS Genetics*. 2023;19(2):e1010596.
58. Jöreskog KG, Goldberger AS. Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*. 1975;70(351):631-9.
59. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904-909.
60. Lam M, Awasthi S, Watson HJ, et al. RICOPILI: Rapid Imputation for COnsortias PIpeLIne. *Bioinformatics*. 2020;36(3):930-933.
61. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-15550.
62. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417-425.
63. Giusti-Rodríguez P, Lu L, Yang Y, Crowley CA, Liu X, Juric I, et al. Using three-dimensional regulatory chromatin interactions from adult and fetal cortex to interpret genetic results for psychiatric disorders and cognitive traits. [Preprint]. bioRxiv. 2018 Oct 31; doi: 10.1101/406330.
64. Wang D, Liu S, Warrell J, et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science*. 2018;362(6420):eaat8464.
65. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46(3):310-315.
66. Wang D, Liu S, Warrell J, et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science*. 2018;362(6420):eaat8464.
67. Hodge RD, Bakken TE, Miller JA, et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature*. 2019;573(7772):61-68.
68. Habib N, Avraham-Davidi I, Basu A, et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods*. 2017;14(10):955-958.
69. La Manno G, Gyllborg D, Codeluppi S, et al. Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. *Cell*. 2016;167(2):566-580.e19.
70. Hochgerner, H., Zeisel, A., Lönnerberg, P., & Linnarsson, S. (2018). Conserved properties of neural stem cells in the adult hippocampus. *Nature neuroscience*, *21*(2), 299-308.

71. Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), 289-300.
72. Grotzinger AD, Rhemtulla M, de Vlaming R, et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat Hum Behav*. 2019;3(5):513-525.
73. Mancuso N, Freund MK, Johnson R, et al. Probabilistic fine-mapping of transcriptome-wide association studies. *Nat Genet*. 2019;51(4):675-682.
74. Singh T, Poterba T, Curtis D, et al. Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature*. 2022;604(7906):509-516.
75. Palmer DS, Howrigan DP, Chapman SB, et al. Exome sequencing in bipolar disorder identifies AKAP11 as a risk gene shared with schizophrenia. *Nat Genet*. 2022;54(5):541-547.
76. Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, *28*(5), 1–26.
77. Fox J (2022). *polycor: Polychoric and Polyserial Correlations*. R package version 0.8-1.
78. Tierney N, Cook D (2023). "Expanding Tidy Data Principles to Facilitate Missing Data Exploration, Visualization and Assessment of Imputations." *Journal of Statistical Software*, 105(7), 1–31.
79. Harrison E, Drake T, Ots R. finalfit: Quickly Create Elegant Regression Results Tables and Plots when Modelling [Internet]. Version 1.0.6. 2023.
80. Kaiser HF. A Second Generation Little Jiffy. *Psychometrika,* 1970;35(4):401-415.
81. Bartlett, M. S. (1950). Tests of Significance in Factor Analysis. British Journal of Statistical Psychology, 3, 77-85.
82. Revelle, W. (2025). psych: Procedures for Personality and Psychological Research. Northwestern University, Evanston, Illinois. R package version 2.5.6.
83. Kassambara, A., & Mundt, F. (2020). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7.
84. Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36.
85. Lishinski A (2025). *lavaanPlot: Path Diagrams for 'Lavaan' Models via 'DiagrammeR'*. R package version 0.8.1.
86. Allardyce, J., Leonenko, G., & Walters, J. T. (2023). The impact of polygenic risk scores on clinical decision making in psychiatry. *The British Journal of Psychiatry*, *222*(1), 1-3.
87. Moshagen, M., & Erdfelder, E. (2015). A New Strategy for Testing Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(1), 54-60.
88. Hothorn, Torsten, et al. "Unbiased Recursive Partitioning: A Conditional Inference Framework." *Journal of Computational and Graphical Statistics*, vol. 15, no. 3, 2006, pp. 651–74. *JSTOR*.
89. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*. 2010;33(1):1-22.
90. Janitza, S., Celik, E., & Boulesteix, A. L. (2018). A computationally fast variable importance test for random forests for high-dimensional data. *Advances in Data Analysis and Classification*, *12*(4), 885-915.
91. Debeer D, Hothorn T, Strobl C (2025). *permimp: Conditional Permutation Importance*. R package version 1.1.0.9000.
92. LeDell E, Petersen M, van der Laan M. Computationally efficient confidence intervals for cross-validated area under the ROC curve estimates. *Electron J Stat*. 2015;9(1):1583-1607.

93. Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software*, *25*(1), 1–18.

8.3 CHAPTER 3

1. Regier DA, Kuhl EA, Kupfer DJ. The DSM-5: classification and criteria changes. *World Psychiatry*. 2013;12(2):92-8.
2. O'Connell KS, Coombes BJ. Genetic contributions to bipolar disorder: current status and future directions. *Psychological Medicine*. 2021;51(13):2156-67.
3. Pavlova B, Perlis RH, Mantere O, Sellgren CM, Isometsä E, Mitchell PB, et al. Prevalence of current anxiety disorders in people with bipolar disorder during euthymia: a meta-analysis. *Psychological Medicine*. 2017;47(6):1107-15.
4. Sanches M, Bauer IE, Galvez JF, Zunta-Soares GB, Soares JC. The management of cognitive impairment in bipolar disorder: current status and perspectives. *American Journal of Therapeutics*. 2015;22(6):477-86.
5. Mignogna KM, Goes FS. Characterizing the longitudinal course of symptoms and functioning in bipolar disorder. *Psychological Medicine*. 2024;54(1):79-89.
6. Grunze H, Born C. The impact of subsyndromal bipolar symptoms on patients functionality and quality of life. *Frontiers in Psychiatry*. 2020;11:510.
7. Altshuler LL, Ventura J, van Gorp WG, Green MF, Theberge DC, Mintz J. Neurocognitive function in clinically stable men with bipolar I disorder or schizophrenia and normal control subjects. *Biological Psychiatry*. 2004;56(8):560-9.
8. Robinson LJ, Thompson JM, Gallagher P, Goswami U, Young AH, Ferrier IN, et al. A meta-analysis of cognitive deficits in euthymic patients with bipolar disorder. *Journal of Affective Disorders*. 2006;93(1-3):105-15.
9. Sparding T, Pålsson E, Joas E, Hansen S, Landén M. Personality traits in bipolar disorder and influence on outcome. *BMC Psychiatry*. 2017;17(1):159.
10. Schulze TG. Genetic research into bipolar disorder: the need for a research framework that integrates sophisticated molecular biology and clinically informed phenotype characterization. *The Psychiatric Clinics of North America*. 2010;33(1):67-82.
11. Saunders EH, Scott LJ, McInnis MG, Burmeister M. Familiality and diagnostic patterns of subphenotypes in the National Institutes of Mental Health bipolar sample. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics*. 2008;147B(1):18-26.
12. Allardyce J, McCreadie RG, Morrison G, van Os J. Do symptom dimensions or categorical diagnoses best discriminate between known risk factors for psychosis? *Social Psychiatry and Psychiatric Epidemiology*. 2007;42(6):429-37.
13. Ruderfer DM, Fanous AH, Ripke S, McQuillin A, Amdur RL, Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Molecular Psychiatry*. 2014;19(9):1017-24.
14. Reininghaus U, Böhnke JR, Hosang G, Farmer A, Burns T, McGuffin P, et al. Evaluation of the validity and utility of a transdiagnostic psychosis dimension encompassing schizophrenia and bipolar disorder. *The British Journal of Psychiatry*. 2016;209(2):107-13.

15. Qiu F, Akiskal HS, Kelsoe JR, Greenwood TA. Factor analysis of temperament and personality traits in bipolar patients: correlates with comorbidity and disorder severity. *Journal of Affective Disorders*. 2017;207:282-90.
16. Allardyce J, Leonenko G, Hamshere M, Pardiñas AF, Forty L, Knott S, et al. Association between schizophrenia-related polygenic liability and the occurrence and level of mood-incongruent psychotic symptoms in bipolar disorder. *JAMA Psychiatry*. 2018;75(1):28-35.
17. Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell*. 2018;173(7):1705-15.e16.
18. Grigoroiu-Serbanescu M, Giaroli G, Thygesen JH, Shenyan O, Bigdeli TB, Bass NJ, et al. Predictive power of the ADHD GWAS 2019 polygenic risk scores in independent samples of bipolar patients with childhood ADHD. *Journal of Affective Disorders*. 2020;265:651-9.
19. Coombes BJ, Markota M, Mann JJ, Colby C, Stahl E, Talati A, et al. Dissecting clinical heterogeneity of bipolar disorder using multiple polygenic risk scores. *Translational Psychiatry*. 2020;10(1):314.
20. Richards AL, Cardno A, Harold G, Craddock NJ, Di Florio A, Jones L, et al. Genetic liabilities differentiating bipolar disorder, schizophrenia, and major depressive disorder, and phenotypic heterogeneity in bipolar disorder. *JAMA Psychiatry*. 2022;79(10):1032-9.
21. Allardyce J, Cardno AG, Gordon-Smith K, Jones L, Di Florio A, Walters JTR, et al. Specificity of polygenic signatures across symptom dimensions in bipolar disorder: an analysis of UK Bipolar Disorder Research Network data. *The Lancet Psychiatry*. 2023;10(8):623-31.
22. Grigoroiu-Serbanescu M, van der Veen T, Bigdeli T, Herms S, Diaconu CC, Neagu AI, et al. Schizophrenia polygenic risk scores, clinical variables and genetic pathways as predictors of phenotypic traits of bipolar I disorder. *Journal of Affective Disorders*. 2024;356:507-18.
23. Song J, Jonsson L, Lu Y, Bergen SE, Karlsson R, Smedler E, et al. Key subphenotypes of bipolar disorder are differentially associated with polygenic liabilities for bipolar disorder, schizophrenia, and major depressive disorder. *Molecular Psychiatry*. 2024;29(7):1941-50.
24. Lawrence JM, Breunig S, Foote IF, Tallis CB, Grotzinger AD. Genomic SEM applied to explore etiological divergences in bipolar subtypes. *Psychological Medicine*. 2024;54(6):1152-9.
25. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Arlington, VA: American Psychiatric Publishing; 2013.
26. Miskowiak KW, Hansen KB, Mariegaard J, Kessing LV. Association between childhood trauma, cognition, and psychosocial function in a large sample of partially or fully remitted patients with bipolar disorder and healthy participants. *International Journal of Bipolar Disorders*. 2023;11(1):31.
27. Judd LL, Akiskal HS, Schettler PJ, Endicott J, Leon AC, Solomon DA, et al. Psychosocial disability in the course of bipolar I and II disorders: a prospective, comparative, longitudinal study. *Archives of General Psychiatry*. 2005;62(12):1322-30.
28. Bennett F, Hodgetts S, Close A, Frye M, Grunze H, Keck P, et al. Predictors of psychosocial outcome of bipolar disorder: data from the Stanley Foundation Bipolar Network. *International Journal of Bipolar Disorders*. 2019;7(1):28.
29. Carballo JJ, Llorente C, Kehrmann L, Flamarique I, Zuddas A, Purper-Ouakil D, et al. Psychosocial risk factors for suicidality in children and adolescents. *European Child & Adolescent Psychiatry*. 2020;29(6):759-76.
30. Rohde P, Lewinsohn PM, Seeley JR, Klein DN, Andrews JA, Small JW. Psychosocial functioning of adults who experienced substance use disorders as adolescents. *Psychology of Addictive Behaviors*. 2007;21(2):155-64.

31. Scott J, Graham A, Yung A, Morgan C, Bellivier F, Etain B. A systematic review and meta-analysis of delayed help-seeking, delayed diagnosis and duration of untreated illness in bipolar disorders. *Acta Psychiatrica Scandinavica*. 2022;146(5):389-405.
32. Dong M, Lu L, Zhang L, Zhang Q, Ungvari GS, Ng CH, et al. Prevalence of suicide attempts in bipolar disorder: a systematic review and meta-analysis of observational studies. *Epidemiology and Psychiatric Sciences*. 2019;29:e63.
33. Léda-Rêgo G, Studart-Bottó P, Abbade P, Rabelo-Da-Ponte FD, Casqueiro JS, Sarmento S, et al. Lifetime prevalence of psychiatric comorbidities in patients with bipolar disorder: a systematic review and meta-analysis. *Psychiatry Research*. 2024;337:115953.
34. Salvi V, Ribuoli E, Servasi M, Orsolini L, Volpe U. ADHD and bipolar disorder in adulthood: clinical and treatment implications. *Medicina (Kaunas, Lithuania)*. 2021;57(5):466.
35. Aas M, Bellivier F, Bettella F, Henry C, Gard S, Kahn JP, et al. Childhood maltreatment and polygenic risk in bipolar disorders. *Bipolar Disorders*. 2020;22(2):174-81.
36. Park YM, Shekhtman T, Kelsoe JR. Interaction between adverse childhood experiences and polygenic risk in patients with bipolar disorder. *Translational Psychiatry*. 2020;10(1):326.
37. Schoeler T, Choi SW, Dudbridge F, Baldwin J, Duncan L, Cecil CM, et al. Multi-polygenic score approach to identifying individual vulnerabilities associated with the risk of exposure to bullying. *JAMA Psychiatry*. 2019;76(7):730-8.
38. Baldwin JR, Caspi A, Meehan AJ, Ambler A, Arseneault L, Fisher HL, et al. Population vs individual prediction of poor health from results of adverse childhood experiences screening. *JAMA Pediatrics*. 2021;175(4):385-93.
39. Yao K, van der Veen T, Thygesen J, Bass N, McQuillin A. Multiple psychiatric polygenic risk scores predict associations between childhood adversity and bipolar disorder. *Journal of Affective Disorders*. 2023;341:137-46.
40. Wang Y, Tsuo K, Kanai M, Neale BM, Martin AR. Challenges and opportunities for developing more generalizable polygenic risk scores. *Annual Review of Biomedical Data Science*. 2022;5:293-320.
41. Mitchell RE, Hartley AE, Walker VM, Gkatzionis A, Yarmolinsky J, Bell JA, et al. Strategies to investigate and mitigate collider bias in genetic and Mendelian randomisation studies of disease progression. *PLoS Genetics*. 2023;19(2):e1010596.
42. Coryell W, Solomon D, Turvey C, Keller M, Leon AC, Endicott J, et al. The long-term course of rapid-cycling bipolar disorder. *Archives of General Psychiatry*. 2003;60(9):914-20.
43. Kupka RW, Luckenbaugh DA, Post RM, Suppes T, Altshuler LL, Keck PE Jr, et al. Comparison of rapid-cycling and non-rapid-cycling bipolar disorder based on prospective mood ratings in 539 outpatients. *The American Journal of Psychiatry*. 2005;162(7):1273-80.
44. Du Rietz E, Coleman J, Glanville K, Choi SW, O'Reilly PF, Kuntsi J. Association of polygenic risk for attention-deficit/hyperactivity disorder with co-occurring traits and disorders. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. 2018;3(7):635-43.
45. Vannucchi G, Medda P, Pallucchini A, Bertelli M, Angst J, Azorin JM, et al. The relationship between attention deficit hyperactivity disorder, bipolarity and mixed features in major depressive patients: evidence from the BRIDGE-II-Mix Study. *Journal of Affective Disorders*. 2019;246:346-54.
46. Dunner DL, Fieve RR. Clinical factors in lithium carbonate prophylaxis failure. *Archives of General Psychiatry*. 1974;30(2):229-33.
47. van Hulzen KJE, Scholz CJ, Franke B, Ripke S, Klein M, McQuillin A, et al. Genetic overlap between attention-deficit/hyperactivity disorder and bipolar disorder: evidence from genome-wide association study meta-analysis. *Biological Psychiatry*. 2017;82(9):634-41.

48. Coombes BJ, Millischer V, Batzler A, Larrabee B, Hou L, Papiol S, et al. Association of attention-deficit/hyperactivity disorder and depression polygenic scores with lithium response: a Consortium for Lithium Genetics study. *Complex Psychiatry*. 2021;7(3-4):80-9.
49. Lin Y, Maihofer AX, Stapp E, Ritchey M, Alliey-Rodriguez N, Anand A, et al. Clinical predictors of non-response to lithium treatment in the Pharmacogenomics of Bipolar Disorder (PGBD) study. *Bipolar Disorders*. 2021;23(8):821-31.
50. Patel KHS, Walters GB, Stefánsson H, Stefánsson K, Degenhardt F, Nothen M, et al. Predicting ADHD in alcohol dependence using polygenic risk scores for ADHD. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics*. 2024;195(4):e32967.
51. Williams J, Farmer AE, Ackenheil M, Kaufmann CA, McGuffin P. A multicentre inter-rater reliability study using the OPCRIT computerized diagnostic system. *Psychological Medicine*. 1996;26(4):775-83.
52. Ge T, Chen CY, Ni Y, Feng YA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications*. 2019;10(1):1776.
53. O'Connell KS, Koromina M, van der Veen T, Boltz T, David FS, Yang JMK, et al. Genomics yields biological and phenotypic insights into bipolar disorder. *Nature*. 2025;639(8056):968-75.
54. Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience*. 2019;22(3):343-52.
55. Trubetskoy V, Pardiñas AF, Qi T, Panagiotaropoulou G, Awasthi S, Bigdeli TB, et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*. 2022;604(7906):502-8.
56. Purves KL, Coleman JRI, Meier SM, Rayner C, Davis KAS, Cheesman R, et al. A major role for common genetic variation in anxiety disorders. *Molecular Psychiatry*. 2020;25(12):3292-303.
57. Demontis D, Walters GB, Athanasiadis G, Walters R, Therrien K, Nielsen TT, et al. Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. *Nature Genetics*. 2023;55(2):198-208.
58. Robins JM. A new approach to causal inference in mortality studies with sustained exposure periods - applications to control of the healthy worker survivor effect. *Mathematical Modeling*. 1986;7:1393-512.
59. van der Wal WM, Geskus RB. ipw: an R package for inverse probability weighting. *Journal of Statistical Software*. 2011;43(13):1-23.
60. Olsson U. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*. 1979;44(4):443-60.
61. Fox J. polycor: Polychoric and Polyserial Correlations [Internet]. Version 0.8-1. 2022. Available from: https://CRAN.R-project.org/package=polycor [Accessed 2025 Jun 27].
62. Tierney N, Cook D. Expanding Tidy Data Principles to Facilitate Missing Data Exploration, Visualization and Assessment of Imputations. J Stat Softw. 2023;105(7):1-31. doi:10.18637/jss.v105.i07.
63. Harrison E, Drake T, Ots R. finalfit: Quickly Create Elegant Regression Results Tables and Plots when Modelling [Internet]. Version 1.0.6. 2023. Available from: https://CRAN.R-project.org/package=finalfit [Accessed 2025 Jun 27].
64. Enders CK. *Applied missing data analysis*. 2nd ed. New York: Guilford Press; 2022.
65. Kuhn M. Building Predictive Models in R Using the caret Package. J Stat Softw. 2008;28(5):1-26. doi:10.18637/jss.v028.i05.
66. Rosseel Y. lavaan: an R package for structural equation modeling. *Journal of Statistical Software*. 2012;48(2):1-36.

67. Revelle W. psych: Procedures for Psychological, Psychometric, and Personality Research [Internet]. Version 2.3.6. Evanston, Illinois: Northwestern University; 2023 [cited 2025 Jun 27]. Available from: https://CRAN.R-project.org/package=psych.
68. Lishinski A. lavaanPlot: Path Diagrams for 'Lavaan' Models via 'DiagrammeR' [Internet]. Version 0.8.1. 2024. Available from: https://CRAN.R-project.org/package=lavaanPlot [Accessed 2025 Jun 27].
69. R Core Team. R: A language and environment for statistical computing. [Computer software]. Version 4.4.2. Vienna, Austria: R Foundation for Statistical Computing; 2024.
70. Carey CE, Shafee R, Wedow R, Elliott A, Palmer DS, Compitello J, et al. Principled distillation of UK Biobank phenotype data reveals underlying structure in human variation. *Nature Human Behaviour*. 2024;8(8):1599-615.
71. Lee S, Tsang A, Kessler RC, Jin R, Sampson N, Andrade L, et al. Rapid-cycling bipolar disorder: cross-national community study. *The British Journal of Psychiatry*. 2010;196(3):217-25.
72. Jöreskog KG, Goldberger AS. Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*. 1975;70(351):631-9.
73. Agnew-Blais JC, Belsky DW, Caspi A, Danese A, Moffitt TE, Polanczyk GV, et al. Polygenic risk and the course of attention-deficit/hyperactivity disorder from childhood to young adulthood: findings From a nationally representative cohort. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2021;60(9):1147-56.
74. Duffy A. The nature of the association between childhood ADHD and the development of bipolar disorder: a review of prospective high-risk studies. *The American Journal of Psychiatry*. 2012;169(12):1247-55.
75. Birmaher B, Axelson D, Goldstein B, Monk K, Kalas C, Obreja M, et al. Psychiatric disorders in preschool offspring of parents with bipolar disorder: the Pittsburgh Bipolar Offspring Study (BIOS). *The American Journal of Psychiatry*. 2010;167(3):321-30.
76. McClelland H, Cleare S, O'Connor RC. Suicide risk in personality disorders: a systematic review. *Current Psychiatry Reports*. 2023;25(9):405-17.
77. McMahon FJ. Prediction of treatment outcomes in psychiatry--where do we stand? *Dialogues in Clinical Neuroscience*. 2014;16(4):455-64.
78. Krug A, Stein F, David FS, Schmitt S, Brosch K, Pfarr JK, et al. Factor analysis of lifetime psychopathology and its brain morphometric and genetic correlates in a transdiagnostic sample. *Translational Psychiatry*. 2024;14(1):235.
79. Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values. *Royal Society Open Science*. 2014;1(3):140216.
80. Oliva V, Fico G, De Prisco M, Gonda X, Rosa AR, Vieta E. Bipolar disorders: an update on critical aspects. *The Lancet Regional Health. Europe*. 2024;48:101135.
81. Meier SM, Pavlova B, Dalsgaard S, Nordentoft M, Mors O, Mortensen PB, et al. Attention-deficit hyperactivity disorder and anxiety disorders as precursors of bipolar disorder onset in adulthood. *The British Journal of Psychiatry*. 2018;213(3):555-60.
82. Nunez NA, Coombes BJ, Romo-Nava F, Bond DJ, Vande Voort J, Croarkin PE, et al. Clinical and genetic correlates of bipolar disorder with childhood-onset attention deficit disorder. *Frontiers in Psychiatry*. 2022;13:884217.
83. Guzman-Parra J, Streit F, Forstner AJ, Strohmaier J, Gonzalez MJ, Gil Flores S, et al. Clinical and genetic differences between bipolar disorder type 1 and 2 in multiplex families. *Translational Psychiatry*. 2021;11(1):31.
84. Miller CJ, Flory JD, Miller SR, Harty SC, Newcorn JH, Halperin JM. Childhood attention-deficit/hyperactivity disorder and the emergence of personality disorders in adolescence: a prospective follow-up study. *The Journal of Clinical Psychiatry*. 2008;69(9):1477-84.

