

The Genetic Relationship Between the Corpus Callosum, Bipolar Disorder and Alcohol Use



Megan Loraine Campbell

CMPMEG002

Thesis Presented for the Degree of
DOCTOR OF PHILOSOPHY
in Psychiatry and Mental Health

University of Cape Town

April 2024

Supervisor: Professor Dan Stein

Co-supervisors: Dr Jaroslav Rokicki & Dr Shareefa Dalvie

Date of Submission: 27 June 2024

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

Plagiarism Declaration

This study was performed from 2021 to 2024 under the supervision of Dr Jaroslav Rokicki, Dr Shareefa Dalvie and Prof. Dan Stein. It is submitted for the Degree of Doctor of Philosophy (PhD) in Psychiatry and Mental Health at the University of Cape Town (UCT). I have used the Harvard style for citation and referencing. Ethics approval was granted by the Human Research Ethics Committee for the Faculty of Health Sciences at the University of Cape Town (HREC REF: 271/2022).

I, Megan L. Campbell, hereby declare that the work on which this thesis is based is my original work (except where acknowledgements indicate otherwise) and where collaborations were involved, it has been indicated. Further, neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university. I authorize the University to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signed

Signed by candidate

Megan L. Campbell

Date: . 27 June 2024

Declaration on the Inclusion of Publications in a PhD Thesis

“I confirm that I have been granted permission by the University of Cape Town’s Doctoral Degrees Board to include the following publication(s) in my PhD thesis, and where co-authorships are involved, my co-authors have agreed that I may include the publication(s):”

List of Publications included in this thesis:

1. Campbell, M.L., Dalvie, S., Shadrin, A., van der Meer, D., O’Connell, K., Frei, O., Andreassen, O.A., Stein, D.J. and Rokicki, J., 2023. Distributed genetic effects of the corpus callosum subregions suggest links to neuropsychiatric disorders and related traits. *Acta Neuropsychiatrica*, pp.1-8.
2. Campbell, M.L., Shadrin, A., van der Meer, D., Tesfaye, M., Smeland, O.B., O’Connell, K.S., Parker, N., Frei, O., Andreassen, O., Stein, D., Dalvie, S. and Rokicki, J., 2023. Regional Genetic Architecture of the Corpus Callosum and its Overlap with Psychiatric and Substance Use Phenotypes. *medRxiv*, pp.2023-12.

Signed

Signed by candidate

Megan L. Campbell

Date: 27 June 2024

Abstract

Background: The corpus callosum (CC) is a brain structure responsible for the integration of information across the two cerebral hemispheres to facilitate language, affective- and sensorimotor function. Morphological alterations in sub-regions of this structure have been observed in various psychiatric disorders, particularly bipolar disorder and alcohol consumption. Although the CC has a high heritability (0.67 for total CC; family study) the genetic architecture of the entire structure, its subregions, and any genetic links with psychiatric disorders remain largely unclear. Therefore, the aim of this study was to investigate the genetic architecture of the CC, and this was accomplished by the following objectives: elucidating the single nucleotide polymorphism (SNP)- based heritability and polygenic architecture of the CC and its subregions, and by investigating the genetic overlap between the CC and its subregions with psychiatric disorders, including bipolar disorder and alcohol consumption.

Methods: Multivariate and univariate genome-wide association studies (GWAS) were conducted utilizing genetic and T1-weighted magnetic resonance imaging (MRI) data of 40,894 individuals from the UK Biobank. The volume of the five subregions of the CC (posterior, mid-posterior, central, mid-anterior and anterior) were extracted using FreeSurfer v7.1. Multivariate GWAS was run combining volumes of all CC subregions. Multivariate GWAS was conducted to boost genetic discovery and to assess the pleiotropic effects across volumes of the five subregions of the CC. Univariate GWAS was conducted across CC subregion volumes and the 28 diffusion weighted imaging modalities. All analyses were co-varied for age, age squared, sex, scanning site, Euler score (a proxy of image quality), total intracranial volume and the first 20 genetic principal components. Annotations of GWAS summary statistics were conducted using OpenTargets and gene-set enrichment analyses were performed using hypergeometric tests. Linkage disequilibrium score regression (LDSC) was

used to determine SNP-based heritability of the CC and subregions (for both volume and diffusion weighted modalities), as well as to assess genetic correlations between these subregions and relevant psychiatric and substance use phenotypes. A causal mixture model framework was employed to determine the polygenicity of each subregion and related psychiatric traits and, leveraging pleiotropy using the conjunctive false discovery rate (cFDR) framework, shared loci between the CC and psychiatric phenotypes were identified.

Results: This study identified seventy independent loci with distributed effects across volumes of the CC subregions, with additional loci showing subregion-specific associations. Univariate GWAS analysis of diffusion metrics revealed a total of 1,520 significant genetic loci associated with diffusion metrics in the subregions of the CC. The polygenicity varied across diffusion modalities and subregions, ranging from 268 to 7,069 trait-influencing variants. The mean number of trait-influencing variants ranged from 1,560 to 2,865 across different diffusion metric groups and 2,385 to 3,108 across subregions. It was found that CC subregion volumes are influenced by approximately 1,400 to 1,800 genetic variants, with approximately 1,800 variants influencing total CC volume. The SNP-based heritability estimates varied, with CC volume showing a relatively high heritability ($h^2_{snp}=0.38$, $SE=0.03$) and subregion volumes ranging from 0.22 ($SE=0.02$; mid-posterior) to 0.37 ($SE=0.03$; posterior). Significant genetic correlations were observed between CC volume and bipolar disorder, major depressive disorder, drinks per week, cannabis use. The diffusion metrics also showed subregion-specific correlations with CC volume, bipolar disorder and alcohol consumption. The genetic overlap between the CC and psychiatric phenotypes differed across subregions, with anterior regions showing almost complete overlap with bipolar disorder and schizophrenia, while the central region exhibited substantial overlap with drinks per week. In conjunctive analysis, I identified 138 significant independent variants between CC subregions and psychiatric

phenotypes, and gene-set enrichment analysis implicated antigen processing, regulation of the extracellular matrix, and hyaluronan metabolism as significant pathways.

Conclusion: These results demonstrate that the CC possesses a polygenic architecture and has a substantial SNP-based heritability across its subregions. Distinct genetic loci were implicated in the CC's anterior and posterior subregions, which is in line with their different functional specializations. There was evidence of significant genetic correlation of the CC with BD, MDD, as well as between the CC and substance use phenotypes (drinks per week and cannabis consumption). Moreover, substantial genetic overlap was observed between the CC with SCZ and BD, and alcohol consumption. These findings of overlapping genetic architectures of the CC and these psychiatric phenotypes suggest the existence of shared biological pathways that may have value in the development of targeted treatment strategies.

Acknowledgements

I would like to express my gratitude to the following individuals and groups whose support and contributions have been instrumental in the completion of my thesis:

To my incredible supervisors- Prof Dan Stein, as always thank you for the myriad of opportunities and the (often I feel, misplaced) belief in my abilities. Dr Shareefa Dalvie, thank you for your gentle mentorship and guidance. Dr Jaroslav Rokicki, where to begin. Thank you for the incredible amount of time invested in this project and myself. So many of the skills gained in the last few years are a direct reflection of your time and effort. You have been an incredible mentor, and I am truly grateful- thank you.

I would like to express my gratitude to Dr Ole Andreassen for his invaluable support and for providing access to resources that were essential to the realization of this project. Dr Alexey Shadrin, thank you for the guidance, statistical support and comedic shows in Jarek's office.

My appreciation also extends to the research participants of the UK Biobank, without which this PhD would not have been possible. The time that you dedicated to this research and your contributions to the field are appreciated.

Thank you to Olivia for being a fantastic office mate, despite being thirteen thousand km's away- thank you for making the coffee and washing my mug afterwards. It's been tough in your absence.

To my folks, needy animals at home, and Angus, I couldn't have done it without you.

Preface

Alterations in the corpus callosum are well documented in many psychiatric disorders. Previous research has demonstrated that both volume and microstructure of the corpus callosum are highly heritable, and there is evidence to suggest a genetic contribution to the corpus callosum. However, there are several gaps in our knowledge regarding the genetic architecture of the corpus callosum and its genetic relationship to related psychiatric phenotypes.

Research Concept and Funding

This thesis aims to use novel statistical techniques to extend the current understanding of the genetic relationship between the corpus callosum and related psychiatric phenotypes. The candidate independently conceived and conducted the research described in this thesis, with guidance and supervision from their primary supervisor, Professor Dan Stein, and their co-supervisors, Dr Jaroslav Rokicki and Dr Shareefa Dalvie. The work presented in this thesis was supported, in part, by National Institute of Mental Health (NIMH: Grant number R01MH120642), and The Research Council of Norway (275054).

Structure of the Thesis

This thesis is presented as three distinct sections: 1) genetic architecture of the corpus callosum and its subregions, 2) the genetic contribution to white matter microstructure of the corpus callosum, and 3) the genetic relationship between the corpus callosum and related psychiatric phenotypes

The thesis begins with a general introductory chapter and an outline of the aims and objectives. The thesis concludes with a chapter dedicated to discussing and synthesizing the results presented in each chapter. The thesis includes two publications, with each intended to contribute to the scientific literature on the subject matter.

Ethical Standards

The work described in this thesis was approved by The University of Cape Town Human Research Ethics Committee (reference number - 271/2022). This thesis is a secondary data analysis but all original study protocols comply with the ethical standards of the relevant national and institutional ethics committee and with the Helsinki Declaration of 1975, as revised in 2008.

Table of Contents

Plagiarism Declaration	2
Declaration on the Inclusion of Publications in a PhD Thesis.....	3
Abstract	4
Acknowledgements	7
Preface	8
List of Figures	13
List of Tables.....	16
Abbreviations	17
1 Introduction	20
1.1 The Corpus Callosum	20
1.1.1 Structure and Function of the Corpus Callosum	20
1.1.2 Heritability of the Corpus Callosum	21
1.2 The Role of the Corpus Callosum in Psychiatric Phenotypes.....	22
1.2.1 The Corpus Callosum in Bipolar Disorder and Alcohol Use	24
1.3 Bipolar Disorder and Alcohol Use	25
1.3.1 The Relationship Between Bipolar Disorder and Alcohol Use	25
1.3.2 Genetic Risk Factors for Bipolar Disorder and Alcohol Use	26
1.4 Novel Statistical Methods to Study Genetics of the Corpus Callosum ad its Role in Psychiatric Phenotypes	27
2 Aims and Objectives.....	31
2.1 Study Rationale.....	31
2.2 Aims and Objectives	31
3 Chapter 3: Original Research Article 1.....	34
<i>Distributed Genetic Effects of the Corpus Callosum Subregions Suggests Links to Neuropsychiatric Disorders and Related Traits</i>	<i>34</i>
Synopsis	34
Authorship Statement	34
3.1 Abstract.....	36
3.2 Introduction	38
3.3 Methods and Materials	39
3.3.1 Processing of Imaging and Genetic Data	39
3.3.2 Genome-wide association analyses: Multivariate and Univariate.....	41
3.3.3 Functional annotation.....	42
3.3.4 SNP-Based Heritability and Genetic Correlation.....	42
3.4 Results	43
3.4.1 Genome-wide association analyses.....	43
3.4.2 Functional annotation.....	47
3.4.3 SNP-based heritability and Genetic Correlation	47
3.5 Discussion.....	50

4	Chapter 4: Original Research Article 2	55
	<i>Genetic Determinants of White Matter Integrity in the Corpus Callosum</i>	55
	<i>Synopsis</i>	55
	<i>Author Statement</i>	55
4.1	<i>Abstract</i>	57
4.2	<i>Introduction</i>	59
4.3	<i>Methods</i>	61
4.3.1	Processing of Imaging and Genetic Data	61
4.4	<i>Results</i>	66
4.4.1	Genome-wide association analyses	66
4.4.2	Polygenicity	69
4.4.3	SNP-based heritability and Genetic Correlation	69
4.4.4	Functional Annotation	73
4.5	<i>Discussion</i>	74
5	Chapter 5: Original Research Article 3	79
	<i>Regional Genetic Architecture of the Corpus Callosum and its Overlap with Psychiatric and Substance Use Phenotypes</i>	79
	<i>Synopsis</i>	79
	<i>Authorship Statement</i>	79
5.1	<i>Abstract</i>	81
5.2	<i>Introduction</i>	83
5.3	<i>Methods</i>	85
5.3.1	Genome-wide Association Studies	85
5.3.2	Analysis of genetic architecture using causal mixture modelling	85
5.3.3	Identification of shared genetic loci with conjunctional false discovery rate	86
5.3.4	Functional annotation of shared loci	87
5.4	<i>Results</i>	87
5.4.1	Analysis of genetic architecture using causal mixture modelling	87
5.4.2	Identification of shared genetic loci with conjunctional false discovery rate	91
5.4.3	Functional annotation of shared loci	92
5.5	<i>Discussion</i>	94
6	Discussion	98
6.1	<i>Main Findings</i>	98
6.2	<i>The Discovery of Genomic Loci Contributing to the Corpus Callosum and its Underlying Biology</i>	99
6.3	<i>The Polygenic Architecture of the Corpus Callosum and its Underlying Heritability</i>	101
6.3.1	The Genetic Relationship Between the Corpus Callosum, Bipolar Disorder and Alcohol Consumption	102
7	Future Directions	104
7.1	<i>Strengths and Limitations</i>	104
7.2	<i>Validation and Harmonization of Segmentation Algorithms</i>	105
7.3	<i>Exploring Other Forms of Genetic Variation in Relation to the Corpus Callosum</i>	105
7.4	<i>Functional Assessment of Identified Loci</i>	106
8	Conclusion	108

References.....	109
Appendices.....	120
<i>Appendix 1: Supplementary Figures and Tables for Chapter 3.....</i>	<i>120</i>
<i>Appendix 2: Supplementary Figures and Tables for Chapter 4.....</i>	<i>125</i>
<i>Appendix 3: Supplementary Figures and Tables for Chapter 5.....</i>	<i>126</i>

List of Figures

Figure 1.1 A) Sagittal section of the brain, illustrating the corpus callosum and the five subregions of the corpus callosum B) The five subregions of the corpus callosum	21
Figure 1.2 Findings from neuroimaging studies of the corpus callosum subregions in bipolar disorder and alcohol use. Findings are summarised from Gifuni et al., (2017), Prunas et al., (2018) and de Souza et al., (2019). These studies indicate some independent effects on subregions of the corpus callosum for alcohol use and bipolar disorder, respectively; however, there is also evidence to suggest that both disorders are associated with reductions in central subregion of the corpus callosum.....	25
Figure 1.3 The four major theories that aim to explain the comorbid relationship between bipolar disorder and substance use. Adapted from Polimanti et al., (2017).....	27
Figure 3.1 Miami plot of the $-\log_{10}(p)$ statistics from the multivariate (top) and univariate (bottom) GWAS of total CC volume. A) Top: The multivariate analysis was conducted using MOSTest and illustrate the distributive genetic architecture of the CC. Bottom: Individual univariate GWAS were conducted for total CC volume and each subregion, indicating both a shared and unique genetic etiology of the CC subregions. B) Top SNPs from the multivariate GWAS show subregion specific effects across the CC. The top ten SNPs from the multivariate GWAS were mapped to the CC subregions using the Z-scores from the respective univariate subregion GWAS. The results indicate a gradient of effect across the posterior and anterior subregions.....	45
Figure 3.2 A) Manhattan plots from univariate GWAS of each CC subregion B) Quantile-Quantile plot (QQ plot) of observed versus expected $-\log_{10}(p\text{-values})$ the GWAS of CC subregions. The shaded regions represent the expected null distribution of $-\log_{10}(p\text{-values})$ under the assumption of no association. C) Volumetric correlations of CC subregions, each square represents the r-squared values	46
Figure 3.3 A) SNP based heritability estimates of total CC volume and volume of the subregions. LDSC was used to estimate the SNP-based heritability of total CC volume and the five subregions. A schematic of the CC and the five subregions, extracted using FreeSurfer, based on the Witelson segmentation. The CC and subregions have been overlaid onto an anatomical reference image. B) Heatmap of genetic correlations between CC subregions and relevant psychiatric and substance use traits each square represents the r-squared values, significant correlations are denoted by *	49

Figure 5.1 Manhattan plots of GWAS across each subregion and diffusion metric. Red dots indicate significant loci with $p < 2.8 \times 10^{-7}$. Significance level is indicated by dashed line. Manhattan plots are plotted in the same order from inner to outer circle: Anterior, mid anterior, central, mid posterior, posterior subregion. Figure illustrating Bayesian rotationally invariant approach, BRIA and diffusion kurtosis imaging, DKI.....67

Figure 5.2 Manhattan plots of GWAS across each subregion and diffusion metric. Red dots indicate significant loci with $p < 2.8 \times 10^{-7}$. Significance level is indicated by dashed line. Manhattan plots are plotted in the same order from inner to outer circle: Anterior, mid anterior, central, mid posterior, posterior subregion. Figure illustrating diffusion tensor imaging, DTI; spherical mean technique, SMT and multicompartmental SMT, SMTmc 68

Figure 5.3 Polygenicity and SNP-based heritability for the CC subregions, across diffusion metrics. A) Barplot showing estimated number of trait influencing variants, that explain 90% of the subregions' heritability, grouped by diffusion metric and stratified by subregion. B) SNP-based heritability estimates of the CC subregions, grouped by diffusion metric.....71

Figure 5.4 Genetic correlations across diffusion metrics of the CC and psychiatric phenotypes for alcohol use disorder, AUD; drinks per week, cannabis, cigarettes per day, bipolar disorder, BD; major depressive disorder, MDD and schizophrenia, SCZ. Results are grouped by CC subregion. Significant results ($p < 0.053.5 \times 10^{-4}$) are marked with *.....72

Figure 5.5 PheWAS results from significant genes, implicated in the univariate GWAS 73

Figure 4.1 Polygenicity of the CC and psychiatric phenotypes illustrated by estimated causal variants. Polygenicity of volumes of the corpus callosum (CC) subregions (anterior, mid anterior, central, mid posterior, posterior), total CC volume, drinks per week, cigarettes per day, smoking cessation, smoking initiation, cannabis, attention deficit hyperactivity disorder (ADHD), autism spectrum disorder (ASD), bipolar disorder (BD), major depressive disorder (MDD), schizophrenia (SCZ).....88

Figure 4.2 Shared trait influencing variants of the CC and psychiatric phenotypes. A-F) The percentage of concordant/ discordant SNPs from the number of shared trait influencing variants between variants (percentages are calculated with respect to total CC variants) G) Heatmap illustrating the estimated percentage of shared variants, calculated with respect to the CC H) Proportion of shared variants, showing the same direction of effect.....90

Figure 4.3 Manhattan plots of conjFDR results illustrated by subregion. A-F) ConjFDR results by subregion. Red dots indicated significant loci with $FDR < 0.05$. Significance level is indicated by dashed red line. Manhattans are plotted in the same order from inner circle to outer circle: Drinks per week, Cigarettes per day, Smoking cessation, Smoking initiation, ASD,

Autism spectrum disorders; ADHD, Attention deficit hyperactivity disorder; BD, Bipolar disorder; MDD, Major depressive disorder; SCZ, Schizophrenia 93

A1: Supplementary Material Figure 1 The differences between A) FreeSurfer and the B) Hofer and Frahm segmentation. FreeSurfer extracts volume, while the Hofer & Frahm segmentation, conducted using C8 software extracts thickness 120

A1: Supplementary Material Figure 2 Overview of the GWAS conducted in this study 120

A1: Supplementary Material Figure 3 Manhattan plots from the univariate GWAS of each CC subregion, co-varying for total CC volume. A) 42 SNPs were significant for total CC volume, B) 31 for volume of the anterior subregion, C) 22 for volume of the mid anterior subregion D) 26 for volume of the central subregion, E) 7 for volume of the mid posterior subregion F) 49 for volume of the posterior subregion 121

A1: Supplementary Material Figure 4 Overlap between the FreeSurfer and Hofer and Frahm segmentation, multivariate analyses. A) Miami plot illustrating the $-\log(p\text{-value})$ multivariate GWAS results from the top) FreeSurfer and bottom) Hofer & Frahm segmentation of the CC subregions B) illustrating the significant, independent SNPs in both segmentations and C) the number of overlapping genes from the gene-based association tests . The multivariate analyses were conducted using MOSTest and illustrates the distributive genetic architecture of the CC. SNPs in red are those that are significant, independent loci in both analyses. Nineteen independent loci were identified in the multivariate GWAS of the Hofer and Frahm segmentation; 58% (7 loci) were also significant in the multivariate GWAS of the FreeSurfer segmentation. In the gene-based association tests, 11 of the 13 significant genes overlapped with results based on the FreeSurfer segmentation. 122

A1: Supplementary Material Figure 5 Manhattan plots of gene-based association tests, conducted through FUMA. A) Anterior, B) Mid anterior, C) Central, D) Mid posterior, E) Posterior 123

A3: Supplementary Material Figure 1 predicted vs observed conditional quantile-quantile plots and log likelihood plots that visualize the parameter estimation procedure 128

List of Tables

Table 5.1 Full list of diffusion metrics examined	63
A1: Supplementary Material Table 1 Significant loci from univariate GWAS	124
A1: Supplementary Material Table 2 Replication of significant loci from univariate GWAS	124
A1: Supplementary Material Table 3 Gene set enrichment analysis results	124
A1: Supplementary Material Table 4 PheWAS results from significant loci and overlap with previous GWAS	124
A1: Supplementary Material Table 5 SNP-based heritability estimates from two segmentation approaches	124
A1: Supplementary Material Table 6 Genetic correlation results between the CC and psychiatric and substance use traits	124
A2: Supplementary Material Table 1 Significant genetic loci from univariate GWAS	125
A2: Supplementary Material Table 2 Genetic correlation results between CC subregions and psychiatric and substance use phenotypes, across 28 diffusion metrics	125
A2: Supplementary Material Table 3 Results from the gene-set enrichment analysis across 28 diffusion metrics	125
A3: Supplementary Material Table 1 Descriptions of the original GWAS	126
A3: Supplementary Material Table 2 Results from the conjunctional FDR analysis of the psychiatric disorders	128
A3: Supplementary Material Table 3 Results from the conjunctional analysis of the substance use traits	128
A3: Supplementary Material Table 4 Biologival processes implicated by the significant loci from the conjunctional FDR analyses	128

Abbreviations

Abbreviation	Term
AD	Axial diffusivity
ADHD	Attention Deficit Hyperactivity Disorder
AK	Axial kurtosis
ASD	Autism Spectrum Disorder
AUD	Alcohol Use Disorder
AWF	Axonal water fraction
axEAD	Axial extra-axonal diffusivity
BD	Bipolar Disorder
BRIA	Bayesian Rotationally Invariant Approach
CC	Corpus Callosum
conjFDR	Conjunctional False Discovery Rate
DAX extra	Extra-axonal axial diffusivity
DAX intra	Intra-axonal axial diffusivity
DKI	Diffusion Kurtosis Imaging
DRAD extra	Extra-axonal radial diffusivity
DTI	Diffusion Tensor Imaging
FA	Fractional Anisotropy
FUMA	Functional Mapping and Annotation of Genome-wide Association Studies
GSEA	Gene-Set Enrichment Analysis
GWAS	Genome-Wide Association Study
h^2_{snp}	SNP-Based Heritability

HWE	Hardy-Weinberg Equilibrium
ICV	Intracranial Volume
LDSC	Linkage Disequilibrium Score Regression
MAF	Minor Allele Frequency
MAGMA	Multi-marker Analysis of GenoMic Annotation
mcSMT	Multi-Compartment Spherical Mean Technique
mcSMT extra	Extra-neurite microscopic mean diffusivity (Multi-Compartment
md	Spherical Mean Technique)
mcSMT extra	Extra-neurite transverse microscopic diffusivity (Multi-Compartment
trans	Spherical Mean Technique)
mcSMT intra	Intra-neurite volume fraction (Multi-Compartment Spherical Mean
	Technique)
MD	Mean diffusivity
MDD	Major Depressive Disorder
MHC	Major Histocompatibility Complex
micro ADC	Microscopical apparent diffusion coefficient
micro AX	Microscopical axial diffusivity
micro FA	Microscopic fractional anisotropy
micro RD	Microscopic radial diffusivity
MiXeR	Mixture Model for Estimating the Number of Causal Variants
MK	Mean kurtosis
MRI	Magnetic Resonance Imaging

PGC	Psychiatric Genomics Consortium
PheWAS	Phenome-Wide Association Study
QC	Quality Control

radEAD	Radial extra-axonal diffusivity
RD	Radial Diffusivity
RK	Radial kurtosis
SCZ	Schizophrenia
SE	Standard Error
SMT	Spherical Mean Technique
SMT FA	Fractional anisotropy (Spherical Mean Technique)
SMT long	Longitudinal diffusion coefficient (Spherical Mean Technique)
SMT mcd	mc SMT diffusion coefficient (Spherical Mean Technique)
SMT md	Mean diffusivity (Spherical Mean Technique)
SMT trans	Transverse diffusion coefficient (Spherical Mean Technique)
SMTmc	Multi-compartment Spherical Mean Technique
SNPs	Single Nucleotide Polymorphisms
SUDs	Substance Use Disorders
V extra	Extra-axonal water fraction
V intra	Intra-axonal water fraction
vCSF	Cerebrospinal fluid fraction
WMTI	White Matter Tract Integrity

1 Introduction

In this thesis, I proposed to investigate the genetic relationships between a white matter structure in the brain, the corpus callosum (CC), and related psychiatric and substance use phenotypes-(attention deficit hyperactivity disorder (ADHD), autism spectrum disorder (ASD), bipolar disorder (BD), major depressive disorder (MDD), schizophrenia (SCZ), cigarettes per day, drinks per week, smoking initiation, smoking cessation and cannabis use with a particular focus on BD and alcohol use. Previous studies have linked alterations in the CC to several psychiatric disorders, particularly BD and alcohol consumption. Additionally, evidence suggests that alcohol consumption affects CC morphometry, particularly in the anterior subregion, regardless of an alcohol use disorder diagnosis. In this introduction, I provide a brief outline of the structure, function, and heritability of the CC. I then discuss the findings of previous studies which have found a link between BD, alcohol consumption and the CC. Lastly, I discuss the use of novel statistical methods to assess the genetic relationship between the CC and related psychiatric and substance use phenotypes.

1.1 The Corpus Callosum

1.1.1 Structure and Function of the Corpus Callosum

The CC (Figure 1.1A) is the largest white matter tract in the brain, and serves as the major interhemispheric commissure, integrating information from the two cerebral hemispheres to facilitate language, affective and sensorimotor function (Wang *et al.*, 2020). The CC is typically described as being made up of five subregions: the genu, rostrum, body, isthmus and splenium regions (Figure 1.1B).

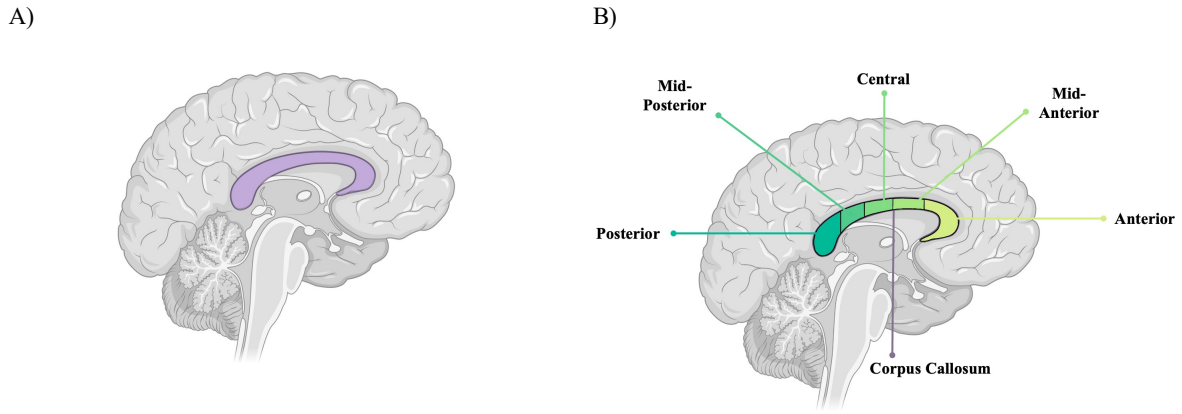


Figure 1.1 A) Sagittal section of the brain, illustrating the corpus callosum and the five subregions of the corpus callosum B) The five subregions of the corpus callosum

Although the CC is one of the few white matter tracts that can be discretely identified by conventional structural images such as T1w and T2w obtained through magnetic resonance imaging (MRI), these modalities do not reveal the morphologically discernible structures relating to the CC subregions described above (Cover *et al.*, 2018). In contrast, studies using diffusion weighted imaging (DWI) tractography have delineated functional subregions of the CC by tracing the CC fibre projections into prefrontal, premotor, primary motor, and primary sensory areas as well as into parietal, temporal and occipital cortical areas (Hofer and Frahm, 2006). Current methods developed to segment the CC have not yet been able to overcome many of the challenges associated with CC segmentation using T1w alone, suggesting that a combined approach utilizing volumetric and DWI is needed to provide insight into functional segmentation of the CC (Cover *et al.*, 2018).

1.1.2 Heritability of the Corpus Callosum

The size of the CC has been shown to be highly heritable in humans, with a heritability of 0.67 for total CC volume and heritability of the subregions ranging from 0.48 (splenium subregion) to 0.85 (genu subregion; Woldehawariat *et al.*, 2014). Diffusion metrics of the

CC such as fractional anisotropy (FA) and radial diffusivity (RD) have also been shown to be under genetic influence. In a study which investigated the contribution of common genetic variants to white matter structure of the brain, RD and FA were found to have the largest number of SNPs associated with the CC, compared to other brain regions examined (Zhao *et al.*, 2020, 2021). Furthermore, an investigation into the heritability of the CC within a population isolate, employing a family study design revealed a notable heritability for FA ($h^2=0.67 \pm 0.22$) and RD ($h^2=0.72 \pm 0.22$). However, given the functional specialization of these areas, it's important to consider the subregions of the CC. Notably, previous work examining the genetic contribution to subcortical structures, including the hippocampus and amygdala, respectively, has shown that subregions of both structures have distinct and shared genetic architectures, providing significant insight into the neurobiology of these structures (van der Meer, Rokicki, *et al.*, 2020; Elvsåshagen *et al.*, 2021; Mufford, van der Meer, *et al.*, 2021; Bahrami *et al.*, 2023; Carrión-Castillo and Boeckx, 2024). However, investigations into the genetic architecture of the CC and its subregions remain limited.

1.2 The Role of the Corpus Callosum in Psychiatric Phenotypes

Neuroimaging studies have provided insight into the structural and functional differences in the CC among individuals with various neuropsychiatric and substance use disorders (SUDs). Studies in BD have identified altered white matter integrity in the CC, potentially affecting interhemispheric communication (Wise *et al.*, 2016). Furthermore, individuals with SCZ show structural abnormalities such as lower thickness and altered shape in the CC, indicating disrupted connectivity between brain regions (Frumin *et al.*, 2002; L. Lu *et al.*, 2011). In BD and SCZ, disrupted interhemispheric communication may contribute to the cognitive and emotional dysregulation observed in these disorders (Frumin *et al.*, 2002; Wise *et al.*, 2016). In individuals with ADHD, there is evidence of decreased volume and alterations

in the microstructure of the CC, indicating disruptions in interhemispheric connectivity (Acer *et al.*, 2017). Similarly, individuals with ASD exhibit differences in CC size, shape, and connectivity, suggesting atypical development of white matter microstructure. Substance use-related traits have also been associated with alterations in CC volume, but findings regarding the effects of alcohol consumption, cigarette smoking, and cannabis use on CC volume are inconsistent (Hampton *et al.*, 2019; Moeller *et al.*, 2004).

Additionally, alterations in white matter integrity, particularly within the CC, have been associated with a range of psychiatric and substance use phenotypes. For instance, subregion-specific alterations of CC integrity have been linked to various psychiatric disorders. Specifically, alterations in FA and RD in the anterior subregion of the CC have been associated with positive symptoms of schizophrenia (Whitford *et al.*, 2010), and lower FA in the posterior and anterior subregions of the CC has consistently been found in BD (Wang *et al.*, 2008; L. Lu *et al.*, 2011; Brown *et al.*, 2021). Moreover, children with ADHD showed widespread decreased FA and higher RD, particularly focused in the CC, with hyperactivity and higher impulsivity associated with higher FA in the CC (Wu *et al.*, 2017). The CC has been suggested to form part of a network underlying emotional regulation in major depression, with lower anisotropy a well-replicated finding (Cole *et al.*, 2012; Liang *et al.*, 2019). Despite the well-established alterations in FA within the CC, there remains a paucity of data on the genetic basis of other multicompartiment DWI metrics and their relationship with neuropsychiatric and substance use traits. These findings underscore the importance of investigating the CC structural integrity using advanced neuroimaging techniques to better understand its role in the pathophysiology of psychiatric and substance use phenotypes.

1.2.1 The Corpus Callosum in Bipolar Disorder and Alcohol Use

As described above, differences in the volume of the CC have been associated with several psychiatric disorders such as mood and SUDs (Rosenbloom, Sullivan and Pfefferbaum, 2003; Newbury and Rosen, 2012; Abramovic *et al.*, 2018; Prunas *et al.*, 2018; de Souza *et al.*, 2019). However, diminished connectivity of the CC may underlie some of the functional abnormalities in BD (Arnone *et al.*, 2008). BD patients have consistently been shown to have significantly smaller CC subregions in comparison to healthy controls (Arnone *et al.*, 2008; Gifuni *et al.*, 2017; Prunas *et al.*, 2018). Volumetric MRI studies have found lower CC volume in individuals with BD and their healthy co-twins, suggesting that these abnormalities may be related to genetic factors (Thompson *et al.*, 2003). Further, volumetric reductions in the CC have been observed in patients with BD, with relatives of these individuals showing intermediate volumetric reductions between controls and BD probands (Walterfang *et al.*, 2009; Francis *et al.*, 2016).

There is evidence to suggest that alcohol consumption affects CC morphometry (de Souza *et al.*, 2019). These effects appear to be subregion specific, whereby the volume of the anterior CC subregion has a negative association to alcohol consumption (i.e., the higher the amount of alcohol consumption, the smaller the anterior CC volume), irrespective of an alcohol use disorder diagnosis (de Souza *et al.*, 2019). It is crucial to differentiate between substance use and SUDs in this context, as the former refers to a broader range of consumption patterns that may not necessarily entail the compulsive, harmful patterns or the clinical severity associated with disorders (Schuckit, 2009). This distinction is significant as even non-disordered levels of consumption can cause changes in the CC, underscoring the importance of examining the full spectrum of alcohol use and its potential effects on brain morphology. While frontal and superior subregions of the CC appear to be most affected by alcohol consumption, the inferior

and posterior CC subregions remain relatively preserved (de Souza *et al.*, 2019). Although overall reduction in CC volume is apparent in both alcohol use and BD, there is also evidence to suggest both distinct and shared differences in CC morphology in these disorders (Figure 1.2). However there are no studies to date, that have investigated the genetic relationship between CC subregions, BD and alcohol use, despite overlapping altered morphometry in this region.

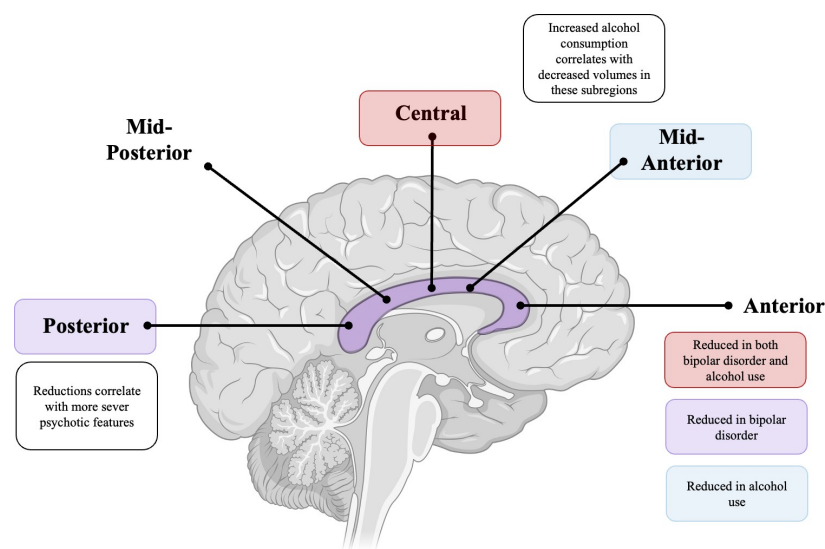


Figure 1.2 Findings from neuroimaging studies of the corpus callosum subregions in bipolar disorder and alcohol use. Findings are summarised from Gifuni *et al.*, (2017), Prunas *et al.*, (2018) and de Souza *et al.*, (2019). These studies indicate some independent effects on subregions of the corpus callosum for alcohol use and bipolar disorder, respectively; however, there is also evidence to suggest that both disorders are associated with reductions in central subregion of the corpus callosum.

1.3 Bipolar Disorder and Alcohol Use

1.3.1 The Relationship Between Bipolar Disorder and Alcohol Use

BD and SUDs are often studied as distinct conditions, however these disorders are often comorbid (Lippard *et al.*, 2017). For example, relative to the general population, individuals with BD show an higher risk of heavy alcohol use (Hartz *et al.*, 2014). In turn, early alcohol use has been found to be predictive of depressive disorders, anxiety disorders, and BD (Hill,

2017; Quattrone *et al.*, 2020). Comorbid alcohol use in BD is associated with more severe illness outcomes, including higher impulsivity, greater severity of mood episodes, cognitive deficits and an increase in the already high risk of suicide attempts (Lippard *et al.*, 2017). Despite the importance of understanding the mechanisms underlying this comorbidity, there has been a paucity of data on the neural and genetic risk factors underlying this comorbid relationship (Lippard *et al.*, 2017).

1.3.2 Genetic Risk Factors for Bipolar Disorder and Alcohol Use

Both BD and alcohol consumption have considerable heritability (0.7 and 0.65, respectively; family studies; McGuffin *et al.*, 2003; Edvardsen *et al.*, 2008; De Moor *et al.*, 2011). The comorbidity between BD symptomatology and alcohol use leads to a significant impairment in affected individuals, however, there is little work examining the shared genetic basis of these phenotypes. The underlying genetic aetiology of this comorbidity is yet to be established however there are four major theories that aim to explain this (Polimanti, Agrawal and Gelernter, 2017; Figure 1.3). First, the “self-medication” hypothesis suggests that individuals with BD develop a SUD because they feel it improves some of their symptoms (Awad and Voruganti, 2015). Second, BD and SUDs are influenced by a common pathophysiology (i.e., shared genetic mechanisms; Walters *et al.*, 2018). Third, in the diathesis-stress model, genetic predisposition interacts with chronic substance use, and this interaction results in the onset of BD (Strakowski and DelBello, 2000). Fourth, impaired social and occupational function due to BD together with exposure to poor social environments coincides with an higher risk of substance use (Strakowski and DelBello, 2000).

A recent genome-wide association study (GWAS) found evidence to suggest pleiotropy and genome-wide correlation between BD and alcohol use (Martínez-Magaña *et al.*, 2021; Bipolar

Disorder Working Group of the Psychiatric Genomics Consortium *et al.*, 2023) supporting the second theory described by Polimanti *et al* (2017) in the paragraph above. Further, this study found individuals with higher BD polygenic risk scores had a higher prevalence of risky alcohol use (Martínez-Magaña *et al.*, 2021). Although this GWAS is the first to describe the comorbid genetic relationship between BD and alcohol use, sample size was limited and focused on individuals of Mexican ancestry only, which may limit its relevance to other population groups.

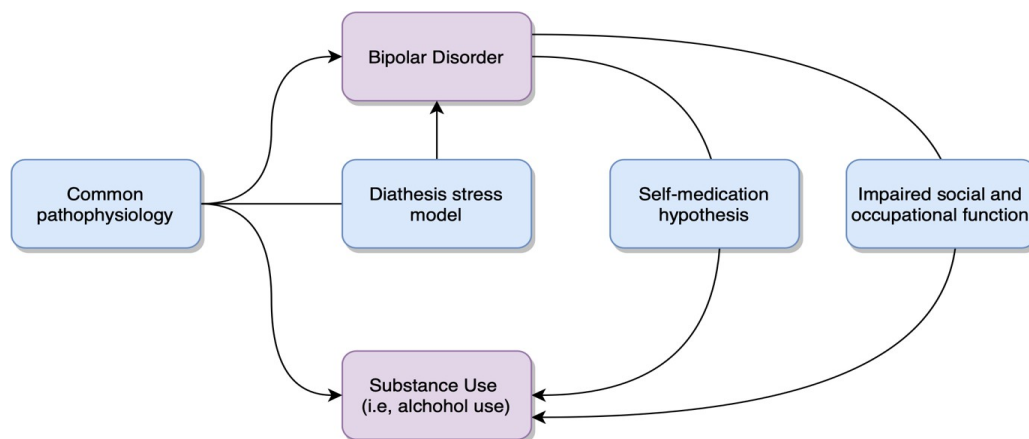


Figure 1.3 The four major theories that aim to explain the comorbid relationship between bipolar disorder and substance use however, it is important to note that these theories are not necessarily mutually exclusive. Adapted from Polimanti *et al.*, (2017)

1.4 Novel Statistical Methods to Study Genetics of the Corpus Callosum and its Role in Psychiatric Phenotypes

A GWAS is a large-scale, hypothesis free approach that involves scanning the genome of many individuals to identify genetic variations associated with a particular disease or trait (Yengo *et al.*, 2018). By comparing genetic variants, particularly SNPs across a diverse population, GWAS can detect markers linked to various conditions (Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium, 2017; Trubetskoy *et al.*, 2022; Bipolar Disorder Working Group of the Psychiatric Genomics Consortium *et al.*, 2023;

Demontis *et al.*, 2023). This methodology has significantly contributed to our understanding of the genetic basis of complex traits, diseases, and phenotypes (Atkinson *et al.*, 2021). Through GWAS, numerous genetic loci associated with a wide range of conditions have been identified, shedding light on disease mechanisms, identifying potential drug targets, and enabling personalized medicine approaches (Ratnakumar *et al.*, 2020). The comprehensive nature of GWAS allows for capturing the polygenic architecture of complex traits, providing insights into the collective impact of multiple genetic variants on trait variability (Shadrin *et al.*, 2020; Hindley *et al.*, 2021; Tesfaye *et al.*, 2023). GWAS has been particularly effective in delineating the polygenic nature of psychiatric disorders and has identified hundreds of risk loci associated with such conditions and brain-related traits, underscoring the value of brain phenotypes as intermediate phenotypes for psychiatric research (van der Meer, Rokicki, *et al.*, 2020; Mufford, van der Meer, *et al.*, 2021; Zhao *et al.*, 2021; Bahrami, Nordengen, *et al.*, 2022; Carrión-Castillo and Boeckx, 2024). By leveraging high-throughput methods and large datasets, GWAS has characterized the polygenic architecture underlying a spectrum of complex traits, from height and BMI to multifaceted diseases such as cancer and psychiatric disorders (Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium, 2017; Pasman *et al.*, 2018; Hatoum *et al.*, 2021; Bipolar Disorder Working Group of the Psychiatric Genomics Consortium *et al.*, 2023; Demontis *et al.*, 2023). These studies have not only identified significant genetic associations but have also highlighted the importance of non-coding genetic variants in influencing traits (Shadrin *et al.*, 2020).

Polygenicity and SNP-based heritability are fundamental concepts in complex trait genetics, each offering unique insights into the genetic architecture of these traits. Polygenicity emphasizes the collective impact of numerous genetic variants distributed across the genome on trait variability, highlighting that complex traits result from the combined influence of many

variants with modest effects (Frei *et al.*, 2019; Hindley *et al.*, 2021; Tesfaye *et al.*, 2023). Methods like MiXeR quantify this distributed genetic influence by estimating the total count of causal variants contributing to a trait's heritability (Frei *et al.*, 2019; Shadrin *et al.*, 2020). On the other hand, SNP-based heritability (h^2_{snp}) focuses on determining the proportion of trait variance attributed to the additive effects of common SNPs dispersed across the genome (B. K. Bulik-Sullivan *et al.*, 2015). Techniques like LD score regression and GCTA-GREML are commonly used to calculate SNP-based heritability, providing insights into the genetic contribution of common variants to trait variability (Yang *et al.*, 2011; B. K. Bulik-Sullivan *et al.*, 2015). The MiXeR framework not only estimates shared and trait-specific causal variants but also provides information on the genetic relationships between complex traits and disorders (Frei *et al.*, 2019). This statistical tool enhances the understanding of the polygenic architecture of complex traits beyond genetic correlation (Frei *et al.*, 2019). Additionally, MiXeR aids in identifying the extent of unique and shared genetic architecture between traits, offering valuable insights into the genetic overlap between different phenotypes (Frei *et al.*, 2019; Shadrin *et al.*, 2020). Polygenicity and SNP-based heritability are essential components in investigating the genetic underpinnings of complex traits. While polygenicity underscores the broad-scale interaction of numerous genetic variants, SNP-based heritability delves into the specific contribution of common SNPs to trait variability. Together, these concepts form a comprehensive framework for studying genetic influences on phenotypes, shedding light on the intricate genetic architecture of complex traits and diseases.

Here, I aimed to quantify the genetic overlap considering polygenicity and mixed effect directions across psychiatric phenotypes and the CC. Genetic overlap and genetic correlation represent distinct yet interconnected aspects of the genetic relationships underlying complex traits. Genetic overlap occurs when genetic variants influence multiple traits simultaneously,

indicating shared genetic factors and pleiotropic effects (Frei *et al.*, 2019; Shadrin *et al.*, 2020). For example, specific variants linked to both height and BD may demonstrate genetic overlap. On the other hand, genetic correlation goes beyond shared genetic variants by quantifying the additive genetic effects between traits and assessing the directionality of these effects (B. Bulik-Sullivan *et al.*, 2015; B. K. Bulik-Sullivan *et al.*, 2015). Positive genetic correlation, for instance, indicates genetic determinants contributing to higher height and higher risk of BD. While genetic overlap emphasizes shared genetic foundations among traits, genetic correlation describes the alignment or divergence of genetic variants in influencing traits (B. Bulik-Sullivan *et al.*, 2015; B. K. Bulik-Sullivan *et al.*, 2015). Moreover, previous studies have highlighted substantial genetic overlap between psychiatric disorders and brain phenotypes, underscoring shared genetic mechanisms influencing brain structure, function, and psychiatric conditions (van der Meer, Rokicki, *et al.*, 2020; Mufford, van der Meer, *et al.*, 2021; Bahrami, Nordengen, *et al.*, 2022; Carrión-Castillo and Boeckx, 2024). These findings emphasize the complex genetic relationships between psychiatric disorders and brain-related traits, shedding light on the shared genetic architecture underlying these conditions. By investigating the joint genetic architecture of psychiatric traits, one can identify common and diverging etiological mechanisms across psychiatric illnesses (Hindley *et al.*, 2021; Tesfaye *et al.*, 2023). In summary, the exploration of genetic overlap and correlation provides valuable insights into the shared genetic underpinnings and interplay of genetic effects among psychiatric phenotypes and brain regions, such as the CC, contributing to a deeper understanding of the genetic architecture of complex traits and disorders.

2 Aims and Objectives

2.1 Study Rationale

The CC is a crucial region of the brain, playing a significant role in integrating information between the two cerebral hemispheres. It is associated with a variety of psychiatric phenotypes. However, our understanding of the genetic foundations of the CC and its relationship to these disorders is still not fully developed. I hypothesized that the application of novel statistical methods could enhance our understanding of the genetic structure of the CC. This could also illuminate the shared genetic architecture with BD and alcohol use. Given the scarcity of data exploring the heritability and genetic structure of the CC, my initial aim was to delineate the contribution of common genetic variation to the CC and determine its SNP-based heritability. Subsequently, this dissertation sought to investigate the genetic overlap between the CC, BD and alcohol use. Additionally, the specificity of the findings were tested by including related neuropsychiatric phenotypes to ensure robustness of the findings. Lastly, I aimed to examine the genetic influence on the white matter integrity of the CC and how these metrics correlate with neuropsychiatric phenotypes. The aims and objectives are outlined in more detail in the following sections.

2.2 Aims and Objectives

This study had three primary aims (*listed 1 to 3*), which were accomplished by the following objectives (*listed i to ix*)

- 1) Investigate the genetic architecture of the CC and its subregions
 - i) Extract the CC and the five subregions
 - ii) Perform GWAS of the CC and its subregions

- iii) Determine the SNP-based heritability of the CC and its subregions
 - iv) Calculate the genetic correlation between the CC and related psychiatric disorders and substance use traits
- 2) Determine the genetic architecture of white matter integrity of the CC
 - i) Perform GWAS of the CC subregions across diffusion modalities
 - ii) Investigate the SNP-based heritability of white matter integrity of the CC and its subregions
 - iii) Determine the genetic relationship between the CC and relevant psychiatric disorders and substance use traits
- 3) Investigate the regional genetic architecture of the CC and its overlap with psychiatric disorders and substance use traits
 - i) Investigate the polygenicity of the CC, calculating how many trait influencing variants contribute to CC volume
 - ii) Delineate the relationship between the CC and neuropsychiatric phenotypes to quantify the genetic overlap, independent of genetic correlation

Aims 1 to 3 listed above are addressed in the following chapters, respectively:

Aim 1: Chapter 3, “*Distributed Genetic Effects of the Corpus Callosum Subregions Suggest Links to Neuropsychiatric Disorders and Related Traits*”.

Aim 2: Chapter 4, “*Genetic Determinants of White Matter Integrity in the Corpus Callosum*”.

Aim 3: Chapter 5, “*Regional Genetic Architecture of the Corpus Callosum and its Overlap with Psychiatric and Substance Use Phenotypes*”.

The future directions and conclusions of the three primary aims are discussed together in a separate concluding chapter (chapters 7 to 9). Additional figures and tables are available in the

appendices (referred to as A), and larger tables and figures are available in the attached supplementary files (referred to as SM).

NOTE: Minor adjustments to the format and structure of the original manuscripts have been made in order to maintain thematic consistency and formatting throughout the thesis.

3 Chapter 3: Original Research Article 1

Distributed Genetic Effects of the Corpus Callosum Subregions Suggests Links to Neuropsychiatric Disorders and Related Traits

Synopsis

In this study, entitled, ‘*Distributed Genetic Effects of the Corpus Callosum Subregions Suggest Links to Neuropsychiatric Disorders and Related Traits*’, I aimed to investigate the genetic architecture of the CC and its subregions. Utilizing MRI data, I segmented the CC into five subregions and total CC volume and conducted a GWAS to investigate the contribution of common genetic variation to the CC.

Campbell, M.L., Dalvie, S., Shadrin, A., van der Meer, D., O’Connell, K., Frei, O., Andreassen, O.A., Stein, D.J. and Rokicki, J., 2023. Distributed genetic effects of the corpus callosum subregions suggest links to neuropsychiatric disorders and related traits. Acta Neuropsychiatrica, pp.1-8.

DOI: <https://doi.org/10.1017/neu.2023.32>

Authorship Statement

I conceptualized the research question and formulated the objectives for this study with input from my doctoral supervisors. I conducted the data analysis under the supervision of J Rokicki and S Dalvie. I wrote the first draft of the manuscript. All co-authors provided feedback prior to submission to the journal. I submitted the manuscript to the journal and I have responded to the reviewers’ feedback with input from all co-authors.

Distributed Genetic Effects of the Corpus Callosum Subregions Suggests Links to Neuropsychiatric Disorders and Related Traits

Megan L. Campbell^{1,2}, Shareefa Dalvie³, Alexey Shadrin⁴, Dennis van der Meer⁵, Kevin O'Connell⁴, Oleksander Frei⁴, Ole A. Andreassen⁴, Dan J. Stein¹, Jaroslav Rokicki^{4,6}

¹ Department of Psychiatry and Mental Health, University of Cape Town, Cape Town, South Africa

² Global Initiative for Neuropsychiatric Genetics Education in Research (GINGER), Harvard T.H. Chan School of Public Health, Boston, MA, USA

³ Division of Human Genetics, Department of Pathology, University of Cape Town, South Africa

⁴ NORMENT Centre for Psychosis Research, Division of Mental Health and Addiction, University of Oslo, and Oslo University Hospital, Oslo, Norway

⁵ School of Mental Health and Neuroscience, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands

⁶ Centre of Research and Education in Forensic Psychiatry (SIFER), Oslo University Hospital, Oslo, Norway

Corresponding author: Megan Campbell

UCT Department of Psychiatry and Mental Health, Groote Schuur Hospital, Anzio Road, Observatory, 7925, Cape Town, South Africa.

Email: CMPMEG002@myuct.ac.za

3.1 Abstract

Background: The CC is a brain structure with a high heritability and potential role in psychiatric disorders. However, the genetic architecture of the CC and the genetic link with psychiatric disorders remains largely unclear. I investigated the genetic architectures of the volume of the CC and its subregions, and the genetic overlap with psychiatric disorders.

Methods: I applied multivariate GWAS to genetic and T1-weighted MRI data of 40,894 individuals from the UK Biobank, aiming to boost genetic discovery and to assess the pleiotropic effects across volumes of the five subregions of the CC (posterior, mid posterior, central, mid anterior and anterior) obtained by FreeSurfer 7.1. Multivariate GWAS was run combining all subregions, co-varying for relevant variables. Gene-set enrichment analyses were performed using MAGMA. Linkage disequilibrium score regression (LDSC) was used to determine SNP-based heritability of total CC volume and volumes of its subregions as well as their genetic correlations with relevant psychiatric traits.

Results: I identified 70 independent loci with distributed effects across the five subregions of the CC ($p < 5 \times 10^{-8}$). Additionally, I identified 33 significant loci in the anterior subregion, 23 in the mid anterior, 29 in the central, 7 in the mid posterior and 56 in the posterior subregion. Gene-set analysis revealed 156 significant genes contributing to volume of the CC subregions ($p < 2.6 \times 10^{-6}$). LDSC estimated the heritability of CC to ($h^2_{SNP} = 0.38$, $SE = 0.03$), and subregions ranging from 0.22 ($SE = 0.02$) to 0.37 ($SE = 0.03$). I found significant genetic correlations of total CC volume with bipolar disorder (BD, $rg = -0.09$, $SE = 0.03$; $p = 5.9 \times 10^{-3}$) and drinks consumed per week ($rg = -0.09$, $SE = 0.02$; $p = 4.8 \times 10^{-4}$), and volume of the mid anterior subregion with BD ($rg = -0.12$, $SE = 0.02$; $p = 2.5 \times 10^{-4}$), major depressive disorder ($rg = -0.12$, $SE = 0.04$; $p = 3.6 \times 10^{-3}$), drinks consumed per week ($rg = -0.13$, $SE = 0.04$; $p = 1.8 \times 10^{-3}$) and cannabis use ($rg = -0.09$, $SE = 0.03$; $p = 8.4 \times 10^{-3}$).

Conclusions: These results demonstrate that the CC has a polygenic architecture implicating multiple genes, and show that CC subregion volumes are heritable. I discovered novel loci not previously associated with the CC and provide evidence for unique genetic etiology of the subregions. I found distinct genetic factors are involved in the development of anterior and posterior subregions, consistent with their divergent functional specialization. Significant genetic correlation between volumes of the CC and bipolar disorder, drinks per week, major depressive disorder and cannabis consumption subregion volumes with psychiatric traits is noteworthy and deserving of further investigation.

3.2 Introduction

The corpus callosum (CC) is the largest white matter tract in the brain, and serves as the major interhemispheric commissure, integrating information from the two cerebral hemispheres to facilitate language, affective and sensorimotor function (Wang et al., 2020). The CC is commonly segmented into five subregions: anterior, mid-anterior, central, mid-posterior and posterior, with each subregion having distinct axonal projections from cortical regions (Hofer and Frahm, 2006). For example, projections from the auditory and somatosensory cortices primarily cross the central subregion, while those from the visual cortex cross the anterior subregion (Hofer and Frahm, 2006). Although the exact boundaries of these subregions are challenging to delineate using MRI (Hofer and Frahm, 2006), their functional importance in various cognitive processes such as language, sensory processing, and emotional regulation is well-established (Cover *et al.*, 2018).