8.4 CHAPTER 4

1. Lee PH, Anttila V, Won H, Feng YCA, Rosenthal J, Zhu Z, et al. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell*. 2019;179(7):1469-82.e11.
2. Grotzinger AD, Mallard TT, Akingbuwa WA, Ip HF, Adams MJ, Lewis CM, et al. Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic and molecular genetic levels of analysis. *Nature Genetics*. 2022;54(5):548-59.
3. Ruderfer DM, Ripke S, McQuillin A, Boocock J, Stahl EA, Pavlides JMW, et al. Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell*. 2018;173(7):1705-15.e16.
4. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511(7510):421-7. doi:10.1038/nature13595.
5. Kalman JL, Loohuis LMO, Vreeker A, McQuillin A, Stahl EA, Ruderfer D, et al. Characterisation of age and polarity at onset in bipolar disorder. *The British Journal of Psychiatry*. 2021;219(6):659-69.
6. Leonenko G, Di Florio A, Allardyce J, Forty L, Knott S, Jones L, et al. A data-driven investigation of relationships between bipolar psychotic symptoms and schizophrenia genome-wide significant genetic loci. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. 2018;177(4):468-75.
7. Calafato MS, Thygesen JH, Ranlund S, Zartaloudi E, Cahn W, Crespo-Facorro B, et al. Use of schizophrenia and bipolar disorder polygenic risk scores to identify psychotic disorders. *The British Journal of Psychiatry*. 2018;213(3):535-41.
8. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature Genetics*. 2019;51(5):793-803.
9. Kalman JL, Papiol S, Forstner AJ, Heilbronner U, Degenhardt F, Strohmaier J, et al. Investigating polygenic burden in age at disease onset in bipolar disorder: findings from an international multicentric study. *Bipolar Disorders*. 2019;21(1):68-75.
10. Allardyce J, Leonenko G, Hamshere M, Pardiñas AF, Forty L, Knott S, et al. Association between schizophrenia-related polygenic liability and the occurrence and level of mood-incongruent psychotic symptoms in bipolar disorder. *JAMA Psychiatry*. 2018;75(1):28-35.
11. Amare AT, Schubert KO, Hou L, Clark SR, Papiol S, Heilbronner U, et al. Association of polygenic score for schizophrenia and HLA antigen and inflammation genes with response to lithium in bipolar affective disorder: a genome-wide association study. *JAMA Psychiatry*. 2018;75(1):65-74.
12. Schubert KO, Thalamuthu A, Amare AT, Frank J, Streit F, Adl M, et al. Combining schizophrenia and depression polygenic risk scores improves the genetic prediction of lithium response in bipolar disorder patients. *Translational Psychiatry*. 2021;11(1):1-8.
13. Trubetskoy V, Pardiñas AF, Qi T, Panagiotaropoulou G, Awasthi S, Bigdeli TB, et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*. 2022;604(7906):502-8.
14. Nurnberger JI, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, et al. Diagnostic interview for genetic studies: rationale, unique features, and training. *Archives of General Psychiatry*. 1994;51(11):849-59.
15. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed. Washington, DC: American Psychiatric Association; 1994.

16. Endicott J, Spitzer RL. A diagnostic interview: the schedule for affective disorders and schizophrenia. *Archives of General Psychiatry*. 1978;35(7):837-44.
17. McGuffin P, Farmer A, Harvey I. A polydiagnostic application of operational criteria in studies of psychotic illness: development and reliability of the OPCRIT system. *Archives of General Psychiatry*. 1991;48(8):764-70.
18. Loh PR, Danecek P, Palamara PF, Fuchsberger C, Reshef YA, Finucane HK, et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nature Genetics*. 2016;48(11):1443-8.
19. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
20. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nature Genetics*. 2016;48(10):1284-7.
21. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*. 2006;38(8):904-9.
22. Grigoroiu-Serbanescu M, Giaroli G, Thygesen JH, Shenyan O, Bigdeli TB, Bass NJ, et al. Predictive power of the ADHD GWAS 2019 polygenic risk scores in independent samples of bipolar patients with childhood ADHD. *Journal of Affective Disorders*. 2020;265:651-9.
23. Euesden J, Lewis CM, O'Reilly PF. PRSice: polygenic risk score software. *Bioinformatics*. 2015;31(9):1466-8.
24. Choi SW, O'Reilly PF. PRSice-2: polygenic risk score software for biobank-scale data. *Gigascience*. 2019;8(7):giz082.
25. R Core Team. R: A language and environment for statistical computing [Computer software]. Version 4.4.2. Vienna, Austria: R Foundation for Statistical Computing; 2024.
26. Pain O, Gillett AC, Austin JC, Folkersen L, Lewis CM. A tool for translating polygenic scores onto the absolute scale using summary statistics. *European Journal of Human Genetics*. 2022;30(3):339-48.
27. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications*. 2019;10(1):1-10.
28. Pain O, Glanville KP, Hagenaars SP, Selzam S, Furtjes AE, Gaspar HA, et al. Evaluation of polygenic prediction methodology within a reference-standardized framework. *PLoS Genetics*. 2021;17(5):e1009021.
29. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015 Dec;4(1):s13742-015.
30. Choi SW, García-González J, Ruan Y, Wu HM, Porras C, Johnson J, et al. PRSet: pathway-based polygenic risk score analyses and software. *PLoS Genetics*. 2023;19(2):e1010624.
31. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Computational Biology*. 2015;11(4):e1004219.
32. Joo JWJ, Hormozdiari F, Han B, Eskin E. Multiple testing correction in linear mixed models. *Genome Biology*. 2016;17:1-18.
33. Lee SH, Goddard ME, Wray NR, Visscher PM. A better coefficient of determination for genetic profile analysis. *Genetic Epidemiology*. 2012;36(3):214-24.
34. Merikangas KR, Jin R, He JP, Kessler RC, Lee S, Sampson NA, et al. Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Archives of General Psychiatry*. 2011;68(3):241-51.
35. Humpston CS, Bebbington P, Marwaha S. Bipolar disorder: prevalence, help-seeking and use of mental health care in England. Findings from the 2014 Adult Psychiatric Morbidity Survey. *Journal of Affective Disorders*. 2021;282:426-33.

36. De Hond AA, Steyerberg EW, Van Calster B. Interpreting area under the receiver operating characteristic curve. *The Lancet Digital Health*. 2022;4(12):e853-5.
37. McMahon FJ. Prediction of treatment outcomes in psychiatry—where do we stand? *Dialogues in Clinical Neuroscience*. 2014;16(4):455-64.
38. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1995;57(1):289-300.
39. Aminoff SR, Onyeka IN, Ødegaard M, Simonsen C, Lagerberg TV, Andreassen OA, et al. Lifetime and point prevalence of psychotic symptoms in adults with bipolar disorders: a systematic review and meta-analysis. *Psychological Medicine*. 2022;52(13):2413-25.
40. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Medicine*. 2020;12(1):1-11.
41. Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, Qiao Z, et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nature Genetics*. 2021;53(6):817-29.
42. Landi I, Kaji DA, Cotter L, Van Vleck T, Belbin G, Preuss M, et al. Prognostic value of polygenic risk scores for adults with psychosis. *Nature Medicine*. 2021;27(9):1576-81.
43. Richards AL, Cardno A, Harold G, Craddock NJ, Di Florio A, Jones L, et al. Genetic liabilities differentiating bipolar disorder, schizophrenia, and major depressive disorder, and phenotypic heterogeneity in bipolar disorder. *JAMA Psychiatry*. 2022;79(10):1032-9.
44. Allardyce J, Cardno AG, Gordon-Smith K, Jones L, Di Florio A, Walters JTR, et al. Specificity of polygenic signatures across symptom dimensions in bipolar disorder: an analysis of UK Bipolar Disorder Research Network data. *The Lancet Psychiatry*. 2023;10(8):623-31.
45. Ruderfer DM, Ripke S, McQuillin A, Boocock J, Stahl EA, Pavlides JM, Mullins N, Charney AW, Ori AP, Loohuis LM, Domenici E. Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. Cell. 2018 Jun 14;173(7):1705-15.
46. Harder A, Nguyen TD, Pasman JA, Mosing MA, Hägg S, Lu Y. Genetics of age-at-onset in major depression. *Translational Psychiatry*. 2022;12(1):124.
47. Lewis CM, Vassos E. Polygenic scores in psychiatry: on the road from discovery to implementation. *The American Journal of Psychiatry*. 2022;179(11):800-6.
48. Pedersen EM, Agerbo E, Plana-Ripoll O, Grove J, Dreier JW, Musliner KL, et al. Accounting for age of onset and family history improves power in genome-wide association studies. *The American Journal of Human Genetics*. 2022;109(3):417-32.
49. Jarskog LF, Glantz LA, Gilmore JH, Lieberman JA. Apoptotic mechanisms in the pathophysiology of schizophrenia. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*. 2005;29(5):846-58.
50. Garey L, Ong W, Patel T, Kanani M, Davis A, Mortimer A, et al. Reduced dendritic spine density on cerebral cortical pyramidal neurons in schizophrenia. *Journal of Neurology, Neurosurgery & Psychiatry*. 1998;65(4):446-53.
51. Ramaker RC, Bowling KM, Lasseigne BN, Hagenauer MH, Hardigan AA, Davis NS, et al. Post-mortem molecular profiling of three psychiatric disorders. *Genome Medicine*. 2017;9:1-12.
52. Whitehurst T, Howes O. The role of mitochondria in the pathophysiology of schizophrenia: a critical review of the evidence focusing on mitochondrial complex one. *Neuroscience & Biobehavioral Reviews*. 2022;132:449-64.
53. Andreazza AC, Shao L, Wang JF, Young LT. Mitochondrial complex I activity and oxidative damage to mitochondrial proteins in the prefrontal cortex of patients with bipolar disorder. *Archives of General Psychiatry*. 2010;67(4):360-8.

54. Xiao X, Zhang C, Grigoroiu-Serbanescu M, Wang L, Li L, Zhou D, et al. The cAMP responsive element-binding (CREB)-1 gene increases risk of major psychiatric disorders. *Molecular Psychiatry*. 2018;23(9):1957-67.
55. Thomas KT, Anderson BR, Shah N, Zimmer SE, Hawkins D, Valdez AN, et al. Inhibition of the schizophrenia-associated microRNA miR-137 disrupts Nrg1α neurodevelopmental signal transduction. *Cell Reports*. 2017;20(1):1-12.
56. Chen BY, Lin JJ, Lu MK, Tan HP, Jang FL, Lin SH. Neurodevelopment regulators miR-137 and miR-34 family as biomarkers for early and adult onset schizophrenia. *npj Schizophrenia*. 2021;7(1):35.
57. Nurnberger JI, Koller DL, Jung J, Edenberg HJ, Foroud T, Guella I, et al. Identification of pathways for bipolar disorder: a meta-analysis. *JAMA Psychiatry*. 2014;71(6):657-64.
58. Benjamin KJM, Chen Q, Jaffe AE, Stolz JM, Collado-Torres L, Huuki-Myers LA, et al. Analysis of the caudate nucleus transcriptome in individuals with schizophrenia highlights effects of antipsychotics and new risk genes. *Nature Neuroscience*. 2022;25(11):1559-68.
59. Perry BI, Zammit S, Jones PB, Khandaker GM. Childhood inflammatory markers and risks for psychosis and depression at age 24: examination of temporality and specificity of association in a population-based prospective birth cohort. *Schizophrenia Research*. 2021;230:69-76.
60. Grigoroiu-Serbanescu M, Diaconu CC, Heilmann-Heimbach S, Neagu AI, Becker T. Association of age-of-onset groups with GWAS significant schizophrenia and bipolar disorder loci in Romanian bipolar I patients. *Psychiatry Research*. 2015;230(3):964-7.
61. Deans PM, Seah C, Johnson J, Gonzalez JG, Townsley K, Cao E, et al. Non-additive effects of schizophrenia risk genes reflect convergent downstream function. *medRxiv*. 2023.
62. Schrode N, Ho SM, Yamamuro K, Dobbyn A, Huckins L, Matos MR, et al. Synergistic effects of common schizophrenia risk variants. *Nature Genetics*. 2019;51(10):1475-85.
63. Kocsis A, Gajwani R, Gross J, Gumley AI, Lawrie SM, Schwannauer M, et al. Altered autonomic function in individuals at clinical high risk for psychosis. *Frontiers in Psychiatry*. 2020;11:580503.
64. Montaquila JM, Trachik BJ, Bedwell JS. Heart rate variability and vagal tone in schizophrenia: a review. *Journal of Psychiatric Research*. 2015;69:57-66.
65. Hannon E, Dempster E, Viana J, Burrage J, Smith AR, Macdonald R, et al. An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. *Genome Biology*. 2016;17:1-16.
66. Hannon E, Dempster EL, Mansell G, Burrage J, Bass N, Bohlken MM, et al. DNA methylation meta-analysis reveals cellular alterations in psychosis and markers of treatment-resistant schizophrenia. *eLife*. 2021;10:e58430.
67. Srivastava A, Chaudhary Z, Qian J, Al Chalabi N, Burhan AM, Fischer CE, et al. Genome-wide methylation analysis of early-onset schizophrenia. *Psychiatric Genetics*. 2022;32(6):214-20.
68. Grigoroiu-Serbanescu M, Wickramaratne PJ, Mihailescu R, Prelipceanu D, Sima D, Codreanu M, et al. Paternal age effect on age of onset in bipolar I disorder is mediated by sex and family history. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. 2012;159(5):567-79.
69. Zhang JP, Robinson D, Yu J, Gallego J, Fleischhacker WW, Kahn RS, et al. Schizophrenia polygenic risk score as a predictor of antipsychotic efficacy in first-episode psychosis. *American Journal of Psychiatry*. 2019;176(1):21-8.
70. Amare AT, Schubert KO, Hou L, Clark SR, Papiol S, Cearns M, Heilbronner U, Degenhardt F, Tekola-Ayele F, Hsu YH, Shekhtman T. Association of polygenic score for major depression with response to lithium in patients with bipolar disorder. Molecular psychiatry. 2021 Jun;26(6):2457-70.

71. Grigoroiu-Serbanescu M, van der Veen T, Bigdeli T, Herms S, Diaconu CC, Neagu AI, et al. Schizophrenia polygenic risk scores, clinical variables and genetic pathways as predictors of phenotypic traits of bipolar I disorder. *Journal of Affective Disorders*. 2024;356:507-18.

8.5 CHAPTER 5

Oliva V, Fico G, De Prisco M, Gonda X, Rosa AR, Vieta E. Bipolar disorders: an update on critical aspects. *The Lancet Regional Health - Europe*. 2024;48:101135.

2. O'Connell KS, Koromina M, van der Veen T, Boltz T, David FS, Yang JMK, et al. Genomics yields biological and phenotypic insights into bipolar disorder. *Nature*. 2025;639(8056):968-75.
3. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature Genetics*. 2019;51(5):793-803.
4. McGuffin P, Rijsdijk F, Andrew M, Sham P, Katz R, Cardno A. The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Archives of General Psychiatry*. 2003;60(5):497-502.
5. Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, Qiao Z, et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nature Genetics*. 2021;53(6):817-29.
6. Lawrence JM, Breunig S, Foote IF, Tallis CB, Grotzinger AD. Genomic SEM applied to explore etiological divergences in bipolar subtypes. *Psychological Medicine*. 2024;54(6):1152-9.
7. Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell*. 2018;173(7):1705-15.e16.
8. Richards AL, Cardno A, Harold G, Di Florio A, Gordon-Smith K, Jones L, et al. Genetic liabilities differentiating bipolar disorder, schizophrenia, and major depressive disorder, and phenotypic heterogeneity in bipolar disorder. *JAMA Psychiatry*. 2022;79(10):1032-9.
9. Grigoroiu-Serbanescu M, Giaroli G, Thygesen JH, Shenyan O, Bigdeli TB, Bass NJ, et al. Predictive power of the ADHD GWAS 2019 polygenic risk scores in independent samples of bipolar patients with childhood ADHD. *Journal of Affective Disorders*. 2020;265:651-9.
10. Geddes JR, Miklowitz DJ. Treatment of bipolar disorder. *The Lancet*. 2013;381(9878):1672-82.
11. Kalman JL, Olde Loohuis LM, Vreeker A, McQuillin A, Stahl EA, Ruderfer D, et al. Characterisation of age and polarity at onset in bipolar disorder. *The British Journal of Psychiatry*. 2021;219(6):659-69.
12. Lefrere A, Godin O, Jamain S, Gard S, Leboyer M, Etain B, et al. Refining criteria for a neurodevelopmental subphenotype of bipolar disorders. *Biological Psychiatry*. 2025;97(8):806-15.
13. Lin PI, McInnis MG, Potash JB, Willour V, MacKinnon DF, DePaulo JR, et al. Clinical correlates and familial aggregation of age at onset in bipolar disorder. *American Journal of Psychiatry*. 2006;163(2):240-6.
14. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: American Psychiatric Association; 1994.
15. Saunders EH, Scott LJ, McInnis MG, Burmeister M. Familiality and diagnostic patterns of subphenotypes in the National Institutes of Mental Health bipolar sample. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. 2008;147B(1):18-26.

16. Dunner DL, Fieve RR. Clinical factors in lithium carbonate prophylaxis failure. *Archives of General Psychiatry*. 1974;30(2):229-33.
17. Miola A, Fountoulakis KN, Baldessarini RJ, Veldic M, Solmi M, Rasgon N, et al. Prevalence and outcomes of rapid cycling bipolar disorder: mixed method systematic meta-review. *Journal of Psychiatric Research*. 2023;164:404-15.
18. Bauer M, Whybrow PC. Thyroid hormone, neural tissue and mood modulation. *World Journal of Biological Psychiatry*. 2001;2(2):59-69.
19. Bauer M, Goetz T, Glenn T, Whybrow PC. The thyroid-brain interaction in thyroid disorders and mood disorders. *Journal of Neuroendocrinology*. 2008;20(10):1101-14.
20. Walshaw PD, Gyulai L, Bauer M, Bauer MS, Calimlim B, Sugar CA, et al. Adjunctive thyroid hormone treatment in rapid cycling bipolar disorder: a double-blind placebo-controlled trial of levothyroxine (L-T4) and triiodothyronine (T3). *Bipolar Disorders*. 2018;20(7):594-603.
21. World Health Organization. *International Statistical Classification of Diseases and Related Health Problems*. 10th ed. Geneva: World Health Organization; 2010.
22. Lam M, Awasthi S, Watson HJ, Goldstein J, Panagiotaropoulou G, Trubetskoy V, et al. RICOPILI: rapid imputation for consortias pipeline. *Bioinformatics*. 2020;36(3):930-3.
23. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*. 2016;48(10):1279-83.
24. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*. 2015;47(3):291-5.
25. Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nature Genetics*. 2018;50(2):229-37.
26. Trubetskoy V, Pardiñas AF, Qi T, Panagiotaropoulou G, Awasthi S, Bigdeli TB, et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*. 2022;604(7906):502-8.
27. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*. 2017;8(1):1826.
28. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Computational Biology*. 2015;11(4):e1004219.
29. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739-40.
30. Watanabe K, Umićević Mirkov M, de Leeuw CA, van den Heuvel MP, Posthuma D. Genetic mapping of cell type specificity for complex traits. *Nature Communications*. 2019;10(1):3222.
31. Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FCP, et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science*. 2018;362(6420):eaat8464.
32. Hodge RD, Bakken TE, Miller JA, Smith KA, Barkan ER, Graybuck LT, et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature*. 2019;573(7772):61-8.
33. Habib N, Avraham-Davidi I, Basu A, Burks T, Kalamzas S, Grijalvo S, et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nature Methods*. 2017;14(10):955-8.
34. La Manno G, Gyllborg D, Codeluppi S, Nishimura K, Salto C, Zeisel A, et al. Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell*. 2016;167(2):566-80.e19.
35. Hochgerner H, Lönnerberg P, Hodge R, Rati M, Murray E, Miches A, et al. STRT-seq-2i: dual-index 5' single cell and nucleus RNA-seq on an addressable microwell array. *Scientific Reports*. 2017;7(1):16327.
36. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics*. 2016;48(3):245-52.

37. de Leeuw C, Werme J, Savage JE, Peyrot WJ, Posthuma D. On the interpretation of transcriptome-wide association studies. *PLoS Genetics*. 2023;19(9):e1010921.
38. Werme J, van der Sluis S, Posthuma D, de Leeuw CA. An integrated framework for local genetic correlation analysis. *Nature Genetics*. 2022;54(3):274-82.
39. Ge T, Chen CY, Ni Y, Feng YA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications*. 2019;10(1):1776.
40. Nurnberger J Jr, Roose SP, Dunner DL, Fieve RR. Unipolar mania: a distinct clinical entity? *American Journal of Psychiatry*. 1979;136(11):1420-3.
41. Mehta S. Unipolar mania: recent updates and review of the literature. *Psychiatry Journal*. 2014;2014:261943.
42. Manchia M, Miola A, Tondo L, Baldessarini RJ. Unipolar Mania: Prevalence and Patient Characteristics. Acta Psychiatrica Scandinavica. 2025 Jun;151(6):680-8.
43. Mallard TT, Linnér RK, Grotzinger AD, et al. Multivariate GWAS of psychiatric disorders and their cardinal symptoms reveal two dimensions of cross-cutting genetic liabilities. *Cell Genom*. 2022;2(6):100140.
44. Grotzinger AD, Werme J, Peyrot WJ, Kunkle ED, de Vlaming R, Mallard TT, et al. The landscape of shared and divergent genetic influences across 14 psychiatric disorders. *medRxiv*. 2025 Jan 14 [Preprint].
45. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*. 2014;46(3):310-5.
46. Song J, Bergen SE, Di Florio A, Karlsson R, Charney A, Ruderfer DM, et al. Genome-wide association study identifies SESTD1 as a novel risk gene for lithium-responsive bipolar disorder. *Molecular Psychiatry*. 2016;21(9):1290-7.
47. International Consortium on Lithium Genetics (ConLi+Gen), Amare AT, Schubert KO, Hou L, Clark SR, Papiol S, et al. Association of polygenic score for schizophrenia and HLA antigen and inflammation genes with response to lithium in bipolar affective disorder: a genome-wide association study. *JAMA Psychiatry*. 2018;75(1):65-74.
48. McQuillin A, Yao K, Nadeem A, van der Veen T, Thygesen J, Jonsson L, et al. Implication of the ADCY1 gene in lithium response in bipolar disorder by genome-wide association meta-analysis. *Research Square*. 2024 Jan 8 [Preprint].
49. Osete JR, Akkouh IA, Ievglevskyi O, Sæther S, Bjørnstad P, O'Connell KS, et al. Transcriptional and functional effects of lithium in bipolar disorder iPSC-derived cortical spheroids. *Molecular Psychiatry*. 2023;28(7):3033-43.
50. Alonso-Merino E, Martín Orozco R, Ruíz-Llorente L, Tovar SA, García-Giménez C, Obregon MJ, et al. Thyroid hormones inhibit TGF-β signaling and attenuate fibrotic responses. *Proceedings of the National Academy of Sciences of the United States of America*. 2016;113(24):E3451-60.
51. Forstner AJ, Awasthi S, Wolf C, Maron E, Erhardt A, Czamara D, et al. Genome-wide association study of panic disorder reveals genetic overlap with neuroticism and depression. *Molecular Psychiatry*. 2021;26(8):4179-90.
52. Gregersen NO, Lescai F, Liang J, Biskopstø M, Egholm M, Wang AG, et al. Whole-exome sequencing implicates DGKH as a risk gene for panic disorder in the Faroese population. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. 2016;171(8):1013-22.
53. Baum AE, Akula N, Cabanero M, Cardona I, Corona W, Klemens B, et al. A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. *Molecular Psychiatry*. 2008;13(2):197-207.