Differences in volume of the CC have been associated with several psychiatric disorders including mood, substance use and developmental disorders (Rosenbloom, Sullivan and Pfefferbaum, 2003; Newbury and Rosen, 2012; Abramovic et al., 2018; Prunas et al., 2018; de Souza et al., 2019). Volumetric reductions in the CC have been observed in bipolar disorder (BD) patients, with their relatives showing intermediate volumetric reductions between controls and BD probands (Walterfang et al., 2009; Francis et al., 2016). There is evidence to suggest that alcohol consumption affects CC morphometry (de Souza et al., 2019). These effects appear to be subregion-specific, with volume of the anterior CC subregion displaying a negative association with alcohol consumption, irrespective of an alcohol use disorder diagnosis (de Souza et al., 2019). Frontal and superior subregions of the CC appear to be most affected by alcohol consumption, with the inferior and posterior CC subregions remain relatively preserved (de Souza et al., 2019). CC volume reductions are also present in autism

spectrum disorder (ASD), with the anterior and mid anterior subregions seemingly most affected (Frazier and Hardan, 2009; Valenti *et al.*, 2020). Differences in volume of the CC subregions have also been noted in attention deficit hyperactivity disorder (ADHD), major depressive disorder (MDD) and schizophrenia (SCZ), although the pattern of alteration in these disorders is less clear (Connaughton *et al.*, 2022; Zhao *et al.*, 2022; Zhou *et al.*, 2022) .

Twin studies have demonstrated a high heritability of CC size, estimated to 0.67 for total CC and 0.48-0.62 for the subregions (Scamvougeras *et al.*, 2003; Woldehawariat *et al.*, 2014). Although twin and family studies have established the heritability of the CC, there has been a lack of large-scale genetic investigations aimed at identifying the contribution of common genetic variants to the morphology of CC. Recent advances in methods for exploring the genetic architecture of the brain have provided insight into the subregion-specific genetic aetiology of multiple brain regions (van der Meer, Rokicki, *et al.*, 2020; Zhao *et al.*, 2020; Mufford, Meer, *et al.*, 2021; Bahrami, Nordengen, *et al.*, 2022; Ou *et al.*, 2023). This work has been valuable in identifying shared and specific genetic loci and pathways across subregions, as well as providing insight into differential genetic overlap with psychiatric disorders. Accordingly, I aimed to determine the genetic influences on CC subregions and total CC volume, assess the heritability of this structure, and investigate the genetic correlation with psychiatric traits and disorders.

3.3 Methods and Materials

3.3.1 Processing of Imaging and Genetic Data

I accessed raw, T1-weighted MRI data from 35,592 genotyped, Caucasian individuals of British ancestry (age range: 45 to 81 years; mean age: 64.3 years, SD: 7.5; 52.3% female) from the UK Biobank (Bycroft *et al.*, 2018; accession code 27412) for the main analysis, and

5,302 individuals of non-Caucasian ancestry for the generalization analyses conducted in independent data (age range: 45 to 81 years; mean age: 62.9, SD: 7.7 years; 46.3% female). Two different methods were used to segment the CC: FreeSurfer v7.1, which is based on the Witelson segmentation (Witelson, 1989), and the division by Hofer and Frahm (2006) implemented with C8 software (Herron, Kang, and Woods, 2012). Both methods divide the CC into five subregions, the differences are illustrated and described in more detail in the Supplementary Material (Appendix 1: Supplementary Material, Figure 1). However, the main difference being that FreeSurfer extracts volume of the CC and each subregion while the C8 software extracts thickness measurements along the medial line in the sagittal plane. Given the widespread use of FreeSurfer for subcortical segmentation, I chose to focus on this method and use the approach proposed by Hofer and Frahm (2006) to validate the findings. To account for possible confounders, neuroCombat was used to pre-residualize for age, age squared, sex, scanning site, Euler score (a proxy of image quality), total intracranial volume (ICV) and the first 20 genetic principal components for each CC subregion and total CC volume (Fortin et al., 2018). These variables were included as covariates in the univariate GWAS. The pre-residualization step was performed for both segmentation methods and for both the European and non-European participants. These residuals were used as the outcome variables in subsequent multivariate analyses.

Standard quality control (QC) procedures were used to process the UK Biobank version 3, imputed genetic data: SNPs with an imputation score of less than 0.5, a minor allele frequency (MAF) of less than 0.005, a missingness rate of more than 10% and those that failed the Hardy-Weinberg test of equilibrium (HWE, $p < 1 \times 10^{-9}$) were removed. A total of 35,593 individuals remained in the European cohort after QC and an additional 5,286 in the non-European cohort.

3.3.2 Genome-wide association analyses: Multivariate and Univariate

The pre-residualized CC subregions were used as the outcome in the multivariate analysis. This was conducted using the multivariate omnibus statistical test (MOSTest), which has been previously applied to a variety of brain morphology traits (van der Meer, Frei, *et al.*, 2020; Bahrami, Nordengen, *et al.*, 2022). MOSTest implements permutation testing to identify genetic effects across multiple phenotypes, resulting in a multivariate GWAS summary statistic across all five CC subregions. Replication of the MOSTest results was conducted using MOSTEST-PolyVertex Score (MOSTEST-PVS; Loughnan *et al.*, 2022) in the non-European sample of 5,210 individuals.

To identify the distinct genetic contribution of every CC subregion, an individual univariate GWAS of each subregion and total CC were performed using Plink v2 (Chang *et al.*, 2015). This approach allows for the determination of the effect direction and enables further follow-up analyses. Additionally, univariate GWAS of the subregions from the Hofer and Frahm segmentation were conducted as an additional replication step. Significant SNPs were defined using a significance threshold of $p < 9.6 \times 10^{-9}$. (Bonferroni, accounting for 5 subregions and total CC). Independent significant genetic variants are identified as variants with $p < 8.3 \times 10^{-9}$ and LD $r^2 < 0.6$ with each other. For each independent significant variant all candidate variants are identified as variants with LD $r^2 \geq 0.6$ with the lead variant. For a given independent significant variant the borders of the genomic locus are defined as min/max positional coordinates over all corresponding candidate variants. Loci are then merged if they are separated by less than 250kb. A subset of independent significant variants with LD $r^2 < 0.1$ are then selected as lead variants. If a locus contains several lead SNPs using this definition, then the one with the most significant p-value is selected as a lead SNP representing the locus. In order to identify the unique genetic contribution to each CC subregion, additional univariate GWAS of each

subregion were conducted, including total CC volume as a covariate. An overview of the GWAS are provided in the Appendix 1: Supplementary Material, Figure 2.

3.3.3 Functional annotation

Functional Mapping and Annotation of Genome-wide Association Studies (FUMA, platform v1.3.7, accessed September 2022) was used to annotate the GWAS results, using the 1000GPhase3 EUR as reference panel and default parameters (Watanabe et al., 2017). I conducted genome-wide gene-based association and gene-set analyses using MAGMA v.1.08 (<http://ctg.cncr.nl/software/magma>) in FUMA. All variants in the major histocompatibility complex (MHC) region (GRCh37: 6:26,000,000–33,000,000) were excluded before running the MAGMA analyses. MAGMA performs multiple linear regression to map the input SNPs to 18,091 protein coding genes and estimates the significance value of each gene. Genes were considered significant if $p < 2.76 \times 10^{-6}$ (Bonferroni correction). These genes were investigated for enrichment of biological processes, tissue, and cell types. In addition to FUMA, I used Open Target Genetics which integrates functional and biological data from multiple sources to highlight functionally implicated genes, conduct phenome-wide association studies and tests for association with previous GWAS (Mountjoy *et al.*, 2021).

3.3.4 SNP-Based Heritability and Genetic Correlation

SNP-based heritability (h^2_{snp}) and genetic correlation were calculated using Linkage Disequilibrium Score Regression (LDSC; B. K. Bulik-Sullivan et al., 2015). All variants in the MHC region (GRCh37: 6:28,477,797–33,448,354) were removed prior to the analysis. Heritability estimates were calculated for subregions from the FreeSurfer segmentation and the Hofer and Frahm segmentation. Genetic correlations were calculated for total CC volume, each subregion and relevant traits which had indicated CC alterations in prior imaging studies and

which had publicly available GWAS. Genetic correlations were calculated for: attention deficit hyperactivity disorder (Stergiakouli *et al.*, 2012), autism spectrum disorder (PGC ASD, 2017), BD (Mullins, Forstner, O’Connell, Coombes, J. R. I. Coleman, *et al.*, 2021), major depressive disorder (MDD; Wray *et al.*, 2018), schizophrenia (SCZ; Liu *et al.*, 2021), the PGC cross-disorder GWAS (Pouget *et al.*, 2019), cigarettes per day, drinks per week, smoking initiation, smoking cessation (Liu *et al.*, 2019) and cannabis use (Pasman *et al.*, 2018).

3.4 Results

3.4.1 Genome-wide association analyses

I found 70 unique genetic loci with distributed associations across the CC, using MOSTest. The polygenic architecture of the CC is shown in Figure 3.1A, illustrating the results of the multivariate analysis. The univariate statistics indicate subregion-specific genetic signals for each of the CC subregions (Figure 3.1). The loci that showed strongest significant effects from the multivariate analysis showed distributed effects with varying patterns, often exhibiting a gradient from posterior to anterior across CC subregions (Figure 3.1B). A list of significant SNPs from the univariate GWAS of total CC volume and the volume of each CC subregion can be found in the supplementary material (Appendix 1: Supplementary Material, Table 1). Figure 3.2A displays the results from the univariate GWAS for the volume of each subregion. The results for the GWAS, co-varying for total CC volume can be found in Appendix 1: Supplementary Material, Figure 3.

Replication of the multivariate results was conducted using MOSTEST-PVS in the 5,286 individuals of non-European ancestry, for the FreeSurfer segmentation. Of the 70 significant, lead SNPs in the multivariate analysis of the European sample, only 33 were available in the individuals of non-European ancestry. Of these 33 SNPs, all had the same direction of effect

in the group of non-European ancestry. Twenty-two of the 33 SNPs are nominally significant ($p < 0.05$) in the replication and six remained significant after correction for multiple testing (Bonferroni, $p < 1.51 \times 10^{-3}$; Appendix 1: Supplementary Material, Table 2). The results for the multivariate analysis of the Hofer and Frahm segmentation can be found in Appendix 1: Supplementary Material, Figure 4.

Lead loci from the univariate, European GWAS were extracted and a linear regression was conducted in the non-European individuals. Of the 195 lead SNPs from the univariate GWAS of total CC and the univariate GWAS of each subregion, 140 were present in the non-European replication sample. Fifty-four of the SNPs are nominally significant ($p < 0.05$) and 5 SNPs remained significant after correction for multiple testing (Bonferroni, $p < 0.05/140 = 3.57 \times 10^{-4}$). Consistent directions of allelic effect were observed in the non-European analysis when compared to the European lead SNPs, with 89% of the SNPs having the same direction of effect (Appendix 1: Supplementary Material, Table 2).

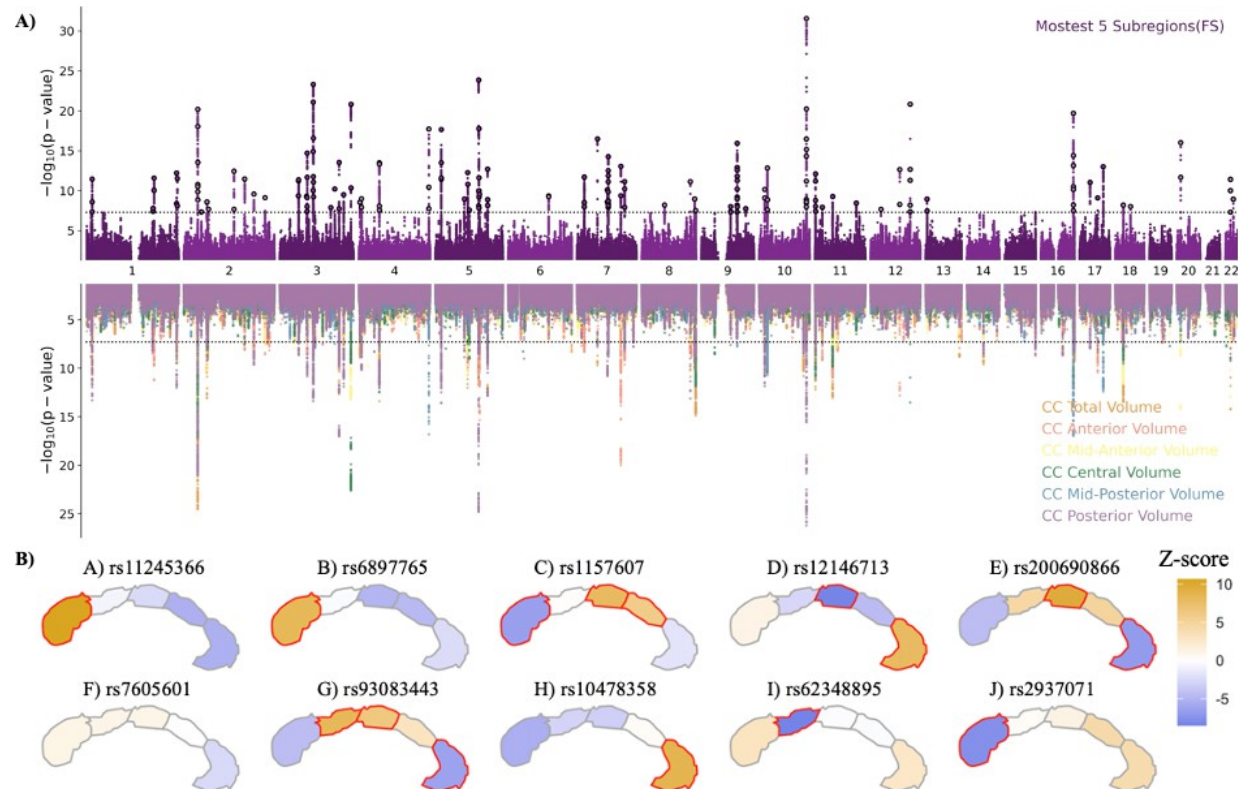


Figure 3.1 Miami plot of the $-\log_{10}(p)$ statistics from the multivariate (top) and univariate (bottom) GWAS of total CC volume. A) Top: The multivariate analysis was conducted using MOSTest and illustrate the distributive genetic architecture of the CC. Bottom: Individual univariate GWAS were conducted for total CC volume and each subregion, indicating both a shared and unique genetic etiology of the CC subregions. B) Top SNPs from the multivariate GWAS show subregion specific effects across the CC. The top ten SNPs from the multivariate GWAS were mapped to the CC subregions using the Z-scores from the respective univariate subregion GWAS. The results indicate a gradient of effect across the posterior and anterior subregions.

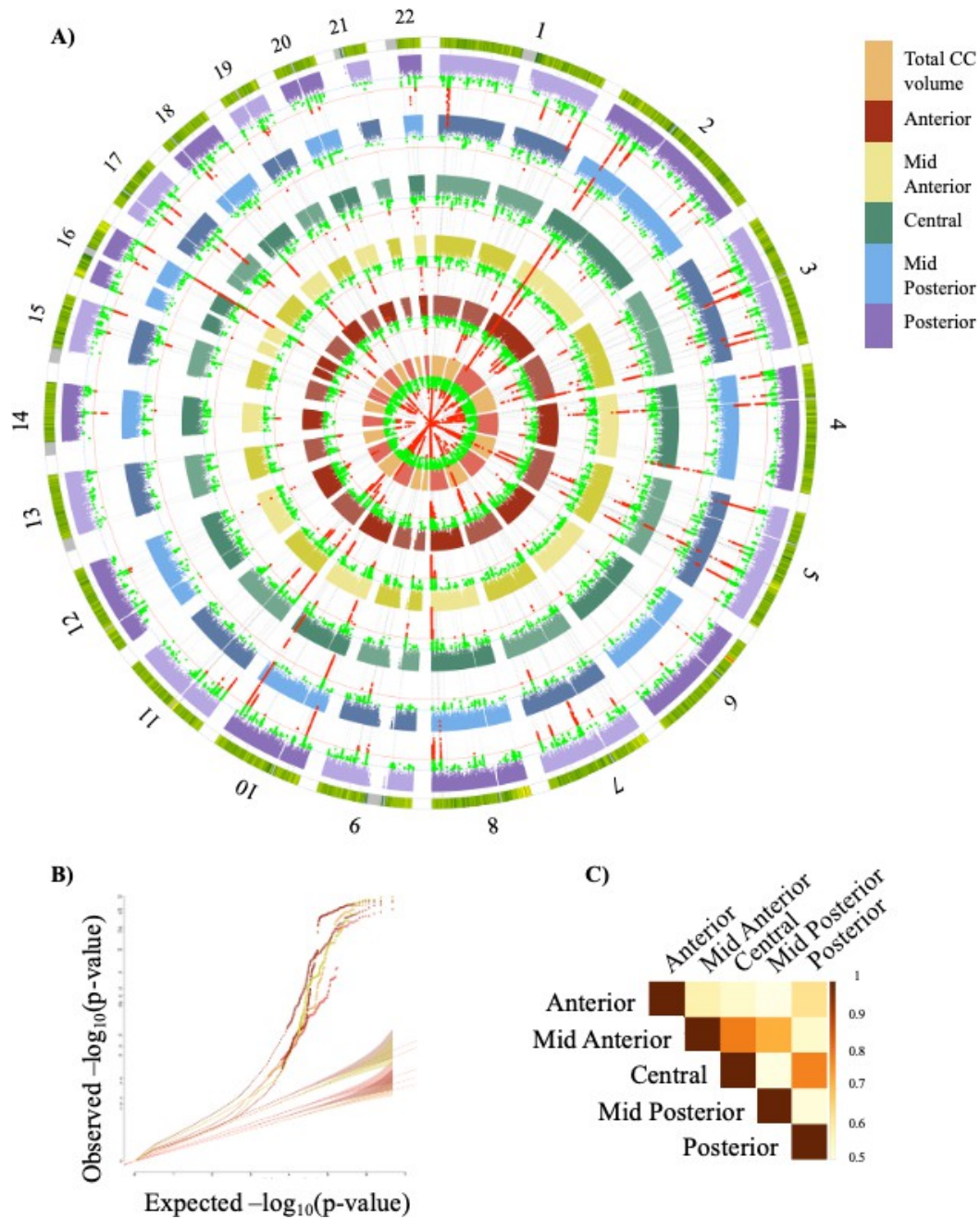


Figure 3.2 A) Manhattan plots from univariate GWAS of each CC subregion B) Quantile-Quantile plot (QQ plot) of observed versus expected $-\log_{10}(p\text{-values})$ the GWAS of CC subregions. The shaded regions represent the expected null distribution of $-\log_{10}(p\text{-values})$ under the assumption of no association. C) Volumetric correlations of CC subregions, each square represents the r^2 values

3.4.2 Functional annotation

Significant gene sets for the GWAS of each subregion and total CC are listed in Appendix 1: Supplementary Material, Table 3. Gene sets identified for the anterior, mid-anterior and central subregion overlapped significantly with previous GWAS of bipolar disorder, alcohol use disorder, Alzheimer's dementia, loneliness and neuroticism (Nagel *et al.*, 2018; Turley *et al.*, 2018; Jansen *et al.*, 2019; Kranzler *et al.*, 2019; Stahl *et al.*, 2019). Gene sets identified for the posterior subregion overlapped with multiple hippocampal phenotypes and implicated cytoskeleton organization (van der Meer, Rokicki, *et al.*, 2020). Manhattan plots from the gene-based association tests, conducted through FUMA, can be found in Appendix 1: Supplementary Material, Figure 4. PheWAS, conducted with Open Target, implicated kidney and liver function (Appendix 1: Supplementary Material, Table 4).

3.4.3 SNP-based heritability and Genetic Correlation

SNP-based heritability was estimated using LDSC for both the FreeSurfer and the Hofer and Frahm segmentation (B. Bulik-Sullivan *et al.*, 2015; B. K. Bulik-Sullivan *et al.*, 2015). The heritability of total CC volume was estimated to be 0.38 ($SE=0.03$) in the FreeSurfer segmentation, with the posterior subregion having the highest heritability estimate ($h^2_{snp}=0.37$, $SE=0.03$) and the mid-posterior subregion having the lowest ($h^2_{snp}=0.22$, $SE=0.03$). The heritability estimates for the Hofer and Frahm segmentation showed the mid-anterior subregion having the highest heritability ($h^2_{snp}=0.32$, $SE=0.03$) and the mid-posterior subregion having the lowest ($h^2_{snp}=0.26$, $SE=0.02$). All heritability estimates for the FreeSurfer segmentation are illustrated in Figure 3A and a full table of heritability estimates for the Hofer and Frahm segmentation can be found in the Appendix 1: Supplementary Material, Table 5.

Using LDSC, I found significant genetic correlations between total CC volume, bipolar disorder (BD, $r_g=-0.09$, $SE=0.03$; $p=5.9\times 10^{-3}$) and drinks per week ($r_g=-0.09$, $SE=0.02$; $p=4.8\times 10^{-4}$). Additionally, I found the significant genetic correlations between volume of the mid anterior subregion BD ($r_g=-0.12$, $SE=0.02$; $p=2.5\times 10^{-4}$), major depressive disorder ($r_g=-0.12$, $SE=0.04$; $p=3.6\times 10^{-3}$), drinks per week ($r_g=-0.13$, $SE=0.04$; $p=1.8\times 10^{-3}$) and cannabis ($r_g=-0.09$, $SE=0.03$; $p=8.4\times 10^{-3}$; Figure 3B). A full table of these results can be found in Appendix 1: Supplementary Material, Table 6.

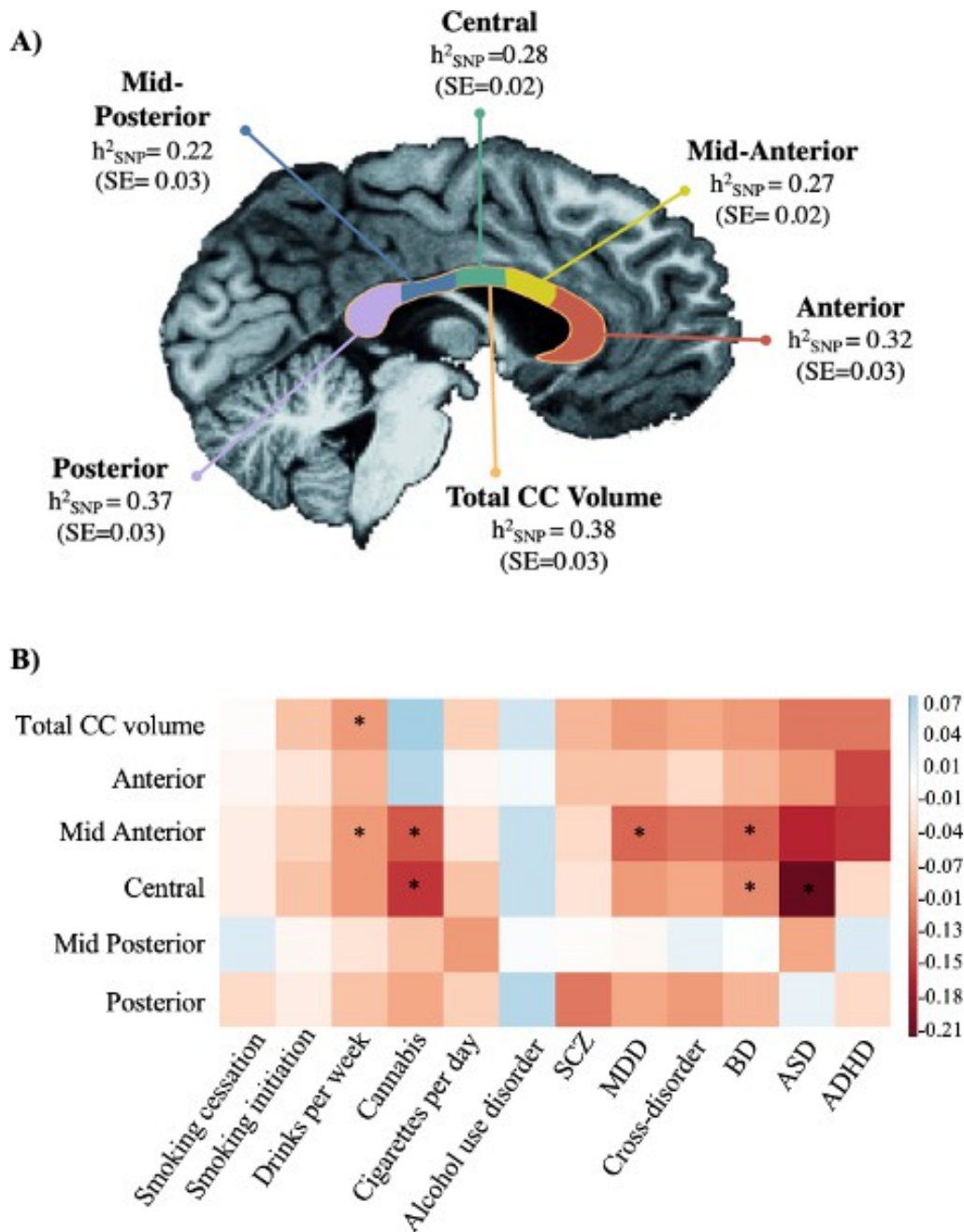


Figure 3.3 A) SNP based heritability estimates of total CC volume and volume of the subregions. LDSC was used to estimate the SNP-based heritability of total CC volume and the five subregions. A schematic of the CC and the five subregions, extracted using FreeSurfer, based on the Witelson segmentation. The CC and subregions have been overlaid onto an anatomical reference image. B) Heatmap of genetic correlations between CC subregions and relevant psychiatric and substance use traits each square represents the r^2 values, significant correlations are denoted by *

3.5 Discussion

Several key findings emerged from this work. First, I identified 70 unique genetic loci contributing to CC volume, with distributed associations across the structure, indicating a polygenic architecture. Second, subregions of the CC also have different genetic architecture, providing evidence that the CC and its subregions have both overlapping and unique genetic contributions. Third, I note significant genetic overlap of CC and CC subregions with neuropsychiatric disorders, including BD, MDD, ASD, as well as cannabis and alcohol consumption.

I identified 70 novel loci that were associated with CC volume, and these loci had distributed effects across the five subregions of the CC. This finding suggests that there is a similar biology underlying the development of these different subregions, and that they may share common genetic pathways. By using multivariate analysis, I was able to boost genetic discovery and gain a more comprehensive understanding of the genetic architecture of the CC and its subregions.

Our univariate GWAS indicate that the five subregions making up the CC have both a shared and distinct genetic etiology, with the loci associated with whole CC volume only partially overlapping with those of the individual CC subregions. I identified a number of novel loci associated with CC subregional volume (33 significant loci in the anterior subregion, 23 in the mid anterior, 29 in the central, 7 in the mid posterior and 56 in the posterior subregion). Although there were no SNPs that overlapped with all subregions, rs755857 was significant in each GWAS except for that of the mid-posterior. Although rs755857 has not been previously associated with any other trait, it is associated with gene *PINX1*. Alterations in *PINX1* have

been associated with gamma glutamyl transferase and serum alkaline phosphatase levels both of which are used as clinical markers of liver function (Barton *et al.*, 2021).

The SNP-based heritability's of the CC subregions ranged from 0.22 (mid posterior) to 0.37 (posterior), with a SNP-based heritability of 0.38 for total CC volume. These findings indicate that the CC and its subregions have a higher SNP-based heritability compared to subcortical brain volumes such as the amygdala and hippocampus (Satizabal *et al.*, 2019; van der Meer, Frei, *et al.*, 2020; Mufford, Meer, *et al.*, 2021; Bahrami, Nordengen, *et al.*, 2022; Ou *et al.*, 2023). The heritability estimates are consistent with previous family reports indicating that the size of the CC and its subregions are under strong genetic influence (Woldehawariat *et al.*, 2014). The anterior and posterior subregions have the highest heritability estimates ($h^2_{snp}=0.32$ and $h^2_{snp}=0.37$, respectively), consistent with previous work Woldehawariat (2014), who found the anterior and posterior subregions of the CC to have the highest heritability estimates. These findings corroborate this pattern of high heritability estimates in the anterior and posterior subregions, and lower estimates in the mid-anterior and mid-posterior subregions.

Gene-set analyses conducted through FUMA indicated a degree of overlap between the subregions, with the exception of the mid-posterior subregion. The mid posterior subregion has a distinct set of significant SNPs, with minimal genetic overlap with the other subregions and no gene sets in common with the other subregions. Additionally, the mid posterior subregion had the lowest number of significant loci and the smallest heritability estimate. This is in line with previous work which shows that the mid posterior subregion is difficult to examine; perhaps due to its smaller size upon extraction, but to date there are no studies that have assessed its heritability in humans. Studies looking at its heritability in other primates, suggest that it has a lower heritability when compared to larger subregions, such as the anterior

subregion (Hopkins *et al.*, 2022). Although I found no evidence to suggest a relationship between increasing subregion volume and the number of significant loci detected, this remains a possibility. The subregion has projections into the posterior parietal and superior temporal cortex which also have lower heritability estimates ($h^2_{snp}=0.07$, $SE=0.02$; $h^2_{snp}=0.12$, $SE=0.02$, respectively) when compared to other cortical regions, which may explain the lower heritability seen for this region (Grasby *et al.*, 2020).

I found significant genetic overlap with previous GWAS of BD, alcohol use disorder and ASD. This is consistent with evidence of CC involvement in these conditions (Prunas *et al.*, 2018; de Souza *et al.*, 2019; Valenti *et al.*, 2020). The genetic correlation results also echoed findings from previous imaging studies. I found significant negative genetic correlation of BD with total CC volume, as well as volumes of the mid anterior and central subregions. Consistently, BD patients have been shown to have significantly smaller CC subregions in comparison to healthy controls (Arnone, Abou-Saleh and Barrick, 2006; Gifuni *et al.*, 2017; Prunas *et al.*, 2018). Volumetric MRI studies have consistently demonstrated lower CC volume in individuals with BD and their healthy co-twins, implying that these abnormalities may be related to genetic factors (Thompson *et al.*, 2003; Bearden *et al.*, 2011). Further, volumetric reductions in the CC have been observed in relatives of patients with BD, showing intermediate volumetric reductions when compared to both controls and BD probands (Walterfang *et al.*, 2009; Francis *et al.*, 2016). These findings add additional evidence to previous work, indicating that altered CC volumes in BD may be due to a shared genetic aetiology. Reductions in total CC volume have also been reported in ASD (Frazier and Hardan, 2009; Valenti *et al.*, 2020). Additionally, there appears to be a gradient of volumetric reduction whereby the anterior subregions show more pronounced effect and the posterior comparatively less affected (Frazier and Hardan,

2009). The genetic correlation results are in line with these findings: there was a significant correlation between ASD and the central subregion.

Previous work has also demonstrated that alcohol consumption is associated with altered CC morphometry, irrespective of an alcohol use disorder diagnosis (de Souza *et al.*, 2019). Here I find significant genetic correlations between the number of drinks per week and total CC, mid anterior and central subregion volumes. Since alterations in the CC are consistently noted in individuals who consume alcohol, the findings suggest that there may be a genetic underpinning to these changes. The exact nature of this relationship warrants further investigation. The top SNP from the multivariate analysis, rs11245366 is associated with the gene *EEF1AKMT2*. PheWAS has found this SNP to be associated with cystatin C, creatine and urate levels (Barton *et al.*, 2021). These phenotypes are related to kidney function and are markers of kidney dysfunction. Abnormalities in renal function have been consistently noted in individuals with agenesis of the CC; additionally, alterations in white matter structure are present in individuals with chronic kidney disease (Franco *et al.*, 1993; Liu *et al.*, 2020). Here I provide preliminary evidence for a shared genetic etiology between these phenotypes. The genetic relationship between the CC and renal phenotypes provides an interesting new path for future investigation

.

Some limitations deserve emphasis. First, this study was limited to individuals residing in the United Kingdom and consisted largely of individuals of European ancestry. I did however, replicate the findings in an independent sample of non-European ancestry. Future genetic work may benefit additional analyses conducted in more diverse cohorts. Second, the age range of individuals in the UK biobank is limited to those above 45 years old. Therefore, it would be important to replicate these findings in a younger sample to ensure generalizability. Third, the

comparison of the results with the literature is challenging due to differences in approach to subregion divisions. Although I implemented two segmentation protocols with consistent results, a robust CC segmentation is needed to improve replicability of research in this field.

In conclusion, these results demonstrate that the CC has a polygenic architecture with multiple genes contributing to CC volume, but also suggest that distinct genetic factors are involved in the development of anterior and posterior subregions, consistent with their divergent functional specialization. The pheWAS findings emphasize the importance of liver and kidney function in CC volume. I provide the first SNP-based evidence that CC and subregion volumes are heritable. Genetic correlations of CC and subregions volumes with specific psychiatric traits is noteworthy, and deserving of further investigation.

4 Chapter 4: Original Research Article 2

Genetic Determinants of White Matter Integrity in the Corpus Callosum

Synopsis

In this study, entitled, ‘*Genetic Determinants of White Matter Integrity in the Corpus Callosum*’, I aimed to investigate the role of common variation in white matter integrity of the CC and its subregions. Unlike chapter 3, which utilized MRI data to segment the CC and extract volumetric measures of the CC, this study utilizes DTI data to specifically assess the white matter integrity of the CC subregions. Through a comprehensive GWAS of 28 diffusion metrics, I have determined the SNP-based heritability, uncovered genetic correlations between the CC and related psychiatric phenotypes and assessed the polygenicity of these traits. This nuanced approach highlights the complex genetic underpinnings that contribute to the structural and functional connectivity of the CC.

Manuscript in preparation

Author Statement

I conceptualized the research question and formulated the objectives for this study with input from my doctoral supervisors. I conducted the data analysis under the supervision of J Rokicki. I wrote the first draft of the manuscript. All co-authors have provided feedback. This manuscript will be submitted following completion of my doctoral thesis. I will submit the manuscript to the journal and I will respond to reviewers’ feedback with input from all co-authors.

Genetic Determinants of White Matter Integrity in the Corpus Callosum

Megan L. Campbell^{1,2}, Alexey Shadrin³, Ivan Ivanovich Maximov³, Dennis van der Meer^{3,4}, Markos Tesfaye^{3,5}, Olav B. Smeland³, Kevin S. O'Connell³, Nadine Parker³, Oleksandr Frei³, Dan J. Stein¹, Ole Andreassen³, Shareefa Dalvie⁶, Jaroslav Rokicki⁷

¹SAMRC Unit on Risk & Resilience in Mental Disorders, Department of Psychiatry and Neuroscience Institute, University of Cape Town, Cape Town, South Africa

²Global Initiative for Neuropsychiatric Genetics Education in Research, Harvard T.H. Chan School of Public Health, Boston, MA, USA

³Precision Psychiatry Group, University of Oslo, Oslo, Norway

⁴School of Mental Health and Neuroscience, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands

⁵NORMENT, Department of Clinical Science, University of Bergen, Bergen, Norway

⁶Division of Human Genetics, Department of Pathology, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

⁷Centre of Research and Education in Forensic Psychiatry (SIFER), Oslo University Hospital, Oslo, Norway

Corresponding author: Megan Campbell

UCT Department of Psychiatry and Mental Health, Groote Schuur Hospital, Anzio Road, Observatory, 7925, Cape Town, South Africa.

Email: CMPMEG002@myuct.ac.za

4.1 Abstract

Background: The CC is a brain structure with a high heritability and hypothesized role in psychiatric disorders. However, the genetic architecture of the CC and the genetic link with psychiatric disorders remains largely unclear. I investigated the genetic architectures white matter structure of the CC and its subregions, and the genetic overlap with psychiatric disorders.

Methods: I conducted univariate GWAS of diffusion MRI data of 40,894 individuals from the UK Biobank, aiming to boost genetic discovery and to assess the pleiotropic effects of 28 diffusion metrics across the five subregions of the CC (posterior, mid posterior, central, mid anterior and anterior). Gene-set enrichment analyses were performed using hypergeometric tests. Linkage disequilibrium score regression (LDSC) was used to determine SNP-based heritability of CC subregions as well as their genetic correlations with relevant psychiatric traits, across 28 diffusion metrics. Polygenicity was assessed using causal mixture modelling.

Results: The univariate GWAS analysis revealed a total of 1,520 significant genetic loci associated with diffusion metrics across the subregions of the CC. Polygenicity estimates varied across modalities and subregions, ranging from 268 to 7,069 trait-influencing variants. The mean number of trait influencing variants ranged from 1,560 to 2,865 across different diffusion metric groups and from 2,385 to 3,108 across subregions. SNP-based heritability estimates showed variations, with the mid-posterior subregion having the lowest mean heritability and the posterior subregion having the highest mean heritability. Significant genetic correlations were observed between alcohol consumption and bipolar disorder with region specific patterns. Gene set enrichment analysis highlighted significant pathways, including the hemidesmosome pathway, pathways related to actin binding and cytoskeletal protein binding.

Conclusion: These results demonstrate that the CC has a polygenic architecture implicating multiple genes, and show that CC subregion microstructure is heritable. Significant genetic

correlations between the CC and bipolar disorder and alcoholic drinks per week are noteworthy and deserving of further investigation.

4.2 Introduction

The corpus callosum (CC), is the largest white matter tract in the brain and serves as the major interhemispheric commissure, integrating information from the two cerebral hemispheres to facilitate language, affective, and sensorimotor function (Wang *et al.*, 2020). The CC is commonly segmented into five subregions: anterior, mid-anterior, central, mid-posterior, and posterior, each with distinct axonal projections from cortical regions (Hofer and Frahm, 2006). For example, projections from the auditory and somatosensory cortices primarily cross the central subregion, while those from the visual cortex cross the anterior subregion (Hofer and Frahm, 2006).

Alterations in white matter integrity, particularly within the CC have been associated with a range of psychiatric and substance use phenotypes. For instance, subregion-specific alterations of CC integrity have been linked to various psychiatric disorders. Specifically, alterations in fractional anisotropy (FA) and radial diffusivity (RD) in the anterior subregion of the CC have been associated with positive symptoms of schizophrenia (Whitford *et al.*, 2010), and lower FA in the posterior and anterior subregions of the CC has been consistently noted in bipolar disorder (BD; Wang *et al.*, 2008; L. H. Lu *et al.*, 2011; Brown *et al.*, 2021). The CC has been suggested to form part of a network underlying emotional regulation in major depression, with lower anisotropy a well-replicated finding (Cole *et al.*, 2012; Liang *et al.*, 2019). Despite the well-established alterations in FA within the CC, there remains a paucity of data on the genetic basis of other diffusion tensor imaging (DTI) metrics and their relationship with neuropsychiatric and substance use traits.

Conventional DTI metrics have been widely used to characterize white matter structure, with measures such as FA, axial diffusivity, mean diffusivity, and RD demonstrating high sensitivity

to age (Westlye *et al.*, 2010; Cox *et al.*, 2016; Beck *et al.*, 2021). However, DTI is constrained by its assumption of Gaussian diffusion and its inability to account for intertwined microstructural features within white matter (Beck *et al.*, 2021; Korbmacher *et al.*, 2023). To address these limitations, I employed additional approaches in this study, including the Bayesian rotationally invariant approach (BRIA; Reisert *et al.*, 2017), diffusion kurtosis imaging (DKI; Jensen *et al.*, 2005), a kurtosis-derived supplement called white matter tract integrity (WMTI; Fieremans, Jensen and Helpert, 2011), the spherical mean technique (SMT; Kaden *et al.*, 2016), and the multi-compartment spherical mean technique (mcSMT; Kaden, Kruggel and Alexander, 2016), alongside DTI. These approaches aim to overcome the limitations of DTI and provide a more comprehensive understanding of white matter microstructure.

Polygenicity emphasizes the combined influence of many genetic variants, while SNP-based heritability focuses on the proportion of trait variance attributed to common SNPs (B. K. Bulik-Sullivan *et al.*, 2015; Frei *et al.*, 2019). Genetic overlap occurs when genetic variants influence multiple traits simultaneously, indicating shared genetic factors (Hindley *et al.*, 2022). On the other hand, genetic correlation goes beyond shared genetic variants by quantifying the additive genetic effects between traits and assessing the directionality of these effects.

Our study aims to investigate the genetic architecture underlying the white matter integrity of the CC, utilising 28 diffusion metrics and elucidating their genetic overlap with neuropsychiatric and substance use phenotypes. By investigating the genetic influences on CC white matter microstructure and its associations with psychiatric and substance use disorders, we aim to contribute a deeper understanding of the genetic underpinnings of brain connectivity and function. Furthermore, the exploration of the genetic overlap between DTI metrics and

neuropsychiatric traits holds promise for shedding new light on the etiology of mental disorders.

4.3 Methods

4.3.1 Processing of Imaging and Genetic Data

I accessed raw, diffusion data from 35,592 genotyped, Caucasian individuals of British ancestry (age range: 45 to 81 years; mean age: 64.3 years, SD: 7.5; 52.3% female) from the UK Biobank (Bycroft *et al.*, 2018; accession code 27412). The CC was extracted and segmented using FreeSurfer's v7.1 subcortical segmentation tool into five subregions: anterior, mid anterior, central, mid posterior and posterior (Fischl *et al.*, 2002). The preprocessing steps for diffusion data were performed according to the guidelines proposed by Maximov *et al.* (2019). This included applying an optimized pipeline that involved corrections for noise (Veraart *et al.*, 2016), Gibbs ringing (Kellner *et al.*, 2016), susceptibility-induced and motion distortions, and eddy current artifacts (Andersson and Sotiropoulos, 2016). Isotropic Gaussian smoothing was conducted using the *fslmaths* function from FSL with a Gaussian kernel size of 1 mm³. Following this, Matlab 2017b (Mathworks, 2017) was used to estimate DTI, DKI, and WMTI metrics. DKI and WMTI metrics were estimated using the multishell data and Matlab code available at <https://github.com/NYU-DiffusionMRI/DESIGNER> (Fieremans, Jensen and Helpert, 2011). SMT and mcSMT metrics were estimated using the original code from <https://github.com/ekaden/smt> (Kaden *et al.*, 2016), while Bayesian estimates were obtained using the original Matlab code from <https://bitbucket.org/reisert/baydiff/src/master/> (Reisert *et al.*, 2017). A total of 28 metrics from 6 diffusion approaches were obtained for each subregion: DTI (Basser, Mattiello and LeBihan, 1994), DKI (Jensen *et al.*, 2005), WMTI (Benitez *et al.*, 2018), SMT (Kaden *et al.*, 2016), SMTmc (Kaden *et al.*, 2016), BRIA (Reisert *et al.*, 2017).

DTI serves as a foundational brain imaging modality that uses a 3D ellipsoid model to characterize the diffusion of water molecules, enabling the inference of tissue structure and the directional orientation of white matter fiber tracts (Voineskos *et al.*, 2012; Voineskos *et al.*, 2012). However, it assumes Gaussian diffusion, which can oversimplify the microstructural properties of brain tissue. Building on the capabilities of DTI, DKI provides an advanced imaging technique that measures non-Gaussian water diffusion in the brain, offering additional insights into tissue complexity, particularly in regions with complex fiber configurations (Brown and Paul, 2019). WMTI, an offshoot of DKI, focuses on quantifying the integrity of white matter tracts by delivering metrics related to the density of axons and the degree of myelination, thus extending beyond what DTI can reveal about white matter health (Deoni *et al.*, 2014). Collectively, these diverse modalities contribute a multifaceted understanding of the brain's intricate microenvironment, each bringing a unique perspective and analytical strength. In contrast to conventional DTI and DKI, the SMT streamlines diffusion signal analysis by averaging the measurements across multiple directions on a sphere, thereby enhancing parameter estimation robustness without resolving discrete fiber orientations (Bruine *et al.*, 2013). Additionally, the SMTmc separates the diffusion signal into distinct compartments, each representing diverse tissue types, which allows for a more acute examination of brain microstructure (Kaden, Kruggel and Alexander, 2016). BRIA employs Bayesian statistics to derive fiber orientations with high robustness, mitigating concerns associated with the rotational variability that can plague diffusion MRI analyses (Reisert *et al.*, 2017). More information on the diffusion metrics can be found in Table 4.1.

Table 4.1 Full list of diffusion metrics examined

Diffusion Approach	Metric	Measurement
Bayesian Rotationally Invariant Approach, (BRIA)	intra-axonal axial diffusivity (DAX intra) extra-axonal radial diffusivity (DRAD extra) microscopic fractional anisotropy (micro FA) extra-axonal axial diffusivity (DAX extra) intra-axonal water fraction (V intra) extra-axonal water fraction (V extra) cerebrospinal fluid fraction (vCSF) microscopic axial diffusivity (micro AX) microscopic radial diffusivity (micro RD) microscopic apparent diffusion coefficient (micro ADC)	Measures the diffusion of water molecules parallel to the axonal fibers within the cellular compartment. Measures the diffusion of water molecules perpendicular to the axonal fibers within the extracellular compartment. Quantifies the degree of diffusion anisotropy, indicating the organization and coherence of tissue microstructures. Measures the diffusion of water molecules parallel to the axonal fibers within the extracellular compartment. Represents the volume fraction of water molecules within the cellular compartment, Represents the volume fraction of water molecules within the extracellular compartment. Quantifies the volume fraction of cerebrospinal fluid within the region of interest. Measures the diffusion of water molecules parallel to the tissue microstructures. Measures the diffusion of water molecules perpendicular to the tissue microstructures. Represents the average diffusion of water molecules within the tissue.
Diffusion Kurtosis Imaging (DKI)	mean kurtosis (MK) radial kurtosis (RK) axial kurtosis (AK)	Represents the average deviation of water diffusion from Gaussian behavior, capturing the complexity of diffusion within the tissue. Measures the deviation of water diffusion perpendicular to the primary diffusion direction. Measures the deviation of water diffusion parallel to the primary diffusion direction.
Diffusion Tensor Imaging (DTI)	fractional anisotropy (FA) axial diffusivity (AD) mean diffusivity (MD) radial diffusivity (RD)	Quantifies the degree of diffusion anisotropy, reflecting the directionality and connectivity of the tissue. Measures the diffusion of water molecules parallel to the primary diffusion direction. Represents the average diffusion of water molecules within the tissue. Measures the diffusion of water molecules perpendicular to the primary diffusion direction.
Spherical Mean Technique (SMT)	fractional anisotropy (SMT FA) mean diffusivity (SMT md) transverse diffusion coefficient (SMT trans) longitudinal diffusion coefficient (SMT long)	Measures the degree of diffusion anisotropy, indicating the organization and coherence of tissue microstructures. Represents the average diffusion of water molecules within the tissue. Represents the diffusion of water molecules perpendicular to the primary diffusion direction, Represents the diffusion of water molecules parallel to the primary diffusion direction.
Multi-compartment Spherical Mean Technique (mcSMT)	extra-neurite mean diffusivity (mcSMT extra md) extra-neurite transverse microscopic diffusivity (mcSMT extra trans) mcSMT diffusion coefficient (SMT md) intra-neurite volume fraction (mcSMT intra)	Measures the average diffusion of water molecules within the extraneurite compartment. Measures the diffusion of water molecules perpendicular to the primary diffusion direction within the extraneurite compartment. Represents the average diffusion of water molecules within the tissue, considering multiple compartments. Represents the volume fraction of neurite compartments within the tissue.
White Matter Tract Integrity (WMTI)	axonal water fraction (AWF) radial extra-axonal diffusivity (radEAD) axial extra-axonal diffusivity (axEAD)	Represents the volume fraction of water molecules within the axonal compartment Measures the diffusion of water molecules perpendicular to the axonal fibers within the extracellular compartment Measures the diffusion of water molecules parallel to the axonal fibers within the extracellular compartment.

To account for possible confounders, age, age squared, sex, euler score as a proxy of data quality, scanning site and the first 20 principal components for each CC subregion were included as covariates in each univariate GWAS. Individuals who were $\pm 4SD$ from the mean were removed from each metric. Standard quality control (QC) procedures were used to process the UK Biobank version 3, imputed genotype data: SNPs with an imputation score of less than 0.5, a minor allele frequency (MAF) of less than 0.005, a missingness rate of more than 10% and those that failed the Hardy-Weinberg test of equilibrium (HWE, $p < 1 \times 10^{-9}$) were removed.

Genome-wide association analyses

To identify the genetic contribution to white matter integrity of each CC subregion, individual univariate GWAS of each subregion and DTI metric were performed using Plink v2 (Chang et al., 2015). Significant SNPs were defined using a significance threshold of $p < 3.5 \times 10^{-10}$ (Bonferroni, accounting for 5 subregions and 28 DTI metrics). Independent significant genetic variants are identified as variants with $p < 3.5 \times 10^{-10}$ and LD $r^2 < 0.6$ with each other. For each independent significant variant all candidate variants are identified as those with LD $r^2 \geq 0.6$.

Loci are then merged if they are separated by less than 250kb. A subset of independent significant variants with LD $r^2 < 0.1$ are then selected as lead variants. If a locus contains several lead SNPs using this definition, then the one with the most significant p-value is selected as a lead SNP representing the locus.

Functional annotation of significant loci

Open Targets Genetics was used to annotate all significant, independent SNPs and conduct phenome-wide association studies (PheWAS; Mountjoy et al., 2021). Open Targets maps SNPs to genes by combining positional information regarding the distance between the variant

and each genes canonical transcription start site, expression quantitative trait loci (QTL), protein QTL, splicing QTL, epigenomic data, and functional prediction (Mountjoy *et al.*, 2021). Open Targets integrates this information and calculates a *Locus2Gene* score, which enables the identification of the gene with the strongest association to a given variant. All variants in the major histocompatibility complex (MHC) region (GRCh37: 6:26,000,000–33,000,000) were excluded prior to annotation. Following annotation, the *Gene2Func* function in Functional Mapping and Annotation of GWAS (FUMA, platform v1.3.7, accessed December 2023; Watanabe *et al.*, 2017) was used to conduct gene-set enrichment analyses, utilising hypergeometric tests. Genes were investigated for enrichment of biological processes, tissue and cell types.

Analysis of genetic architecture using causal mixture modelling

The proportion of causal variants (polygenicity) and the average explained variance per causal variant (discoverability) were estimated using MiXeR v1.2 (Frei *et al.*, 2019). The univariate mixture model considers MAF, sample size, LD structure and genomic inflation to derive estimates of heritability, polygenicity, and discoverability using maximum likelihood estimation. For each univariate MiXeR analysis, 20 independent runs were performed using randomly selected SNPs. I report the average number of trait-influencing variants and corresponding standard deviations across these 20 runs.

SNP-Based Heritability and Genetic Correlation

SNP-based heritability (h^2_{snp}) and genetic correlation were calculated using Linkage Disequilibrium Score Regression (LDSC; B. Bulik-Sullivan *et al.*, 2015). Heritability estimates were calculated for each subregion and each DTI metric separately. LDSC estimates heritability by correlating genetic covariance with the linkage disequilibrium structure of the genome, but does not provide p-values without extensive resampling or modeling. In large

GWAS datasets, like ours, the heritability estimates are precise and statistically significant, so here we report magnitude. Genetic correlations were calculated between each subregion and substance use and psychiatric

phenotypes. Genetic correlations were calculated for: BD (Mullins, Forstner, O’Connell, Coombes, J. R. Coleman, *et al.*, 2021), MDD (Wray *et al.*, 2018), SCZ (Trubetskoy *et al.*, 2022; Trubetskoy *et al.*, 2022), alcohol use disorder (AUD; N=274,424; Kranzler *et al.*, 2019), alcoholic drinks per week (N=941,280), number of cigarettes per day (N=337,334; Liu *et al.*, 2019) and cannabis use (Pasman *et al.*, 2018). Bonferroni correction was utilized to correct for multiple testing ($p < 0.05/28 \text{ metrics} * 5 \text{ subregions} * 14 \text{ disorders}$)

4.4 Results

4.4.1 Genome-wide association analyses

I found a total of 1,520 independent SNPs to be significantly associated with diffusion metrics across the subregions of the CC, using univariate GWAS. Figure 4.1 and Figure 4.2, illustrate the results of each univariate analysis, organised by the modality group: DTI, DKI, BRIA, WMTI, SMT and SMTmc. A list of significant SNPs from the univariate GWAS of diffusion metrics across each subregion can be found in the A2 Supplementary Material Table 1.

BRIA

A) DAX Extra

B) DAX Intra

C) DRAD Extra

D) Micro ADC

E) Micro Ax

F) Micro FA

G) Micro RD

H) V CSF

I) V Intra

J) V Extra

DKI

K) Radial kurtosis

L) Mean kurtosis

M) Axial kurtosis

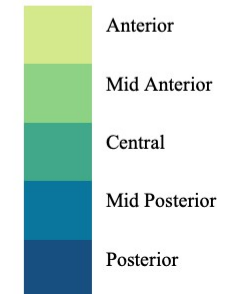


Figure 4.1 Manhattan plots of GWAS across each subregion and diffusion metric. Red dots indicate significant loci with $p < 2.8 \times 10^{-7}$. Significance level is indicated by dashed line. Manhattan plots are plotted in the same order from inner to outer circle: Anterior, mid anterior, central, mid posterior, posterior subregion. Figure illustrating Bayesian rotationally invariant approach, BRIA and diffusion kurtosis imaging, DKI.

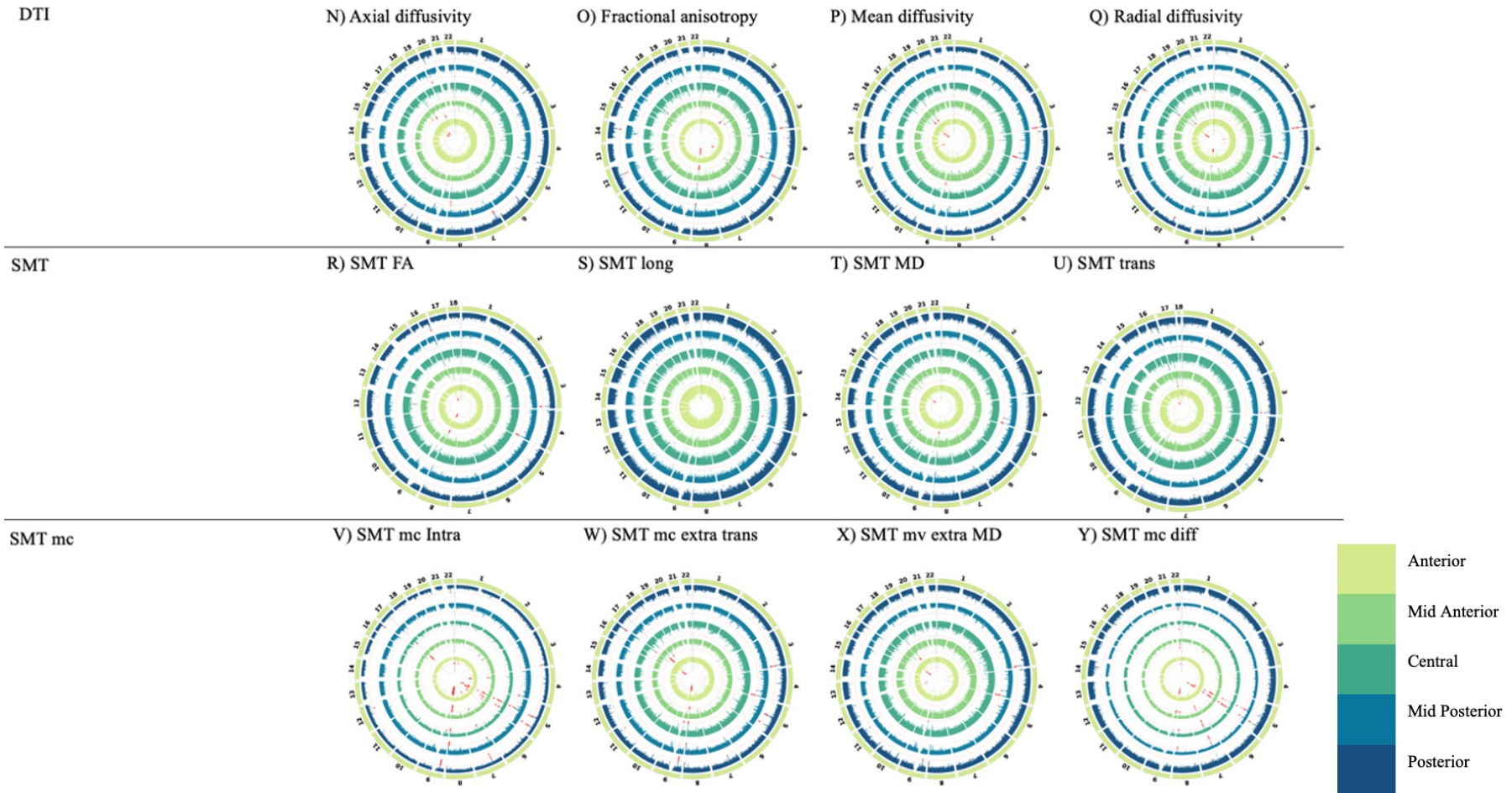


Figure 4.2 Manhattan plots of GWAS across each subregion and diffusion metric. Red dots indicate significant loci with $p < 2.8 \times 10^{-7}$. Significance level is indicated by dashed line. Manhattan plots are plotted in the same order from inner to outer circle: Anterior, mid anterior, central, mid posterior, posterior subregion. Figure illustrating diffusion tensor imaging, DTI; spherical mean technique, SMT and multicompartmental SMT, SMTmc.