54. Grotzinger AD, Mallard TT, Akingbuwa WA, Ip HF, Adams MJ, Lewis CM, et al. Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic and molecular genetic levels of analysis. *Nature Genetics*. 2022;54(5):548-59.
55. Koromina M, Ravi A, Panagiotaropoulou G, O'Connell KS, Tsetsos F, Savage JE, et al. Fine-mapping genomic loci refines bipolar disorder risk genes. *medRxiv*. 2024;2024.02.12.24302716.
56. Eriksson PS, Perfilieva E, Björk-Eriksson T, Alborn AM, Nordborg C, Peterson DA, et al. Neurogenesis in the adult human hippocampus. *Nature Medicine*. 1998;4(11):1313-7.
57. Nebert DW, Liu Z. SLC39A8 gene encoding a metal ion transporter: discovery and bench to bedside. *Human Genomics*. 2019;13(Suppl 1):51.
58. Mealer RG, Jenkins BG, Chen CY, Faundez V, Cubillos C, Shaw AD, et al. The schizophrenia risk locus in SLC39A8 alters brain metal transport and plasma glycosylation. *Scientific Reports*. 2020;10(1):13162.
59. Mealer RG, Williams SE, Noel M, Lencer R, Cutler DJ, Faundez V, et al. The schizophrenia-associated variant in SLC39A8 alters protein glycosylation in the mouse brain. *Molecular Psychiatry*. 2022;27(3):1405-15.
60. Witt SH, Streit F, Jungkunz M, Frank J, Awasthi S, Reinbold CS, et al. Genome-wide association study of borderline personality disorder reveals genetic overlap with bipolar disorder, major depression and schizophrenia. *Translational Psychiatry*. 2017;7(6):e1155.
61. Streit F, Awasthi S, Hall ASM, Niarchou M, Marouli E, Babajide O, et al. Genome-wide association study of borderline personality disorder identifies six loci and highlights shared risk with mental and somatic disorders. *medRxiv*. 2024 Nov 12 [Preprint].
62. Goes FS, Hamshere ML, Seifuddin F, Pirooznia M, Belmonte-de-Abreu P, Medeiros H, et al. Genome-wide association of mood-incongruent psychotic bipolar disorder. *Translational Psychiatry*. 2012;2(10):e180.
63. McMahon FJ, Akula N, Schulze TG, Muglia P, Tozzi F, Detera-Wadleigh SD, et al. Meta-analysis of genome-wide association data identifies a risk locus for major mood disorders on 3p21.1. *Nature Genetics*. 2010;42(2):128-31.
64. Wilson RI, Nicoll RA. Endogenous cannabinoids mediate retrograde signalling at hippocampal synapses. *Nature*. 2001;410(6828):588-92.
65. Tao R, Li C, Jaffe AE, Madison JM, Pochareddy S, Shin JH, et al. Cannabinoid receptor CNR1 expression and DNA methylation in human prefrontal cortex, hippocampus and caudate in brain development and schizophrenia. *Translational Psychiatry*. 2020;10(1):158.
66. Chang H, Li L, Peng T, Cui K, Li X, Li M, et al. Identification of a bipolar disorder vulnerable gene CHDH at 3p21.1. *Molecular Neurobiology*. 2017;54(7):5166-76.
67. Lüscher B, Shen Q, Sahir N. The GABAergic deficit hypothesis of major depressive disorder. *Molecular Psychiatry*. 2011;16(4):383-406.
68. Poels EM, Kegeles LS, Kantrowitz JT, Slifstein M, Javitt DC, Lieberman JA, et al. Imaging glutamate in schizophrenia: review of findings and implications for drug discovery. *Molecular Psychiatry*. 2014;19(1):20-9.
69. O'Connell KS, Sønderby IE, Frei O, van der Meer D, Athanasiu L, Schork AJ, et al. Association between complement component 4A expression, cognitive performance and brain imaging measures in UK Biobank. *Psychological Medicine*. 2021 Mar 3 [Epub ahead of print].
70. Borbye-Lorenzen N, Zhu Z, Agerbo E, Brikell I, Li Z, Pedersen CB, et al. The correlates of neonatal complement component 3 and 4 protein concentrations with a focus on psychiatric and autoimmune disorders. *Cell Genomics*. 2023;3(12):100457.
71. Hörbeck E, Jonsson L, Malwade S, Karlsson R, Pålsson E, Sigström R, et al. Dissecting the impact of complement component 4A in bipolar disorder. *Brain, Behavior, and Immunity*. 2024;116:150-9.

72. Zeng J, Xue A, Jiang L, Wu Y, Campos AI, Tredget Z, et al. Widespread signatures of natural selection across human complex traits and functional genomic categories. *Nature Communications*. 2021;12(1):1164.

8.6 CHAPTER 6

McIntyre RS, Berk M, Brietzke E, Goldstein BI, López-Jaramillo C, Kessing LV, et al. Bipolar disorders. *The Lancet (London, England)*. 2020;396(10265):1841-56.

2. Aghababaie-Babaki P, Malekpour MR, Mohammadi E, Saeedi Moghaddam S, Rashidi MM, Ghanbari A, et al. Global, regional, and national burden and quality of care index (QCI) of bipolar disorder: a systematic analysis of the Global Burden of Disease Study 1990 to 2019. *International Journal of Social Psychiatry*. 2023;69(8):1958-70.
3. Carvalho AF, Firth J, Vieta E. Bipolar disorder. *New England Journal of Medicine*. 2020;383(1):58-66.
4. Geddes JR, Miklowitz DJ. Treatment of bipolar disorder. *The Lancet*. 2013;381(9878):1672-82.
5. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders, (DSM-5®)*. 5th ed. Washington, DC; 2013.
6. Reed GM, First MB, Kogan CS, Hyman SE, Gureje O, Gaebel W, et al. Innovations and changes in the ICD-11 classification of mental, behavioural and neurodevelopmental disorders. *World Psychiatry*. 2019;18(1):3-19.
7. Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, Qiao Z, et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nature Genetics*. 2021;53(6):817-29.
8. Sklar P, Ripke S, Scott LJ, Andreassen OA, Cichon S, Craddock N, et al. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genetics*. 2011;43(10):977-85.
9. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature Genetics*. 2019;51(5):793-803.
10. Lam M, Awasthi S, Watson HJ, Goldstein J, Panagiotaropoulou G, Trubetskoy V, et al. RICOPILI: rapid imputation for consortias pipeline. *Bioinformatics*. 2020;36(3):930-3.
11. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*. 2016;48(10):1279-83.
12. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190-1.
13. Chen W, Wu Y, Zheng Z, Qi T, Visscher PM, Zhu Z, et al. Improved analyses of GWAS summary statistics by reducing data heterogeneity and errors. *Nature Communications*. 2021;12(1):7117.
14. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics*. 2011;88(1):76-82.
15. Yang J, Ferreira T, Morris AP, Medland SE, Madden PA, Heath AC, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics*. 2012;44(4):369-75, S1-3.

16. Demontis D, Walters GB, Athanasiadis G, Walters R, Therrien K, Nielsen TT, et al. Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. *Nature Genetics*. 2023;55(2):198-208.
17. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications*. 2019;10(1):1-10.
18. R Core Team. R: A language and environment for statistical computing [Computer software]. Version 4.4.2. Vienna, Austria: R Foundation for Statistical Computing; 2024.
19. Lee SH, Goddard ME, Wray NR, Visscher PM. A better coefficient of determination for genetic profile analysis. *Genetic Epidemiology*. 2012;36(3):214-24.
20. McGrath JJ, Al-Hamzawi A, Alonso J, Altwaijri Y, Andrade LH, Bromet EJ, et al. Age of onset and cumulative risk of mental disorders: a cross-national analysis of population surveys from 29 countries. *The Lancet Psychiatry*. 2023;10(9):668-81.
21. O'Connell KS, Koromina M, van der Veen T, Boltz T, David FS, Yang JMK, et al. Genomics yields biological and phenotypic insights into bipolar disorder. *Nature*. 2025;639(8056):968-75.
22. Merikangas KR, Jin R, He JP, Kessler RC, Lee S, Sampson NA, et al. Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Archives of General Psychiatry*. 2011;68(3):241-51.
23. Jonsson L, Song J, Joas E, Pålsson E, Landén M. Polygenic scores for psychiatric disorders associate with year of first bipolar disorder diagnosis: a register-based study between 1972 and 2016. *Psychiatry Research*. 2024;339:116081.
24. Zimmerman M, Ruggero CJ, Chelminski I, Young D. Is bipolar disorder overdiagnosed? *The Journal of Clinical Psychiatry*. 2008;69(6):935-40.
25. Zimmerman M, Ruggero CJ, Chelminski I, Young D. Psychiatric diagnoses in patients previously overdiagnosed with bipolar disorder. *The Journal of Clinical Psychiatry*. 2010;71(1):26-31.
26. Bryc K, Durand EY, Macpherson JM, Reich D, Mountain JL. The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *American Journal of Human Genetics*. 2015;96(1):37-53.
27. Duncan L, Shen H, Gelaye B, Meijsen J, Ressler K, Feldman M, et al. Analysis of polygenic risk score usage and performance in diverse human populations. *Nature Communications*. 2019;10(1):1-9.
28. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Medicine*. 2020;12(1):1-11.
29. Wang X, Walker A, Revez JA, Ni G, Adams MJ, McIntosh AM, Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM. Polygenic risk prediction: why and when out-of-sample prediction R2 can exceed SNP-based heritability. The American Journal of Human Genetics. 2023 Jul 6;110(7):1207-15.

8.7 CHAPTER 7

1. Wang X, Walker A, Revez JA, et al. Polygenic risk prediction: why and when out-of-sample prediction R2 can exceed SNP-based heritability. *The American Journal of Human Genetics*. 2023;110:1207-15.
2. Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nature Genetics*. 2018;50(2):229-37.

3. Mullins N, Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, Qiao Z, et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nature Genetics*. 2021;53(6):817-29.
4. O'Connell KS, Koromina M, van der Veen T, Boltz T, David FS, Yang JM, Lin KH, Wang X, Coleman JR, Mitchell BL, McGrouther CC. Genomics yields biological and phenotypic insights into bipolar disorder. *Nature*. 2025 Jan 22:1-2.
5. Meng X, Navoly G, Giannakopoulou O, Levey DF, Koller D, Pathak GA, et al. Multi-ancestry genome-wide association study of major depression aids locus discovery, fine mapping, gene prioritization and causal inference. *Nature Genetics*. 2024;56(2):222-33.
6. Adams MJ, Streit F, Meng X, Awasthi S, Adey BN, Choi KW, Chundru VK, Coleman JR, Ferwerda B, Foo JC, Gerring ZF. Trans-ancestry genome-wide study of depression identifies 697 associations implicating cell types and pharmacotherapies. *Cell*. 2025 Feb 6;188(3):640-52.
7. Peterson RE, Kuchenbaecker K, Walters RK, Chen CY, Popejoy AB, Periyasamy S, et al. Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. *Cell*. 2019;179(3):589-603.
8. Ruan Y, Lin YF, Feng YA, Chen CY, Lam M, Guo Z, et al. Improving polygenic prediction in ancestrally diverse populations. *Nature Genetics*. 2022;54(5):573-80.
9. Zimmerman M, Ruggero CJ, Chelminski I, Young D. Is bipolar disorder overdiagnosed? *The Journal of Clinical Psychiatry*. 2008;69(6):935-40.
10. Zimmerman M, Ruggero CJ, Chelminski I, Young D. Psychiatric diagnoses in patients previously overdiagnosed with bipolar disorder. *The Journal of Clinical Psychiatry*. 2010;71(1):26-31.
11. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics*. 2014;23(R1):R89-98.
12. Denault WRP, Bohlin J, Page CM, Burgess S, Jugessur A. Cross-fitted instrument: a blueprint for one-sample Mendelian randomization. *PLoS Computational Biology*. 2022;18(8):e1010268.
13. Yeung SL, Luo S, Iwagami M, Goto A. Introduction to Mendelian randomization. *Annals of Clinical Epidemiology*. 2025 Jan 1;7(1):27-37.
14. Chen J, Duan W, Liu P, Long C, Li A, Zhang X, et al. Schizophrenia, bipolar disorder and major depressive disorder are probably not risk factors for cardiovascular disease: a Mendelian randomized study. *Journal of Affective Disorders*. 2025;377:184-96.
15. Lu T, Chen Y, Yoshiji S, Ilboudo Y, Forgetta V, Zhou S, et al. Circulating metabolite abundances associated with risks of bipolar disorder, schizophrenia, and depression: a Mendelian randomization study. *Biological Psychiatry*. 2024;96(10):782-91.
16. Stacey D, Benyamin B, Lee SH, Hyppönen E. A metabolome-wide Mendelian randomization study identifies dysregulated arachidonic acid synthesis as a potential causal risk factor for bipolar disorder. *Biological Psychiatry*. 2024;96(6):455-62.
17. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460(7256):748-52.
18. Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Statistics in medicine*. 2017 May 20;36(11):1783-802.
19. Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *International Journal of Epidemiology*. 2019;48(3):713-27.

20. Sanderson E, Smith GD, Windmeijer F, Bowden J. Corrigendum to: an examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *International Journal of Epidemiology*. 2020;49(3):1057.
21. Iakunchykova O, Leonardsen EH, Wang Y. Genetic evidence for causal effects of immune dysfunction in psychiatric disorders: where are we? *Translational Psychiatry*. 2024;14(1):63.
22. Shi R, Morrison J. Multivariable Mendelian randomization adjusting for heritable confounding analyzes the causal effects of C-reactive protein on multiple diseases. *bioRxiv*. 2025 Jan 8 [Preprint].
23. Chesnaye NC, Stel VS, Tripepi G, Dekker FW, Fu EL, Zoccali C. An introduction to inverse probability of treatment weighting in observational research. *Clinical Kidney Journal*. 2022;15(1):14-20.
24. Skelly AC, Dettori JR, Brodt ED. Assessing bias: the importance of considering confounding. *Evidence-Based Spine-Care Journal*. 2012;3(1):9-12.
25. Mitchell RE, Hartley AE, Walker VM, Gkatzionis A, Yarmolinsky J, Bell JA, et al. Strategies to investigate and mitigate collider bias in genetic and Mendelian randomisation studies of disease progression. *PLoS Genetics*. 2023;19(2):e1010596.
26. Smeland OB, Bahrami S, Frei O, Shadrin A, O'Connell K, Savage J, et al. Genome-wide analysis reveals extensive genetic overlap between schizophrenia, bipolar disorder, and intelligence. *Molecular Psychiatry*. 2020;25(4):844-53.
27. Birmaher B, Merranko J, Hafeman D, Goldstein BI, Diler R, Levenson JC, et al. A longitudinal study of psychiatric disorders in offspring of parents with bipolar disorder from preschool to adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2021;60(11):1419-29.
28. Mahmoud O, Dudbridge F, Davey Smith G, Munafo M, Tilling K. A robust method for collider bias correction in conditional genome-wide association studies. *Nature Communications*. 2022;13(1):619.
29. Mallard TT, Karlsson Linnér R, Grotzinger AD, Sanchez-Roige S, Seok M, Karcher MH, et al. Multivariate GWAS of psychiatric disorders and their cardinal symptoms reveal two dimensions of cross-cutting genetic liabilities. *Cell Genomics*. 2022;2(6):100140.
30. Conomos MP, Nelson SC, Choudhury A, Conti D, Fullerton SM, Gogarten SM, et al. The PRIMED Consortium: reducing disparities in polygenic risk assessment. *American Journal of Human Genetics*. 2024;111(12):2594-606.
31. Vassos E, Collier DA, Holden S, Patch C, Rujescu D, St Clair D, et al. Penetrance for copy number variants associated with schizophrenia. *Human Molecular Genetics*. 2010;19(17):3477-81.
32. Bergen SE, Ploner A, Howrigan D, O'Donovan MC, Smoller JW, Sullivan PF, et al. Joint contributions of rare copy number variants and common SNPs to risk for schizophrenia. *The American Journal of Psychiatry*. 2019;176(1):29-35.
33. McGrouther CC, Rangan AV, Di Florio A, Elman JA, Schork NJ, Kelsoe J. Heterogeneity analysis provides evidence for a genetically homogeneous subtype of bipolar-disorder. *PLoS One*. 2025;20(1):e0314288.
34. Laksshman S, Bhat RR, Viswanath V, Li X. DeepBipolar: Identifying genomic mutations for bipolar disorder via deep learning. *Human mutation*. 2017 Sep;38(9):1217-24.
35. Ozdemir OB, Chen R, Li R. A deep ensemble encoder network method for improved polygenic risk score prediction. *medRxiv*. 2024 Jul 31 [Preprint].
36. Williams J, Chen T, Hua X, Wong W, Yu K, Kraft P, et al. Integrating common and rare variants improves polygenic risk prediction across diverse populations. *medRxiv*. 2024 Nov 5 [Preprint].
37. Kang JH, Lee Y, Kim DJ, Kim JW, Cheon MJ, Lee BC. Polygenic risk and rare variant gene clustering enhance cancer risk stratification for breast and prostate cancers. *Communications Biology*. 2024;7(1):1-10.

38. Aldisi R, Hassanin E, Sivalingam S, Buness A, Klinkhammer H, Mayr A, et al. Gene-based burden scores identify rare variant associations for 28 blood biomarkers. *BMC Genetics*. 2023;24(1):50.
39. Wang B, Irizar H, Thygesen JH, Zartaloudi E, Austin-Zimmerman I, Bhat A, Harju-Seppänen J, Pain O, Bass N, Gkofa V, Alizadeh BZ. Psychosis endophenotypes: a gene-set-specific polygenic risk score analysis. 2023.
40. Hörbeck E, Jonsson L, Malwade S, Karlsson R, Pålsson E, Sigström R, et al. Dissecting the impact of complement component 4A in bipolar disorder. *Brain, Behavior, and Immunity*. 2024;116:150-9.
41. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
42. Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, et al. Multi-omics factor analysis—a framework for unsupervised integration of multi-omics data sets. *Molecular Systems Biology*. 2018;14(6):e8124.
43. Trastulla L, Dolgalev G, Moser S, Jiménez-Barrón LT, Andlauer TFM, von Scheidt M, et al. Distinct genetic liability profiles define clinically relevant patient strata across common diseases. *Nature Communications*. 2024;15(1):5534.
44. Jacobs FMJ, van der Heide LP, Wijchers PJEC, Burbach JPH, Hoekman MFM, Smidt MP. FoxO6, a novel member of the FoxO class of transcription factors with distinct shuttling dynamics. *The Journal of Biological Chemistry*. 2003;278(38):35959-67.
45. Witt SH, Streit F, Jungkunz M, Frank J, Awasthi S, Reinbold CS, et al. Genome-wide association study of borderline personality disorder reveals genetic overlap with bipolar disorder, major depression and schizophrenia. *Translational Psychiatry*. 2017;7(6):e1155.
46. Hindley G, Frei O, Shadrin AA, Cheng W, O'Connell KS, Icick R, et al. Charting the landscape of genetic overlap between mental disorders and related traits beyond genetic correlation. *American Journal of Psychiatry*. 2022;179(11):833-43.
47. Streit F, Awasthi S, Hall AS, Niarchou M, Marouli E, Babajide O, et al. Genome-wide association study of borderline personality disorder identifies six loci and highlights shared risk with mental and somatic disorders. *medRxiv*. 2024 Nov 12 [Preprint].

9 Appendix

9.1 Molecular mechanisms associated with bipolar disorders etiology

Table 68. Molecular mechanisms associated with bipolar disorders etiology

The table is an extract of SNP results publicly available at the NHGRI-EBI GWAS Catalog downloaded at https://www.ebi.ac.uk/gwas/. See 8.1.1.

SNP	PMID	REGION	Date	Mapped gene(s)	SNP Replication
rs1012053	17486107	13q14.11	08/05/2007	*DGKH*	Pisanu(*DGKG*);Cross-Disorder(*DGKI*)
rs1006737	20351715	12p13.33	30/03/2010	*CACNA1C*	Smoller(23453885), Ruderfer(24280982)
rs12576775	21926972	11q14.1	18/09/2011	*TENM4*	Ikeda(28115744), Smoller(23453885)
rs4765913	21926972	12p13.33	18/09/2011	*CACNA1C*	Charney(28072414)
rs4650608	22182935	1p31.1	20/12/2011	*IFI44_ ADGRL4*	Ruderfer(24280982)
rs9834970	22182935	3p22.2	20/12/2011	*HSPD1P6_ LINC02033*	CrossDisorder(31835028), Stahl(31043756), Mullins(34002096), Peyrot(33686288), Ruderfer(24280982), Gong(36753304), Wang(34159505), Ikeda(28115744), Yao(33479212), Li(33263727), Charney(28072414), Hou(27329760), Wu(32606422)
rs2535629	23453885	3p21.1	27/02/2013	*ITIH3*	Amare(30626913)
rs12576775	23453885	11q14.1	27/02/2013	*TENM4*	Ikeda(28115744), Sklar(21926972)
rs1006737	23453885	12p13.33	27/02/2013	*CACNA1C*	Liu(20351715), Ruderfer(24280982)
rs2710323	24166486	3p21.1	29/10/2013	*ITIH1*	Wang(34159505), Wu(32606422)
rs17693963	24166486	6p22.1	29/10/2013	*GPR89P_ RSL24D1P1*	Ruderfer(24280982), Wang(38154582)
rs9834970	24280982	3p22.2	26/11/2013	*HSPD1P6_ LINC02033*	CrossDisorder (31835028), Stahl(31043756), Mullins(34002096), Peyrot(33686288), Gong(36753304), Wang(34159505),

SNP	PMID	REGION	Date	Mapped gene(s)	SNP Replication
					Ikeda(28115744), Yao(33479212), Li(33263727), Charney(28072414), Chen(22182935), Hou(27329760), Wu(32606422)
rs1006737	24280982	12p13.33	26/11/2013	*CACNA1C*	Smoller(23453885), Liu(20351715)
rs17693963	24280982	6p22.1	26/11/2013	*GPR89P_ RSL24D1P1*	Sleiman(24166486), Wang(38154582)
rs4650608	24280982	1p31.1	26/11/2013	*IFI44_ ADGRL4*	Chen(22182935)
rs10994415	24618891	10q21.2	11/03/2014	*ANK3*	Mullins(34002096)
rs12202969	24618891	6q16.1	11/03/2014	*MIR2113_ EIF4EBP2P3*	Mullins(34002096)
rs9834970	27329760	3p22.2	21/06/2016	*HSPD1P6_ LINC02033*	CrossDisorder(31835028), Stahl(31043756), Mullins(34002096), Peyrot(33686288), Ruderfer(24280982), Gong(36753304), Wang(34159505), Ikeda(28115744), Yao(33479212), Li(33263727), Charney(28072414), Chen(22182935), Wu(32606422)
rs9834970	28072414	3p22.2	10/01/2017	*HSPD1P6_ LINC02033*	CrossDisorder (31835028), Stahl(31043756), Mullins(34002096), Peyrot(33686288), Ruderfer(24280982), Gong(36753304), Wang(34159505), Ikeda(28115744), Yao(33479212), Li(33263727), Chen(22182935), Hou(27329760), Wu(32606422)
rs2302417	28072414	3p21.1	10/01/2017	*ITIH1*	Li(33263727), Stahl(31043756)
rs4765913	28072414	12p13.33	10/01/2017	*CACNA1C*	Sklar(21926972)
rs12576775	28115744	11q14.1	24/01/2017	*TENM4*	Sklar(21926972), Smoller(23453885)
rs9834970	28115744	3p22.2	24/01/2017	*HSPD1P6_ LINC02033*	CrossDisorder(31835028), Stahl(31043756), Mullins(34002096), Peyrot(33686288), Ruderfer(24280982), Gong(36753304), Wang(34159505),

SNP	PMID	REGION	Date	Mapped gene(s)	SNP Replication
					Yao(33479212), Li(33263727), Charney(28072414), Chen(22182935), Hou(27329760), Wu(32606422)
rs7405404	29121268	16p13.12	09/11/2017	*TMF1P1_ ERCC4*	Cross-Disorder(31835028)
rs2535629	30626913	3p21.1	09/01/2019	*ITIH3*	Smoller(23453885)
rs3804640	31043756	3q13.12	01/05/2019	*CD47*	Peyrot(33686288), Li(33263727)
rs35958438	31043756	15q14	01/05/2019	*LINC02694*	Mullins(34002096)
rs489337	31043756	11q13.1	01/05/2019	*PACS1*	Mullins(34002096)
rs2388334	31043756	6q16.1	01/05/2019	*MIR2113_ EIF4EBP2P3*	Peyrot(33686288), Li(33263727), Yu(38858783), Gong(36753304), Cross-Disorder (31835028)
rs11624408	31043756	14q32.2	01/05/2019	*BCL11B*	Wu(32606422), Wang(34159505)
rs71395455	31043756	15q25.2	01/05/2019	*ZSCAN2-AS1, ZSCAN2*	Yao(33479212), Wang(34159505), Cross-Disorder (31835028)
rs113779084	31043756	7p21.3	01/05/2019	*THSD7A*	Mullins(34002096)
rs10896090	31043756	11q13.2	01/05/2019	*PACS1*	Li(33263727)
rs4447398	31043756	15q15.2	01/05/2019	*STARD9*	Mullins(34002096), Peyrot(33686288)
rs2305929	31043756	2p23.2	01/05/2019	*BABAM2, MRPL33*	Wang(38154582)
rs10994318	31043756	10q21.2	01/05/2019	*ANK3*	Li(33263727)
rs9834970	31043756	3p22.2	01/05/2019	*HSPD1P6_ LINC02033*	CrossDisorder (31835028), Mullins(34002096), Peyrot(33686288), Ruderfer(24280982), Gong(36753304), Wang(34159505), Ikeda(28115744), Yao(33479212), Li(33263727), Charney(28072414), Chen(22182935), Hou(27329760), Wu(32606422)
rs11557713	31043756	18q21.33	01/05/2019	*ZCCHC2*	Li(33263727)
rs6130764	31043756	20q13.12	01/05/2019	*WFDC5_ WFDC12*	Li(33263727)
rs2302417	31043756	3p21.1	01/05/2019	*ITIH1*	Li(33263727), Charney(28072414)
rs17183814	31043756	2q24.3	01/05/2019	*SCN2A*	Li(33263727),