4.4.2 Polygenicity

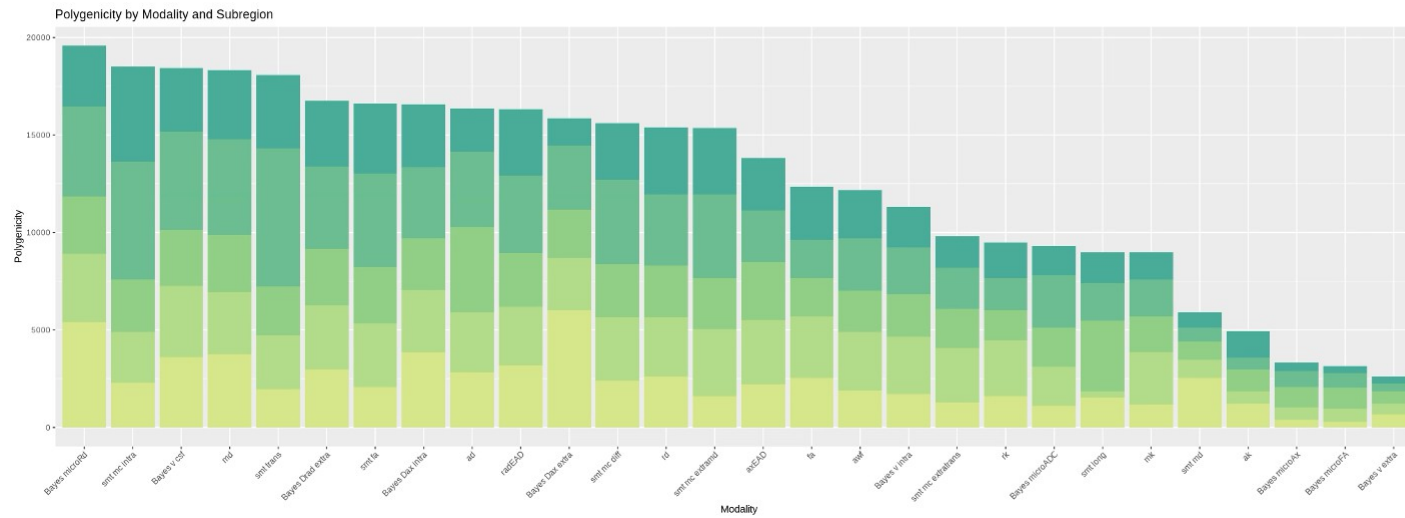
In the univariate MiXeR analysis, I estimated that polygenicity relating to white matter integrity of the CC ranges from 268 ± 133 (reported as: number of trait influencing variants $\pm$ SE) variants for Bayesian microscopic FA in the posterior subregion to 7069 ± 178 variants for SMT transverse diffusion coefficient (SMT trans) in the central subregion (reported numbers of trait-influencing variants are truncated to variants with strongest effects which cumulatively account for 90% of the phenotype's heritability; Figure 2). DKI had the lowest mean number of trait influencing variants (1560 ± 163), while DTI had the highest, with 2865 ± 178 trait influencing variants. By subregion, the mean number of trait-influencing variants ranged from 2385 ± 223 in the anterior subregion to 3108 ± 327 in the central subregion (Figure 4.3A).

4.4.3 SNP-based heritability and Genetic Correlation

SNP-based heritability was estimated using LDSC for all modalities by each subregion (Figure 4.3B). The mid posterior subregion had the lowest mean heritability estimate across all modalities ($h^2_{snp}=22.28\%$, $SE=1.16$) while the posterior had the highest mean heritability ($h^2_{snp}=24.34\%$, $SE=1.14$). The SMT metrics had the lowest mean heritability across diffusion metric groups ($h^2_{snp}=15.04\%$, $SE=1.14$), the remaining modalities all had similar heritability estimates ranging from $\sim 24\%$ to $\sim 26\%$ ($h^2_{snp(BRIA)}=26.5\%$, $SE=0.6$; $h^2_{snp(DKI)}=24.31\%$, $SE=1.82$; $h^2_{snp(DTI)}=24.39\%$, $SE=0.5$; $h^2_{snp(SMTmc)}=24.09\%$, $SE=1.40$; $h^2_{snp(WMTI)}=24.76\%$, $SE=1.40$). Heritability estimates ranged from 2.02% ($SE=1.32$) for the longitudinal diffusion coefficient in the mid posterior subregion to 33.24% ($SE=2.83$) for intra-axonal axial diffusivity in the mid anterior subregion. A full table of these results can be found in A2 Supplementary Material Table 2.

Using LDSC, I found significant genetic correlations between drinks per week in the anterior and mid anterior subregions, and between BD and the anterior subregion and the posterior subregion, across multiple metrics (Figure 4.4).

A)



B)

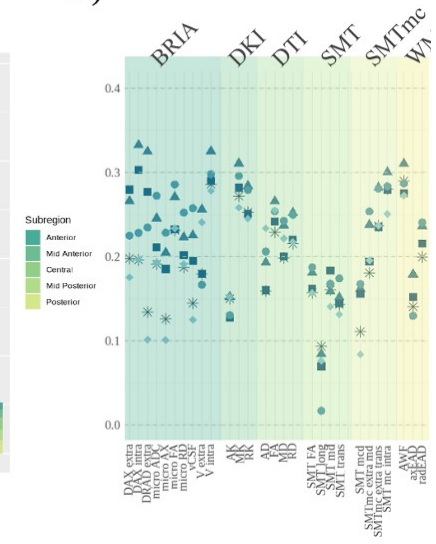


Figure 4.3 Polygenicity and SNP-based heritability of the CC subregions, across diffusion metrics. A) Barplot showing estimated number of trait influencing variants, that explain 90% of the subregions' heritability, grouped by diffusion metric and stratified by subregion. B) SNP-based heritability estimates of the CC subregions, grouped by diffusion metric.

4.4.4 Functional Annotation

The hypergeometric gene-set tests revealed several significant pathways associated with alterations in the white matter integrity of the CC (A2 Supplementary Material Table 3). In the DTI gene-set enrichment analysis, the hemidesmosome pathway was found to be significantly associated ($p=1.8\times10^{-2}$; Bonferroni corrected). In the SMTmc, multiple pathways related to actin binding and cytoskeletal protein binding were identified, including actin binding ($p=1.15\times10^{-2}$), actin filament binding ($p=1.15\times10^{-2}$), cytoskeletal protein binding ($p=2.68\times10^{-2}$) and hemidesmosome pathway ($p=1.46\times10^{-2}$). Expression across the lifespan for genes from SMTmc and BRIA GWAS implicated significant downregulation at age 4 years of age and significant up-regulation at age 23. PheWAS analysis implicated substance use, BD and other psychiatric phenotypes (Figure 4.5)

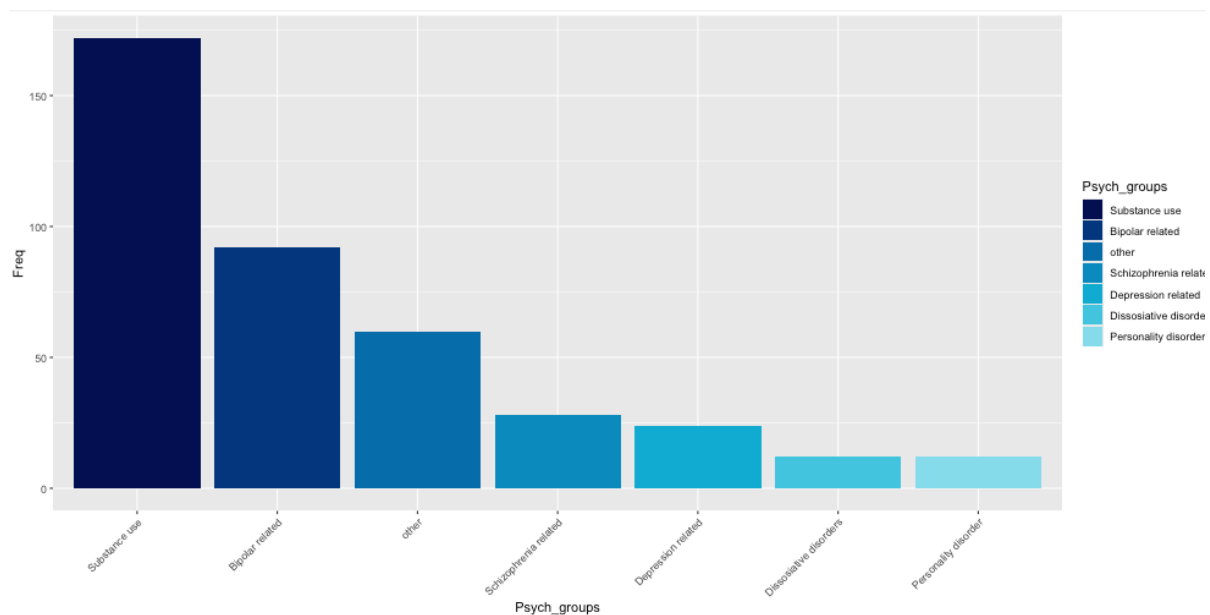


Figure 4.5 PheWAS results from significant genes, implicated in the univariate GWAS

4.5 Discussion

Different brain regions are interconnected through white matter nerve fibers; this fundamental property subserves functional networks involved in cognition and behavior, particularly important in psychiatric phenotypes. In over 30,000 adults from the UK Biobank, I found that variation in white matter microstructure of the CC is polygenic and a substantial proportion of its heritability can be attributed to common variants. The CC is especially influenced by genes that are (i) related to the hemidesmosome pathway, actin binding and cytoskeletal organisation; (ii) active in the developing brain at age four and 23; and (iii) significantly correlated to both BD and alcohol consumption. Thus, genetic variation may account for alterations in white matter microstructure of the CC, contributing to the observed interindividual variation in adult white matter connectivity.

The mean SNP-based heritability of WM integrity metrics in CC subregions ranged from 22.28% ($SE=1.16$; mid posterior) to 24.34% ($SE=1.14$; posterior). The findings here are in line with previous work, examining SNP-based heritability of CC volume, indicating a pattern of high heritability estimates in the anterior and posterior subregions, and lower estimates in the mid-anterior and mid-posterior subregions. The functional roles of different subregions in neural connectivity and cognitive processes might influence their genetic determinants. Subregions with more complex roles may have higher polygenicity and heritability whereas those involved in more conserved functions may have a lower polygenicity and a higher SNP based heritability. Additionally, these findings indicate that the CC and its subregions may have a particularly important conserved function, while other subcortical structures may be more responsive to the environment (Satizabal *et al.*, 2019; van der Meer *et al.*, 2020; Mufford *et al.*, 2021; Bahrami *et al.*, 2022; Ou *et al.*, 2023).

The gene-set analysis of diffusion metrics revealed several pathways that may contribute to alterations of white matter in the CC. One pathway identified in the DTI GSEA was the hemidesmosome pathway. Hemidesmosomes are specialized cell-matrix adhesion complexes that play a crucial role in stabilizing cell adhesion and tissue integrity. Therefore, alterations in

this pathway could potentially impact the structural integrity of the CC. In the SMTmc GSEA, several pathways related to actin binding and cytoskeletal protein binding were identified. These pathways included genes such as *EGFR*, *PLEC*, *NEBL*, *MYO16*, and *NAV3*, all of which are involved in various aspects of actin filament binding and cytoskeletal organization. Actin filaments provide structural support and are essential for maintaining cell morphology and stability (Bosch *et al.*, 2014). Dysregulation of the actin cytoskeleton has been implicated in the regulation of neuronal plasticity, synapse formation, and dendritic spine morphology, which are crucial processes for normal brain function (Bosch *et al.*, 2014). Thus, dysregulation of these pathways could disrupt the organization and stability of white matter tracts within the CC, leading to alterations in white matter integrity (Szikora *et al.*, 2017). Disruptions in cellular structure have been observed in psychiatric disorders, therefore, the genes within these pathways could be potential targets for further investigation and therapeutic interventions (Szikora *et al.*, 2017).

The expression analysis across the lifespan for the genes identified in the SMTmc analysis and the BRIA GWAS revealed significant downregulation at the age of four years and significant up-regulation at the age of 23. At four years of age, the CC undergoes significant growth and maturation, with variations in size and microstructure impacting neurodevelopmental outcomes (Li *et al.*, 2011); at 23, the CC continues to increase in size with higher myelination becoming prominent during this time period (Yates and Juraska, 2007). This suggests that these genes may play a crucial role in the developmental processes and maturation of the CC, highlighting their potential involvement and implications in neurodevelopmental trajectories.

I found significant genetic correlation between the anterior and posterior subregion and BD, across multiple modalities (Bayes microFA, Bayes microADC and MD). Microscopic

fractional anisotropy (micro FA), derived from DTI, assesses anisotropy at a microscopic level, providing information about the orientation of microstructural tissue elements (Reisert *et al.*, 2017). It is sensitive to tissue microarchitecture and useful for studying subtle changes in tissue organization and integrity. Additionally, BRIA microADC, a metric derived from BRIA, provides information about the microscopic apparent diffusion coefficient, offering insights into the movement of water molecules within tissue at a microscopic scale (Reisert *et al.*, 2017). The observed positive genetic correlation between Bayes microADC and BD, along with the negative genetic correlation between microFA and BD, in the anterior and posterior subregions of the CC, suggests alterations in cellular density, structural integrity, and extracellular space, with less organized microstructures in these subregions. These findings align with prior studies that have reported changes in white matter microstructure and lower FA values in the CC of individuals with BD, indicating disruptions in axonal density and myelination (Bellani *et al.*, 2009; Yang *et al.*, 2019). The positive correlation with Bayes microADC could imply changes in water diffusion characteristics, potentially reflecting higher cellular density or alterations in cellular membrane properties. Conversely, the negative correlation with microFA may reflect decreased myelination or disrupted axonal integrity in the CC. These genetic correlations highlight the potential involvement of genetic factors in influencing microstructural characteristics in the CC associated with BD. The less organized microstructures observed in the CC could contribute to the structural disconnectivity reported in previous studies, indicating compromised interhemispheric communication in individuals with BD (Caetano *et al.*, 2008). Additionally, the anterior subregions of the CC play an important role in connecting the prefrontal cortices, areas integral to emotional regulation- disruptions in the connectivity of this region, could contribute to the observed emotional dysregulation noted in BD (Wise *et al.*, 2016).

Positive genetic correlations were found between drinks per week and Bayes MicroADC, Bayes cerebrospinal fluid fraction, MD, Radial Extra-Axonal Diffusivity (RadEAD), RD, and SMTmc extra neurone transverse microscopic diffusivity and SMT MD. These measures reflect information about cellular density, extracellular space, tissue organization, and diffusion properties within the CC (Korbmacher *et al.*, 2023). These findings are in line with existing literature that highlights the role of microstructural integrity of the CC, particularly in the frontal subregions, in alcohol consumption. Alcohol consumption has been associated with alterations in the microstructure of the CC, particularly affecting axonal density, orientation, and myelination (Pfefferbaum and Sullivan, 2004; Bellis *et al.*, 2008). Furthermore, the significant negative genetic correlations between drinks per week and Bayes microFA, FA, and SMT FA in the anterior and mid anterior subregions in this study indicate a potential adverse relationship between alcohol and axonal integrity, myelination, and white matter microstructural coherence. These findings align with previous research suggesting that alcohol consumption can disrupt the microstructure of the CC, leading to arrested growth and potential atrophy, particularly when combined with thiamine deficiency (He *et al.*, 2007).

Some limitations deserve emphasis. First, this study was limited to individuals residing in the United Kingdom and consisted of individuals of European ancestry. Future genetic work may benefit from validation of these findings in more diverse cohorts. Including non-European ancestry individuals in GWAS research is crucial for several reasons. Broadly, it helps to address the underrepresentation of non-European populations in mental health research, ultimately reducing negative health outcomes due to medications not being effective in specific populations and improving our understanding of underlying disease pathology. This has important social implications, promoting equity and inclusivity in healthcare research and treatment. Specifically incorporating diverse populations enhances the generalizability of GWAS findings. A more diverse dataset allows us to draw conclusions that are applicable to a

broader range of individuals, improving the relevance and applicability of the research across different demographics. Although prediction is beyond the scope of the current study, future research could use these results to incorporate polygenic scores and assess whether the European CC GWAS can predict volumetric changes in non-European populations. Tools such as PRSs have been developed to facilitate trans-ancestry analyses and improve the accuracy and utility of genetic predictions across diverse populations. Second, the age range of individuals in the UK biobank is limited to those above 45 years old. Therefore, it would be important to replicate these findings in a younger sample to ensure generalizability. Third, the comparison of results with the literature is challenging due to differences in diffusion metrics used across studies. Nevertheless, I extracted information across multiple metrics, with evidence of convergent findings.

In conclusion, this investigation into the genetic underpinnings of the CC sheds light on its complex architecture. The identification of genetic loci associated with CC morphology and integrity, across various diffusion metrics elucidates both shared and distinct genetic pathways underlying CC development and function. The observed high SNP-based heritability of CC subregions, particularly in the anterior and posterior segments, underscores the genetic basis of CC variability. Pathway analyses reveal candidate molecular mechanisms, such as hemidesmosome and cytoskeletal pathways, contributing to variation in structural integrity and cellular organization in CC. Furthermore, the genetic correlations with BD and alcohol consumption highlight the role of shared genetic factors contributing to both CC microstructural changes and mental disorders. Overall, these findings highlight the potential significance of these pathways in understanding the genetic architecture underlying alterations in the CC, with potential implications for the development of targeted interventions in psychiatry.

5 Chapter 5: Original Research Article 3

Regional Genetic Architecture of the Corpus Callosum and its Overlap with Psychiatric and Substance Use Phenotypes

Synopsis

In this study, entitled, ‘*Regional Genetic Architecture of the Corpus Callosum and its Overlap with Psychiatric and Substance Use Phenotypes*’, I aimed to characterise patterns of genetic overlap when accounting for polygenicity and mixed effect directions between the CC and neuropsychiatric disorders. I applied MiXeR to investigate the genetic overlap between these phenotypes and conjFDR to identify loci that exhibit joint associations with the CC and psychiatric phenotypes.

Campbell, M.L., Shadrin, A., van der Meer, D., Tesfaye, M., Smeland, O.B., O'Connell, K.S., Parker, N., Frei, O., Andreassen, O., Stein, D., Dalvie, S., and Rokicki J., 2023. *Regional Genetic Architecture of the Corpus Callosum and its Overlap with Psychiatric and Substance Use Phenotypes*. medRxiv, pp.2023-12.

DOI: <https://doi.org/10.1101/2023.12.03.23299351>

Authorship Statement

I conceptualized the research question and formulated the objectives for this study with input from my doctoral supervisors. I conducted the data analysis under the supervision of J Rokicki and S Dalvie. I wrote the first draft of the manuscript. All co-authors provided feedback prior to submission to the journal. I submitted the manuscript to the journal and I have responded to the reviewers' feedback with input from all co-authors.

Regional Genetic Architecture of the Corpus Callosum and its Overlap with Psychiatric and Substance Use Phenotypes

Megan L. Campbell^{1,2}, Alexey Shadrin³, Dennis van der Meer^{3,4}, Markos Tesfaye^{3,5}, Olav B. Smeland³, Kevin S. O'Connell³, Nadine Parker³, Oleksandr Frei³, Ole Andreassen³, Dan J. Stein¹, Shareefa Dalvie⁶, Jaroslav Rokicki^{5,7}

¹SAMRC Unit on Risk & Resilience in Mental Disorders, Department of Psychiatry and Neuroscience Institute, University of Cape Town, Cape Town, South Africa

²Global Initiative for Neuropsychiatric Genetics Education in Research, Harvard T.H. Chan School of Public Health, Boston, MA, USA

³Norwegian Centre for Mental Disorders Research (NORMENT), KG Jebsen Centre for Neurodevelopmental Disorders, University of Oslo, Oslo, Norway

⁴School of Mental Health and Neuroscience, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands

⁵NORMENT, Department of Clinical Science, University of Bergen, Bergen, Norway

⁶Division of Human Genetics, Department of Pathology, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

⁷Centre of Research and Education in Forensic Psychiatry (SIFER), Oslo University Hospital, Oslo, Norway

Corresponding author: Megan Campbell

UCT Department of Psychiatry and Mental Health, Groote Schuur Hospital, Anzio Road, Observatory, 7925, Cape Town, South Africa.

Email: CMPMEG002@myuct.ac.za

5.1 Abstract

The CC is a brain structure with a high heritability and a potential role in psychiatric and substance use traits. However, the genetic relationship of the CC, and its subregions, with psychiatric phenotypes remains largely unclear. I aimed to investigate the genetic architecture of the CC and its subregions and its overlap with psychiatric phenotypes using causal mixture model analysis. I employed summary statistics from GWAS of the CC ($n=35,592$), psychiatric (attention deficit hyperactivity disorder, ADHD; autism spectrum disorder, ASD; bipolar disorder, BD; major depressive disorder, MDD; schizophrenia, SCZ) and substance use phenotypes (drinks per week, smoking cessation, smoking initialization, number of cigarettes per day and cannabis use) phenotypes, and performed bivariate causal mixture model analyses with MiXeR to investigate the genetic overlap between these phenotypes. Conjunctive false discovery rate (conjFDR) was used to identify genetic loci that exhibit joint associations with the CC and psychiatric or substance use phenotypes. Open Targets was used to annotate significant, independent variants and MAGMA was used to conduct gene-set enrichment analyses. The polygenicity of the CC subregions ranges from 1.4k variants in the posterior subregion to 1.8k variants in the mid posterior subregion, with approximately 1.8k variants influencing total CC volume. The extent of genetic overlap between CC and psychiatric and substance use phenotypes varied across CC subregion; loci contributing to anterior regions showing almost complete overlap with those involved in BD and SCZ (96% and 98%, respectively), while a distinct pattern emerged for MDD indicating fewer variants overlapping with central and mid anterior subregions (47% and 38%, respectively), loci contributing to the central region exhibited substantial overlap with drinks per week (87%). Loci contributing to the CC subregions exhibited lower overlap with ASD and ADHD (26% to 48%). The conjunctive analysis of CC subregions and psychiatric phenotypes revealed a total of 138 significant independent variants, with gene-set enrichment analyses implicating antigen

processing, regulation of the extracellular matrix and hyaluronan metabolism. I observed significant genetic overlap between the CC and SCZ, BD, and drinks per week, identifying loci linked to brain-related traits with specific overlap patterns found across conditions. Overlapping genetic architectures suggest shared biological pathways, and the potential value of targeting such pathways for the purposes of treatment.

5.2 Introduction

The CC is the largest white matter tract in the brain and serves as the major interhemispheric commissure, integrating information from the two cerebral hemispheres to facilitate language, affective and sensorimotor function (Wang *et al.*, 2020). The CC is commonly segmented into five subregions: anterior, mid-anterior, central, mid-posterior and posterior, with each subregion having distinct axonal projections from cortical regions (Witelson, 1989; Fischl *et al.*, 2002; Hofer and Frahm, 2006). Previous work examining the heritability of the volume of CC and its subregions has revealed a polygenic architecture, and suggested genetic overlap with neuropsychiatric disorders. However, the extent to which the genetic architecture of the volume of CC overlaps with common neuropsychiatric- and substance use phenotypes is largely unknown (Campbell *et al.*, 2023; Gadewar *et al.*, 2023).

Neuroimaging studies have provided insight into the structural and functional differences in the CC among individuals with various neuropsychiatric and substance use disorders. Studies in bipolar disorder (BD) have identified altered white matter integrity in the CC, potentially affecting interhemispheric communication (Wise *et al.*, 2016). Furthermore, individuals with schizophrenia (SCZ) show structural abnormalities such as lower thickness and altered shape in the CC, indicating disrupted connectivity between brain regions (Frumin *et al.*, 2002; L. H. Lu *et al.*, 2011). In BD and SCZ, disrupted interhemispheric communication may contribute to the cognitive and emotional dysregulation observed in these disorders (Frumin *et al.*, 2002; Wise *et al.*, 2016). In individuals with attention deficit hyperactivity disorder (ADHD), there is evidence of decreased volume and alterations in the microstructure of the CC, indicating disruptions in interhemispheric connectivity (Acer *et al.*, 2017). Similarly, individuals with autism spectrum disorder (ASD) exhibit differences in CC size, shape, and connectivity, suggesting atypical development of white matter microstructure (Frazier and Hardan, 2009).

Substance use-related traits have also been associated with alterations in CC volume, but findings regarding the effects of alcohol consumption, cigarette smoking, and cannabis use on CC volume are inconsistent (Moeller *et al.*, 2004; Hampton, Hanik and Olson, 2019).

Despite the known genetic contributions to both the CC and to psychiatric and substance use phenotypes, the extent of genetic overlap of the CC and its subregions with these phenotypes remains unclear. I previously found significant genetic correlations between total CC volume, BD and drinks per week, suggestive of genetic overlap between these traits (Campbell *et al.*, 2023). However, genome-wide genetic correlations do not fully characterize genetic overlap since they cannot detect an overlap when shared trait-influencing variants have a balanced mixture of concordant and opposing directions of effect, thus cancelling out genome-wide correlation (Hindley *et al.*, 2021; Karadag *et al.*, 2023). In this study, the aims are twofold: first, I aim to provide a comprehensive description of the genetic architecture of the CC, by investigating the polygenicity determining the number of trait influencing variants. Second, I seek to further delineate the relationship between the CC and relevant phenotypes, utilizing the causal mixture model framework (MiXeR) to quantify genetic overlap independent of genetic correlations (Frei *et al.*, 2019; Hindley *et al.*, 2021; Bahrami, Hindley, *et al.*, 2022; Karadag *et al.*, 2023; Tesfaye *et al.*, 2023). Additionally, I employ conjunctive false discovery rate (conjFDR) to identify loci jointly associated with both phenotypes (Andreassen *et al.*, 2013). This approach allows us to gain a more nuanced understanding of the shared genetic components between the CC and psychiatric phenotypes, while addressing the limitations of conventional genetic correlation analyses.

5.3 Methods

5.3.1 Genome-wide Association Studies

I compared genome-wide association studies (GWAS) of five subregions of the CC and total CC volume (Campbell *et al.*, 2023; conducted using data from the UK Biobank (accession code 27412) with GWAS of the following psychiatric phenotypes: ADHD (Demontis *et al.*, 2019), ASD (PGC ASD, 2017), BD (Mullins, Forstner, O’Connell, Coombes, J. R. Coleman, *et al.*, 2021), MDD (Wray *et al.*, 2018), SCZ (Trubetskoy *et al.*, 2022) and substance use traits: drinks per week (N=941,280), number of cigarettes per day (N=337,334), smoking cessation (N=547,219) and smoking initiation (N=1,232,091; Liu *et al.*, 2019) and cannabis use (Pasman *et al.*, 2018). In all GWAS datasets except that of the CC, I excluded samples from the UK Biobank from phenotypes to avoid potential sample overlap as required by conjFDR. This was not possible for the GWAS of cannabis use as it consists predominantly of individuals from the UK Biobank. Since MiXeR accounts for sample overlap, I used the whole samples without the removal of overlapping samples. A full table of these GWAS can be found in the Appendix 3: Supplementary Material, Table 1.

5.3.2 Analysis of genetic architecture using causal mixture modelling

Univariate MiXeR analyses were conducted to estimate the number of single nucleotide polymorphisms (SNPs) influencing each trait, accounting for linkage disequilibrium (LD). To investigate the genetic overlap between CC volume and the aforementioned phenotypes, causal mixture models were applied using MiXeR. Bivariate MiXeR analyses were performed to estimate the number of SNPs shared between each pair of phenotypes and the number of trait-specific SNPs associated with either of these phenotypes. For each univariate and bivariate MiXeR analysis 20 independent runs were performed using randomly selected SNPs. I report the average number of trait-influencing variants and corresponding standard deviations across

these 20 runs. The difference between the Akaike Information Criterion (AIC) of the comparative model and that of the fitted model was calculated to serve as a measure of the relative model fit. A positive AIC difference indicates that the fitted model is suitable, indicating better model discrimination (Hindley *et al.*, 2022). Additionally, model fit was evaluated based on predicted versus observed conditional quantile-quantile plots (QQ-plots) and log-likelihood plots that visualize the parameter estimation procedure (Appendix 3: Supplementary Material, Table 2).

From the number of trait influencing variants, I calculated the percentage of SNPs (with respect to the number of shared variants) with the same direction of effect (concordant) and those with opposite direction of effect (discordant). Discordant directions of SNP effects refer to situations where the same genetic variant has opposite effects on two different traits. For example, a specific SNP might increase the risk of one trait (like a psychiatric disorder) while simultaneously decreasing the risk or expression of another trait (like corpus callosum volume). This indicates that the genetic influences on these two traits are opposing each other. While, negative genetic correlation is a broader construct. It represents the overall relationship between the genetic influences on two traits across the entire genome. A negative genetic correlation means that, on average, genetic variants that increase the risk of one trait tend to decrease the risk of the other trait. This is an aggregate measure that considers all genetic variants simultaneously.

5.3.3 Identification of shared genetic loci with conjunctive false discovery rate

To identify loci associated with the CC and those associated with conditional phenotypes, I employed the pleiotropy-informed conjFDR method (accessed at <https://github.com/precimed/pleiofdr>; Andreassen *et al.*, 2013). The conjFDR combines

GWAS summary statistics of two phenotypes to identify genetic loci that exhibit joint associations with both phenotypes. The conjFDR statistic, calculated from two conditional FDR statistics obtained by swapping the roles of primary and conditional phenotypes, identifies SNPs associated with both the primary and the conditional phenotypes. It yields an estimate of the posterior probability that a SNP is null for either or both traits, provided that the p-values for both phenotypes combined are equal to or smaller than the p-values for each individual trait. A conjFDR threshold of 5% was applied to identify statistically significant loci associated with these phenotypes.

SNPs within the extended major histocompatibility complex (MHC, GRCh37: 6:26,000,000–33,000,000) region were excluded prior to fitting the conjFDR models to mitigate potential

bias due to the intricate LD structure of the MHC. Random pruning of SNPs in 500 iterations was performed to avoid the inflation of the results due to LD-dependency. For every iteration a single SNP was randomly selected to represent each LD block (clump of variants with an r^2 value greater than 0.1), and their respective p-values were used to compute the empirical cumulative distribution functions. The results were then averaged across all iterations.

5.3.4 Functional annotation of shared loci

Open Targets Genetics was used to annotate all significant, independent SNPs (Mountjoy *et al.*, 2021). Open Targets maps SNPs to genes by combining positional information regarding the distance between the variant and each genes canonical transcription start site, expression quantitative trait loci (QTL), protein QTL, splicing QTL, epigenomic data, and functional prediction (Mountjoy *et al.*, 2021). Open Targets integrates this information and calculates variants Locus2Gene score, which enables the identification of the gene with the strongest association with the variant. All variants in the major histocompatibility complex (MHC) region (GRCh37: 6:26,000,000–33,000,000) were excluded prior to annotation. Following annotation, the Gene2Func function in Functional Mapping and Annotation of GWAS (FUMA, platform v1.3.7, accessed October 2023) was used to conduct gene-set enrichment analyses, utilising MAGMA v.1.08 (; de Leeuw *et al.*, 2015; Watanabe *et al.*, 2017). Genes were investigated for enrichment of biological processes, tissue and cell types.

5.4 Results

5.4.1 Analysis of genetic architecture using causal mixture modelling

In univariate MiXeR analysis, I estimated that CC volume has approximately $1.8 \times 10^3 \pm 88$ trait influencing variants (here and below reported numbers of trait-influencing variants are

truncated to variants with strongest effects which cumulatively account for 90% of phenotype's heritability; Figure 5.1). The numbers of trait influencing variants for the CC subregions were similar: anterior (1700±145), mid anterior (1600±90), central (1500±109), mid posterior (1800±124) and posterior (1400±70). Additionally, I found the CC to be less polygenic than substance use phenotypes: drinks per week 6400±508, smoking cessation 2600±465, smoking initiation 9600±316, cannabis consumption 8100±911 and cigarettes per day 900±475 and psychiatric phenotypes: ADHD 3900±158, ASD 1200±1590, BD 8600±307, MDD 2100±2323, SCZ 9600±254.

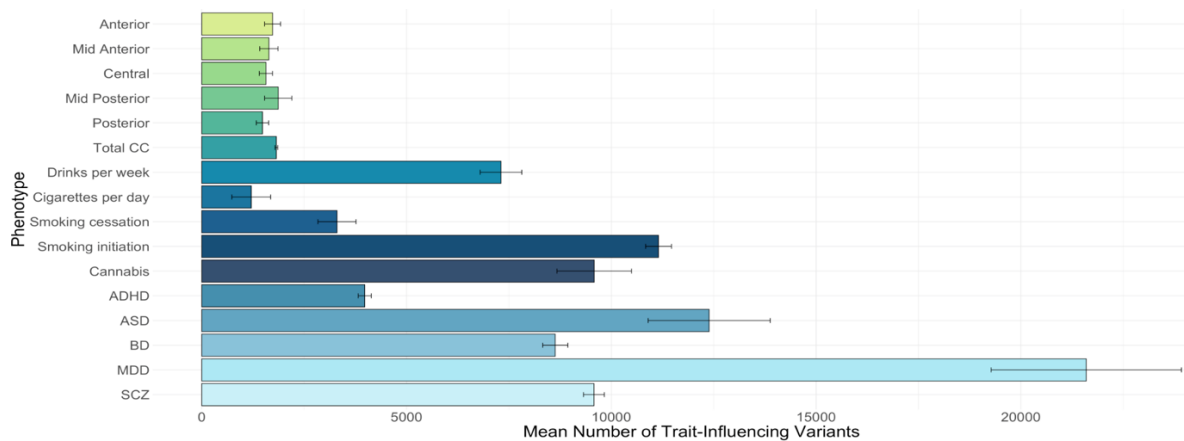


Figure 5.1 Polygenicity of the CC and psychiatric phenotypes illustrated by estimated causal variants. Polygenicity of volumes of the corpus callosum (CC) subregions (anterior, mid anterior, central, mid posterior, posterior), total CC volume, drinks per week, cigarettes per day, smoking cessation, smoking initiation, cannabis, attention deficit hyperactivity disorder (ADHD), autism spectrum disorder (ASD), bipolar disorder (BD), major depressive disorder (MDD), schizophrenia (SCZ).

The mid anterior and anterior subregions consistently showed a greater proportion of shared variants with psychiatric disorders, ranging from 91% to 97% for the anterior subregion (percentage calculated with respect to the CC) compared to other CC subregions (Figure 2). BD and SCZ exhibited similar patterns of overlap with the CC, with the highest proportion of overlap observed in the anterior subregion (91% and 97%, respectively) and the lowest in the mid posterior subregion for both disorders (20% and 15%, respectively). Ninety-five percent

of the variants influencing the anterior subregion overlapped with MDD. However, MDD had a lower proportion of overlap with the central and mid anterior subregions (47% and 38%, respectively). For both ADHD and ASD, the proportion of overlapping variants for all subregions was less than 50%, except for ASD and the anterior subregion, which exhibited a 95% overlap in shared causal variants. When examining the directions of SNP effects in shared causal variants, only BD and ASD displayed more than 50% of SNPs with a consistent direction of effect. Specifically, 80% of the shared causal variants between the mid posterior subregion and BD had a consistent direction of effect, while 64% of the shared variants between the posterior subregion and ASD exhibited concordance.

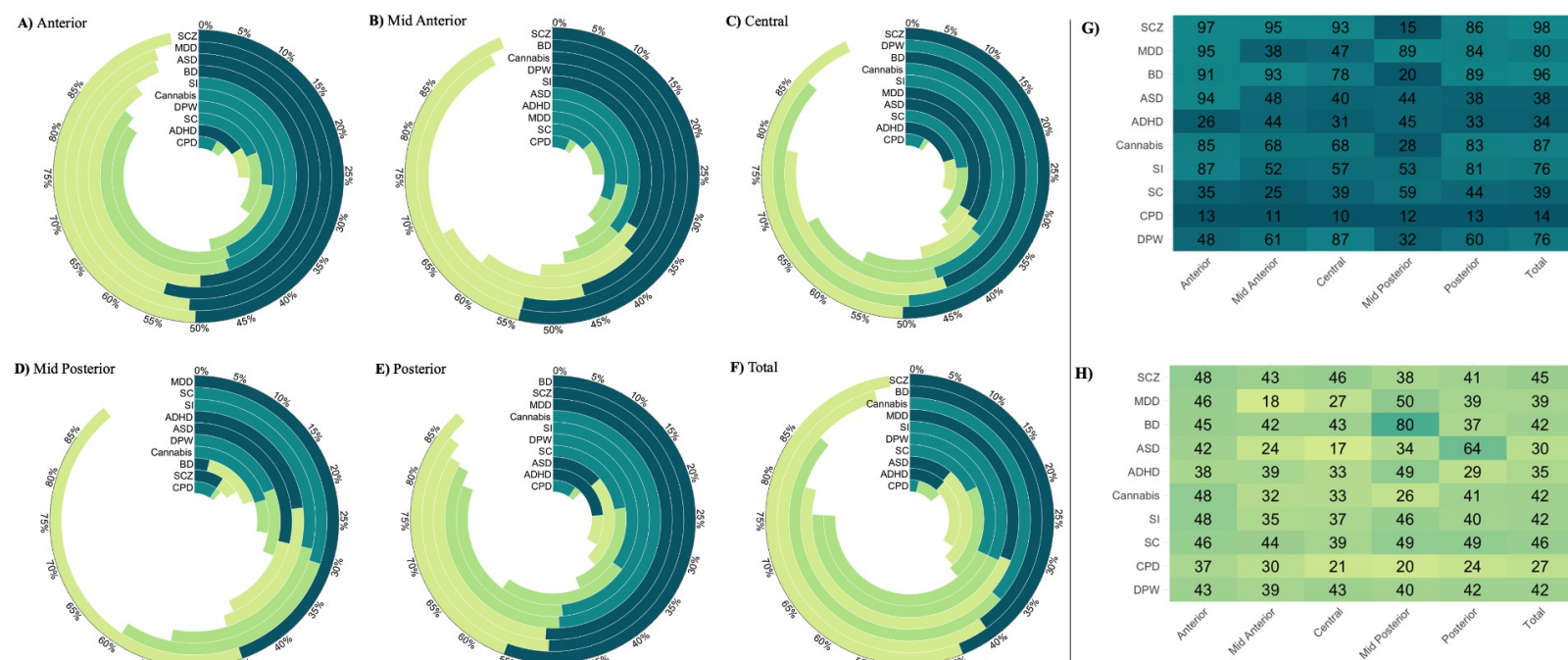


Figure 5.2 Shared trait influencing variants of the CC, covaried for intracranial volume, and psychiatric phenotypes. A-F) The percentage of concordant/ discordant SNPs from the number of shared trait influencing variants between variants (percentages are calculated with respect to total CC variants) G) Heatmap illustrating the estimated percentage of shared variants, calculated with respect to the CC H) Proportion of shared variants, showing the same direction of effect

In contrast to the psychiatric disorders, the substance use phenotypes exhibited a more variable genetic overlap with the CC and its subregions. Drinks per week had its highest overlap with the central subregion (87%), with lower overlap noted in the mid anterior and posterior subregions (61% and 60%, respectively) and the lowest proportion of overlap with the mid posterior subregion (32%). Cigarettes per day had the lowest proportion of shared variants, compared to the other psychiatric phenotypes, across all CC subregions (<20. However, smoking cessation showed slightly higher overlap across subregions compared to cigarettes per day, with the highest overlap in the mid posterior subregion at 59%. Smoking initiation had a higher degree of overlap across all subregions, with the highest overlap in the anterior and posterior subregions (87% and 81%, respectively); the mid posterior, central, and mid anterior subregions had lower overlap, all below 60%. Among the substance use phenotypes, cannabis use had the highest overlap across the CC, with the highest overlap in the anterior subregion (85%) and the lowest in the mid posterior subregion (28%). In terms of SNP effects, the majority of shared causal variants were discordant (>50% discordant), across all comparisons. For cigarettes per day, across all CC subregions, over 70% of the shared variants had discordant directions of effect. Drinks per week, smoking cessation, and smoking initiation exhibited a similar pattern, with approximately 60% of variants indicating mixed effect directions across all CC subregions. For variants shared between cannabis use and the mid posterior, central and mid anterior subregions, ~65% exhibited discordant directions of effect while slightly lower rates of discordance were noted for the anterior and posterior subregions (52% and 59%, respectively; Figure 2A-F, H).

5.4.2 Identification of shared genetic loci with conjunctive false discovery rate

I conducted conjFDR analysis using volumes of the CC and relevant psychiatric phenotypes to identify genetic loci shared between these phenotypes. The conjunctive analysis of CC

subregions and psychiatric phenotypes revealed 110 significant loci (37 ADHD, 7 ASD, 44 BD, 6 MDD, 165 SCZ; conjFDR<0.05; Appendix 3: Supplementary Material, Table 4; Figure 3). Twenty-eight unique loci were identified between the CC and substance use phenotypes (conjFDR<0.05; four drinks per week, six cigarettes per day, 20 smoking initiation, two smoking cessation; Appendix 3: Supplementary Material, Table 5).

5.4.3 Functional annotation of shared loci

Open Targets Genetics was used to annotate significant, independent SNPs (Mountjoy et al., 2021). Annotated genes were then used as the input for the Gene2Func function in FUMA to conduct gene-set enrichment analyses. All loci identified in the conjDFR analyses were intergenic or intronic SNPs. Only three comparisons had significant biological processes implicated from the gene-set analyses (Appendix 3: Supplementary Material, Table 6): anterior subregion and ADHD (39 processes), central subregion and smoking initiation (five processes), the posterior subregion and SCZ (one process) and total CC volume and SCZ (four processes). The top biological processes from ADHD and the anterior subregion implicated antigen processing ($p=7\times 10^{-7}$; Bonferroni corrected). Negative regulation of the extracellular matrix was implicated by the central subregion and smoking initiation ($p<0.05$; Bonferroni corrected) and hyaluronan metabolism was implicated in both the posterior subregion and SCZ and total CC volume and SCZ.

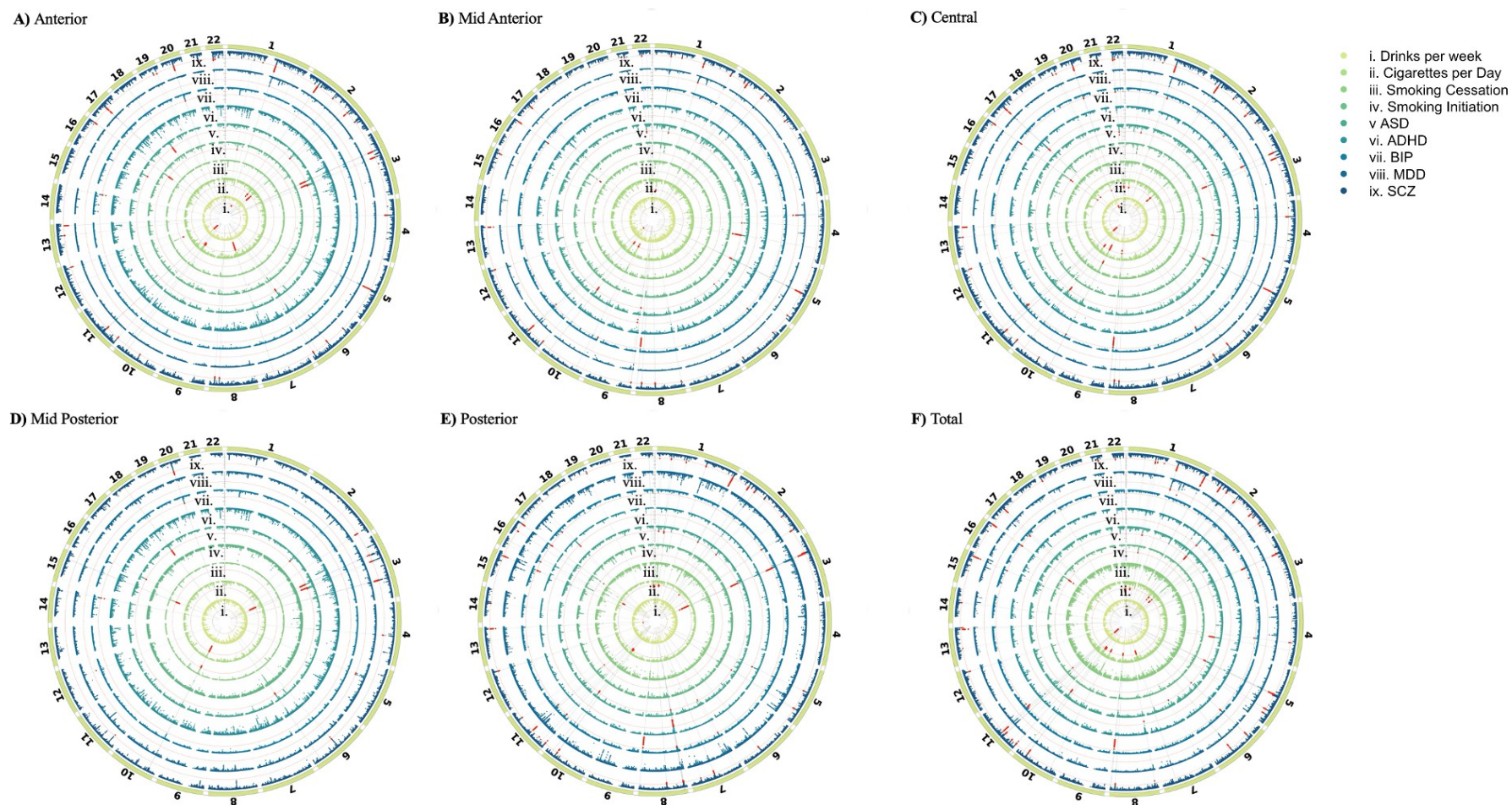


Figure 5.3 Manhattan plots of conjFDR results illustrated by subregion. A-F) ConjFDR results by subregion. Red dots indicated significant loci with $FDR < 0.05$. Significance level is indicated by dashed red line. Manhattans are plotted in the same order from inner circle to outer circle: Drinks per week, Cigarettes per day, Smoking cessation, Smoking initiation, ASD, Autism spectrum disorders; ADHD, Attention deficit hyperactivity disorder; BD, Bipolar disorder; MDD, Major depressive disorder; SCZ, Schizophrenia.

5.5 Discussion

I found the CC to be less polygenic than the psychiatric phenotypes examined here, with approximately ~1800 variants influencing total CC volume, and ~1400 (posterior) to ~1800 (mid-posterior) influencing CC subregions. The extent of genetic overlap between CC and psychiatric phenotypes varied across CC subregion; variants contributing to anterior regions showed almost complete overlap with those involved in BD and SCZ, while a distinct pattern emerged for MDD, indicating fewer variants overlapping with central subregions, variants contributing to the central region overlapped with alcohol consumption. Finally, I note similar patterns of genetic overlap between SCZ, BD and cannabis use with the CC.

The CC shows lower polygenicity compared to the psychiatric phenotypes examined in this study, with roughly 1800 variants affecting the total CC volume and about 1400 to 1800 variants influencing different subregions of the CC. This is in line with previous work; brain regions exhibit lower polygenicity compared to substance use and psychiatric phenotypes due to several factors (Mufford et al., 2023; Tesfaye et al., 2023; van der Meer et al., 2022). These regions are measured using precise methods like FreeSurfer, leading to consistent measurements, whereas psychiatric diagnoses can introduce biases related to cultural context, increasing phenotypic variance (Iheanacho et al., 2014; Lewis-Fernández and Aggarwal, 2013). Additionally, brain regions have conserved functions critical to survival, resulting in lower variability and consequently, lower polygenicity (van der Meer et al., 2022). This conservation means fewer significant deviations, as severe abnormalities are less likely to be observed in studies. However, given the high polygenicity of psychiatric disorders, the substantial overlap observed suggests that specific, impactful genetic variants within the CC may be particularly relevant to the pathology of these disorders.

Cannabis use and SCZ exhibited the similar patterns of overlap with the CC, emphasizing the complex interplay between these phenotypes. The observed genetic overlap here suggests

shared risk variants, potentially influencing how individuals may be predisposed to both conditions (Vaissiere *et al.*, 2020). Cannabis use has also been associated with changes in the CC microstructural organization, particularly in patients with first episode psychosis (Rigucci *et al.*, 2016). Substance use behaviors are closely linked to the brain's reward and motivation pathways, which involve regions such as the prefrontal cortex and striatum. These regions are connected by the CC, and genetic factors influencing these pathways may also impact CC structure. This shared neurobiological basis can lead to the observed genetic overlap. A genetic propensity to use substances may place an individual in environments that alter brain morphology. Conversely, a genetic predisposition to altered CC subregions may result in an individual being more likely to engage in risky or substance use behaviors. This bidirectional influence highlights the complex interplay between genetics, environment, and brain structure. Disruptions in this structure, influenced by genetics but exacerbated by cannabis use, may underlie some of the cognitive dysfunction inherent to SCZ.

The finding that BD and SCZ have similar genetic overlap with the anterior subregion, is consistent with a range of previous work. First, it is increasingly clear that there is extensive genetic overlap between BD and SCZ (Smeland *et al.*, 2020; Hindley *et al.*, 2022). Second, the anterior subregions of the CC play a crucial role in connecting the prefrontal cortices, areas integral to emotion regulation. Disruptions in the connectivity of this region could contribute to the observed emotional dysregulation inherent to psychotic spectrum disorders (Wise *et al.*, 2016).

The mid-posterior subregion had the lowest number of overlapping trait-influencing variants with all traits. However, a notable exception emerges in the case of MDD, where there is almost complete genetic overlap with this region. Tracts originating from the mid-posterior subregion terminate in the occipital lobe, which has been consistently implicated in individuals with MDD (Maciag *et al.*, 2010; Kihira *et al.*, 2020; Song *et al.*, 2021). It is possible that changes in the connectivity or integrity of this subregion could contribute to the differences noted in the occipital lobe in individuals with MDD and could underlie some of the visual processing deficits and altered neurochemistry noted in the disorder.

It is noteworthy that there was limited genetic overlap contributing to CC with those contributing to ADHD and ASD. Although individuals with ASD have been shown to have lower overall CC volume, the lack of evidence to suggest genetic overlap is in line with prior work (Fame, MacDonald and Macklis, 2011; Campbell *et al.*, 2023). The lower number of significant loci identified after conditioning on ADHD and ASD, may be due to the power of these GWAS. It is worth noting that MiXeR is robust to power fluctuations, suggesting that the observed smaller number of loci still indicates overlap. MiXeR reliably captures subthreshold signals, reinforcing the potential genetic connection between ADHD, ASD and the CC. It is worth to highlight this observation, especially as GWAS sample sizes constantly keep increasing, as this may enhance the overlap over time.

The central subregion presents a distinct pattern, with a consistently lower proportion of overlap with all substance use phenotypes, except for drinks per week, which has its highest proportion of overlap within this region. Of these shared variants, only 43% had a concordant direction of effect- indicating mixed genetic effects between the central subregion and alcohol

consumption. These findings are consistent with work showing that higher alcohol consumption is associated with smaller central subregion volumes (González-Reimers *et al.*, 2019). While the overall overlap of CC and substance use phenotypes is slightly lower than in psychotic disorders, there was substantial overlap with anterior and posterior subregions. This trend is most pronounced in the case of cannabis and smoking initiation traits.

Our gene-set enrichment analyses revealed a noteworthy association between the extracellular matrix (ECM) in both SCZ and smoking initiation and the CC. The ECM, a complex network of macromolecules, plays a pivotal role in maintaining the structural integrity of the CC. In the conjunctural analysis of SCZ and the CC, the findings implicated negative regulation of the ECM, aligning with previous studies that have reported alterations in ECM components in patients with SCZ (Pantazopoulos *et al.*, 2020). These alterations can lead to neurophysiological dysregulation and progressive neuronal impairment. The conjunctural analyses of the CC and smoking initiation, highlighted the significance of hyaluronan metabolism, a key component of the ECM. Hyaluronan contributes to the ECM's viscoelastic properties, influencing cell behaviour and potentially modulates the neurobiological mechanisms underlying smoking initiation. These findings underscore the potential importance of the ECM in the pathophysiology of both SCZ and smoking initiation, and its role in the structural and functional dynamics of the CC. The ECM is crucial for structural support and cell signaling. Alterations in ECM components can affect synaptic plasticity, impacting cognitive functions and contributing to psychiatric symptoms seen in disorders like BD and SCZ. Hyaluronan is integral to the ECM, influencing cell proliferation, migration, and inflammation. Abnormal hyaluronan metabolism has been linked to psychiatric disorders, potentially affecting brain structure and function. Further research into these associations could provide valuable insights into potential therapeutic targets. Some limitations deserve emphasis. I used summary statistics from predominantly European GWAS, which limits the

generalizability of these findings. Further work is needed with more diverse samples, and trans-ancestry tools. Additionally, since the conjunctural analyses require non-overlapping samples, I was not able to execute the analyses for all phenotypes (e.g. cannabis consumption). Additionally, some of the null findings may be due to low power in the original GWAS which may underestimate the shared genetic architecture between phenotypes however, MiXeR is robust to variations in sample size and should produce consistent estimates for GWASs of the same phenotype with different sample sizes assuming that the GWAS is powerful enough (Frei et al., 2019; Hindley et al., 2022). The latter is assessed in our analysis by the AIC criterion. We do not report the results when the model is not able to provide robust results according to the AIC criterion. MiXeR analysis is based on GWAS summary statistics and therefore depends on the quality on the summary statistics and inherits all limitations of the standard GWAS (Frei et al., 2019; Hindley et al., 2022). So while an increase in sample size may improve discoverability, and allow us to uncover additional biological pathways implicated in the pathology of the disorder, since that was not the focus of this study, it should not alter these results (Frei et al., 2019; Hindley et al., 2022).

Genetic correlation often underrepresents the shared genetic component between two traits, which is why we included bivariate MiXeR in our analysis. Global genetic correlation may underrepresent the shared genetic component because there can be many localised negative and positive correlations that result in an overall null effect (Hindley et al., 2022). Our interpretation of the largely discordant effects is that it reinforces the complexity of these traits. The interaction of these variants may result in complex phenotypes. For example, while one genomic region may convey risk for a psychiatric disorder, it may not convey risk to altered CC volumes. Although there is a large component of shared genetic 'risk', there also remains some distinct genetic component, and of course, the interaction with the environment must be considered. We did not investigate different pathways implicated by variants with concordant or discordant relationships between the two phenotypes because MiXeR provides a 'bird's eye view' of the genetic

relationship between two traits and does not identify specific SNPs. Unlike conjFDR, which can identify specific pleiotropic loci, MiXeR provides i) the proportion of SNPs that are trait-influencing, ii) the genetic overlap between two traits, and iii) the proportion of genetic variants with concordant directions of effect. This is based on a model of true discovery, meaning that as sample sizes become large enough to discover all variants associated with a trait of interest, this is what we will see (Hindley et al., 2022)..

In conclusion, I describe distinct patterns of genetic overlap, shedding light on the potential shared biological pathways and mechanisms underlying these associations. The genetic overlap identified between the CC, BD and SCZ aligns with previous research highlighting extensive genetic commonalities between these psychiatric disorders. The involvement of the anterior CC in emotion regulation and connectivity to prefrontal cortices further underscores the significance of these findings in understanding emotional dysregulation in psychotic spectrum disorders. Although replication is needed, overlapping genetic architectures suggest the involvement of shared biological pathways, and the potential value of targeting such pathways for the purposes of treatment.

6 Discussion

6.1 Main Findings

This thesis leveraged novel statistical genetic methodologies to enhance our understanding of the CC and its genetic relationship with mental disorders and substance use phenotypes. I utilized both multivariate and univariate GWAS techniques to dissect the genetic architecture of the overall volume of the CC and its individual subregions, uncovering multiple novel variants with small, yet distributed effects, in addition to those identified through univariate analyses. Distinct variants were implicated in the development of CC, resulting in regional patterns with clear anterior-posterior variation and corresponding functional relevance. Distinct variants were implicated in the development of the CC's anterior and posterior subregions, in alignment with their functional distinctions. CC volumes exhibited substantial h^2_{snp} , with variations observed across subregions. A significant genetic correlation was found between the volumes of specific CC subregions and BD, alcohol and cannabis consumption, and MDD, underscoring the genetic complexity of the CC. Subsequent analysis revealed varying degrees of polygenicity within CC subregions. Bivariate investigations indicated a broad genetic overlap between CC subregion volumes and many psychiatric phenotypes, which was observed independent of genome-wide genetic correlation—a sign of widespread mixed effect directions among the shared variants. Leveraging this pleiotropy, I pinpointed common variants associated with the CC, its subregion volumes, and psychiatric phenotypes, revealing the involvement in biological pathways involved in antigen processing, extracellular matrix, and hyaluronan metabolism. Further exploration into the genetic architecture of white matter integrity within the CC, assessed across 28 diffusion imaging modalities, led to the identification of several genetic loci, highlighting actin binding and cytoskeletal protein binding. Estimates of polygenicity and h^2_{snp} across these imaging

modalities and CC subregions demonstrated similar genetic overlap patterns, with significant genetic correlations observed for BD and alcohol consumption across multiple modalities, indicating subregion-specific genetic influences.