SNP	PMID	REGION	Date	Mapped gene(s)	SNP Replication
					Peyrot(33686288), Mullins(34002096)
rs11647445	31043756	16p13.2	01/05/2019	*GRIN2A*	Wu(32606422), Li(33263727)
rs10744560	31043756	12p13.33	01/05/2019	*CACNA1C-IT3, CACNA1C*	Wu(32606422), Li(33263727), Peyrot(33686288)
rs10035291	31043756	5q14.1	01/05/2019	*SSBP2*	Li(33263727)
rs7122539	31043756	11q13.2	01/05/2019	*PC*	Mullins(34002096)
rs112114764	31043756	17q21.31	01/05/2019	*HDAC5*	Li(33263727)
rs10455979	31043756	6q27	01/05/2019	*RPS6KA2*	Mullins(34002096)
rs111444407	31043756	19p13.11	01/05/2019	*NCAN*	Li(33263727), Peyrot(33686288), Wang(38154582), Wu(32606422)
rs12575685	31043756	11q13.4	01/05/2019	*SHANK2*	Mullins(34002096)
rs11724116	31043756	4q32.2	01/05/2019	*RPS14P7_FSTL5*	Peyrot(33686288)
rs61867293	31926635	10q25.1	01/11/2019	*SORCS3*	Cross-Disorder (31835028)
rs12552	31926635	13q14.3	01/11/2019	*OLFM4*	Cross-Disorder (31835028)
rs7531118	31926635	1p31.1	01/11/2019	*LINC02796*	Cross-Disorder (31835028)
rs2514218	31926635	11q23.2	01/11/2019	*DRD2_TMPRSS5*	Cross-Disorder(31835028), Wu(32606422), Yao(33479212)
rs34215985	31926635	4p13	01/11/2019	*SLC30A9*	Cross-Disorder (31835028)
rs11682175	31926635	2p16.1	01/11/2019	*EIF2S2P7_ACTG1P22*	Yao(33479212), Wang(34159505), Wu(32606422)
rs102275	31926635	11q12.2	01/11/2019	*TMEM258*	Gong(36753304)
rs12958048	31926635	18q21.2	01/11/2019	*TCF4*	Cross-Disorder(31835028)
rs116755193	31926635	5q23.2	01/11/2019	*LINC02240*	Cross-Disorder(31835028)
rs4526442	31926635	9p13.2	01/11/2019	*ZCCHC7*	Cross-Disorder(31835028)
rs915057	31926635	14q23.2	01/11/2019	*SYNE2, ESR2*	Cross-Disorder(31835028)
rs10149470	31926635	14q32.33	01/11/2019	*RNU7-160P_BAG5*	Cross-Disorder(31835028)
rs1002656	31926635	1p34.3	01/11/2019	*FTLP18_GRIK3*	Cross-Disorder(31835028)
rs1226412	31926635	2q24.1	01/11/2019	*LINC01876*	Cross-Disorder(31835028)
rs79879286	31926635	7p15.3	01/11/2019	*GSDME_OSBPL3*	Cross-Disorder(31835028)
rs1518367	31926635	2q33.1	01/11/2019	*PLCL1*	Cross-Disorder(31835028)
rs1806153	31926635	11p13	01/11/2019	*PAUPAR*	Cross-Disorder(31835028)
rs58235352	31926635	12q24.31	01/11/2019	*ACADS_SPPL3*	Cross-Disorder(31835028)
rs1516725	31754094	3q27.2	21/11/2019	*ETV5, DGKG*	See Baum(*DGKH*);Cross-Disorder(*DGKI)*

SNP	PMID	REGION	Date	Mapped gene(s)	SNP Replication
rs34215985	31835028	4p13	01/12/2019	*SLC30A9*	Coleman(31926635)
rs10043984	31835028	5q31.2	01/12/2019	*KDM3B*	Mullins(34002096),Wu (32606422)
rs1226412	31835028	2q24.1	01/12/2019	*LINC01876*	Coleman(31926635)
rs2522831	31835028	7q21.11	01/12/2019	*PCLO*	Yao(33479212)
rs79879286	31835028	7p15.3	01/12/2019	*GSDME_ OSBPL3*	Coleman(31926635)
rs6125656	31835028	20q13.13	01/12/2019	*KCNB1*	Wu(32606422)
rs111294930	31835028	5q33.1	01/12/2019	*LINC01470*	Yao(33479212)
rs174592	31835028	11q12.2	01/12/2019	*FADS2*	Wang(38154582), Mullins(34002096)
rs2693698	31835028	14q32.2	01/12/2019	*BCL11B*	Mullins(34002096)
rs59979824	31835028	2q32.3	01/12/2019	*PCGEM1_ SLC44A3P1*	Wu(32606422)
rs1518367	31835028	2q33.1	01/12/2019	*PLCL1*	Coleman(31926635)
rs7531118	31835028	1p31.1	01/12/2019	*LINC02796*	Coleman(31926635)
rs11887562	31835028	2p24.1	01/12/2019	*LINC03116, LINC01830*	Yao(33479212)
rs2910032	31835028	5q33.1	01/12/2019	*LINC01470*	Wu(32606422)
rs12704290	31835028	7q21.12	01/12/2019	*GRM3, GRM3-AS1*	Wu(32606422), Yao(33479212)
rs71395455	31835028	15q25.2	01/12/2019	*ZSCAN2-AS1, ZSCAN2*	Stahl(31043756), Yao(33479212), Wang(34159505)
rs28681284	31835028	15q25.1	01/12/2019	*CHRNA3*	Yao(33479212), Wu(32606422)
rs12958048	31835028	18q21.2	01/12/2019	*TCF4*	Coleman(31926635)
rs4481150	31835028	3p21.1	01/12/2019	*ITIH3*	Yao(33479212)
rs6694545	31835028	1p35.2	01/12/2019	*LINC01756_ LINC01648*	Wang(34159505)
rs1806153	31835028	11p13	01/12/2019	*PAUPAR*	Coleman(31926635)
rs35346733	31835028	3p26.3	01/12/2019	*CNTN4*	Wu(32606422), Yao(33479212)
rs915057	31835028	14q23.2	01/12/2019	*SYNE2, ESR2*	Coleman(31926635)
rs12668848	31835028	7p22.3	01/12/2019	*MAD1L1*	Wang(38154582), Mullins(34002096)
rs10149470	31835028	14q32.33	01/12/2019	*RNU7-160P_ BAG5*	Coleman(31926635)
rs12898460	31835028	15q14	01/12/2019	*LINC02694*	Peyrot(33686288)
rs12474906	31835028	2p23.2	01/12/2019	*RBKS, MRPL33*	Wu(32606422), Wang(34159505), Yao(33479212)
rs4702	31835028	15q26.1	01/12/2019	*FURIN*	Mullins(34002096), Yao(33479212), Wu(32606422)
rs2388334	31835028	6q16.1	01/12/2019	*MIR2113_ EIF4EBP2P3*	Peyrot(33686288), Li(33263727), Stahl(31043756), Yu(38858783), Gong(36753304)

SNP	PMID	REGION	Date	Mapped gene(s)	SNP Replication
rs35774874	31835028	11q25	01/12/2019	*SNX19_ RN7SL167P*	Yao(33479212)
rs4298967	31835028	12p13.33	01/12/2019	*CACNA1C, CACNA1C-IT3*	Wang(34159505),Yao (33479212),Wu (32606422)
rs116755193	31835028	5q23.2	01/12/2019	*LINC02240*	Coleman(31926635)
rs4380187	31835028	2q32.1	01/12/2019	*ZNF804A_ ELF2P4*	Yao(33479212),Wu (32606422), Wang(34159505)
rs12552	31835028	13q14.3	01/12/2019	*OLFM4*	Coleman(31926635)
rs778353	31835028	2q37.1	01/12/2019	*NGEF*	Wu(32606422)
rs4526442	31835028	9p13.2	01/12/2019	*ZCCHC7*	Coleman(31926635)
rs9834970	31835028	3p22.2	01/12/2019	*HSPD1P6_ LINC02033*	Stahl(31043756), Mullins(34002096), Peyrot(33686288), Ruderfer(24280982, Gong(36753304), Wang(34159505), Ikeda(28115744), Yao(33479212), Li(33263727), Charney(28072414), Chen(22182935), Hou(27329760), Wu(32606422)
rs7405404	31835028	16p13.12	01/12/2019	*TMF1P1_ ERCC4*	Amare(29121268)
rs2514218	31835028	11q23.2	01/12/2019	*DRD2_ TMPRSS5*	Coleman(31926635), Wu(32606422), Yao(33479212)
rs1002656	31835028	1p34.3	01/12/2019	*FTLP18_ GRIK3*	Coleman(31926635)
rs55648125	31835028	6p12.3	01/12/2019	*TFAP2B_ RPS17P5*	Wang(34159505), Yao(33479212)
rs80256351	31835028	2p16.1	01/12/2019	*ACTG1P22_ VRK2*	Wu(32606422), Yao(33479212)
rs61867293	31835028	10q25.1	01/12/2019	*SORCS3*	Coleman(31926635)
rs4619651	31835028	2q11.2	01/12/2019	*LMAN2L_ CNNM4*	Mullins(34002096), Yao(33479212), Wu(32606422), Peyrot(33686288)
rs760648	31835028	22q13.2	01/12/2019	*TCF20*	Wu(32606422)
rs58235352	31835028	12q24.31	01/12/2019	*ACADS_ SPPL3*	Coleman(31926635)
rs7785663	31835028	7q33	01/12/2019	*DGKI*	SeeBaum(*DGKH*); Pisanu(*DGKG*)
rs4619651	32606422	2q11.2	30/06/2020	*LMAN2L_ CNNM4*	Mullins(34002096), Yao(33479212), CrossDisorder (31835028), Peyrot(33686288)
rs111444407	32606422	19p13.11	30/06/2020	*NCAN*	Li(33263727), Peyrot(33686288),

SNP	PMID	REGION	Date	Mapped gene(s)	SNP Replication
					Stahl(31043756), Wang(38154582)
rs11167136	32606422	8q24.3	30/06/2020	*TSNARE1*	Yao(33479212)
rs80256351	32606422	2p16.1	30/06/2020	*ACTG1P22_ VRK2*	CrossDisorder (31835028), Yao(33479212)
rs4702	32606422	15q26.1	30/06/2020	*FURIN*	Mullins(34002096), Yao(33479212), Cross-Disorder (31835028)
rs2710323	32606422	3p21.1	30/06/2020	*ITIH1*	Sleiman(24166486), Wang(34159505)
rs11647445	32606422	16p13.2	30/06/2020	*GRIN2A*	Li(33263727), Stahl(31043756)
rs10043984	32606422	5q31.2	30/06/2020	*KDM3B*	Mullins(34002096), Cross-Disorder (31835028)
rs13236223	32606422	7q34	30/06/2020	*BRAF_ CCT4P1*	Yao(33479212)
rs58120505	32606422	7p22.3	30/06/2020	*MAD1L1*	Wang(34159505)
rs59979824	32606422	2q32.3	30/06/2020	*PCGEM1_ SLC44A3P1*	Cross-Disorder (31835028)
rs4298967	32606422	12p13.33	30/06/2020	*CACNA1C, CACNA1C-IT3*	CrossDisorder (31835028), Wang(34159505), Yao(33479212)
rs760648	32606422	22q13.2	30/06/2020	*TCF20*	Cross-Disorder(31835028)
rs2514218	32606422	11q23.2	30/06/2020	*DRD2_ TMPRSS5*	Coleman(31926635), Cross-Disorder (31835028), Yao(33479212)
rs13135092	32606422	4q24	30/06/2020	*SLC39A8*	Yao(33479212)
rs2910032	32606422	5q33.1	30/06/2020	*LINC01470*	Cross-Disorder(31835028)
rs35346733	32606422	3p26.3	30/06/2020	*CNTN4*	CrossDisorder (31835028), Yao(33479212)
rs12805133	32606422	11q13.2	30/06/2020	*SPTBN2*	Yao(33479212)
rs11682175	32606422	2p16.1	30/06/2020	*EIF2S2P7_ ACTG1P22*	Coleman(31926635), Yao(33479212), Wang(34159505)
rs169738	32606422	6p21.31	30/06/2020	*Metazoa_SRP _BAK1*	Wang(34159505)
rs6434928	32606422	2q33.1	30/06/2020	*SF3B1_ RNU6-1029P*	Yao(33479212)
rs2339519	32606422	2p24.1	30/06/2020	*LINC03116, LINC01830*	Yao(33479212)
rs12474906	32606422	2p23.2	30/06/2020	*RBKS, MRPL33*	Wang(34159505), Yao(33479212), Cross-Disorder (31835028)
rs12704290	32606422	7q21.12	30/06/2020	*GRM3, GRM3-AS1*	CrossDisorder (31835028), Yao(33479212)

SNP	PMID	REGION	Date	Mapped gene(s)	SNP Replication
rs11624408	32606422	14q32.2	30/06/2020	*BCL11B*	Wang(34159505), Stahl(31043756)
rs28681284	32606422	15q25.1	30/06/2020	*CHRNA3*	Cross-Disorder (31835028), Yao(33479212)
rs740417	32606422	12p13.33	30/06/2020	*CACNA1C*	Yao(33479212)
rs10497655	32606422	2q32.1	30/06/2020	*MIR548AE1_ ZNF804A*	Wang(34159505)
rs6125656	32606422	20q13.13	30/06/2020	*KCNB1*	Cross-Disorder(31835028)
rs4380187	32606422	2q32.1	30/06/2020	*ZNF804A_ ELF2P4*	Yao(33479212), Wang(34159505), Cross-Disorder (31835028)
rs10744560	32606422	12p13.33	30/06/2020	*CACNA1C-IT3, CACNA1C*	Li(33263727), Peyrot(33686288), Stahl(31043756)
rs6922815	32606422	6p22.1	30/06/2020	*VN1R10P_ ZNF204P*	Wang(34159505)
rs75836205	32606422	8p12	30/06/2020	*RPL6P22_ RPL10AP3*	Wang(34159505)
rs778353	32606422	2q37.1	30/06/2020	*NGEF*	Cross-Disorder (31835028)
rs9834970	32606422	3p22.2	30/06/2020	*HSPD1P6_ LINC02033*	CrossDisorder (31835028), Stahl(31043756), Mullins(34002096), Peyrot(33686288), Ruderfer(24280982), Gong(36753304), Wang(34159505), Ikeda(28115744), Yao(33479212), Li(33263727), Charney(28072414), Chen(22182935), Hou(27329760)
rs62533709	33169155	9p13.2	10/11/2020	*PAX5*	Yu(38858783)
rs994280	33169155	2q33.1	10/11/2020	*SPATS2L*	Wang(38154582)
rs7969091	33263727	12q13.12	02/12/2020	*RHEBL1_ DHH*	Peyrot(33686288)
rs2388334	33263727	6q16.1	02/12/2020	*MIR2113_ EIF4EBP2P3*	Peyrot(33686288), Stahl(31043756), Yu(38858783), Gong(36753304), Cross-Disorder (31835028)
rs3804640	33263727	3q13.12	02/12/2020	*CD47*	Stahl(31043756), Peyrot(33686288)
rs9834970	33263727	3p22.2	02/12/2020	*HSPD1P6_ LINC02033*	CrossDisorder (31835028), Stahl(31043756), Mullins(34002096), Peyrot(33686288), Ruderfer(24280982),

SNP	PMID	REGION	Date	Mapped gene(s)	SNP Replication
					Gong(36753304), Wang(34159505), Ikeda(28115744), Yao(33479212), Charney(28072414), Chen(22182935), Hou(27329760), Wu(32606422)
rs12672003	33263727	7p15.3	02/12/2020	*PALS2*	Mullins(34002096)
rs2302417	33263727	3p21.1	02/12/2020	*ITIH1*	Stahl(31043756), Charney(28072414)
rs17183814	33263727	2q24.3	02/12/2020	*SCN2A*	Peyrot(33686288), Stahl(31043756), Mullins(34002096)
rs10896090	33263727	11q13.2	02/12/2020	*PACS1*	Stahl(31043756)
rs10035291	33263727	5q14.1	02/12/2020	*SSBP2*	Stahl(31043756)
rs10994318	33263727	10q21.2	02/12/2020	*ANK3*	Stahl(31043756)
rs11557713	33263727	18q21.33	02/12/2020	*ZCCHC2*	Stahl(31043756)
rs11647445	33263727	16p13.2	02/12/2020	*GRIN2A*	Wu(32606422), Stahl(31043756)
rs10744560	33263727	12p13.33	02/12/2020	*CACNA1C-IT3, CACNA1C*	Wu(32606422), Peyrot(33686288), Stahl(31043756)
rs6130764	33263727	20q13.12	02/12/2020	*WFDC5_ WFDC12*	Stahl(31043756)
rs111444407	33263727	19p13.11	02/12/2020	*NCAN*	Peyrot(33686288), Stahl(31043756), Wang(38154582), Wu(32606422)
rs112114764	33263727	17q21.31	02/12/2020	*HDAC5*	Stahl(31043756)
rs12704290	33479212	7q21.12	21/01/2021	*GRM3, GRM3-AS1*	Wu (32606422), Cross-Disorder (31835028)
rs35774874	33479212	11q25	21/01/2021	*SNX19_ RN7SL167P*	Cross-Disorder(31835028)
rs4380187	33479212	2q32.1	21/01/2021	*ZNF804A_ ELF2P4*	Wu(32606422), Wang(34159505), Cross-Disorder (31835028)
rs28681284	33479212	15q25.1	21/01/2021	*CHRNA3*	Cross-Disorder(31835028), Wu(32606422)
rs484201	33479212	11q13.1	21/01/2021	*MACROD1*	Wang(38154582)
rs71395455	33479212	15q25.2	21/01/2021	*ZSCAN2-AS1, ZSCAN2*	Stahl(31043756), Wang(34159505), Cross-Disorder (31835028)
rs9834970	33479212	3p22.2	21/01/2021	*HSPD1P6_ LINC02033*	CrossDisorder (31835028), Stahl(31043756), Mullins(34002096), Peyrot(33686288),

SNP	PMID	REGION	Date	Mapped gene(s)	SNP Replication
					Ruderfer(24280982), Gong(36753304), Wang(34159505), Ikeda(28115744), Li(33263727), Charney(28072414), Chen(22182935), Hou(27329760), Wu(32606422)
rs55648125	33479212	6p12.3	21/01/2021	*TFAP2B_ RPS17P5*	Wang(34159505), Cross-Disorder (31835028)
rs4619651	33479212	2q11.2	21/01/2021	*LMAN2L_ CNNM4*	Mullins(34002096), Wu(32606422), Cross-Disorder (31835028),Peyrot (33686288)
rs80256351	33479212	2p16.1	21/01/2021	*ACTG1P22_ VRK2*	Wu(32606422), Cross-Disorder (31835028)
rs13236223	33479212	7q34	21/01/2021	*BRAF_ CCT4P1*	Wu(32606422)
rs11167136	33479212	8q24.3	21/01/2021	*TSNARE1*	Wu(32606422)
rs4702	33479212	15q26.1	21/01/2021	*FURIN*	Mullins(34002096), Cross-Disorder (31835028), Wu(32606422)
rs4298967	33479212	12p13.33	21/01/2021	*CACNA1C, CACNA1C-IT3*	Cross-Disorder(31835028), Wang(34159505), Wu(32606422)
rs7001340	33479212	8p11.23	21/01/2021	*LETM2, FGFR1*	Wang(34159505)
rs13135092	33479212	4q24	21/01/2021	*SLC39A8*	Wu(32606422)
rs12563424	33479212	1p21.3	21/01/2021	*ALG14_ TLCD4*	Peyrot(33686288), Wang(34159505)
rs12154473	33479212	7p22.3	21/01/2021	*MAD1L1*	Huang(35912095), Mullins(34002096), Wang(38154582)
rs111294930	33479212	5q33.1	21/01/2021	*LINC01470*	Cross-Disorder (31835028)
rs11887562	33479212	2p24.1	21/01/2021	*LINC03116, LINC01830*	Cross-Disorder (31835028)
rs2514218	33479212	11q23.2	21/01/2021	*DRD2_ TMPRSS5*	Coleman(31926635), Cross-Disorder (31835028), Wu(32606422)
rs4481150	33479212	3p21.1	21/01/2021	*ITIH3*	Cross-Disorder(31835028)
rs740417	33479212	12p13.33	21/01/2021	*CACNA1C*	Wu(32606422)
rs2522831	33479212	7q21.11	21/01/2021	*PCLO*	Cross-Disorder(31835028)
rs12805133	33479212	11q13.2	21/01/2021	*SPTBN2*	Wu(32606422)
rs35346733	33479212	3p26.3	21/01/2021	*CNTN4*	Wu(32606422), Cross-Disorder (31835028)

SNP	PMID	REGION	Date	Mapped gene(s)	SNP Replication
rs11682175	33479212	2p16.1	21/01/2021	*EIF2S2P7_ ACTG1P22*	Coleman(31926635), Wang(34159505), Wu(32606422)
rs12474906	33479212	2p23.2	21/01/2021	*RBKS, MRPL33*	Wu(32606422), Wang(34159505), Cross-Disorder (31835028)
rs68188794	33479212	6p22.1	21/01/2021	*ZSCAN16-AS1, ZSCAN16*	Wang(34159505)
rs6434928	33479212	2q33.1	21/01/2021	*SF3B1_ RNU6-1029P*	Wu(32606422)
rs72692857	33479212	1q21.2	21/01/2021	*OTUD7B_ RPL6P31*	Wang(34159505)
rs2339519	33479212	2p24.1	21/01/2021	*LINC03116, LINC01830*	Wu(32606422)
rs10744560	33686288	12p13.33	08/03/2021	*CACNA1C-IT3, CACNA1C*	Wu(32606422), Li(33263727), Stahl(31043756)
rs11724116	33686288	4q32.2	08/03/2021	*RPS14P7_ FSTL5*	Stahl(31043756)
rs3804640	33686288	3q13.12	08/03/2021	*CD47*	Stahl(31043756) ,Li(33263727)
rs111444407	33686288	19p13.11	08/03/2021	*NCAN*	Li(33263727), Stahl(31043756), Wang(38154582), Wu(32606422)
rs7969091	33686288	12q13.12	08/03/2021	*RHEBL1_ DHH*	Li(33263727)
rs28565152	33686288	5p15.31	08/03/2021	*ADCY2*	Mullins(34002096)
rs2388334	33686288	6q16.1	08/03/2021	*MIR2113_ EIF4EBP2P3*	Li(33263727) ,Stahl(31043756), Yu(38858783), Gong(36753304), Cross-Disorder (31835028)
rs4447398	33686288	15q15.2	08/03/2021	*STARD9*	Stahl(31043756), Mullins(34002096)
rs4619651	33686288	2q11.2	08/03/2021	*LMAN2L_ CNNM4*	Mullins(34002096), Yao(33479212), Wu(32606422), Cross-Disorder (31835028)
rs2011503	33686288	19p13.11	08/03/2021	*MAU2*	Wang(34159505)
rs9834970	33686288	3p22.2	08/03/2021	*HSPD1P6_ LINC02033*	CrossDisorder(31835028), Stahl(31043756), Mullins(34002096), Ruderfer(24280982), Gong(36753304), Wang(34159505), Ikeda(28115744), Yao(33479212), Li(33263727), Charney(28072414),