In this final chapter, I will summarize the findings by major themes that arise from this thesis' three empirical chapters: **i)** The discovery of genomic loci contributing to the CC and its subregions and its underlying biology; **ii)** Polygenicity and heritability of the CC; **iii)** The genetic relationship between the CC and psychiatric phenotypes. I will discuss these themes in the context of their implications for understanding the genetic architecture of the CC and relevant psychiatric phenotypes. I will then consider the potential clinical implications of these findings and discuss relevant strengths and weaknesses, before addressing the future directions of this research.

6.2 The Discovery of Genomic Loci Contributing to the Corpus Callosum and its Underlying Biology

This study aimed to investigate the genetic architecture of the CC, in this section we discuss the discovery of novel genetic loci associated with CC volume and white matter integrity and their implications for the underlying biology of the CC.

I identified 70 novel loci associated with CC volume (chapter 3). These loci exhibit distributed effects across the five subregions of the CC. This finding suggests that there is a similar biology underlying the development of these different subregions, potentially involving common genetic pathways. Multivariate analysis was able to substantially boost the genetic discovery efforts, gaining a more comprehensive understanding of the intricate genetic architecture influencing volume of the CC and its subregions. The univariate GWAS reveal that five

subregions have both shared and distinct genetic etiology. While some loci, associated with whole CC volume partially overlap with those of individual subregions, others remain unique to specific subregions. Notably, rs755857 was significant in all GWAS except for the mid-posterior subregion. Although this SNP has no prior associations with other traits, it is linked to the PINX1 gene, which has implications for liver function markers such as gamma glutamyl transferase and serum alkaline phosphatase levels (Barton *et al.*, 2021).

I found variation in white matter microstructure of the CC to be polygenic and a substantial proportion of its heritability can be attributed to common variants. Univariate GWAS analysis of diffusion metrics revealed a total of 1,520 significant genetic loci associated with diffusion metrics in the subregions of the CC. The gene-set analysis of diffusion metrics revealed several pathways that may contribute to alterations of white matter in the CC. One pathway was the hemidesmosome pathway. Hemidesmosomes are specialized cell-matrix adhesion complexes that play a crucial role in stabilizing cell adhesion and tissue integrity. Therefore, alterations in this pathway could potentially impact the structural integrity of the CC. Several pathways related to actin binding and cytoskeletal protein binding were also identified.. Actin filaments provide structural support and are essential for maintaining cell morphology and stability (Bosch *et al.*, 2014). Dysregulation of the actin cytoskeleton has been implicated in the regulation of neuronal plasticity, synapse formation, and dendritic spine morphology, which are crucial processes for normal brain function (Bosch *et al.*, 2014). Thus, dysregulation of these pathways could disrupt the organization and stability of white matter tracts within the CC, leading to alterations in white matter integrity (Szikora *et al.*, 2017). Disruptions in cellular structure have been observed in psychiatric disorders, therefore, the genes within these pathways could be potential targets for further investigation and therapeutic interventions (Szikora *et al.*, 2017).

6.3 The Polygenic Architecture of the Corpus Callosum and its Underlying Heritability

The genetic architecture of the CC and its subregions is complex and polygenic. This work has identified numerous unique genetic loci contributing to CC volume and white matter integrity. Notably, we conducted the first GWAS of white matter integrity across the five CC subregions, revealing 1,520 genetic loci associated with 28 diffusion metrics (Chapter 4). In contrast, our GWAS focused on CC volume identified 70 unique genetic loci (as discussed in Chapter 3). These findings underscore the novelty of our research and suggest a shared biological basis underlying the development of different CC subregions, hinting at common genetic pathways that shape this critical brain structure.

The heritability estimates of the CC and its subregions are high, indicating a strong genetic influence. The SNP-based heritability of total CC volume was estimated at 0.38 (Chapter 3), while the mean SNP-based heritability of white matter integrity metrics in CC subregions ranged from 22.28% (mid posterior) to 24.34% (posterior; Chapter 4). These estimates are consistent with previous family reports indicating that the size of the CC and its subregions are under strong genetic influence (Woldehawariat *et al.*, 2014; Chapter 3).

In conclusion, the polygenic architecture of the CC and its subregions is complex and involves numerous genetic loci. The genetic contributions to the CC and its subregions are both overlapping and unique, and the heritability estimates indicate a strong genetic influence. Despite this complexity, the CC is less polygenic than substance use phenotypes and psychiatric phenotypes. These findings provide a comprehensive understanding of the genetic

architecture of the CC and its subregions, and highlight the need for further research to elucidate the specific genetic pathways involved.

6.3.1 The Genetic Relationship Between the Corpus Callosum and Psychiatric Phenotypes

In Chapters 3,4 and 5 I found evidence to suggest a significant genetic relationship between the CC and many psychiatric phenotypes. ASD is associated with decreased total CC volume in prior imaging studies, yet the genetic overlap between the CC and ASD remains minimal. However, significant genetic correlation between the central subregion was observed with ASD, supporting previous volumetric findings (Frazier and Hardan, 2009; Fame, MacDonald and Macklis, 2011; Valenti *et al.*, 2020). In SCZ, notable genetic overlap emerges between cannabis use and the CC, with gene-set enrichment analyses implicating the ECM as a shared pathological feature influencing both SCZ and smoking initiation (Pantazopoulos *et al.*, 2020; Vaissiere *et al.*, 2020). MDD was also observed to have almost complete genetic overlap with the mid-posterior subregion of the CC, suggesting that changes in this area may impact the disorder's associated visual processing and neurochemical discrepancies (Maciag *et al.*, 2010; Kihira *et al.*, 2020; Song *et al.*, 2021). These collective findings illuminate the intricate genetic interrelations between the CC and various psychiatric and behavioral conditions, providing a nuanced understanding of their aetiology and potential points of intervention.

Previous work consistently implicates the CC in BD and alcohol consumption. The findings from this study provide further evidence supporting this robust relationship. The significant genetic overlap between BD and the CC, especially within the mid anterior and central subregions, reflects neuroimaging studies that show smaller CC volumes in BD patients

relative to control groups. Such volume reductions suggest alterations in cellular composition and structural integrity, as indicated by microFA and microADC studies (Prunas et al., 2018; de Souza et al., 2019; Valenti et al., 2020; Arnone, Abou-Saleh and Barrick, 2006; Gifuni et al., 2017; Reisert et al., 2017). Similarly, genetic correlations between alcohol consumption and CC microstructure, particularly in the mid anterior and central subregions, highlight how alcohol may contribute to morphometric changes, potentially leading to CC ultrastructure disruption and atrophy (de Souza et al., 2019; Korbmacher et al., 2023; He et al., 2007). The findings presented here align with those from previous imaging studies. Moreover, the findings documented in Chapters 3 and 4 indicate similar patterns of genetic overlap between BD, alcohol consumption and the CC, across imaging modalities.

7 Future Directions

In the following chapter, I will outline the strengths and weakness of this study before focusing on three critical avenues for advancing our understanding of the CC and its genetic underpinnings.

7.1 Strengths and Limitations

Some limitations deserve emphasis. First, this study was the largest genetic investigation in to the genetic architecture of the CC, it is limited to individuals residing in the United Kingdom and consisted largely of individuals of European ancestry (Chapter 3 and 4). I did however, replicate these findings in an independent sample of non-European ancestry. Future genetic work may benefit additional analyses conducted in more diverse cohorts. Second, the age range of individuals in the UK biobank is limited to those above 45 years old. Therefore, it would be important to replicate these findings in a younger sample to ensure generalizability. Third, the comparison of the results with the literature is challenging due to differences in approach to subregion divisions. Although I implemented two segmentation protocols with consistent results, a robust CC segmentation is needed to improve replicability of research in this field (Chapter 3). Additionally, since the conjunctural analyses require non-overlapping samples, I was not able to execute the analyses for all phenotypes (e.g. cannabis consumption; Chapter 5). Finally, since not all GWAS were equally powered, some of the null findings may be due to low power in the original GWAS which may underestimate the shared genetic architecture between phenotypes (Chapter 5).

7.2 Validation and Harmonization of Segmentation Algorithms

Validation and harmonization of segmentation algorithms for the CC is essential for ensuring accurate and reliable neuroimaging analyses. Various studies have focused on refining segmentation techniques to enhance the precision of delineating the CC and its subregions. For instance, a tractography-based segmentation approach has been shown to identify lower interhemispheric connectivity in SCZ, emphasizing the importance of accurate CC segmentation. To date, there are no standardised segmentation protocols to segment the CC in an automated fashion, allowing for replicable, large-scale analysis. In this study we utilised FreeSurfer and C8 software. While I found a degree of similarity between these two segmentation protocols (chapter 3), it remains imperative for there to be a standardised segmentation protocols to allow for comparison and meta-analysis across studies. These studies (chapter 3 and 4) underscore the significance of validating and harmonizing segmentation algorithms to ensure consistent and accurate delineation of the CC facilitating reliable neuroanatomical analyses and advancing the understanding of brain structure and function.

7.3 Exploring Other Forms of Genetic Variation in Relation to the Corpus Callosum

The full extent of genetic variation impacting the characteristics of the CC and its subregions may not be fully covered by SNP-based heritability measures (Rosas et al., 2010). Therefore, it is essential to take into account additional types of genetic variation, including non-additive factors such as gene-environment interactions, and rare genetic variations like copy number variations (CNVs; Schependorn et al., 2016). CNVs, which are known to have a greater impact compared to SNPs, have been linked to various psychiatric and brain-related disorders, indicating their potential significance in explaining the "missing" heritability and providing deeper insights into the physiological aspects of CC volume and its implications for

related disorders (Nazem-Zadeh et al., 2012). Integrating data from genomics, neuroimaging, and other functional analyses, such as gene expression, may aid in the understanding the genetic architecture relating to the structure and function of the CC (Yaldizli et al., 2013).

7.4 Functional Assessment of Identified Loci

The clinical relevance of the genetic variations and identified genes connected to the CC is not well understood, which presents a challenge for interpreting their functional significance. The complex nature of LD in the human genome, coupled with design constraints of microarray technology, complicates the task of pinpointing the causal genetic variants (Cano-Gamez and Trynka, 2020). A large proportion of these associated variants reside in intergenic regions which are thought to play a role in gene transcription levels, possibly due to their proximity to regulatory elements (Maurano et al., 2012) and locations where transcription factors bind (Musunuru et al., 2010). Effectively attributing these intergenic regions to specific genes is challenging—they are either mapped to numerous genes or mapping is unreliable, making it difficult to discern the exact genes that contribute to disease and to the affected tissue types (Cano-Gamez and Trynka, 2020). To advance the translational impact of such research, strategies such as single-cell sequencing are vital for understanding gene involvement.

Statistical methods, such as fine mapping, are also useful in understanding the causal variant associated with a phenotype and could be useful in the case of the CC. This process involves advanced statistical techniques such as Bayesian approaches and functional annotations of genetic variants to prioritize and pinpoint the variants most likely responsible for observed associations (Chen et al., 2015). By utilizing fine mapping algorithms like Probabilistic Identification of Causal SNPs (PICS), one can estimate the probability of each SNP being a causal variant based on the observed association patterns at the locus (Farh et al., 2014).

Through the detailed analysis of genetic information within implicated loci, fine mapping studies can prioritize variants according to their probability of causality, aiding in the identification of specific causal variants within GWAS loci (Kichaev et al., 2016). This approach is crucial for understanding the genetic architecture underlying complex traits like the CC and distinguishing causal variants from neighboring genetic variations (Chen et al., 2015). Therefore future work employing fine mapping techniques, could enhance our understanding of the genetic factors influencing the CC and related psychiatric phenotypes.

Mendelian Randomization (MR) is another method in genetic epidemiology used to determine causal relationships between genetic risk factors, like CC volume alterations, and disease outcomes, such as psychiatric disorders (Burgess et al., 2015). MR leverages genetic variants as instrumental variables, taking advantage of their random assortment, which mimics the randomization process in controlled trials (Burgess et al., 2013). By using genetic proxies associated with CC volume, MR can provide robust evidence of causality, offering insights into the biological mechanisms underlying these disorders. This approach is clinically valuable, as it helps identify specific genetic targets for therapeutic interventions, improves risk prediction for psychiatric disorders, and guides clinical decision-making by distinguishing causal from non-causal associations (Burgess et al., 2015). Incorporating MR into future research will significantly enhance our understanding of the genetic underpinnings of psychiatric disorders and facilitate the development of targeted treatments.

8 Conclusion

This thesis has significantly advanced our understanding of the CC and its genetic associations with psychiatric disorders and substance use phenotypes. Leveraging novel statistical genetic methodologies, I identified 70 novel loci associated with CC volume, revealing distributed effects across the five subregions. These findings imply a shared underlying biology governing the development of distinct CC areas, potentially involving common genetic pathways. Multivariate analysis further enhanced genetic discovery efforts, providing a comprehensive view of the intricate genetic architecture influencing CC volume. Additionally, high heritability estimates underscore the strong genetic influence on CC characteristics. Notably, the CC's polygenic architecture, while complex, remains less pronounced than that of substance use phenotypes and psychiatric conditions. Collectively, these insights deepen our understanding of the CC's genetic basis and highlight potential therapeutic targets for disorders where cellular structure disruptions play a crucial role.

References

- Abramovic, L. *et al.* (2018) 'White matter disruptions in patients with bipolar disorder', *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology*, 28(6), pp. 743–751. Available at: <https://doi.org/10.1016/j.euroneuro.2018.01.001>.
- Acer, N. *et al.* (2017) 'Anatomical Characterization Of ADHD Using An Atlas-based Analysis: A Diffusion Tensor Imaging Study', *The Eurobiotech Journal*, 1(1), pp. 46–56. Available at: <https://doi.org/10.24190/issn2564-615x/2017/01.08>.
- Andersson, J.L. and Sotiropoulos, S.N. (2016) 'An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging', *Neuroimage*, 125, pp. 1063–1078.
- Andreassen, O.A. *et al.* (2013) 'Improved Detection of Common Variants Associated with Schizophrenia and Bipolar Disorder Using Pleiotropy-Informed Conditional False Discovery Rate', *PLOS Genetics*, 9(4), p. e1003455. Available at: <https://doi.org/10.1371/journal.pgen.1003455>.
- Arnone, D. *et al.* (2008) 'Meta-analysis of magnetic resonance imaging studies of the corpus callosum in bipolar disorder', *Acta Psychiatrica Scandinavica*, 118(5), pp. 357–362. Available at: <https://doi.org/10.1111/j.1600-0447.2008.01229.x>.
- Arnone, D., Abou-Saleh, M.T. and Barrick, T.R. (2006) 'Diffusion tensor imaging of the corpus callosum in addiction', *Neuropsychobiology*, 54(2), pp. 107–113.
- Atkinson, E.G. *et al.* (2021) 'Tractor uses local ancestry to enable the inclusion of admixed individuals in GWAS and to boost power', *Nature Genetics*, 53(2), pp. 195–204. Available at: <https://doi.org/10.1038/s41588-020-00766-y>.
- Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium (2017) 'Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia', *Molecular Autism*, 8, p. 21. Available at: <https://doi.org/10.1186/s13229-017-0137-9>.
- Awad, A.G. and Voruganti, L.L.N.P. (2015) 'Revisiting the “self-medication” hypothesis in light of the new data linking low striatal dopamine to comorbid addictive behavior', *Therapeutic Advances in Psychopharmacology*, 5(3), pp. 172–178. Available at: <https://doi.org/10.1177/2045125315583820>.
- Bahrani, S., Hindley, G., *et al.* (2022) 'Dissecting the shared genetic basis of migraine and mental disorders using novel statistical tools', *Brain*, 145(1), pp. 142–153. Available at: <https://doi.org/10.1093/brain/awab267>.
- Bahrani, S., Nordengen, K., *et al.* (2022) 'Distributed genetic architecture across the hippocampal formation implies common neuropathology across brain disorders', *Nature Communications*, 13(1), p. 3436. Available at: <https://doi.org/10.1038/s41467-022-31086-w>.

Bahrami, S. *et al.* (2023) ‘Unveiling the Genetic Landscape of Basal Ganglia: Implications for Common Brain Disorders’. Available at: <https://doi.org/10.1101/2023.07.26.23293206>.

Barton, A.R. *et al.* (2021) ‘Whole-exome imputation within UK Biobank powers rare coding variant association and fine-mapping analyses’, *Nature genetics*, 53(8), pp. 1260–1269. Available at: <https://doi.org/10.1038/s41588-021-00892-1>.

Basser, P.J., Mattiello, J. and LeBihan, D. (1994) ‘MR diffusion tensor spectroscopy and imaging’, *Biophysical journal*, 66(1), pp. 259–267.

Bearden, C.E. *et al.* (2011) ‘Mapping Corpus Callosum Morphology in Twin Pairs Discordant for Bipolar Disorder’, *Cerebral Cortex*, 21(10), pp. 2415–2424. Available at: <https://doi.org/10.1093/cercor/bhr030>.

Beck, D. *et al.* (2021) ‘White matter microstructure across the adult lifespan: A mixed longitudinal and cross-sectional study using advanced diffusion models and brain-age prediction’, *NeuroImage*, 224, p. 117441. Available at: <https://doi.org/10.1016/j.neuroimage.2020.117441>.

Bipolar Disorder Working Group of the Psychiatric Genomics Consortium, 23andMe Research Team *et al.* (2023) ‘Genetic diversity enhances gene discovery for bipolar disorder’. medRxiv, p. 2023.10.07.23296687. Available at: <https://doi.org/10.1101/2023.10.07.23296687>.

Bosch, M. *et al.* (2014) ‘Structural and molecular remodeling of dendritic spine substructures during long-term potentiation’, *Neuron*, 82(2), pp. 444–459. Available at: <https://doi.org/10.1016/j.neuron.2014.03.021>.

Brown, J.R. *et al.* (2021) ‘Lower white matter microstructure in bipolar disorder with and without psychosis’, *Bipolar Disorders*, 23(8), pp. 801–809. Available at: <https://doi.org/10.1111/bdi.13055>.

Brown, W. and Paul, L. (2019) ‘The neuropsychological syndrome of agenesis of the corpus callosum’, *Journal of the International Neuropsychological Society*, 25(3), pp. 324–330. Available at: <https://doi.org/10.1017/s135561771800111x>.

Bruï ne, F. *et al.* (2013) ‘Tractography of white-matter tracts in very preterm infants: a 2-year follow-up study’, *Developmental Medicine & Child Neurology*, 55(5), pp. 427–433. Available at: <https://doi.org/10.1111/dmcn.12099>.

Bulik-Sullivan, B. *et al.* (2015) ‘An atlas of genetic correlations across human diseases and traits’, *Nature genetics*, 47(11), pp. 1236–1241.

Bulik-Sullivan, B.K. *et al.* (2015) ‘LD Score regression distinguishes confounding from polygenicity in genome-wide association studies’, *Nature genetics*, 47(3), pp. 291–295.

Bycroft, C. *et al.* (2018) ‘The UK Biobank resource with deep phenotyping and genomic data’, *Nature*, 562(7726), pp. 203–209.

Campbell, M.L. *et al.* (2023) ‘Distributed Genetic Effects of the Corpus Callosum Subregions Suggest Links to Neuropsychiatric Disorders and Related Traits’. medRxiv, p. 2023.05.13.23289938. Available at: <https://doi.org/10.1101/2023.05.13.23289938>.

Carrión-Castillo, A. and Boeckx, C. (2024) ‘Insights into the genetic architecture of cerebellar lobules derived from the UK Biobank’, *Scientific Reports*, 14(1), p. 9488. Available at: <https://doi.org/10.1038/s41598-024-59699-9>.

Cole, J.H. *et al.* (2012) ‘White matter abnormalities and illness severity in major depressive disorder’, *British Journal of Psychiatry*, 201(1), pp. 33–39. Available at: <https://doi.org/10.1192/bjp.bp.111.100594>.

Connaughton, M. *et al.* (2022) ‘White matter microstructure in children and adolescents with ADHD’, *NeuroImage: Clinical*, 33, p. 102957. Available at: <https://doi.org/10.1016/j.nicl.2022.102957>.

Cover, G.S. *et al.* (2018) ‘Computational methods for corpus callosum segmentation on MRI: A systematic literature review’, *Computer Methods and Programs in Biomedicine*, 154, pp. 25–35. Available at: <https://doi.org/10.1016/j.cmpb.2017.10.025>.

Cox, S.R. *et al.* (2016) ‘Ageing and brain white matter structure in 3,513 UK Biobank participants’, *Nature Communications*, 7(1), p. 13629. Available at: <https://doi.org/10.1038/ncomms13629>.

De Moor, M. *et al.* (2011) ‘Heritability of Problem Drinking and the Genetic Overlap with Personality in a General Population Sample’, *Frontiers in Genetics*, 2, p. 76. Available at: <https://doi.org/10.3389/fgene.2011.00076>.

Demontis, D. *et al.* (2019) ‘Genome-wide association study implicates CHRNA2 in cannabis use disorder’, *Nature Neuroscience*, 22(7), pp. 1066–1074. Available at: <https://doi.org/10.1038/s41593-019-0416-1>.

Demontis, D. *et al.* (2023) ‘Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains’, *Nature Genetics*, 55(2), pp. 198–208. Available at: <https://doi.org/10.1038/s41588-022-01285-8>.

Deoni, S. *et al.* (2014) ‘White matter maturation profiles through early childhood predict general cognitive ability’, *Brain Structure and Function*, 221(2), pp. 1189–1203. Available at: <https://doi.org/10.1007/s00429-014-0947-x>.

Edvardsen, J. *et al.* (2008) ‘Heritability of bipolar spectrum disorders. Unity or heterogeneity?’, *Journal of affective disorders*, 106(3), pp. 229–240.

Elvsåshagen, T. *et al.* (2021) ‘The genetic architecture of the human thalamus and its overlap with ten common brain disorders’, *Nature Communications*, 12(1), p. 2909. Available at: <https://doi.org/10.1038/s41467-021-23175-z>.

Fame, R.M., MacDonald, J.L. and Macklis, J.D. (2011) ‘Development, Specification, and Diversity of Callosal Projection Neurons’, *Trends in neurosciences*, 34(1), pp. 41–50. Available at: <https://doi.org/10.1016/j.tins.2010.10.002>.

Fan, Z. *et al.* (2019) ‘Alcohol Consumption Can be a “Double-Edged Sword” for Chronic Kidney Disease Patients’, *Medical Science Monitor : International Medical Journal of Experimental and Clinical Research*, 25, pp. 7059–7072. Available at: <https://doi.org/10.12659/MSM.916121>.

- Fieremans, E., Jensen, J.H. and Helpert, J.A. (2011) 'White matter characterization with diffusional kurtosis imaging', *NeuroImage*, 58(1), pp. 177–188. Available at: <https://doi.org/10.1016/j.neuroimage.2011.06.006>.
- Fischl, B. *et al.* (2002) 'Whole Brain Segmentation', *Neuron*, 33(3), pp. 341–355. Available at: [https://doi.org/10.1016/S0896-6273\(02\)00569-X](https://doi.org/10.1016/S0896-6273(02)00569-X).
- Fortin, J.-P. *et al.* (2018) 'Harmonization of cortical thickness measurements across scanners and sites', *NeuroImage*, 167, pp. 104–120. Available at: <https://doi.org/10.1016/j.neuroimage.2017.11.024>.
- Francis, A.N. *et al.* (2016) 'Callosal Abnormalities Across the Psychosis Dimension: Bipolar Schizophrenia Network on Intermediate Phenotypes', *Biological Psychiatry*, 80(8), pp. 627–635. Available at: <https://doi.org/10.1016/j.biopsych.2015.12.026>.
- Franco, I. *et al.* (1993) 'Genitourinary malformations associated with agenesis of the corpus callosum', *The Journal of Urology*, 149(5), pp. 1119–1121. Available at: [https://doi.org/10.1016/s0022-5347\(17\)36313-9](https://doi.org/10.1016/s0022-5347(17)36313-9).
- Frazier, T.W. and Hardan, A.Y. (2009) 'A Meta-Analysis of the Corpus Callosum in Autism', *Biological Psychiatry*, 66(10), pp. 935–941. Available at: <https://doi.org/10.1016/j.biopsych.2009.07.022>.
- Frei, O. *et al.* (2019) 'Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation', *Nature Communications*, 10(1), p. 2417. Available at: <https://doi.org/10.1038/s41467-019-10310-0>.
- Frumin, M. *et al.* (2002) 'Shape Differences In the Corpus Callosum In First-episode Schizophrenia And First-episode Psychotic Affective Disorder', *AJP*, 5(159), pp. 866–868. Available at: <https://doi.org/10.1176/appi.ajp.159.5.866>.
- Gadewar, S.P. *et al.* (2023) 'A Comprehensive Corpus Callosum Segmentation Tool for Detecting Callosal Abnormalities and Genetic Associations from Multi Contrast MRIs'. arXiv. Available at: <https://doi.org/10.48550/arXiv.2305.01107>.
- Gifuni, A.J. *et al.* (2017) 'Corpus callosum volumes in bipolar disorders and suicidal vulnerability', *Psychiatry Research: Neuroimaging*, 262, pp. 47–54. Available at: <https://doi.org/10.1016/j.psychres.2017.02.002>.
- González-Reimers, E. *et al.* (2019) 'Chapter 15 - Effects of Alcohol on the Corpus Callosum', in V.R. Preedy (ed.) *Neuroscience of Alcohol*. Academic Press, pp. 143–152. Available at: <https://doi.org/10.1016/B978-0-12-813125-1.00015-5>.
- Grasby, K.L. *et al.* (2020) 'The genetic architecture of the human cerebral cortex', *Science*, 367(6484). Available at: <https://doi.org/10.1126/science.aay6690>.
- Hampton, W.H., Hanik, I.M. and Olson, I.R. (2019) 'Substance Abuse and White Matter: Findings, Limitations, and Future of Diffusion Tensor Imaging Research', *Drug and alcohol dependence*, 197, pp. 288–298. Available at: <https://doi.org/10.1016/j.drugalcdep.2019.02.005>.

Hartz, S.M. *et al.* (2014) ‘Comorbidity of Severe Psychotic Disorders With Measures of Substance Use’, *JAMA Psychiatry*, 71(3), p. 248. Available at: <https://doi.org/10.1001/jamapsychiatry.2013.3726>.

Hatoum, A.S. *et al.* (2021) ‘The Addiction Genetic Factor a(g): A Unitary Genetic Vulnerability Characterizes Substance Use Disorders and Their Associations with Common Correlates’, *medRxiv*, p. 2021.01.26.21250498. Available at: <https://doi.org/10.1101/2021.01.26.21250498>.

Hill, K.P. (2017) ‘Cannabis Use and Risk for Substance Use Disorders and Mood or Anxiety Disorders’, *JAMA*, 317(10), p. 1070. Available at: <https://doi.org/10.1001/jama.2016.19706>.

Hindley, G. *et al.* (2021) ‘Characterising the shared genetic determinants of bipolar disorder, schizophrenia and risk-taking’, *Translational Psychiatry*, 11(1), pp. 1–10. Available at: <https://doi.org/10.1038/s41398-021-01576-4>.

Hindley, G. *et al.* (2022) ‘Charting the Landscape of Genetic Overlap Between Mental Disorders and Related Traits Beyond Genetic Correlation’, *The American Journal of Psychiatry*, 179(11), pp. 833–843. Available at: <https://doi.org/10.1176/appi.ajp.21101051>.

Hofer, S. and Frahm, J. (2006) ‘Topography of the human corpus callosum revisited—Comprehensive fiber tractography using diffusion tensor magnetic resonance imaging’, *NeuroImage*, 32(3), pp. 989–994. Available at: <https://doi.org/10.1016/j.neuroimage.2006.05.044>.

Hopkins, W.D. *et al.* (2022) ‘Heritability in corpus callosum morphology and its association with tool use skill in chimpanzees (*Pan troglodytes*): Reproducibility in two genetically isolated populations’, *Genes, Brain and Behavior*, 21(2), p. e12784. Available at: <https://doi.org/10.1111/gbb.12784>.