SNP	PMID	REGION	Date	Mapped gene(s)	SNP Replication
					Chen(22182935), Hou(27329760), Wu(32606422)
rs12898460	33686288	15q14	08/03/2021	*LINC02694*	Cross-Disorder(31835028)
rs17183814	33686288	2q24.3	08/03/2021	*SCN2A*	Li(33263727), Stahl(31043756), Mullins(34002096)
rs12563424	33686288	1p21.3	08/03/2021	*ALG14_ TLCD4*	Yao(33479212), Wang(34159505)
rs12202969	34002096	6q16.1	17/05/2021	*MIR2113_ EIF4EBP2P3*	Muhleisen(24618891)
rs4676412	34002096	2q37.3	17/05/2021	*GPR35, CAPN10*	Wang(38154582)
rs113779084	34002096	7p21.3	17/05/2021	*THSD7A*	Stahl(31043756)
rs12668848	34002096	7p22.3	17/05/2021	*MAD1L1*	Wang(38154582),Cross-Disorder(31835028)
rs7122539	34002096	11q13.2	17/05/2021	*PC*	Stahl(31043756)
rs35958438	34002096	15q14	17/05/2021	*LINC02694*	Stahl(31043756)
rs12575685	34002096	11q13.4	17/05/2021	*SHANK2*	Stahl(31043756)
rs9834970	34002096	3p22.2	17/05/2021	*HSPD1P6_ LINC02033*	CrossDisorder (31835028), Stahl(31043756), Peyrot(33686288), Ruderfer(24280982), Gong(36753304), Wang(34159505), Ikeda(28115744), Yao(33479212), Li(33263727), Charney(28072414), Chen(22182935), Hou(27329760), Wu(32606422)
rs112481526	34002096	4q27	17/05/2021	*BLTP1*	Wang(38154582)
rs28455634	34002096	16p13.2	17/05/2021	*HAPSTR1_ RPL21P119*	Wang(38154582)
rs4619651	34002096	2q11.2	17/05/2021	*LMAN2L_ CNNM4*	Yao(33479212), Wu(32606422),Cross-Disorder(31835028), Peyrot(33686288)
rs2693698	34002096	14q32.2	17/05/2021	*BCL11B*	Cross-Disorder (31835028)
rs12672003	34002096	7p15.3	17/05/2021	*PALS2*	Li(33263727)
rs174592	34002096	11q12.2	17/05/2021	*FADS2*	Cross-Disorder(31835028), Wang(38154582)
rs10043984	34002096	5q31.2	17/05/2021	*KDM3B*	Cross-Disorder(31835028), Wu(32606422)
rs17183814	34002096	2q24.3	17/05/2021	*SCN2A*	Li(33263727), Peyrot(33686288), Stahl(31043756)

SNP	PMID	REGION	Date	Mapped gene(s)	SNP Replication
rs12154473	34002096	7p22.3	17/05/2021	*MAD1L1*	Huang(35912095), Wang(38154582), Yao(33479212)
rs489337	34002096	11q13.1	17/05/2021	*PACS1*	Stahl(31043756)
rs28565152	34002096	5p15.31	17/05/2021	*ADCY2*	Peyrot(33686288)
rs112219496	34002096	19p13.11	17/05/2021	*NCAN*	Wang(38154582)
rs10994415	34002096	10q21.2	17/05/2021	*ANK3*	Muhleisen(24618891)
rs10455979	34002096	6q27	17/05/2021	*RPS6KA2*	Stahl(31043756)
rs4447398	34002096	15q15.2	17/05/2021	*STARD9*	Stahl(31043756), Peyrot(33686288)
rs13195402	34002096	6p22.2	17/05/2021	*BTN2A1*	Wang(38154582)
rs4702	34002096	15q26.1	17/05/2021	*FURIN*	Yao(33479212), Cross-Disorder (31835028), Wu(32606422)
rs6694545	34159505	1p35.2	16/06/2021	*LINC01756_ LINC01648*	Cross-Disorder (31835028)
rs11682175	34159505	2p16.1	16/06/2021	*EIF2S2P7_ ACTG1P22*	Coleman(31926635), Yao(33479212), Wu(32606422)
rs68188794	34159505	6p22.1	16/06/2021	*ZSCAN16-AS1, ZSCAN16*	Yao(33479212)
rs72692857	34159505	1q21.2	16/06/2021	*OTUD7B_ RPL6P31*	Yao(33479212)
rs12474906	34159505	2p23.2	16/06/2021	*RBKS, MRPL33*	Wu(32606422), Yao(33479212), Cross-Disorder (31835028)
rs169738	34159505	6p21.31	16/06/2021	*Metazoa_SRP _BAK1*	Wu(32606422)
rs10497655	34159505	2q32.1	16/06/2021	*MIR548AE1_ ZNF804A*	Wu(32606422)
rs4380187	34159505	2q32.1	16/06/2021	*ZNF804A_ ELF2P4*	Yao(33479212), Wu(32606422), Cross-Disorder (31835028)
rs6922815	34159505	6p22.1	16/06/2021	*VN1R10P_ ZNF204P*	Wu(32606422)
rs11624408	34159505	14q32.2	16/06/2021	*BCL11B*	Wu(32606422), Stahl(31043756)
rs71395455	34159505	15q25.2	16/06/2021	*ZSCAN2-AS1, ZSCAN2*	Stahl(31043756), Yao(33479212), Cross-Disorder(31835028)
rs9834970	34159505	3p22.2	16/06/2021	*HSPD1P6_ LINC02033*	CrossDisorder (31835028), Stahl(31043756), Mullins(34002096), Peyrot(33686288), Ruderfer(24280982), Gong(36753304), Ikeda(28115744), Yao(33479212), Li(33263727),

SNP	PMID	REGION	Date	Mapped gene(s)	SNP Replication
					Charney(28072414), Chen(22182935), Hou(27329760), Wu(32606422)
rs55648125	34159505	6p12.3	16/06/2021	*TFAP2B_ RPS17P5*	Cross-Disorder(31835028), Yao(33479212)
rs2710323	34159505	3p21.1	16/06/2021	*ITIH1*	Sleiman(24166486), Wu(32606422)
rs12563424	34159505	1p21.3	16/06/2021	*ALG14_ TLCD4*	Peyrot(33686288), Yao(33479212)
rs58120505	34159505	7p22.3	16/06/2021	*MAD1L1*	Wu(32606422)
rs7001340	34159505	8p11.23	16/06/2021	*LETM2, FGFR1*	Yao(33479212)
rs2011503	34159505	19p13.11	16/06/2021	*MAU2*	Peyrot(33686288)
rs4298967	34159505	12p13.33	16/06/2021	*CACNA1C, CACNA1C-IT3*	CrossDisorder (31835028), Yao(33479212), Wu(32606422)
rs75836205	34159505	8p12	16/06/2021	*RPL6P22_ RPL10AP3*	Wu(32606422)
rs12154473	35912095	7p22.3	15/07/2022	*MAD1L1*	Mullins(34002096), Wang(38154582), Yao(33479212)
rs2388334	36753304	6q16.1	08/02/2023	*MIR2113_ EIF4EBP2P3*	Peyrot(33686288), Li(33263727), Stahl(31043756), Yu(38858783), Cross-Disorder (31835028)
rs9834970	36753304	3p22.2	08/02/2023	*HSPD1P6_ LINC02033*	CrossDisorder (31835028), Stahl(31043756), Mullins(34002096), Peyrot(33686288), Ruderfer(24280982), Wang(34159505), Ikeda(28115744), Yao(33479212), Li(33263727), Charney(28072414), Chen(22182935), Hou(27329760), Wu(32606422)
rs102275	36753304	11q12.2	08/02/2023	*TMEM258*	Coleman(31926635)
rs12668848	38154582	7p22.3	26/12/2023	*MAD1L1*	Mullins(34002096),Cross-Disorder(31835028)
rs28455634	38154582	16p13.2	26/12/2023	*HAPSTR1_ RPL21P119*	Mullins(34002096)
rs112481526	38154582	4q27	26/12/2023	*BLTP1*	Mullins(34002096)
rs2305929	38154582	2p23.2	26/12/2023	*BABAM2, MRPL33*	Stahl(31043756)
rs174592	38154582	11q12.2	26/12/2023	*FADS2*	Cross-Disorder(31835028) ,Mullins(34002096)

SNP	PMID	REGION	Date	Mapped gene(s)	SNP Replication
rs112219496	38154582	19p13.11	26/12/2023	*NCAN*	Mullins(34002096)
rs484201	38154582	11q13.1	26/12/2023	*MACROD1*	Yao(33479212)
rs994280	38154582	2q33.1	26/12/2023	*SPATS2L*	Bigdeli(33169155)
rs12154473	38154582	7p22.3	26/12/2023	*MAD1L1*	Huang(35912095), Mullins(34002096), Yao(33479212)
rs17693963	38154582	6p22.1	26/12/2023	*GPR89P_ RSL24D1P1*	Sleiman(24166486), Ruderfer(24280982)
rs13195402	38154582	6p22.2	26/12/2023	*BTN2A1*	Mullins(34002096)
rs111444407	38154582	19p13.11	26/12/2023	*NCAN*	Li(33263727), Peyrot(33686288), Stahl(31043756), Wu(32606422)
rs4676412	38154582	2q37.3	26/12/2023	*GPR35, CAPN10*	Mullins(34002096)
rs62533709	38858783	9p13.2	11/06/2024	*PAX5*	Bigdeli(33169155)
rs2388334	38858783	6q16.1	11/06/2024	*MIR2113_ EIF4EBP2P3*	Peyrot(33686288), Li(33263727), Stahl(31043756), Gong(36753304), Cross-Disorder (31835028)

9.2 Key Gene Associations

ADCY2 (Adenylate Cyclase 2) (Thesis Chapters 1, 5) Associated with BD1 (from *SCN2A* locus, Subphenotype-BD-MTAG) and lithium response.

AGER (Advanced Glycosylation End-product Specific Receptor) (Chapter 5) A top-ranked gene from the BD-SCZ MTAG gene-based analysis, indicating a role in shared severe psychiatric illness.

AKAP11 (A-kinase Anchoring Protein 11) (Chapter 1) Identified as a shared risk gene for BD and Schizophrenia from exome sequencing (WES) studies.

ANK3 (Ankyrin-G) (Chapters: 1, 5) A robust and consistently replicated risk gene for BD, implicated in neuronal excitability pathways.

ANKRD44 (Ankyrin Repeat Domain 44) (Thesis Chapters: 5) Associated with BD2 (highest CADD score) and shared across multiple BD subphenotypes.

BTN1A1 (Butyrophilin Subfamily 1 Member A1) (Chapter 5) A core gene consistently ranking at the top for almost every BD-only subphenotype in gene-based analysis.

C4A (Complement C4A) (Chapters 5) Associated with several BD subphenotypes, with expression noted in the amygdala and hypothalamus. Implicated in psychosis and shared immune-related pathways.

CACNA1C (Calcium Voltage-Gated Channel Subunit Alpha1 C) (Chapters 1, 4, 5) One of the strongest and most consistently replicated risk genes for BD. Implicated in pathways for psychosis in BD1 and associated with nearly all subphenotypes in MTAG analyses.

CHDH (Choline Dehydrogenase) (Chapter 5) Mentioned as an established risk locus for BD.

CNR1 (Cannabinoid Receptor 1) (Chapter 5) Showed its highest association with the Suicide Attempt (SA) subphenotype.

CREB3L4 (CREB3 Like Transcription Factor 4) (Chapters: 4) Implicated in psychosis in BD1 through pathway analysis ("Mitochondrion" and "*ZNF318*" pathways).

DAOA (D-Amino Acid Oxidase Activator) (Chapter 1) Associated with BD susceptibility in early candidate gene studies.

DCC (DCC Netrin 1 Receptor) (Chapter 5) A novel association shared across the RC, UM, PD, and OCD sub-group, suggesting a role for altered axonal guidance.

DGKH (Diacylglycerol Kinase Eta) (Chapters 1, 5) Implicated in the lithium-sensitive PI pathway and associated with BD. The broader gene *DGKI* was associated with all 11 subphenotypes in the BD-SCZ MTAG.

DISC1 (Disrupted in Schizophrenia 1) (Chapter 1) Associated with Schizoaffective disorder, bipolar type.

DRD2 (Dopamine Receptor D2) (Chapter 5) Credibly associated almost exclusively with the psychosis-spectrum subphenotypes in TWAS analysis.

FADS1 / FADS2 (Fatty Acid Desaturase 1/2) (Chapters 1, 5) Consistently linked to BD, with *FADS1* showing negative cerebellar expression across subphenotypes in TWAS analysis. A top shared gene in BD-only MTAGs.

FEN1 (Flap Endonuclease 1) (Chapter 5) A core gene consistently ranking at the top for almost every BD-only subphenotype, implicated in DNA repair.

FOXO6 (Forkhead Box O6) (Chapter 5) Associated with most BD subphenotypes *except* BD1, suggesting a role in non-psychotic presentations.

FURIN (Furin Paired Basic Amino Acid Cleaving Enzyme) (Chapters 4, 5) Associated with BD and implicated in psychosis in BD1 through immune system pathways.

GABBR1 / GABBR2 (Gamma-Aminobutyric Acid Type B Receptor Subunit 1/2) (Chapters 4, 5) Implicated in pathways for psychosis in BD1 and SZA. *GABBR1* associated with AlcSUD with BD.

GLYCTK (Glycine C-Acetyltransferase) (Chapter 4) Showed extremely strong protective associations in the amygdala across numerous BD-SCZ subphenotypes in TWAS analysis, pointing to glycine metabolism as a key pathway.

GNL3 (Guanine Nucleotide-binding Protein-like 3) (Chapter 5) Showed pervasive and extremely strong positive associations across almost all subphenotypes and brain regions in TWAS analysis.

GRIN2A (Glutamate Ionotropic Receptor NMDA Type Subunit 2A) (Chapter 5) Associated with multiple subphenotypes in the BD-SCZ MTAG, particularly the comorbidity dimension (OCD, UM, PD, etc.).

GRM7 (Glutamate Metabotropic Receptor 7) (Chapter 1) Associated with BD and related personality traits in early GWAS.

HIST1H gene family (e.g., *HIST1H2BK*) (Chapter 5) Consistently a top hit within the chr6p22 positional gene set, highlighting the critical importance of histone structure and chromatin organization.

HLA-DMA (Major Histocompatibility Complex, Class II, DM Alpha) (Chapter 5) Showed a strong, protective association in the BD-SCZ MTAG context, providing a specific neuro-immune link between BD and SCZ.

ITIH1 / ITIH3 / ITIH4 (Inter-Alpha-Trypsin Inhibitor Heavy Chain family) (Chapters 1, 4, 5) A locus robustly associated with BD and SCZ. Implicated in RC and PD subphenotypes, and a top shared gene in BD-only MTAGs.

MAD1L1 (Mitotic Arrest Deficient 1 Like 1) (Chapters 1, 4, 5) A consistently linked gene for BD and SCZ, associated with psychosis in BD1 and emerging as a top pleiotropic gene in BD-SCZ MTAG analyses.

NCAN (Neurocan) (Chapters 1, 5) A well-established risk gene for BD, associated with mania.

NEK4 (NIMA Related Kinase 4) (Chapter 1) Associated with BD in a large PGC-led GWAS.

NRG1 (Neuregulin 1) (Chapter 1) Associated with BD in early candidate gene studies.

NT5C2 (5'-Nucleotidase, Cytosolic 2) (Chapters 4, 5) Associated with psychosis in BD1 and the BD2 subphenotype.

PACS1 (Phosphofurin Acidic Cluster Sorting Protein 1) (Chapter 5) Uniquely associated with BD1 in TWAS analysis, suggesting a role in neuronal protein trafficking.

PBRM1 (Polybromo 1) (Chapter 5) Associated with BD1 and mood-incongruent psychosis; also noted as a key gene in SCHEMA rare-variant enrichment.

SCN2A (Sodium Voltage-Gated Channel Alpha Subunit 2) (Chapter 5) A novel, deleterious variant was identified as a strong marker for BD1 and Psychosis subphenotypes.

SLC39A8 (Solute Carrier Family 39 Member 8) (Chapter 5) A highly pleiotropic and deleterious variant was identified as a novel locus for seven subphenotypes, suggesting a core biological mechanism involving metal ion transport.

SMAD3 (SMAD Family Member 3) (Chapter 5) A novel association specific to Panic Disorder and Rapid Cycling, providing a potential link to thyroid-interacting pathways.

SP4 (Sp4 Transcription Factor) (Chapter 5) Implicated in BD pathobiology via TWAS analysis across multiple subphenotypes.

TCF4 (Transcription Factor 4) (Chapter 5) A key gene driving the significant enrichment of BD-SCZ credible gene sets with established rare-variant risk genes from the SCHEMA consortium.

TMEM258 (Transmembrane Protein 258) (Chapter 5) A core gene consistently ranking at the top for almost every BD-only subphenotype in gene-based analysis.

TRANK1 (Tetratricopeptide Repeat and Ankyrin Repeat Containing 1) (Chapter 1, 5) A well-established and highly pleiotropic risk locus for BD, associated with almost all subphenotypes in MTAG analyses.

ZEB2 (Zinc Finger E-Box Binding Homeobox 2) (Chapter 5) A key gene driving the significant enrichment of BD-SCZ credible gene sets with established rare-variant risk genes from the SCHEMA consortium.

9.3 Transdiagnostic Profiles of BD Subphenotypes

This overview synthesises the primary transdiagnostic genetic associations for the key bipolar disorder (BD) subphenotypes as investigated and discussed within this thesis.

1. Bipolar Disorder I (BD1) Presents as a severe, psychosis-spectrum illness.

- Primary Genetic Overlap: Shows a very strong genetic correlation with Schizophrenia (SCZ) (*rG* ~.71) and is genetically almost indistinguishable from the Psychosis subphenotype (*rG* ~.94) (Chapter 5). The SCZ Polygenic Risk Score (PRS) was a strong predictor of BD1 status and its features (Chapter 4).
- Secondary Overlaps: Has a weaker genetic correlation with Major Depressive Disorder (MDD) (*rG* ~.30).
- Key Genetic Features: SBayesS analysis confirms its genetic architecture overlaps most with SCZ. It is specifically associated with deleterious variants in genes related to neuronal excitability, such as *SCN2A* (Chapter 5).

2. Bipolar Disorder II (BD2) Presents with a genetic profile aligned more with internalizing and affective/attentional disorders.

- Primary Genetic Overlap: Shows its strongest genetic correlation with MDD (*rG* ~.65) and a strong correlation with ADHD (*rG* ~.42).
- Key Genetic Features: SBayesS analysis showed BD2's genetic architecture overlaps most strongly with Anxiety disorders. It genomically clusters with the "Comorbidity" group (PD, OCD, RC, UM) (Chapter 5).

3. Schizoaffective Disorder, Bipolar Type (SZA) Acts as a genetic bridge between BD and SCZ.

- Primary Genetic Overlap: Shares substantial genetic risk with both SCZ and BD1, clustering with them in genomic analyses (Chapter 5).
- Key Genetic Features: Shows one of a very high SNP-based heritability, similar to BD1. The inclusion of SCZ genetics in MTAG analyses massively amplified the number of shared loci with the Psychosis subphenotype, confirming its intermediate genetic position.

4. Psychosis (as a feature within BD) This feature is a key marker of the shared biology between severe BD and SCZ.

- Primary Genetic Overlap: Very high correlation with BD1 and strongly predicted by SCZ PRS (Chapter 4).
- Key Genetic Features: Associated with a unique, deleterious variant in the neuronal sodium channel gene *SCN2A* (Chapter 5). The genetic signal for the neuro-immune gene *HLA-DMA* was only robustly significant when SCZ data was included (Chapter 5), highlighting a specific shared pathway.

5. Rapid Cycling (RC) Presents a unique genetic profile suggestive of severe, multi-faceted dysregulation.

- Primary Genetic Overlap: Clusters with the "Comorbidity and Mood Instability" group (PD, OCD, SA, UM). Has a specific shared novel genetic locus (*SMAD3*) with Panic Disorder.
- Contrasting Overlap: Shows an inverse relationship with SCZ PRS (Chapter 4), suggesting its genetic drivers are distinct from the core psychosis spectrum.
- Key Genetic Features: Correlates positively with ADHD and Anxiety PRS (Chapter 3). It exhibited the most pronounced signature of negative selection, suggesting its architecture may be influenced by rarer, more highly penetrant variants (Chapter 5).

6. Suicide Attempt (SA) Shares genetic architecture with both mood and externalizing/impulsive disorders.

- Primary Genetic Overlap: Strongest external correlations are with MDD, Anxiety, and PTSD. Within the BD subphenotypes, it has a strong genetic correlation with AlcSUD (rG ~.80) (Chapter 5).
- Key Genetic Features: The *CNR1* (cannabinoid receptor) gene showed its highest association with the SA subphenotype (Chapter 5).

7. Alcohol/Substance Use Disorder (AlcSUD) Presents with a profile linked to impulsivity and executive dysfunction.

- Primary Genetic Overlap: Its strongest external correlation is with ADHD. It also clusters with the other "Comorbidity" subphenotypes like SA and RC.
- Key Genetic Features: A novel association with the neurodevelopmental gene *MAD1L1* was identified for the AlcSUD subphenotype in the BD-SCZ MTAG (Chapter 5).

8. Panic Disorder (PD) & Obsessive-Compulsive Disorder (OCD) These anxiety-related subphenotypes show a remarkably strong and specific shared genetic link.

- Primary Genetic Overlap: While showing only moderate global correlation, LAVA analysis revealed that OCD and PD share 30 significant local genetic loci, the strongest local link found between any pair in the comorbidity cluster (Chapter 5).
- Key Genetic Features: A shared novel association with the neurodevelopmental guidance gene *DCC* (along with RC and UM) points to a common vulnerability in brain development for this internalizing/anxious dimension.

9. Unipolar Mania (UM) Presents with a distinct genetic profile that validates its unique position in psychiatric nosology.

- Primary Genetic Overlap: Genomically clusters with the "Comorbidity" group (RC, PD, OCD).
- Key Genetic Features: PRS analysis showed it had the highest predictive power of all subphenotypes (R^2-Liability = 12.4%), suggesting a "purer" genetic signal for mania that is less confounded by the genetic liabilities for depression and psychosis found in BD1 (Chapter 5). It also has unique loci not shared with other subphenotypes, such as one near *YWHAE.*

9.4 Detailed Cohort Descriptions

This section provides detailed information on each cohort contributing to the study, including ascertainment procedures, diagnostic methods, and inclusion/exclusion criteria. For details on the references included below see O'Connell et al., (2025)[Chapter 1, 55].

======== PGC1 Samples ========

Rietschel, M; Nöthen, MM, Cichon, S | 21926972 [PGC1] | BOMA-Germany I | bip_bonn_eur

Cases for the BOMA-Bipolar Study were ascertained from consecutive admissions to the inpatient units of the Department of Psychiatry and Psychotherapy at the University of Bonn and at the Central Institute for Mental Health in Mannheim, University of Heidelberg, Germany. DSM-IV lifetime diagnoses of bipolar I disorder were assigned using a consensus best-estimate procedure, based on all available information, including a structured interview with the SCID and SADS-L, medical records, and the family history method. In addition, the OPCRIT[6] checklist was used for the detailed polydiagnostic documentation of symptoms. Controls were ascertained from three population-based studies in Germany (PopGen, KORA, and Heinz-Nixdorf-Recall Study). The control subjects were not screened for mental illness. Study protocols were reviewed and approved in advance by Institutional Review Boards of the participating institutions. All subjects provided written informed consent.

Corvin, A | 18711365 [PGC1] | Ireland | bip_dub1_eur

Samples were collected as part of a larger study of the genetics of psychotic disorders in the Republic of Ireland, under protocols approved by the relevant IRBs and with written informed consent that permitted repository use. Cases were recruited from Hospitals and Community psychiatric facilities in Ireland by a psychiatrist or psychiatric nurse trained to use the SCID. Diagnosis was based on the structured interview supplemented by case note review and collateral history where available. All diagnoses were reviewed by an independent reviewer. Controls were ascertained with informed consent from the Irish GeneBank and represented blood donors who met the same ethnicity criteria as cases. Controls were not specifically screened for psychiatric illness.

Blackwood, D | 18711365 [PGC1] | Edinburgh, UK | bip_edi1_eur

This sample comprised Caucasian individuals contacted through the inpatient and outpatient services of hospitals in South East Scotland. A BD-I diagnosis was based on an interview with the patient using the SADS-L supplemented by case note review and frequently by information from medical staff, relatives and caregivers. Final diagnoses, based on DSM-IV criteria, were reached by consensus between two trained psychiatrists. Ethnically matched controls from the same region were recruited through the South of Scotland Blood Transfusion Service. Controls were not directly screened to exclude those with a personal or family history of psychiatric illness. The study was approved by the Multi-Centre Research Ethics Committee for Scotland and patients gave written informed consent for the collection of DNA samples for use in genetic studies.

Kelsoe, J | 21926972 [PGC1] | USA (GAIN) | bip_gain_eur

Genetic Association Information Network (GAIN)/ The Bipolar Genome Study (BiGS) The BD sample was collected under the auspices of the NIMH Genetics Initiative for BD (http://zork.wustl.edu/nimh/), genotyped as part of GAIN and analyzed as part of a larger GWAS conducted by the BiGS consortium. Approximately half of the GAIN sample was collected as multiplex families or sib pair families (waves 1-4), the remainder were collected as individual cases (wave 5). Subjects were ascertained at 12 sites: Indiana University, John Hopkins University, the NIMH Intramural Research Program, Washington University at St. Louis, University of Pennsylvania, University of Chicago, Rush Medical School, University of Iowa, University of California, San Diego, University of California, San Francisco, Howard University, and University of Michigan. All investigations were carried out after the review of protocols by the IRB at each participating institution. At all sites, potential cases were identified from screening admissions to local treatment facilities and through publicity programs or advocacy groups. Potential cases were evaluated using the DIGS[7], FIGS[8], and information from relatives and medical records. All information was reviewed through a best estimate diagnostic procedure by two independent and non-interviewing clinicians and a consensus best-estimate diagnosis was reached. In the event of a disagreement, a third review was done to break the tie. Controls were from the NIMH Genetic Repository sample obtained by Dr. P. Gejman through a contract to Knowledge Networks, Inc. Only individuals with complete or near-complete psychiatric questionnaire data who did not fulfill diagnostic criteria for major depression and denied a history of psychosis or BD were included as controls for BiGS analyses. Controls were matched for gender and ethnicity to the cases.

Scott, L; Myer, RM; Boehnke, M | 19416921 [PGC1] | Michigan, USA (Pritzker and NIMH) | bip_mich_eur

The Pritzker Neuropsychiatric Disorders Research Consortium (NIMH/Pritzker) case and control samples were from the NIMH Genetics Initiative Genetics Initiative Repository. Cases were diagnosed according to DMS-III or DSM-IV criteria using diagnostic interviews and/or medical record review. Cases with low confidence diagnoses were excluded. From each wave 1-5 available non-Ashkenazi European-origin family, two BD1 siblings were included when possible and the proband was preferentially included if available (n=946 individuals in 473 sibling pairs); otherwise, a single BD1 case was included (n=184). The bipolar sibling pairs were retained within the NIMH/Pritzker sample when individuals in more than one study were uniquely assigned to a study set. Controls had non-Ashkenazi European origin, were aged 20-70 years and reported no diagnosis with or treatment for BD or schizophrenia, and that they had not heard voices that others could not hear. Individuals with suspected major depression were excluded based on answers to questions related to depressive mood. NIMH controls were further selected as the best match(es) to NIMH cases based on self-reported ancestry.

Sklar, P; Smoller, J | 18317468 [PGC1] | USA (STEP1) | bip_stp1_eur

The Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD) was a seven-site, national U.S., longitudinal cohort study designed to examine the effectiveness of treatments and their impact on the course of BD that enrolled 4,361 participants who met DSM-IV criteria for BD1, BD2, bipolar not otherwise specified (NOS), schizoaffective manic or bipolar type, or cyclothymic disorder based on diagnostic interviews. From the parent study, 2,089 individuals who were over 18 years of age with BD1 and BD2 diagnoses consented to the collection of blood samples for DNA. BD samples with a consensus diagnosis of BD1 were selected for inclusion in STEP1. Two groups of controls samples from the NIMH repository were used. One comprised DNA samples derived from US Caucasian anonymous cord blood donors. The second were controls who completed the online self-administered psychiatric screen and were ascertained as described above, by Knowledge Networks Inc. For the second sample of controls only those without a history of schizophrenia, psychosis, BD or major depression with functional impairment were used.

Sklar, P; Smoller, J | 18711365 [PGC1] | USA (STEP2) | bip_stp2_eur
The STEP2 sample included BD-1 and BD-2 samples from the STEP-BD study described above along with BD-2 subjects from UCL study also described above. The controls samples for this study were from the NIMH repository as described above for the STEP1 study.
Andreassen, OA | PMID:21926972 [PGC1], PMID:20451256 | Norway (TOP) | bip_top7_eur
In the TOP study (Tematisk omrade psykoser), cases of European ancestry, born in Norway, were recruited from psychiatric hospitals in the Oslo region. Patients were diagnosed according to the ICD9 and further ascertainment details have been reported. Healthy control subjects were randomly selected from statistical records of persons from the same catchment area as the patient groups. The control subjects were screened by interview and with the Primary Care Evaluation of Mental Disorders (PRIME-MD). None of the control subjects had a history of moderate/severe head injury, neurological disorder, mental retardation or an age outside the age range of 18-60 years. Healthy subjects were excluded if they or any of their close relatives had a lifetime history of a severe psychiatric disorder. All participants provided written informed consent and the human subjects protocol was approved by the Norwegian Scientific-Ethical Committee and the Norwegian Data Protection Agency.
McQuillin, A; Gurling, H | 18317468 [PGC1] | UCL (University College London), London, UK | bip_uclo_eur
The UCL sample comprised Caucasian individuals who were ascertained and received clinical diagnoses of bipolar 1 disorder according to UK National Health Service (NHS) psychiatrists at interview using the categories of the International Classification of Disease version 1. In addition, bipolar subjects were included only if both parents were of English, Irish, Welsh or Scottish descent and if three out of four grandparents were of the same descent. All volunteers read an information sheet approved by the Metropolitan Medical Research Ethics Committee who also approved the project for all NHS hospitals. Written informed consent was obtained from each volunteer. The UCL control subjects were recruited from London branches of the National Blood Service, from local NHS family doctor clinics and from university student volunteers. All control subjects were interviewed with the SADS-L to exclude all psychiatric disorders.
Craddock, N, Jones, I, Jones, L | 17554300 | WTCCC | bip_wtcc_eur_sr-qc
Cases were all over the age of 17 yr, living in the UK and of European descent. Recruitment was undertaken throughout the UK and included individuals who had been in contact with mental health services and had a lifetime history of high mood. After providing written informed consent, participants were interviewed by a trained psychologist or psychiatrist using a semi-structured lifetime diagnostic psychiatric interview (Schedules for Clinical Assessment in Neuropsychiatry) and available psychiatric medical records were reviewed. Using all available data, best-estimate life-time diagnoses were made according to the RDC[12]. In the current study I included cases with a lifetime diagnosis of RDC bipolar 1 disorder, bipolar 2 disorder or schizo-affective disorder, bipolar type.
Controls were recruited from two sources: the 1958 Birth Cohort study and the UK Blood Service (blood donors) and were not screened for history of mental illness.
All cases and controls were recruited under protocols approved by the appropriate IRBs. All subjects gave written informed consent.

======== PGC2 Samples ========
Adolfsson, R | Not published | Umeå, Sweden | bip_ume4_eur
Clinical characterization of the patients included the Mini-International Neuropsychiatric Interview (MINI[11]), the Diagnostic Interview for Genetic Studies (DIGS[7]), the Family Interview for Genetic Studies (FIGS[8]) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN)[12]. The final diagnoses were made according to the DSM-IV-TR and determined by consensus of 2 research

psychiatrists. The unrelated Swedish control individuals, consisting of a large population-based sample representative of the general population of the region, were randomly selected from the 'Betula study'.

Alda, M; Smoller, J | Not published | Nova Scotia, Canada; I2B2 controls | bip_hal2_eur

The case samples were recruited from patients longitudinally followed at specialty mood disorders clinics in Halifax and Ottawa (Canada). Cases were interviewed in a blind fashion with the Schedule of Affective Disorders and Schizophrenia-Lifetime version (SADS-L)[13] and consensus diagnoses were made according to DSM-IV[14] and Research Diagnostic Criteria (RDC)[15]. Protocols and procedures were approved by the local Ethics Committees and written informed consent was obtained from all patients before participation in the study. Control subjects were drawn from the I2B2 (Informatics for Integrating Biology and the Bedside) project[16]. The study consists of de-identified healthy individuals recruited from a healthcare system in the Boston, MA, US area. The de-identification process meant that the Massachusetts General Hospital Institutional Review Board elected to waive the requirement of seeking informed consent as detailed by US Code of Federal Regulations, Title 45, Part 46, Section 116 (46.116).

Andreassen, OA | Not published | Norway (TOP) | bip_top8_eur

The TOP8 bipolar disorder cases and controls were ascertained in the same way as the bip_top7_eur (TOP7) samples described above and recruited from hospitals across Norway.

Biernacka, JM; Frye, MA | 27769005 | Mayo Clinic, USA | bip_may1_eur

Bipolar cases were drawn from the Mayo Clinic Bipolar Biobank[17]. Enrolment sites included Mayo Clinic, Rochester, Minnesota; Lindner Center of HOPE/University of Cincinnati College of Medicine, Cincinnati, Ohio; and the University of Minnesota, Minneapolis, Minnesota. Enrolment at each site was approved by the local Institutional Review Board, and all participants consented to use of their data for future genetic studies. Participants were identified through routine clinical appointments, from in-patients admitted in mood disorder units, and recruitment advertising. Participants were required to be between 18 and 80 years old and be able to speak English, provide informed consent, and have DSM-IV-TR diagnostic confirmation of type 1 or 2 bipolar disorder or schizoaffective bipolar disorder as determined using the SCID. Controls were selected from the Mayo Clinic Biobank[18]. Potential controls with ICD9 codes for bipolar disorder, schizophrenia or related diagnoses in their electronic medical record were excluded.

Rietschel, M; Nöthen, MM; Schulze, TG; Reif, A; Forstner, AJ | 24618891 | BOMA-Germany II | bip_bmg2_eur

Cases were recruited from consecutive admissions to psychiatric in-patient units at the University Hospital Würzburg. All cases received a lifetime diagnosis of BD according to the DSM-IV criteria using a consensus best-estimate procedure based on all available information, including semi-structured diagnostic interviews using the Association for Methodology and Documentation in Psychiatry[23], medical records and the family history method. In addition, the OPCRIT system was used for the detailed poly diagnostic documentation of symptoms.

Control subjects were ascertained from the population-based Heinz Nixdorf Recall (HNR) Study[24]. The controls were not screened for a history of mental illness. Study protocols were reviewed and approved in advance by Institutional Review Boards of the participating institutions. All subjects provided written informed consent.

Rietschel, M; Nöthen, MM; Schulze, TG; Bauer, M; Forstner, AJ; Müller-Myhsok, B | 24618891 | BOMA-Germany III | bip_bmg3_eur[25]

Cases were recruited at the Central Institute of Mental Health in Mannheim, University of Heidelberg, and other collaborating psychiatric hospitals in Germany. All cases received a lifetime diagnosis of BD according to the DSM-IV criteria using a consensus best-estimate procedure based on all available information including structured diagnostic interviews using the AMDP, Composite International

Diagnostic Screener (CID-S)[26], SADS-L and/or SCID, medical records, and the family history method. In addition, the OPCRIT system was used for the detailed poly diagnostic documentation of symptoms. Controls were selected randomly from a Munich-based community sample and recruited at the Max-Planck Institute of Psychiatry. They were screened for the presence of anxiety and mood disorders using the CID-S. Only individuals without mood and anxiety disorders were collected as controls. Study protocols were reviewed and approved in advance by Institutional Review Boards of the participating institutions. All subjects provided written informed consent.

Hauser, J; Lissowska, J; Forstner, AJ | 24618891 | BOMA-Poland | bip_bmpo_eur

Cases were recruited at the Department of Psychiatry, Poznan University of Medical Sciences, Poznan, Poland. All cases received a lifetime diagnosis of BD according to the DSM-IV criteria on the basis of a consensus best-estimate procedure and structured diagnostic interviews using the SCID. Controls were drawn from a population-based case-control sample recruited by the Cancer-Center and Institute of Oncology, Warsaw, Poland and a hospital-based case-control sample recruited by the Nofer Institute of Occupational Medicine, Lodz, Poland. The Polish controls were produced by the International Agency for Research on Cancer (IARC) and the Centre National de Génotypage (CNG) GWAS Initiative for a study of upper aerodigestive tract cancers. The controls were not screened for a history of mental illness. Study protocols were reviewed and approved in advance by Institutional Review Boards of the participating institutions. All subjects provided written informed consent.

Rietschel, M; Nöthen, MM; Rivas, F; Mayoral, F; Kogevinas, M; others | 24618891 | BOMA-Spain | bip_bmsp_eur

Cases were recruited at the mental health departments of the following five centers in Andalusia, Spain: University Hospital Reina Sofia of Córdoba, Provincial Hospital of Jaen; Hospital of Jerez de la Frontera (Cádiz); Hospital of Puerto Real (Cádiz); Hospital Punta Europa of Algeciras (Cádiz); and Hospital Universitario San Cecilio (Granada). Diagnostic assessment was performed using the SADS-L; the OPCRIT; a review of medical records; and interviews with first and/or second degree family members using the Family Informant Schedule and Criteria (FISC)[27]. Consensus best estimate BD diagnoses were assigned by two or more independent senior psychiatrists and/or psychologists, and according to the RDC, and the DSM-IV. Controls were Spanish subjects drawn from a cohort of individuals recruited in the framework of the European Community Respiratory Health Survey (ECRHS, http://www.ecrhs.org/). The controls were not screened for a history of mental illness. Study protocols were reviewed and approved in advance by Institutional Review Boards of the participating institutions. All subjects provided written informed consent.

Fullerton, J.M.; Mitchell, P.B.; Schofield, P.R.; Martin N.G.; Cichon, S. | 24618891 | BOMA-Australia | bip_bmau_eur

Cases were recruited at the Mood Disorder Unit, Prince of Wales Hospital in Sydney. All cases received a lifetime diagnosis of BD according to the DSM-IV criteria on the basis of a consensus best-estimate procedure[19] and structured diagnostic interviews using the DIGS, FIGS, and the SCID. Controls were parents of unselected adolescent twins from the Brisbane Longitudinal Twin Study. The controls were not screened for a history of mental illness. Study protocols were reviewed and approved in advance by Institutional Review Boards of the participating institutions. All subjects provided written informed consent.

Grigoroiu-Serbanescu, M; Nöthen, MM | 21353194 | BOMA-Romania | bip_rom3_eur

Cases were recruited from consecutive admissions to the Obregia Clinical Psychiatric Hospital, Bucharest, Romania. Patients were administered the DIGS[28] and FIGS[8] interviews. Information was also obtained from medical records and close relatives. The diagnosis of BP-I was assigned according

to DSM-IV criteria using the best estimate procedure. All patients had at least two hospitalized illness episodes. Population-based controls were evaluated using the DIGS to exclude a lifetime history of major affective disorders, schizophrenia, schizoaffective disorders, and other psychoses, obsessive-compulsive disorder, eating disorders, and alcohol or drug addiction.

Kelsoe, J; Sklar, P; Smoller, J | [PGC1 Replication] | USA (FAT2; FaST, BiGS, TGEN) | bip_fat2_eur

Cases were collected from individuals at the 11 U.S. sites described for the GAIN sample. Eligible participants were age 18 or older meeting DSM-IV criteria for BD-I or BD-II by consensus diagnosis based on interviews with the Affective Disorders Evaluation (ADE) and MINI. All participants provided written informed consent and the study protocol was approved by IRBs at each site. Collection of phenotypic data and DNA samples were supported by NIMH grants MH063445 (JW Smoller); MH067288 (PI: P Sklar), MH63420 (PI: V Nimgaonkar) and MH078151, MH92758 (PI: J. Kelsoe). The control samples were NIMH controls that were using the methods described in that section. The case and control samples were independent of those included in the GAIN sample.

Kirov, G | 25055870 | Bulgarian trios | bip_butr_eur

All cases were recruited in Bulgaria from psychiatric inpatient and outpatient services. Each proband had a history of hospitalisation and was interviewed with an abbreviated version of the SCAN. Consensus best-estimate diagnoses were made according to DSM-IV criteria by two researchers. All participants gave written informed consent and the study was approved by local ethics committees at the participating centers.

Kirov, G | 25055870 | UK trios | bip_uktr_eur

The BD subjects were recruited from lithium clinics and interviewed in person by a senior psychiatrist, using the abbreviated version of the SCAN. Consensus best-estimate diagnoses were made based on the interview and hospital notes. Ethics committee approval for the study was obtained from the relevant research ethics committees and all individuals provided written informed consent for participation.

Landén, M; Sklar, P | [ICCBD] | Sweden (ICCBD) | bip_swa2_eur

The BD subjects were identified using the Swedish National Quality Register for Bipolar Disorders (BipoläR) and the Swedish National Patient Register (using a validated algorithm[29] requiring at least two hospitalizations with a BD diagnosis). A confirmatory telephone interview with a diagnostic review was conducted. Additional subjects were recruited from the St. Göran Bipolar Project (Affective Center at Northern Stockholm Psychiatry Clinic, Sweden), enrolling new and ongoing patients diagnosed with BD using structured clinical interviews. Diagnoses were made according to the DSM-IV criteria (BipoläR and St. Göran Bipolar Project) and ICD-10 (National Patient Register). The control subjects used were the same as for the SCZ analyses described above. All ascertainment procedures were approved by the Regional Ethical Committees in Sweden.

Landén, M; Sklar, P | [ICCBD] | Sweden (ICCBD) | bip_swei_eur

The cases and controls in the bip_swei_eur sample were recruited using the same ascertainment methods described for the bip_swa2_eur sample.

Leboyer, M |[30]; [PGC1 replication] | France | bip_fran_eur

Cases with BD1 or BD2 and control samples were recruited as part of a large study of genetics of BD in France (Paris-Creteil, Bordeaux, Nancy) with a protocol approved by relevant IRBs and with written informed consent. Cases of French descent for more than 3 generations were assessed by a trained psychiatrist or psychologist using structured interviews supplemented by medical case notes, mood scales and self-rating questionnaire assessing dimensions.

Li, Q | 24166486; 27769005 | USA (Janssen), SAGE controls | bip_jst5_eur

The study included unrelated patients with bipolar 1 disorder from 6 clinical trials (IDs: NCT00253162, NCT00257075, NCT00076115, NCT00299715, NCT00309699, and NCT00309686). Participant recruitment was conducted by Janssen Research & Development, LLC (formerly known as Johnson &

Johnson Pharmaceutical Research & Development, LLC) to assess the efficacy and safety of risperidone. Bipolar cases were diagnosed according to DSM-IV-TR criteria. The diagnosis of bipolar disorder was confirmed by the Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL) in NCT00076115, by the SCID in NCT00257075 and NCT00253162, or by the MINI in NCT00299715 and NCT00309699, and NCT00309686, respectively. Additional detailed descriptions of these clinical trials can be found at ClinicalTrials.gov. Only patients of European ancestry with matching controls were included in the current analysis. Controls subjects were drawn from the Study of Addiction: Genetics and Environment (SAGE, dbGaP Study Accession: phs000092.v1.p1). Control subjects did not have alcohol dependence or drug dependence diagnoses; however, mood disorders were not an exclusion criterion.

Craddock, N; Jones, I; Jones, L | [ICCBD] | Cardiff and Worcester, UK (ICCBD-BDRN) | bip_icuk_eur

Cases were all over the age of 17 yr, living in the UK and of European descent. Cases were recruited via systematic and not systematic methods as part of the Bipolar Disorder Research Network project (www.bdrn.org), provided written informed consent and were interviewed using a semi-structured diagnostic interview, the Schedules for Clinical Assessment in Neuropsychiatry. Based on the information gathered from the interview and case notes review, best-estimate lifetime diagnosis was made according to DSM-IV. Inter-rater reliability was formally assessed using 20 randomly selected cases (mean κ Statistic = .85). In the current study I included cases with a lifetime diagnosis of DSM-IV bipolar disorder or schizo-affective disorder, bipolar type. The BDRN study has UK National Health Service (NHS) Research Ethics Committee approval and local Research and Development approval in all participating NHS Trusts/Health Boards.Controls were part of the Wellcome Trust Case Control Consortium common control set, which comprised healthy blood donors recruited from the UK Blood Service and samples from the 1958 British Birth Cohort. Controls were not screened for a history of mental illness. All cases and controls were recruited under protocols approved by the appropriate IRBs. All subjects gave written informed consent.

Ophoff, RA | Not Published | Netherlands | bip_ucla_eur

The case sample consisted of inpatients and outpatients recruited through psychiatric hospitals and institutions throughout the Netherlands. Cases with DSM-IV bipolar disorder, determined after interview with the SCID, were included in the analysis. Controls were collected in parallel at different sites in the Netherlands and were volunteers with no psychiatric history after screening with the (MINI[11]). Ethical approval was provided by UCLA and local ethics committees and all participants gave written informed consent.

Paciga, S | [PGC1] | USA (Pfizer) | bip_pf1e_eur

This sample comprised Caucasian individuals recruited into one of three Geodon (ziprasidone) clinical trials (NCT00141271, NCT00282464, NCT00483548). Subjects were diagnosed by a clinician with a primary diagnosis of Bipolar 1 Disorder, most recent episode depressed, with or without rapid cycling, without psychotic features, as defined in the DSM-IV-TR (296.5x) and confirmed by the MINI (version 5..0). Subjects also were assessed as having a HAM-D-17 total score of >20 at the screening visit. The trials were conducted in accordance with the protocols, International Conference on Harmonization of Good Clinical Practice Guidelines, and applicable local regulatory requirements and laws. Patients gave written informed consent for the collection of blood samples for DNA for use in genetic studies.

Pato, C | [ICCBD] | Los Angeles, USA (ICCBD-GPC)| bip_usc2_eur

Genomic Psychiatry Consortium (GPC) cases and controls were collected via the University of Southern California healthcare system, as previously described[31]. Using a combination of focused, direct interviews and data extraction from medical records, diagnoses were established using the OPCRIT and were based on DSM-IV-TR criteria. Age and gender-matched controls were ascertained

from the University of Southern California health system and assessed using a validated screening instrument and medical records.

======== PGC2 Followup Samples ========

Kelsoe, J | [PGC1] | USA (BiGS/TGEN1) | TGEN1_eur

Cases and controls for this sample were ascertained using the same procedures applied for the bip_gain_eur sample described above. These samples formed a distinct PCA cluster from the samples described above and were therefore analysed separately.

Li, Q | 24166486 | various Eastern Europe, shared T. Esku controls | JJ_EAST_eur

The cases were drawn from the same six clinical studies described for bip_jst5_eur except that only patients of east European ancestry with matching controls were included in this cohort. Most of the Eastern European controls were from the Estonian Biobank project (EGCUT)[32] and were ancestrally matched with cases.

Schulze, T | [ConLiGen] | Germany | BIP_KFO_eur

The KFO sample was derived from the Clinical Research Group 241 (KFO241 consortium; www.kfo241.de) and the PsyCourse consortium (www.psycourse.de). The samples form part of a multi-site German/Austrian longitudinal study. Diagnoses were made according to DSM-IV. German Red Cross controls were collected by the Central Institute for Mental Health in Mannheim, University of Heidelberg, Germany. Volunteers who gave blood to the Red Cross were asked whether they would be willing to participate in genetic studies of psychiatric disorders. Control subjects were not selected on the basis of mental health screening.

======== External studies PGC3 ========

Stefánsson, H | [PGC1 replication] | Iceland (deCODE genetics) | deCODE

The Icelandic sample consisted of 2,908 subjects with BD (1661 SNP typed) and 344,848 controls (141,854 SNP typed). DNA was isolated from blood samples provided by patients and controls that were recruited throughout Iceland. Approval for the study was granted by the National Bioethics Committee of Iceland and the Icelandic Data Protection Authority and informed consent was obtained for all participants providing a sample for the study. Diagnoses were assigned according to Research Diagnostic Criteria[38] through the use of the SADS-L[39] for 303 subjects. DSM-IV BD diagnoses were obtained through the use of the Composite International Diagnostic Interview (CIDI-Auto) for 82 subjects. The remaining BD subjects were diagnosed by ICD 9 or ICD 10 at Landspitali University Hospital in the years 1987-2018. Controls were recruited as a part of various genetic programs at deCODE and were not screened for psychiatric disorders. Whole genome sequencing was performed on samples from 541 BD cases and 26,014 controls. Two types of imputations were performed; into SNP-typed individuals based on long-range phasing, followed by a familial imputation step into un-typed relatives of SNP-typed individuals. Cases of bipolar I disorder were defined using ICD-10 codes 31.1 and 31.2 and ICD-9 codes 296.0 and 296.2. Cases of bipolar II disorder were defined using the ICD-10 code 31.0 in the absence of ICD-10 codes F31.1 and F31.2 and ICD-9 codes 296.0 and 296.2.

Milani L | 24518929 | Estonia (Estonian Biobank) | EstonianBiobank

The Estonian Biobank (EstBB) is a population-based cohort of 200,000 participants with a rich variety of phenotypic and health-related information collected for each individual[32]. At recruitment, all participants signed a consent to allow follow-up linkage of their electronic health records (EHR), thereby providing a longitudinal collection of phenotypic information. Health records have been extracted from the national Health Insurance Fund Treatment Bills (from 2004), Tartu University Hospital (from 2008), and North Estonia Medical Center (from 2005). The diagnoses are coded in ICD-10 format and drug dispensing data include drug ATC codes, prescription status and purchase date (if available). For the current study, cases of bipolar disease were determined by searching the EHRs for

data on F31* ICD-10 diagnosis. All remaining participants who did not have any ICD-10 F* group diagnoses were defined as controls. Cases with bipolar I disorder were those with ICD codes of F31.1 and F31.2.

Zwart JA | Unpublished | Norway (the Trøndelag Health Study) | HUNT

The HUNT sample consisted of 905 subjects with BD and 41,914 population controls[41]. Patients and controls were of European ancestry and were recruited from the Nord-Trøndelag County, Norway. Diagnoses were assigned according to ICD-9 or ICD-1. The controls included individuals not diagnosed with substance use disorders, schizophrenia, bipolar disorder, major depressive disorder, anxiety disorders, eating disorders, personality disorders, or ADHD in hospitals (ICD-9 or ICD-10) or general practice (ICPC2). They also were >40 years of age, had low self-reported levels of anxiety and depression (HADS-A and HADS-D < 11), and reported no use of antidepressants, anxiolytics, or hypnotics. Approval for the study was granted by the Data Inspectorate of Norway, the Health Directorate and the Regional Committee for Medical and Health Research Ethics. Cases of bipolar I disorder were those with ICD codes of F31.1, F31.2 or F31.6 and individuals with an ICD-9 code of 295 or ICD-10 codes F20-F29 were excluded. Cases of bipolar II disorder were those with ICD codes of F31.8 and individuals with an ICD-9 code of 295 or ICD-10 codes F20-F29, F31.1-.2 or F31.6 were excluded.

======== PGC PsychChip Samples ========

Pato, C | Not published | [PGC Psychchip] | gpcw1

The cases and controls in this study were ascertained in the same manner as those described above for bip_usc2_eur.

Reif, A | Not published | [PGC Psychchip] | germ1

Cases were recruited in the same manner as those described above for BOMA-Germany II | bip_bmg2_eur. Control subjects were healthy participants who were recruited from the community of the same region as cases. They were of Caucasian descent and fluent in German. Exclusion criteria were manifest or lifetime DSM-IV axis I disorder, severe medical conditions, intake of psychoactive medication as well as alcohol abuse or abuse of illicit drugs. Absence of DSM-IV axis I disorder was ascertained using the German versions of the Mini International Psychiatric Interview. IQ was above 85 as ascertained by the German version of the Culture Fair Intelligence Test 2[44]. Study protocols were reviewed and approved by the ethical committee of the Medical Faculty of the University of Würzburg. All subjects provided written informed consent.

Serretti, A, Vieta E, Ribases M | Not published | [PGC Psychchip] | spsp3

The sample includes 267 BD subjects (Spanish Wave2 Serretti PsychChip QC Summary), of which 180 Spanish and 87 Italian. Spanish sample: 180 subjects were enrolled in a naturalistic cohort study, consecutively admitted to the out-patient Bipolar Disorders Unit, Hospital Clinic, University of Barcelona. This is a systematic cross-sectional analysis deeply described in a previous paper on the same sample investigating rs10997870 SIRT1 gene variant[45]. Inclusion criteria were a diagnosis of bipolar disorder (type 1 or 2) according to DSM-IV TR criteria and age of 18 years or older. The study was approved by the local ethical committee and carried out in accordance with the ethical standards laid down in the Declaration of Helsinki. Signed informed consent was obtained from all participants after a detailed and extensive description of the study and patient's confidentiality was preserved. The current and lifetime diagnoses of mental disorders were formulated by independent senior psychiatrists (diagnostic concordance: Kappa=.80) according to DSM-IV TR clinical criteria and confirmed through the semi-structured interviews for Axis I disorders according to DSM IV TR criteria (SCID I). Furthermore, all available clinical data coming from follow-up at our unit and collateral information concerning illness history were cross-referred in order to ensure accuracy and obtain complete clinical information. Specific psychopathological dimensions were assessed by means of rating scales and

clinical questionnaires administered by clinicians, adequately trained to enhance inter-rater reliability. Mood episodes were defined according to DSM-IV TR criteria and their severity was measured through the administration of the 21-item Hamilton Depression Rating Scale (HDRS-21, Spanish version). The most severe depressive episode was defined on the basis of the severity at the HDRS (total score > 14) and clinical judgment. Italian sample: 87 subjects with bipolar depression were enrolled into the study when admitted at the Department of Psychiatry, University of Bologna, Italy. A description of the subjects has been previously reported when analyzing clinical features[46]. Inclusion criteria were a diagnosis of bipolar disorder, most recent episode depressive as assessed by DSM-IV-TR criteria; Young Mania Rating Scale (YMRS) score <12; Hamilton Depression Rating Scale (HAM-D) <12. Exclusion criteria were presence of a bipolar disorder, most recent episode manic or hypomanic; presence of severe medical conditions; presence of moderate to severe dementia (Mini Mental State Examination score <20). The following scales were administered biweekly during the hospitalization: HAM-D, Hamilton Anxiety Rating Scale (HAM-A), YMRS and Dosage Record and Treatment Emergent Symptom Scale (DOTES). Written informed consent was obtained for each patient recruited. The study protocol was approved by the local Ethical Committee and it has been performed in accordance with the ethical standards laid down in the 1975 Declaration of Helsinki.

The Spanish controls were part of the Mental-Cat clinical sample or the INSchool population-based cohort. A total of 1,774 controls from the Mental-Cat cohort (6.5% males) were evaluated and recruited prospectively from a restricted geographic area at the Hospital Universitari Vall d'Hebron of Barcelona (Spain) and consisted of unrelated healthy blood donors. The INSchool sample consisting of 771 children (76.2% males) from schools in Catalonia were involved for screening using the Achenbach System of Empirically Based Assessment (ASEBA) with the Child Behavior Checklist CBCL/4-18 (completed by parents or surrogates), the Teacher Report Form TRF/5-18 (completed by teachers and other school staff) and the Youth Self-Report YSR/11-18 (completed by youths); the Strengths and Difficulties Questionnaire (SDQ) and the Conner's ADHD Rating Scales (Parents and Teachers). Genomic DNA samples were obtained either from peripheral blood lymphocytes by the salting out procedure or from saliva using the Oragene DNA Self-Collection Kit (DNA Genotek, Kanata, Ontario Canada). DNA concentrations were determined using the Pico- Green dsDNA Quantitation Kit (Molecular Probes, Eugene, OR) and genotyped with the Illumina Infinium PsychArray-24 v1.1 at the Genomics Platform of the Broad Institute. The study was approved by the Clinical Research Ethics Committee (CREC) of Hospital Universitari Vall d'Hebron, all methods were performed in accordance with the relevant guidelines and regulations and written informed consent was obtained from participant parents before inclusion into the study. Detailed information has been published previously[47].

Perlis, R; Sklar, P; Smoller, J, Goes F, Mathews CA, Waldman I | Not published | [PGC Psychchip] | usaw4

Perlis, R; Sklar, P; Smoller, J: EHR data were obtained from a health care system of more than 4.6 million patients[48] spanning more than 20 years. Experienced clinicians reviewed charts to identify text features and coded data consistent or inconsistent with a diagnosis of bipolar disorder. Natural language processing was used to train a diagnostic algorithm with 95% specificity for classifying bipolar disorder. Filtered coded data were used to derive three additional classification rules for case subjects and one for control subjects. The positive predictive value (PPV) of EHR-based bipolar disorder and subphenotype diagnoses was calculated against diagnoses from direct semistructured interviews of 190 patients by trained clinicians blind to EHR diagnosis. The PPV of bipolar disorder defined by natural language processing was .86. Coded classification based on strict filtering achieved a value of .84, but classifications based on less stringent criteria performed less well. No EHR-classified control subject received a diagnosis of bipolar disorder on the basis of direct interview (PPV=1.0). For most subphenotypes, PPV exceeded .8. The EHR-based classifications were used to accrue bipolar disorder cases and controls for genetic analyses. Samples were genotyped on the Psychchip array.

Goes, FS: Cases represented independent probands from a European American family sample that was collected at Johns Hopkins University from 1988-201. Families had at least 2 additional relatives with a major mood disorder (defined as bipolar disorder type 1, bipolar type 2 or recurrent major depressive disorder). Diagnostic interviews were performed using the Schedule for Affective Disorders and Schizophrenia-Lifetime Version (*N*=81) and the Diagnostic Instrument for Genetics Studies (*N*=161). All cases underwent best-estimate diagnostic procedures. After genotyping quality control there were 242 cases, of which 240 were diagnosed as bipolar disorder type 1 and 2 as schizoaffective disorder, bipolar type. Diagnoses were based on DSM-III and DSM-IV criteria. Probands from this sample have been previously studied in family based linkage and exome studies.[49–51]

Mathews CA: Control samples were ascertained as part of ongoing genetic and neurophysiological studies of hoarding, obsessive compulsive and tic disorders. Controls reported no current or lifetime history of mania or hypomania at the time of ascertainment. Sixty-two of the 104 controls were screened for psychiatric illness using the Structured Clinical Interview for DSM-IV TR diagnoses and diagnoses of bipolar disorder, lifetime or current, were ruled out through a best estimate consensus diagnosis. Other psychiatric diagnoses were not excluded. The remaining 42 participants were not formally screened but reported no lifetime or current history of bipolar disorder, obsessive compulsive, hoarding, or tic disorders. Samples were genotyped on the Psychchip array. Ethical approvals were obtained from the University of Florida Human Subjects Review Board.

Waldman I: Control samples were ascertained as part of an ongoing genetic study of ADHD and other Externalizing disorders (I.e., Oppositional Defiant Disorder and Conduct Disorder). Controls reported no current diagnoses of Externalizing or Internalizing disorders at the time of ascertainment. Controls were assessed for psychiatric conditions using the Emory Diagnostic Rating Scale (EDRS)[52], a questionnaire that assessed parent ratings of symptoms of common DSM-IV Externalizing and Internalizing disorders (e.g., Major Depressive Disorder and various anxiety disorders). Samples were genotyped on the Psychchip array. Ethical approvals were obtained from the Emory University and University of Arizona Human Subjects Review Boards.

Baune, BT; Dannlowski, U | Not published | [PGC Psychchip] | bdtrs

The Bipolar Disorder treatment response Study (BP-TRS) comprises BD inpatient cases and screened controls of Caucasian background. Psychiatric diagnosis of bipolar disorders was ascertained using SCID or MINI 6.0 using DSM-IV criteria in a face-to-face interview by a trained psychologist / psychiatrist for both cases and controls. Healthy controls were included if no current or lifetime psychiatric diagnosis was identified. Cases were included if current or lifetime diagnosis of bipolar disorder was ascertained by structured diagnostic interview. Cases and controls are of similar age range (>=18 yrs of age) and were collected from the same geographical areas. Other assessments including symptom ratings, psychiatric history, treatment history, treatment response was based on interview and carried out by trained psychologists/psychiatrists. Samples were genotyped on the Psychchip array. Ethical approval was obtained from the University of Münster Human Ethics Committee, Münster, Germany.

Ophoff R, Posthuma D, Lochner C, Franke B | Not published | [PGC Psychchip] | dutch

Ophoff R: Cases and controls were collected using the same protocol as described above for the “ucla” sample.

Lochner C: Controls include South African Caucasian population based-controls ascertained from blood banks and controls recruited through university campuses and newspaper advertisements, who underwent a psychiatric interview and had no current or lifetime psychiatric disorder 53,[54].

Franke B: The controls included are healthy individuals from the Dutch part of the International Multicenter ADHD Genetics (IMAGE) project 55,[56].

Posthuma D: Data were provided for 960 unscreened Dutch population controls from the Netherlands Study of Cognition, Environment and Genes (NESCOG)[57]. The study was approved by the institutional review board of Vrije Universiteit Amsterdam and participants provided informed consent.

Gawlik M | Not published | [PGC Psychchip] | gawli

Patients were recruited at the Department of Psychiatry, Psychosomatics and Psychotherapy, University of Würzburg, Germany. Diagnosis according to DSM-IV (Diagnostic and Statistical Manual of Mental Disorders-fourth edition) was made by the best estimate lifetime diagnosis method, based on all available information, including medical records, and the family history method.

Fullerton J, Mitchell PB, Schofield PR, Green MJ, Weickert CS, Weickert TW, The Australian Schizophrenia Research Bank | Not published | [PGC Psychchip] | neuc1

The NeuRA collection comprised BD cases from three cohorts ascertained in Australia: the bipolar high risk study[58] (*n*=97), the Imaging Genetics in Psychosis Study (IGP; *n*=47)[59] and a clinic sample (*n*=109) recruited via the Sydney Bipolar Disorders Clinic. The clinic sample used the same ascertainment procedures as described for the bip_bmau_eur sample. The bipolar high risk study is a collaborative study with 4 US and one Australian groups, with young participants aged 12-3. The IGP sample was recruited from outpatient services of the South Eastern Sydney-Illawarra Area Health Service (SESIAHS), the Sydney Bipolar Disorders Clinic and the Australian Schizophrenia Research Bank. Healthy controls were sourced from the high risk, IGP and the Cognitive and Affective Symptoms of Schizophrenia Intervention (CASSI) trial[61] studies, and were recruited from the community, had no personal lifetime history of a DSM-IV Axis-I diagnosis as determined by psychiatric interview, and no history of psychotic disorders among first-degree biological relatives. Additional controls were recruited as part of the strategy to develop an Australian Schizophrenia Research Biobank for studies into the genetics of this disease. The ascertainment of these controls has been previously described[62].

Landen M, Hillert J, Alfredsson L | Not published | [PGC Psychchip] | swed1

The cases in the swed1 sample were recruited using the same ascertainment methods described for the bip_swa2_eur sample. Population-based healthy controls, randomly selected from the Swedish national population register, were collected as part of two case-control studies of multiple sclerosis: GEMS (Genes and Environment in Multiple Sclerosis) and EIMS (Epidemiological Investigation of Multiple Sclerosis)[63].

Di Florio A, McQuillin A, McIntosh A, Breen G | Not published | [PGC Psychchip] | ukwa1

McQuillin A: BD cases were recruited using the same protocol as the bip_uclo_eur described above. A subset (*n*=448) of the control subjects were random UK blood donors obtained from the ECACC DNA Panels (https://www.phe-culturecollections.org.uk/products/dna/hrcdna/hrcdna.jsp). The remaining control subjects (*n*=814) had been screened for an absence of mental illness in using the same protocol as the bip_uclo_eur described above.

Di Florio A: Cases were recruited across the United Kingdom in the same manner as described for the bip_wtcc_eur and bip_icuk_eur samples.

McIntosh AM: BD cases were recruited from the clinical case loads of treating psychiatrists from Edinburgh and across the central belt of Scotland. Controls were identified from non-genetic family members and from the extended networks of the participants themselves. All participants were of European ancestry and diagnosis was confirmed using an established battery developed for ICCCBD.

Breen G: Controls were drawn from blood donors to the UK Motor Neuron Disease Association DNA Biobank[64]

Perlis, R; Sklar, P; Smoller, J, Nievergelt C, Kelsoe J | Not published | [PGC Psychchip] | usaw5

Kelsoe, J: The Pharmacogenomics of Bipolar Disorder (PGBD) study was a prospective assessment of lithium response in BDI patients. The goal was to identify genes for lithium response. Subjects were recruited from clinics at 11 international sites and followed for up to 2.5 years. Diagnosis was obtained by DIGS interview and medical records reviewed by blind experienced clinicians. As the comparison

was between lithium responders and non-responders, no controls were collected. All subjects provided written informed consent.
Perlis R: Cases of bipolar disorder were Individuals treated with lithium drawn from the Partners Healthcare electronic health record (EHR) database, which spans two large academic medical centers, Massachusetts General Hospital and Brigham and Women's Hospital in addition to community and specialty outpatient clinics[65]. Any patients aged 18 years or older with at least one lithium prescription between 2006 and 2013 based on e-prescribing data were included. The Partners Institutional Review Board approved all aspects of this study. Individuals with a diagnosis of schizophrenia based on ICD9 codes were excluded.
Smoller J: Cases and controls were recruited in the same manner as described above for "usaw4".

======== PGC3 Samples ========

Ferentinos P, Dikeos D, Patrinos G | Not published | Greece (Attikon General Hospital) | greek

All adult patients with a DSM-IV-TR/DSM-5 diagnosis of bipolar disorder hospitalized at the inpatient unit or followed-up at the specialized 'Affective disorders and Suicide' outpatient clinic of the 2nd Department of Psychiatry, National and Kapodistrian University of Athens, Attikon General Hospital, Athens, Greece from 2012 to 2017 were recruited for the current study. Patients were referred to the specialized 'Affective disorders and Suicide' outpatient clinic either from the inpatient unit after hospitalization or from the community. Diagnosis was established and demographic (age, gender, family status, profession, employment status, education) and relevant clinical features (e.g. age at onset, polarity of first and most recent episode, number of lifetime depressive and manic/hypomanic episodes, number of hospitalizations, lifetime suicidality, lifetime psychosis) were extracted through a M.I.N.I.-5..0-based semi-structured diagnostic interview, which was administered during patients' initial clinical assessment and regularly updated ever since, interviews of primary caregivers and inspection of medical records. Lifetime presence of any DSM-IV-TR axis I psychiatric comorbidities (dysthymia, panic disorder, agoraphobia, social phobia, generalized anxiety disorder, obsessive-compulsive disorder, post-traumatic stress disorder, alcohol and substance abuse and dependence, anorexia nervosa, bulimia nervosa) was similarly extracted. Family history of major psychiatric disorders and suicidality in first and second degree relatives was recorded with a specific questionnaire based on the Family Interview for Genetic Studies. Medical comorbidities were recorded with the Cumulative Illness Rating Scale, completed on the basis of interview with patient and primary caregivers, inspection of patient's medical records and laboratory exams (basic or specific, if considered necessary). Presence of selected medical diseases was specifically recorded.
Control (unaffected) participants were a convenient sample drawn from the same geographic area as case participants, either within health care facilities or as community volunteers. All of them went through a brief clinical interview including items on psychiatric and medical history, psychiatric family history, past and current medical or psychiatric therapies, and a brief mental state examination. Only participants found to be free of lifetime major mental disorders (MDD, BD, schizophrenia, or other psychotic disorders) and with no family history of major mental disorder in their first-degree relatives were recruited as controls.
All cases and controls were native Greek speakers. All participants provided written informed consent before being included in the study and the study protocol was approved by the Research Ethics Committee of Attikon General Hospital.

Andreassen, OA | Not published | Norway (TOP) | norgs

The NORGS bipolar disorder cases and controls were ascertained in the same way as the bip_top7_eur (TOP7) samples described above and recruited from hospitals across Norway.

Andreassen, OA | Not published | Norway (TOP) | noroe

The MONROE bipolar disorder cases and controls were ascertained in the same way as the bip_top7_eur (TOP7) samples described above and recruited from hospitals across Norway.

Reininghaus EZ | Not published| Austria (Medical University of Graz) | graza

Univ. Prof. DDr. Eva Reininghaus, Priv.Doz. DDr. Susanne Bengesser, Priv.Doz. Dr. Nina Dalkner, Dr. Frederike Fellendorf and further team members of the special outpatient's department for bipolar affective disorders at the Department of Psychiatry and Psychotherapeutic Medicine, Medical University of Graz, Austria: Cases with bipolar affective disorder (type I and II) and healthy controls were recruited at the Department of Psychiatry and Psychotherapeutic Medicine at the Medical University of Graz (MUG), Austria. Study protocols were approved by the ethics committee of the Medical University of Graz. Patients and healthy controls gave written informed consent and the study was conducted according to the declaration of Helsinki. All patients received a clinical interview by a psychiatrist or psychologist and a diagnosis according to DSM-IV with the SCID-I (Structured clinical interview). Healthy controls did not have a history of a psychiatric disorder. Furthermore, healthy controls did not have any first or second degree relatives with a psychiatric disorder. The PGC-Graz sample (*n*= 244; 114 males, 130 females) includes 167 cases with bipolar disorder and 77 healthy controls genotyped with Omniexpress 1.2 by Illumina.

Grigoroiu-Serbanescu M | 31791676; 26806518 | Romania (BOMA-Romania) | bmtron

This sample includes the BOMA-Romania sample and additional cases from the ConLiGen-Romania sample. For the BOMA-Romania sample, unrelated BP-I patients were recruited from consecutive admissions in the Obregia Psychiatric Hospital of Bucharest, Romania. All participants provided written informed consent following a detailed explanation of the study aims and procedures. The study was performed in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). All participants were of Romanian descent according to self-reported ancestry. Genealogical information about parents and all four grandparents was obtained through direct interview of the subjects.

The patients were investigated with the Diagnostic Interview for Genetic Studies (DIGS)[28] and the Family Interview for Genetic Studies (FIGS)[8] The diagnosis of BP-I was assigned according to DSM-IV criteria on the basis of both the DIGS and medical records. Patients were included in the sample if they had at least two documented hospitalized illness episodes (one manic/mixed and one depressive or two manic episodes) and no residual mood incongruent psychotic symptoms during remissions. This information was also confirmed by first degree relatives for 64% of the cases. The illness age-of-onset was defined as the age at which the proband first met DSM-IV criteria for a manic, mixed, or major depressive episode. Family history of psychiatric illness was obtained with FIGS administered both to the patients and to all available relatives.

Cases in the ConLiGen-Romania study were ascertained in the same manner as for BOMA-Romania. Cases were required to have taken lithium for at least two years and lithium treatment response was evaluated with the Alda scale[66]**.**

Population-based controls were evaluated using the DIGS and FIGS to screen for a lifetime history of major affective disorders, schizoaffective disorders, SCZ and other psychoses, obsessive-compulsive disorder, eating disorders, and alcohol or drug addiction. Unaffected individuals were included as controls in the present study.

======== PGC4 Samples ========

Grigoroiu-Serbanescu M | PMID : 31791676| Romania (BOMA-Romania) | rom4

Cases were recruited from consecutive admissions to the Obregia Clinical Psychiatric Hospital, Bucharest, Romania. Patients were administered the DIG 28 and FIGS[8] interviews. Information was also obtained from medical records and close relatives. The diagnosis of BP-I was assigned according to DSM-IV-R criteria using the best estimate procedure. All patients had at least two hospitalized

illness episodes. Population-based controls were evaluated using the DIGS to exclude a lifetime history of major affective disorders, schizophrenia, schizoaffective disorders, and other psychoses, obsessive-compulsive disorder, eating disorders, and alcohol or drug addiction.

McQuillin A | PMID: 37643680 | UCL (University College London), London, UK | amq1
Case and controls were collected using the protocol described above for bip_uclo_eur.

Squassina A, | PMID: 21961650 | Italy | ital1
Patients with bipolar I or bipolar II disorder were recruited at the outpatient unit (Lithium Clinic) of the Clinical Psychopharmacology Centre at the Department of Biomedical Science, Section of Neuroscience & Clinical Pharmacology, University of Cagliari, University Hospital Agency of Cagliari, Italy. Clinical assessments followed a strict procedure. After providing informed consent, participants were interviewed using one of the structured or semistructured interviews SADS-L. Clinical diagnosis was confirmed by DSM-IV criteria. I also used available medical records, narrative summaries of all interviews, and details such as baseline assessments, clinical course, response to treatment, treatment adherence, psychiatric and medical comorbidities, history of suicidal behavior, and symptom profiles in OPCRIT format.[6]
For uniform evaluation of treatment response, I used all available information including data from clinical records, diagnostic interviews, and prospective follow-up assessed by NIMH Life- Chart Method[67]. I used the Alda scale to assess lithium response[66].

Manchia M, Carpiniello B, Squassina A | PMID: 35566641 | Italy | ital2
The case samples were recruited among patients attending the outpatient clinic of the community mental health center of the Unit of Clinical Psychiatry within the University Hospital of Cagliari, Italy. Patients were enrolled in the genetic study if they met the following inclusion criteria: diagnosis of either Bipolar I or Bipolar II disorder according to DSM 5[68] criteria validated through the Italian version of the SCID-5-CV (Structured Clinical Interview for DSM-5 Clinical Version); being in euthymic phase. All patients provided a written consent form regarding the use of their biological and clinical data for research purposes. Blood samples were gathered at the beginning of the study along with the relevant demographic and biometric data. All the clinical documents are stored in an anonymized database, accessible only by authorized personnel.
The recruited subjects were phenotypically characterized with the use of the following standardized tests:

- Brief Assessment of Cognition in Affective Disorders (BACA)
- Brief Assessment of Cognition in Schizophrenia to assess baseline cognitive capacities
- Hamilton Depression Rating Scale (HDRS)
- Young Mania Rating Scale (YMRS)
- Hamilton Anxiety Rating Scale (HAM-A)
- Barratt Impulsivity scale (BIS)
- Clinical Global Impression Scale – Severity (CGI-S)
- Alda score for Lithium response (clinical response defined as a score >7)
- OPCRIT

Tondo L, Squassina A | PMID: 20348464 | Italy | ital3
Our sample population encompasses a cohort of patients followed at the Mood Disorder Lucio Bini Center in Cagliari (Italy), a specialized outpatient clinic for the diagnosis, treatment and research of affective disorders. Since the founding of this outpatient clinic in 1977, all demographic and clinical information about patients have been recorded systematically by means of semi-structured initial and

follow-up interviews, a life chart, extensive clinical evaluation and repeated assessments with standard rating scales for mood such as the Hamilton Depression Rating Scale (HDRS)[69], and Young Mania Rating Scale[70], typically every 4–6 weeks. Diagnoses were updated to meet the Diagnostic and Statistical Manual of Mental Disorders (DSM)-5 criteria[68] after the year 2013. Written informed consent was obtained for collection and analysis of patient data to be presented anonymously in aggregate form, in accordance with the requirements of Italian law and following review by a local ethical committee. Required data were entered into a computerized database in coded form to protect subject identity.
Patients were included in the study if they had at least 12 months of treatment with lithium and if they had a diagnosis of bipolar disorder (BD) or major depressive disorder (MDD) according to DSM-5. The clinical response to lithium treatment was characterized using the "Retrospective Criteria of Long-Term Treatment Response in Research Subjects with Bipolar Disorder" scale, also known as Alda Scale[66].

Alda M | Not published | Nova Scotia, Canada | hal3

The case samples were recruited from patients longitudinally followed at a specialty mood disorders clinic in Halifax (Canada). Cases were interviewed in a blind fashion with the Schedule of Affective Disorders and Schizophrenia-Lifetime version (SADS-L)[13] by pairs of clinician researchers (psychiatrists and/or nurses). The interviews together with medical records were subsequently reviewed in a blind fashion by a panel of senior clinical researchers. Consensus diagnoses were made according to DSM-IV[14] and Research Diagnostic Criteria (RDC)[15] Protocols and procedures were approved by the local Ethics Committees and written informed consent was obtained from all patients before participation in the study.

======== External Samples PGC4 ========

Genomic Psychiatry Cohort (GPC) (USA) | 33169155

Details of ascertainment and diagnosis, genotyping and quality control have been described in detail previously[82]. Briefly, cases were ascertained using the Diagnostic Interview for Psychosis and Affective Disorders (DI-PAD), a semi-structured clinical interview administered by mental health professionals, which was developed specifically for the GPC study. Individuals reporting no lifetime symptoms indicative of psychosis or mania and who have no first-degree relatives with these symptoms are included as control participants.

Key Statistical Equations

Equations for the following methods that were prominent in this thesis:

1. Polygenic Risk Score (PRS) (basic summation formula, see Chapter 1 also)
2. Linear Regression
3. Logistic Regression (model equation and link to Odds Ratio)
4. Odds Ratio (OR) (derived from logistic regression)
5. Nagelkerke R^2
6. Area Under the ROC Curve (AUC)
7. Inverse Probability Weighting (IPW)
8. Exploratory Factor Analysis (EFA) / Confirmatory Factor Analysis (CFA)
9. LD Score Regression (LDSC)
10. Convert Nagelkerke's R^2 to Liability Scale R
11. Fisher's Exact Test

For more complex algorithms including MTAG, PRS-CS, SEM, Random Forest, PRSet, LAVA, SBayesS, FUMA, MAGMA, and TWAS, providing a single, comprehensive equation is often not feasible. These are typically complex statistical frameworks or software packages involving multiple steps or algorithms. For these, I recommend consulting the original publications or specialized statistical texts for the full mathematical details referenced below.

---

1. Polygenic Risk Score (PRS) - Basic Summation

The basic (weighted) polygenic risk score for an individual j (PRS_j) is typically calculated as the sum of risk alleles an individual possesses, weighted by the effect size of each variant:

$$PRS_j = \sum_{i=1}^{N} \beta_i \times dosage_{ij}$$

Where:

- N is the number of SNPs included in the score.

- β_i is the effect size (e.g., log odds ratio or beta coefficient from GWAS) of variant *i*.
- $dosage_{ij}$ is the number of copies of the risk allele for SNP *i* in individual *j* (can be 0, 1, or 2 for hard-called genotypes, or a value between 0 and 2 for imputed dosages).

---

2. Linear Regression (Simple)

The equation for a simple linear regression line, predicting a dependent variable *Y* from an independent variable *X*, is:

$$\hat{Y} = \beta_0 + \beta_1 X$$

Where:

- $\hat{Y}$ is the predicted value of the dependent variable.
- β_0 is the intercept (the predicted value of *Y* when *X*=0).
- β_1 is the slope (the change in *Y* for a one-unit change in *X*). For multiple linear regression with *k* predictors $(X_1, X_2, \ldots X_k)$: $\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$

---

3. Logistic Regression

Logistic regression models the probability of a binary outcome (e.g., case/control status). The relationship is often expressed in terms of the log-odds (logit) of the outcome:

$$logit\big(P(Y=1)\big) = \ln\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

here:

- $P(Y = 1)$ is the probability of the outcome occurring.
- ln is the natural logarithm.
- $\left(\frac{P(Y=1)}{1-P(Y=1)}\right)$ is the odds of the outcome.
- β_0 is the intercept (log-odds when all Xi=0).
- β_i are the coefficients representing the change in log-odds for a one-unit change in X_i.

Alternatively, the probability itself can be expressed as:

$$P(Y=1) = \frac{e^{(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k)}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k)}}$$

---

4. Odds Ratio (OR) from Logistic Regression

For a one-unit increase in a predictor X_i in a logistic regression model, the odds ratio is given by exponentiating the coefficient β_i:

$$OR_i = e^{\beta_i}$$

This OR represents how the odds of the outcome Y=1 change for each one-unit increase in X_i, holding other predictors constant.

---

5. Nagelkerke R^2 (Pseudo R-squared)

Nagelkerke R^2 is a pseudo R-squared statistic used to assess the goodness of fit for models with categorical outcomes, like logistic regression. It is a modification of the Cox & Snell R^2 to achieve a maximum value of 1.

The Cox & Snell R^2 is calculated as:

$$R^2_{CS} = 1 - \left(\frac{L_0}{L_{model}}\right)^{\frac{2}{n}}$$

Then, Nagelkerke R^2 is:

$$R^2_{Nag} = \frac{R^2_{CS}}{1 - L_0^{\frac{2}{n}}}$$

Where:

- L_0 is the likelihood of the null model (model with only an intercept).
- L_{model} is the likelihood of the fitted model (with predictors).
- n is the number of observations.

---

6. Area Under the ROC Curve (AUC)

The Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) is a measure of a binary classifier's ability to distinguish between classes. The ROC curve plots the True Positive Rate (TPR or Sensitivity) against the False Positive Rate (FPR or 1-Specificity) at various threshold settings.

$$TPR = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

$$FPR = \frac{False\ Positives}{False\ Positives + True\ Negatives}$$

The AUC is the area under this plotted curve. While there isn't a single simple "equation" for AUC that's as straightforward as a regression equation (it's often calculated numerically, e.g., using the trapezoidal rule or by its statistical interpretation as the probability that a randomly chosen positive instance is ranked higher than a randomly chosen negative instance), a common method to calculate it involves summing the areas of trapezoids under the ROC curve segments. Alternatively, AUC = (Percent Concordant + 0.5 * Percent Tied)/100.

---

7. Inverse Probability Weighting (IPW)

The core idea of IPW is to weight each individual in a study by the inverse of their probability of receiving the exposure (or treatment) they received, conditional on measured confounders. This creates a pseudo-population where the exposure is independent of the measured confounders.

For estimating the Average Treatment Effect (ATE) under unconfoundedness ($Y(d) \perp D|X$) and positivity ($0 < P(D = 1|X) < 1$), the ATE can be expressed using IPW as:

$$ATE = E\left[\frac{D.Y}{P(D = 1|X)}\right] - E\left[\frac{(1 - D).Y}{P(D = 0|X)}\right]$$

Or, for a mean outcome $E[Y^d]$ for a potential treatment *d*:

$$E[Y^d] = \mathrm{E}\left[\frac{1(D=d).Y}{P(D=d|X)}\right]$$

Where:

- *D* is the treatment/exposure indicator (1 if treated, 0 if not).
- *Y* is the observed outcome.
- Y(*d*) is the potential outcome if treatment d was received.
- *X* are the measured confounders.
- *P(D=d|X)* is the propensity score, the probability of receiving treatment d given confounders X.
- 1(·) is the indicator function.

These are often estimated using sample averages with estimated propensity scores. Different estimators like the Horvitz-Thompson or Hájek estimator exist.

---

8. Confirmatory Factor Analysis (CFA) - Basic Measurement Model Equation

A common representation of the measurement model in CFA for a vector of observed variables x is:

$X=\Lambda\xi+\delta$

Where:

- X is a p×1 vector of p observed variables (indicators).
- Λ (Lambda) is a $p \times k$ matrix of factor loadings, representing the relationship between each observed variable and each latent factor.
- ξ (xi) is a $k \times 1$ vector of k latent factors (unobserved constructs).
- δ (delta) is a $p \times 1$ vector of unique variances or measurement errors for each observed variable.

The model-implied covariance matrix ($\Sigma(\theta)$) is then:

$\Sigma(\theta)=\Lambda\Psi\Lambda'+\Theta_\delta$

Where:

- Ψ (Psi) is the $k \times k$ covariance matrix of the latent factors.
- Θ_δ (Theta-delta) is the $p \times p$ covariance matrix of the measurement errors (often diagonal, assuming uncorrelated errors). CFA aims to test how well this model-implied covariance matrix reproduces the observed sample covariance matrix S.

---

9. LD Score Regression (LDSC)

The core equation for univariate LD Score regression relates the chi-squared (χ2) statistic of a SNP *j* from a GWAS to its LD score l_j:

$$E\left[\chi_j^2 | l_j\right] = N\frac{h^2}{M} l_j + Na + 1$$

Where:

- $E\left[\chi_j^2 | l_j\right]$ is the expected χ^2 statistic for SNP *j* given its LD score.
- *N* is the sample size of the GWAS.
- h^2 is the (SNP-based) heritability of the trait.
- *M* is the number of SNPs used to estimate h^2 (often the number of common SNPs in the reference panel).
- $l_j = \sum_k r_{jk}^2$ is the LD score of SNP *j*, calculated by summing the squared correlations (r^2) between SNP *j* and all other SNPs k in a reference panel (typically within a defined window).
- *a* measures the contribution of confounding biases, such as cryptic relatedness and population stratification. The intercept of the regression of χ^2 statistics on l_j (minus 1, scaled by N for some forms) estimates *a*.

For cross-trait LDSC, the equation looks at the product of Z-scores for two traits:

$$E\left[z_{1j} z_{2j} | l_j\right] = \frac{\sqrt{N_1 N_2}\,\rho_g}{M} l_j + \frac{\rho_N}{\sqrt{N_1 N_2}} S_j$$

(where S_j can be 1 or related to sample overlap, and ρ_g is genetic covariance, ρ_N is environmental correlation/sample overlap). The genetic correlation r_g is then derived from ρ_g and the heritabilities of the two traits.

---

10. Conversion of Nagelkerke's R^2 (R_N^2) to Liability Scale R^2 (R_L^2) (Lee et al., 2012)

This formula is used for converting Nagelkerke's R^2 from a logistic regression model (observed scale) to the proportion of variance explained on an underlying continuous liability scale, particularly in case-control studies.

The formula is:

$$R_L^2 = \frac{C \cdot R_N^2}{1 + C \cdot \theta \cdot R_N^2}$$

Where:

- R_L^2 = Proportion of variance explained on the liability scale.
- R_N^2 = Nagelkerke's R^2 (from the logistic regression model).
- K = Population prevalence of the disease/phenotype.
- P = Proportion of cases in the study sample.

And the components C and θ are derived from the following intermediate calculations:

1. zK = qnorm(1-K)
    - This is the Z-score (quantile) from a standard normal distribution corresponding to the threshold defined by the population prevalence K. (In Excel, this can be calculated as NORM.S.INV(1-K)).
2. t = dnorm(zK)
    - This is the height (probability density function, PDF) of the standard normal distribution at the threshold zK. (In Excel, NORM.S.DIST(NORM.S.INV(1-K), FALSE)).
3. $c = \frac{K^2(1-K)^2}{t^2 \cdot P(1-P)}$
    - This is a scaling constant derived from the properties of the truncated normal distribution for cases and controls.
4. $e = 1 - P^{2P} \cdot (1-P)^{2(1-P)}$
    - This is a specific scaling factor used in the Lee et al. (2012) formulation.
5. C=$c \cdot e$
    - This is the overall scaling coefficient for R_N^2 in the numerator of the main conversion formula.
6. $i = \frac{t}{K}$
    - This term i represents the mean liability of affected individuals (cases) above the threshold zK, assuming the overall population mean liability is 0 and variance is 1.

7. $\theta = i \cdot \left(\frac{P-K}{1-K}\right) \cdot \left(i \cdot \left(\frac{P-K}{1-K}\right) - \mathrm{z}K\right)$
    - This term θ accounts for ascertainment (case/control sampling) and the properties of the underlying liability distribution.

---

12. **Fisher's Exact Test** is a statistical significance test used for analyzing contingency tables, especially when sample sizes are small. For a 2x2 contingency table:

The formula is:

$$P = \frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{a!\,b!\,c!\,d!\,N!}$$

Where:

- a, b, c and d: the cell counts
- N = a + b + c + d is the total number of observations. The P-value for Fisher's Exact Test is obtained by summing the probabilities of the observed table and all other possible tables that are "more extreme" (i.e., have a lower or equal probability under the null hypothesis) while maintaining the same marginal totals.

Methods with Complex Algorithmic/Statistical Frameworks (Not easily summarized by a single equation here):

- Exploratory Factor Analysis (EFA): While based on the same common factor model as CFA, EFA is a data-driven technique to uncover latent structure, and its process involves various extraction (e.g., principal axis factoring, maximum likelihood) and rotation (e.g., varimax, promax) methods, each with its own mathematical basis.
- Structural Equation Modeling (SEM): A very broad framework that simultaneously models relationships among multiple observed and latent variables. It involves systems of linear equations and covariance structure modelling. The specific equations depend entirely on the model being tested.
- PRS-CS / PRS-CS-auto: (See Chapter 1 for basic summation) These methods use a Bayesian regression framework with continuous shrinkage priors to estimate SNP effect sizes for PRS. The underlying mathematics involves posterior distributions and Bayesian inference, which are not captured by a single equation.
- Multi-Trait Analysis of GWAS (MTAG): This is a meta-analytic method that combines summary statistics from GWAS of different traits to boost power for discovering loci for a focal trait, accounting for genetic correlations. Its derivation involves matrix operations and generalises least squares.
- Random Forest Models: An ensemble learning method based on constructing multiple decision trees. Its "equation" is the aggregated prediction of many trees, not a simple formula.
- PRSet: This involves permutation testing and aggregating PRS effects for gene sets.
- LAVA (Local Analysis of [Co]Variant Association), SBayesS, FUMA (Functional Mapping and Annotation of GWAS), MAGMA (Multi-marker Analysis of GenoMic Annotation),

TWAS (Transcriptome-Wide Association Studies): These are sophisticated bioinformatic and statistical tools or pipelines that involve multiple analytical steps, algorithms, and often external databases.

References

1. Polygenic Risk Score (PRS) - Basic Summation

- Wang, Y., Tsuo, K., Shang, L., & Li, X. (2022). Calculating Polygenic Risk Scores (PRS) in UK Biobank: A Practical Guide for Epidemiologists. *Frontiers in Genetics*, *13*, 818574. DOI: 10.3389/fgene.2022.818574. Available at: https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2022.818574/full

2. Linear Regression (Simple)

- Newcastle University. (n.d.). *Simple Linear Regression - Numeracy, Maths and Statistics - Academic Skills Kit*. Retrieved June 1, 2025, from https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/regression-and-correlation/simple-linear-regression.html

3. Logistic Regression

- The Carpentries Incubator. (n.d.). *An introduction to logistic regression*. Retrieved June 1, 2025, from https://carpentries-incubator.github.io/logistic-regression-public-health/02-introLogistic/index.html

4. Odds Ratio (OR) from Logistic Regression

- Datatab. (n.d.). *Odds Ratio in Logistic Regression: Interpreting Relationships*. Retrieved June 1, 2025, from https://datatab.net/tutorial/odds-ratio-logistic-regression

5. Nagelkerke R^2 (Pseudo R-squared)

- Number Analytics. (n.d.). *How to Interpret Nagelkerke R-Squared Metrics in Regression Models*. Retrieved June 1, 2025, from https://www.numberanalytics.com/blog/how-to-interpret-nagelkerke-r-squared-metrics-in-regression-models

6. Area Under the ROC Curve (AUC)

- Evidently AI. (n.d.). *How to explain the ROC curve and ROC AUC score?* Retrieved June 1, 2025, from https://www.evidentlyai.com/classification-metrics/explain-roc-curve
- ListenData. (2014, August 18). *A Complete Guide to Area Under Curve (AUC)*. Retrieved June 1, 2025, from https://www.listendata.com/2014/08/learn-area-under-curve-auc.html

7. Inverse Probability Weighting (IPW)

- Sant'Anna, P. H. C. (n.d.). *Lecture 15: Inverse Probability Weighted Estimators*. Retrieved June 1, 2025, from https://psantanna.com/Econ520/Slides/15-ipw/15slides.html

- Hernán, M. A., & Robins, J. M. (2020). *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC. (Note: Specific link not provided, typically a book purchase or library access). Link to publisher's page: https://www.routledge.com/Causal-Inference-What-If/Hernan-Robins/p/book/9781420076165

8. Confirmatory Factor Analysis (CFA) - Basic Measurement Model Equation

- Number Analytics. (n.d.). *Guide to Conducting Confirmatory Factor Analysis (CFA) Correctly*. Retrieved June 1, 2025, from https://www.numberanalytics.com/blog/guide-to-conducting-confirmatory-factor-analysis-cfa-correctly
- UCLA: OARC Stats. (n.d.). *Confirmatory Factor Analysis (CFA) in R with lavaan*. Retrieved June 1, 2025, from https://stats.oarc.ucla.edu/r/seminars/rcfa/

9. LD Score Regression (LDSC)

- Bulik-Sullivan, B. K., Loh, P. R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., ... & Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, *47*(3), 291-295. DOI: 10.1038/ng.3211. Available at: https://www.nature.com/articles/ng.3211
- Cloufield, GWASTutorial. (n.d.). *LD score regression by LDSC*. Retrieved June 1, 2025, from https://cloufield.github.io/GWASTutorial/08_LDSC/

10. Conversion of Nagelkerke's R^2 to Liability Scale R^2 (Lee et al., 2012)

- Lee, S. H., Goddard, M. E., Wray, N. R., & Visscher, P. M. (2012). A better coefficient of determination for genetic profile analysis. *Genetic Epidemiology*, *36*(3), 214-224. DOI: 10.1002/gepi.21614. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/gepi.21614
- Lee, S. H., Wray, N. R., Goddard, M. E., & Visscher, P. M. (2011). Estimating Missing Heritability for Disease from Genome-wide Association Studies. *American Journal of Human Genetics*, *88*(3), 294-305. DOI: 10.1016/j.ajhg.2011.02.002. Available at: https://www.cell.com/ajhg/fulltext/S0002-9297(11)00065-9

- Exploratory Factor Analysis (EFA)
    - Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272-299. DOI: 10.1037/1082-989X.4.3.272. Available at: https://psycnet.apa.org/record/1999-05172-003
    - Gorsuch, R. L. (1983). *Factor Analysis* (2nd ed.). Lawrence Erlbaum Associates. (Note: Typically, a book purchase or library access).
- Structural Equation Modeling (SEM)
    - Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford publications.
    - Bollen, K. A. (1989). *Structural Equations with Latent Variables*. John Wiley & Sons.
- PRS-CS / PRS-CS-auto
    - Ge, T., Chen, C.-Y., Ni, Y., Feng, Y. C. A., & Smoller, J. W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature*

Communications, 10(1), 1776. DOI: 10.1038/s41467-019-09718-5. Available at: https://www.nature.com/articles/s41467-019-09718-5

- Multi-Trait Analysis of GWAS (MTAG)
  - Turley, P., Walters, R. K., Maghzian, O., Okbay, A., Lee, J. J., Fontana, M. A., ... & Benjamin, D. J. (2018). Multi-trait analysis of genome-wide association summary statistics. *Nature Genetics, 50*(2), 229-237. DOI: 10.1038/s41588-017-0009-4. Available at: https://www.nature.com/articles/s41588-017-0009-4
- Random Forest Models
  - Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5-32. DOI: 10.1023/A:1010933404324. Available at: https://link.springer.com/article/10.1023/A:1010933404324
- PRSet
  - Euesden, J., Lewis, C. M., & O'Reilly, P. F. (2015). PRSice: Polygenic Risk Score software. *Bioinformatics, 31*(9), 1466-1468. DOI: 10.1093/bioinformatics/btu848. Available at: https://academic.oup.com/bioinformatics/article/31/9/1466/240999
  - Choi SW, García-González J, Ruan Y, Wu HM, Porras C, et al. (2023) PRSet: Pathway-based polygenic risk score analyses and software. PLOS Genetics 19(2): e1010624. https://doi.org/10.1371/journal.pgen.1010624
- LAVA (Local Analysis of [Co]Variant Association)
  - Werme, J., van der Sluis, S., Posthuma, D., & de Leeuw, C. A. (2022). LAVA: Local Analysis of Variant Association. *Bioinformatics, 38*(10), 2965-2967. DOI: 10.1093/bioinformatics/btac117. Available at: https://academic.oup.com/bioinformatics/article/38/10/2965/6526077
- SBayesS
  - Zeng, J., Xue, A., Jiang, L., Lloyd-Jones, L. R., Wu, Y., Wang, H., ... & Yang, J. (2021). Widespread signatures of natural selection across human complex traits and functional genomic categories. *Nature Communications, 12*(1), 1019. DOI: 10.1038/s41467-021-21446-3. Available at: https://www.nature.com/articles/s41467-021-21446-3
- FUMA (Functional Mapping and Annotation of GWAS)
  - Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature Communications, 8*(1), 1826. DOI: 10.1038/s41467-017-01261-5. Available at: https://www.nature.com/articles/s41467-017-01261-5
- MAGMA (Multi-marker Analysis of GenoMic Annotation)
  - de Leeuw, C. A., Mooij, J. M., Heskes, T., & Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. *PLoS Computational Biology, 11*(4), e1004219. DOI: 10.1371/journal.pcbi.1004219. Available at: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004219
- TWAS (Transcriptome-Wide Association Studies)
  - Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B. W., ... & Pasaniuc, B. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics, 48*(3), 245-252. DOI: 10.1038/ng.3506. Available at: https://www.nature.com/articles/ng.3506
  - Gamazon, E. R., Wheeler, H. E., Shah, K. P., Mozaffari, S. V., Aquino-Michaels, K., Carroll, R. J., ... & Cox, N. J. (2015). A gene-based association method for mapping

traits using reference transcriptome data. *Nature Genetics, 47*(9), 1091-1098. DOI: 10.1038/ng.3367. Available at: https://www.nature.com/articles/ng.3367