Jansen, I.E. *et al.* (2019) ‘Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk’, *Nature Genetics*, 51(3), pp. 404–413. Available at: <https://doi.org/10.1038/s41588-018-0311-9>.

Jensen, J.H. *et al.* (2005) ‘Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging’, *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 53(6), pp. 1432–1440.

Kaden, E. *et al.* (2016) ‘Multi-compartment microscopic diffusion imaging’, *NeuroImage*, 139, pp. 346–359. Available at: <https://doi.org/10.1016/j.neuroimage.2016.06.002>.

Kaden, E., Kruggel, F. and Alexander, D.C. (2016) ‘Quantitative mapping of the per-axon diffusion coefficients in brain white matter’, *Magnetic Resonance in Medicine*, 75(4), pp. 1752–1763. Available at: <https://doi.org/10.1002/mrm.25734>.

Karadag, N. *et al.* (2023) ‘Identification of novel genomic risk loci shared between common epilepsies and psychiatric disorders’, *Brain: A Journal of Neurology*, 146(8), pp. 3392–3403. Available at: <https://doi.org/10.1093/brain/awad038>.

Kellner, E. *et al.* (2016) ‘Gibbs-ringing artifact removal based on local subvoxel-shifts’, *Magnetic resonance in medicine*, 76(5), pp. 1574–1581.

- Kihira, S. *et al.* (2020) ‘Regional parieto-occipital hypoperfusion on arterial spin labeling associates with major depressive disorder’, *The Open Neuroimaging Journal*, 13(1), pp. 30–36. Available at: <https://doi.org/10.2174/1874440002013010030>.
- Korbmacher, M. *et al.* (2023) ‘Brain-wide associations between white matter and age highlight the role of fornix microstructure in brain ageing’, *Human Brain Mapping*, 44(10), pp. 4101–4119. Available at: <https://doi.org/10.1002/hbm.26333>.
- Kranzler, H.R. *et al.* (2019) ‘Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations’, *Nature Communications*, 10(1), p. 1499. Available at: <https://doi.org/10.1038/s41467-019-09480-8>.
- de Leeuw, C.A. *et al.* (2015) ‘MAGMA: generalized gene-set analysis of GWAS data’, *PLoS computational biology*, 11(4), p. e1004219.
- Liang, S. *et al.* (2019) ‘White Matter Abnormalities in Major Depression Biotypes Identified by Diffusion Tensor Imaging’, *Neuroscience Bulletin*, 35(5), pp. 867–876. Available at: <https://doi.org/10.1007/s12264-019-00381-w>.
- Lippard, E.T.C. *et al.* (2017) ‘Brain circuitry associated with the development of substance use in bipolar disorder and preliminary evidence for sexual dimorphism in adolescents’, *Journal of Neuroscience Research*, 95(1–2), pp. 777–791. Available at: <https://doi.org/10.1002/jnr.23901>.
- Liu, J. *et al.* (2021) ‘Genome-wide association study followed by trans-ancestry meta-analysis identify 17 new risk loci for schizophrenia’, *BMC Medicine*, 19, p. 177. Available at: <https://doi.org/10.1186/s12916-021-02039-9>.
- Liu, M. *et al.* (2019) ‘Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use’, *Nature Genetics*, 51(2), pp. 237–244. Available at: <https://doi.org/10.1038/s41588-018-0307-5>.
- Liu, M. *et al.* (2020) ‘White Matter Microstructure Changes and Cognitive Impairment in the Progression of Chronic Kidney Disease’, *Frontiers in Neuroscience*, 14. Available at: <https://www.frontiersin.org/articles/10.3389/fnins.2020.559117> (Accessed: 2 April 2023).
- Loughnan, R.J. *et al.* (2022) ‘Generalization of cortical MOSTest genome-wide associations within and across samples’, *NeuroImage*, 263, p. 119632. Available at: <https://doi.org/10.1016/j.neuroimage.2022.119632>.
- Lu, L. *et al.* (2011) ‘White Matter Microstructure In Untreated First Episode Bipolar Disorder With Psychosis: Comparison With Schizophrenia’, *Bipolar Disorders*, 7–8(13), pp. 604–613. Available at: <https://doi.org/10.1111/j.1399-5618.2011.00958.x>.
- Lu, L.H. *et al.* (2011) ‘White matter microstructure in untreated first episode bipolar disorder with psychosis: comparison with schizophrenia’, *Bipolar Disorders*, 13(7–8), pp. 604–613. Available at: <https://doi.org/10.1111/j.1399-5618.2011.00958.x>.
- Maciąg, D. *et al.* (2010) ‘Lower density of calbindin immunoreactive gabaergic neurons in the occipital cortex in major depression: relevance to neuroimaging studies’, *Biological Psychiatry*, 67(5), pp. 465–470. Available at: <https://doi.org/10.1016/j.biopsych.2009.10.027>.

Martínez-Magaña, J.J. *et al.* (2021) ‘Genome-wide association study of psychiatric and substance use comorbidity in Mexican individuals’, *Scientific Reports*, 11(1), p. 6771. Available at: <https://doi.org/10.1038/s41598-021-85881-4>.

McGuffin, P. *et al.* (2003) ‘The heritability of bipolar affective disorder and the genetic relationship to unipolar depression’, *Archives of general psychiatry*, 60(5), pp. 497–502.

van der Meer, D., Rokicki, J., *et al.* (2020) ‘Brain scans from 21,297 individuals reveal the genetic architecture of hippocampal subfield volumes’, *Molecular Psychiatry*, 25(11), pp. 3053–3065. Available at: <https://doi.org/10.1038/s41380-018-0262-7>.

van der Meer, D., Frei, O., *et al.* (2020) ‘Understanding the genetic determinants of the brain with MOSTest’, *Nature Communications*, 11(1), p. 3512. Available at: <https://doi.org/10.1038/s41467-020-17368-1>.

Moeller, F. *et al.* (2004) ‘Lower Anterior Corpus Callosum White Matter Integrity Is Related To Higher Impulsivity and Lower Discriminability In Cocaine-dependent Subjects: Diffusion Tensor Imaging’, *Neuropsychopharmacol*, 3(30), pp. 610–617. Available at: <https://doi.org/10.1038/sj.npp.1300617>.

Mountjoy, E. *et al.* (2021) ‘An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci’, *Nature Genetics*, 53(11), pp. 1527–1533. Available at: <https://doi.org/10.1038/s41588-021-00945-5>.

Mufford, M.S., van der Meer, D., *et al.* (2021) *The Genetic Architecture of Amygdala Nuclei*. preprint. Genetic and Genomic Medicine. Available at: <https://doi.org/10.1101/2021.06.30.21258615>.

Mufford, M.S., Meer, D. van der, *et al.* (2021) *The Genetic Architecture of Amygdala Nuclei*, p. 2021.06.30.21258615. Available at: <https://doi.org/10.1101/2021.06.30.21258615>.

Mullins, N., Forstner, A.J., O’Connell, K.S., Coombes, B., Coleman, J.R., *et al.* (2021) ‘Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology’, *Nature Genetics*, pp. 1–13.

Mullins, N., Forstner, A.J., O’Connell, K.S., Coombes, B., Coleman, J.R.I., *et al.* (2021) ‘Genome-wide association study of over 40,000 bipolar disorder cases provides new insights into the underlying biology’, *medRxiv*, p. 2020.09.17.20187054. Available at: <https://doi.org/10.1101/2020.09.17.20187054>.

Nagel, M. *et al.* (2018) ‘Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways’, *Nature Genetics*, 50(7), pp. 920–927. Available at: <https://doi.org/10.1038/s41588-018-0151-7>.

Newbury, A. and Rosen, G. (2012) ‘Genetic, Morphometric, and Behavioral Factors Linked to the Midsagittal Area of the Corpus Callosum’, *Frontiers in Genetics*, 3, p. 91. Available at: <https://doi.org/10.3389/fgene.2012.00091>.

Ou, Y.-N. *et al.* (2023) ‘The genetic architecture of human amygdala volumes and their overlap with common brain disorders’, *Translational Psychiatry*, 13, p. 90. Available at: <https://doi.org/10.1038/s41398-023-02387-5>.

Pantazopoulos, H. *et al.* (2020) ‘Molecular signature of extracellular matrix pathology in schizophrenia’, *European Journal of Neuroscience*, 53(12), pp. 3960–3987. Available at: <https://doi.org/10.1111/ejn.15009>.

Pasman, J.A. *et al.* (2018) ‘GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal effect of schizophrenia liability’, *Nature neuroscience*, 21(9), pp. 1161–1170.

PGC ASD (2017) ‘Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia’, *Molecular Autism*, 8, p. 21. Available at: <https://doi.org/10.1186/s13229-017-0137-9>.

Polimanti, R., Agrawal, A. and Gelernter, J. (2017) ‘Schizophrenia and substance use comorbidity: a genome-wide perspective’, *Genome Medicine*, 9(1), p. 25. Available at: <https://doi.org/10.1186/s13073-017-0423-3>.

Pouget, J.G. *et al.* (2019) ‘Cross-disorder analysis of schizophrenia and 19 immune-mediated diseases identifies shared genetic risk’, *Human Molecular Genetics*, 28(20), pp. 3498–3513. Available at: <https://doi.org/10.1093/hmg/ddz145>.

Prunas, C. *et al.* (2018) ‘Diffusion imaging study of the Corpus Callosum in bipolar disorder’, *Psychiatry Research: Neuroimaging*, 271, pp. 75–81. Available at: <https://doi.org/10.1016/j.psychresns.2017.11.001>.

Quattrone, D. *et al.* (2020) ‘Daily use of high-potency cannabis is associated with more positive symptoms in first-episode psychosis patients: the EU-GEI case–control study’, *Psychological Medicine*, pp. 1–9. Available at: <https://doi.org/10.1017/S0033291720000082>.

Ratnakumar, A. *et al.* (2020) ‘Protein-protein interactions uncover candidate “core genes” within omnigenic disease networks’, *PLOS Genetics*, 16(7), p. e1008903. Available at: <https://doi.org/10.1371/journal.pgen.1008903>.

Reisert, M. *et al.* (2017) ‘Disentangling micro from mesostructure by diffusion MRI: A Bayesian approach’, *NeuroImage*, 147, pp. 964–975. Available at: <https://doi.org/10.1016/j.neuroimage.2016.09.058>.

Rigucci, S. *et al.* (2016) ‘Effect of high-potency cannabis on corpus callosum microstructure’, *Psychological Medicine*, 46(4), pp. 841–854. Available at: <https://doi.org/10.1017/S0033291715002342>.

Rosenbloom, M., Sullivan, E.V. and Pfefferbaum, A. (2003) ‘Using Magnetic Resonance Imaging and Diffusion Tensor Imaging to Assess Brain Damage in Alcoholics’, *Alcohol Research & Health*, 27(2), pp. 146–152.

Satizabal, C.L. *et al.* (2019) ‘Genetic architecture of subcortical brain structures in 38,851 individuals’, *Nature Genetics*, 51(11), pp. 1624–1636. Available at: <https://doi.org/10.1038/s41588-019-0511-y>.

Schuckit, M.A. (2009) ‘Alcohol-use disorders’, *The Lancet*, 373(9662), pp. 492–501. Available at: [https://doi.org/10.1016/S0140-6736\(09\)60009-X](https://doi.org/10.1016/S0140-6736(09)60009-X).

Shadrin, A.A. *et al.* (2020) ‘Phenotype-specific differences in polygenicity and effect size distribution across functional annotation categories revealed by AI-MiXeR’, *Bioinformatics*, 36(18), pp. 4749–4756. Available at: <https://doi.org/10.1093/bioinformatics/btaa568>.

Smeland, O.B. *et al.* (2020) ‘Discovery of shared genomic loci using the conditional false discovery rate approach’, *Human genetics*, 139(1), pp. 85–94.

Song, X. *et al.* (2021) ‘Reduction of higher-order occipital gaba and impaired visual perception in acute major depressive disorder’, *Molecular Psychiatry*, 26(11), pp. 6747–6755. Available at: <https://doi.org/10.1038/s41380-021-01090-5>.

de Souza, R.S.M. *et al.* (2019) ‘Anterior to Midposterior Corpus Callosum Subregions Are Volumetrically Lower in Male Alcoholics but Only the Anterior Segment Is Associated to Alcohol Use’, *Frontiers in Psychiatry*, 10, p. 196. Available at: <https://doi.org/10.3389/fpsy.2019.00196>.

Stahl, E.A. *et al.* (2019) ‘Genome-wide association study identifies 30 loci associated with bipolar disorder’, *Nature genetics*, 51(5), pp. 793–803.

Stergiakouli, E. *et al.* (2012) ‘Investigating the contribution of common genetic variants to the risk and pathogenesis of ADHD’, *The American Journal of Psychiatry*, 169(2), pp. 186–194. Available at: <https://doi.org/10.1176/appi.ajp.2011.11040551>.

Strakowski, S.M. and DelBello, M.P. (2000) ‘The co-occurrence of bipolar and substance use disorders’, *Clinical Psychology Review*, 20(2), pp. 191–206. Available at: [https://doi.org/10.1016/S0272-7358\(99\)00025-2](https://doi.org/10.1016/S0272-7358(99)00025-2).

Szikora, S. *et al.* (2017) ‘The formin daam is required for coordination of the actin and microtubule cytoskeleton in axonal growth cones’, *Development*, 144(16), p. e1.1–e1.1. Available at: <https://doi.org/10.1242/dev.157982>.

Tesfaye, M. *et al.* (2023) ‘Shared genetic architecture between irritable bowel syndrome and psychiatric disorders reveals molecular pathways of the gut-brain axis’, *Genome Medicine*, 15(1), p. 60. Available at: <https://doi.org/10.1186/s13073-023-01212-4>.

Thompson, P.M. *et al.* (2003) ‘Mapping structural alterations of the corpus callosum during brain development and degeneration.’

Trubetsky, V. *et al.* (2022) ‘Mapping genomic loci implicates genes and synaptic biology in schizophrenia’, *Nature*, 604(7906), pp. 502–508. Available at: <https://doi.org/10.1038/s41586-022-04434-5>.

Turley, P. *et al.* (2018) ‘Multi-trait analysis of genome-wide association summary statistics using MTAG’, *Nature Genetics*, 50(2), pp. 229–237. Available at: <https://doi.org/10.1038/s41588-017-0009-4>.

Vaissiere, J. *et al.* (2020) ‘Exploring phenotypic and genetic overlap between cannabis use and schizotypy’, *Twin Research and Human Genetics*, 23(4), pp. 221–227. Available at: <https://doi.org/10.1017/thg.2020.68>.

Valenti, M. *et al.* (2020) ‘Abnormal Structural and Functional Connectivity of the Corpus Callosum in Autism Spectrum Disorders: a Review’, *Review Journal of Autism and*

Developmental Disorders, 7(1), pp. 46–62. Available at: <https://doi.org/10.1007/s40489-019-00176-9>.

Veraart, J. *et al.* (2016) ‘Denoising of diffusion MRI using random matrix theory’, *Neuroimage*, 142, pp. 394–406.

Voineskos, A.N. *et al.* (2012) ‘Age-related decline in white matter tract integrity and cognitive performance: a dti tractography and structural equation modeling study’, *Neurobiology of Aging*, 33(1), pp. 21–34. Available at: <https://doi.org/10.1016/j.neurobiolaging.2010.02.009>.

Walterfang, M. *et al.* (2009) ‘Corpus Callosum Size and Shape in Established Bipolar Affective Disorder’, *Australian & New Zealand Journal of Psychiatry*, 43(9), pp. 838–845. Available at: <https://doi.org/10.1080/00048670903107534>.

Walters, R.K. *et al.* (2018) ‘Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders’, *Nature Neuroscience*, 21(12), pp. 1656–1669. Available at: <https://doi.org/10.1038/s41593-018-0275-1>.

Wang, F. *et al.* (2008) ‘Abnormal anterior cingulum integrity in bipolar disorder determined through diffusion tensor imaging’, *British Journal of Psychiatry*, 193(2), pp. 126–129. Available at: <https://doi.org/10.1192/bjp.bp.107.048793>.

Wang, P. *et al.* (2020) ‘The Organization of the Human Corpus Callosum Estimated by Intrinsic Functional Connectivity with White-Matter Functional Networks’, *Cerebral Cortex*, 30(5), pp. 3313–3324. Available at: <https://doi.org/10.1093/cercor/bhz311>.

Watanabe, K. *et al.* (2017) ‘Functional mapping and annotation of genetic associations with FUMA’, *Nature communications*, 8(1), pp. 1–11.

Westlye, L.T. *et al.* (2010) ‘Life-span changes of the human brain white matter: diffusion tensor imaging (DTI) and volumetry’, *Cerebral Cortex (New York, N.Y.: 1991)*, 20(9), pp. 2055–2068. Available at: <https://doi.org/10.1093/cercor/bhp280>.

Whitford, T.J. *et al.* (2010) ‘Corpus callosum abnormalities and their association with psychotic symptoms in patients with schizophrenia’, *Biological Psychiatry*, 68(1), pp. 70–77. Available at: <https://doi.org/10.1016/j.biopsych.2010.03.025>.

Wise, T. *et al.* (2016) ‘Voxel-based Meta-analytical Evidence Of Structural Disconnectivity In Major Depression and Bipolar Disorder’, *Biological Psychiatry*, 4(79), pp. 293–302. Available at: <https://doi.org/10.1016/j.biopsych.2015.03.004>.

Witelson, S.F. (1989) ‘Hand and sex differences in the isthmus and genu of the human corpus callosum: a postmortem morphological study’, *Brain*, 112(3), pp. 799–835. Available at: <https://doi.org/10.1093/brain/112.3.799>.

Woldehawariat, G. *et al.* (2014) ‘Corpus Callosum Size Is Highly Heritable in Humans, and May Reflect Distinct Genetic Influences on Ventral and Rostral Regions’, *PLOS ONE*, 9(6), p. e99980. Available at: <https://doi.org/10.1371/journal.pone.0099980>.

Wray, N.R. *et al.* (2018) ‘Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression’, *Nature Genetics*, 50(5), pp. 668–681. Available at: <https://doi.org/10.1038/s41588-018-0090-3>.

Wu, Z.-M. *et al.* (2017) ‘White Matter Microstructural Alterations in Children with ADHD: Categorical and Dimensional Perspectives’, *Neuropsychopharmacology*, 42(2), pp. 572–580. Available at: <https://doi.org/10.1038/npp.2016.223>.

Yang, J. *et al.* (2011) ‘Genomic inflation factors under polygenic inheritance’, *European Journal of Human Genetics*, 19(7), pp. 807–812.

Zhao, B. *et al.* (2020) *Common genetic variation influencing human white matter microstructure*. preprint. Genetics. Available at: <https://doi.org/10.1101/2020.05.23.112409>.

Zhao, B. *et al.* (2021) ‘Large-scale GWAS reveals genetic architecture of brain white matter microstructure and genetic overlap with cognitive and mental health traits (n = 17,706)’, *Molecular Psychiatry*, 26(8), pp. 3943–3955. Available at: <https://doi.org/10.1038/s41380-019-0569-z>.

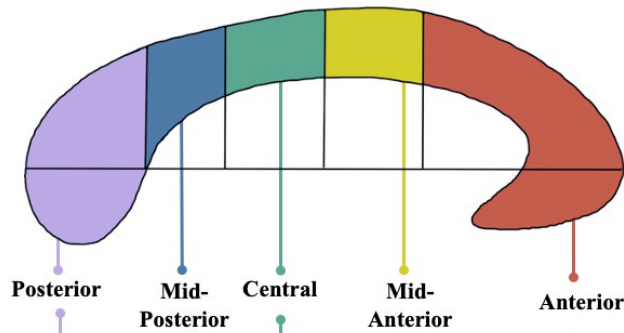
Zhao, G. *et al.* (2022) ‘A Comparative Multimodal Meta-analysis of Anisotropy and Volume Abnormalities in White Matter in People Suffering From Bipolar Disorder or Schizophrenia’, *Schizophrenia Bulletin*, 48(1), pp. 69–79. Available at: <https://doi.org/10.1093/schbul/sbab093>.

Zhou, L. *et al.* (2022) ‘Alterations in white matter microarchitecture in adolescents and young adults with major depressive disorder: A voxel-based meta-analysis of diffusion tensor imaging’, *Psychiatry Research: Neuroimaging*, 323, p. 111482. Available at: <https://doi.org/10.1016/j.psychresns.2022.111482>.

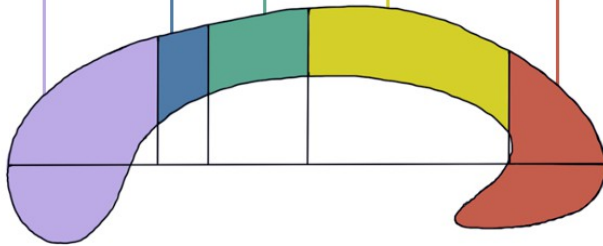
Appendices

Appendix 1: Supplementary Figures and Tables for Chapter 3

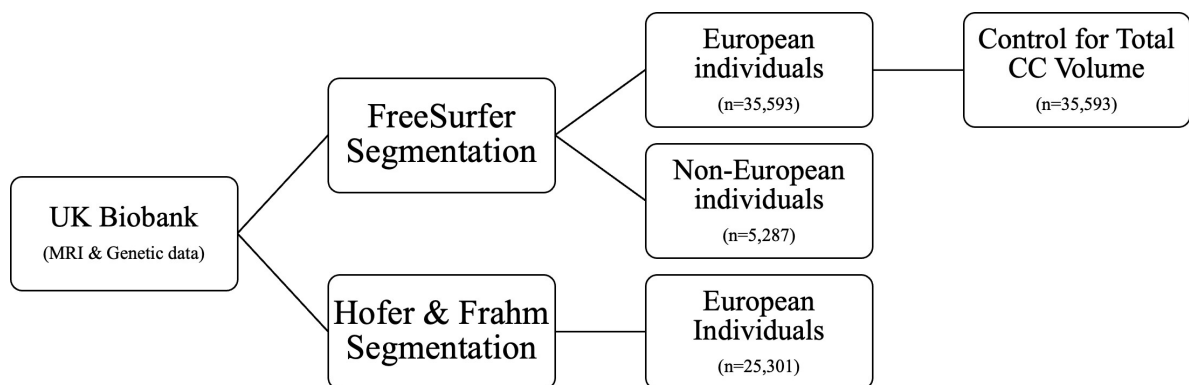
A.)



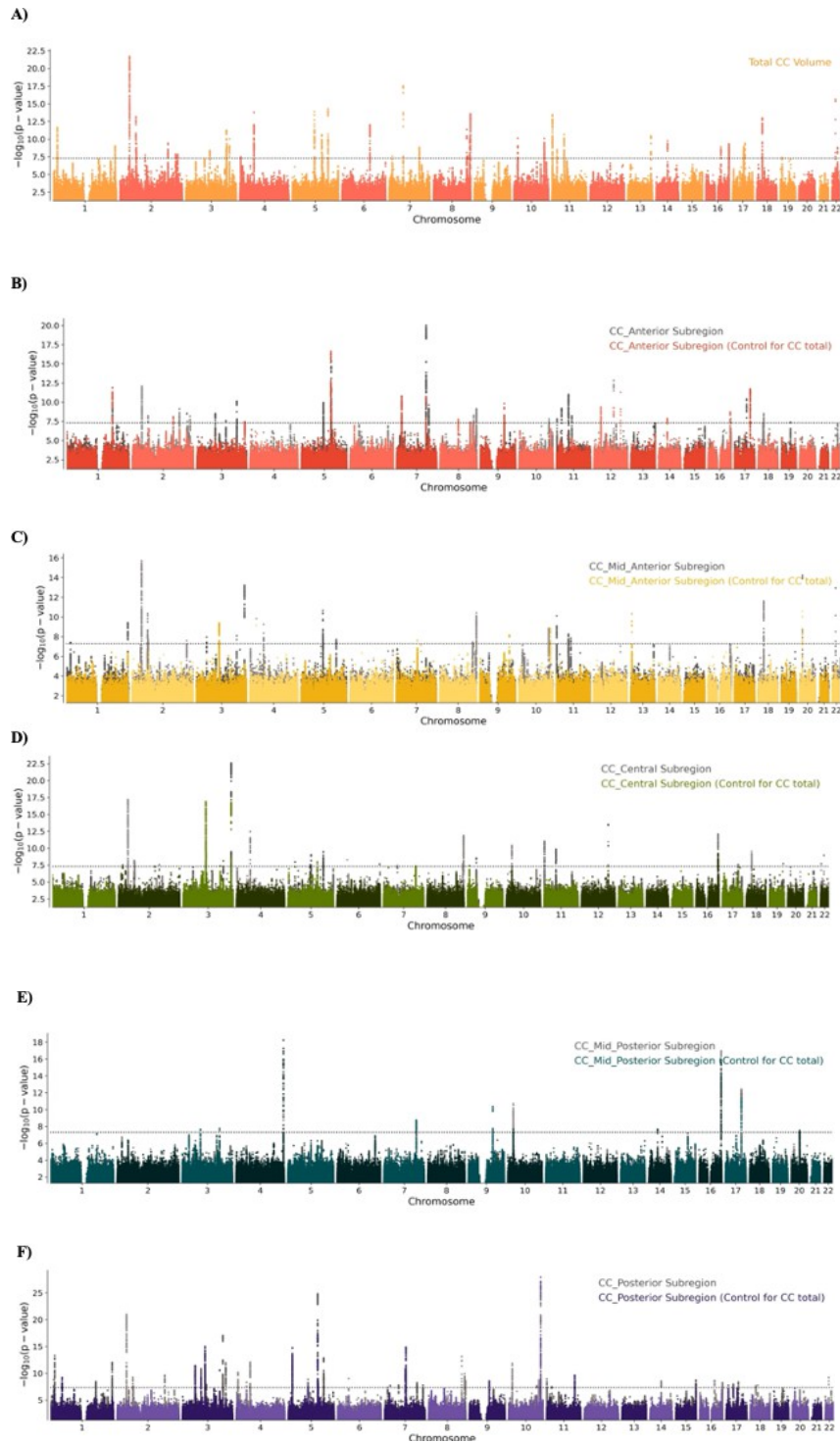
B.)



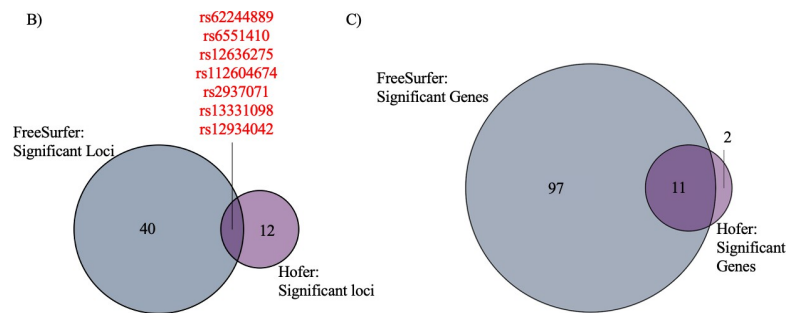
A1: Supplementary Material Figure 1 The differences between A) FreeSurfer and the B) Hofer and Frahm segmentation. FreeSurfer extracts volume, while the Hofer & Frahm segmentation, conducted using C8 software extracts thickness



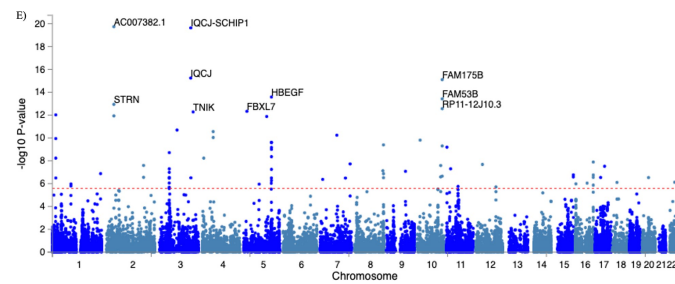
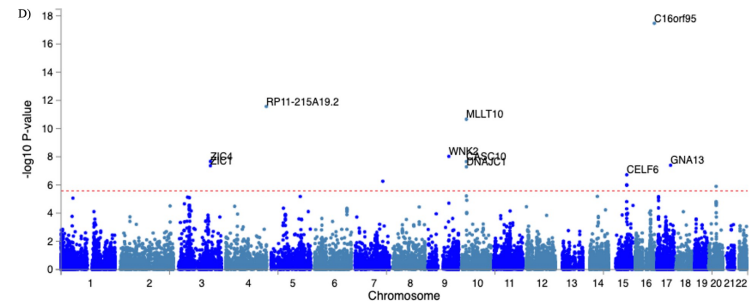
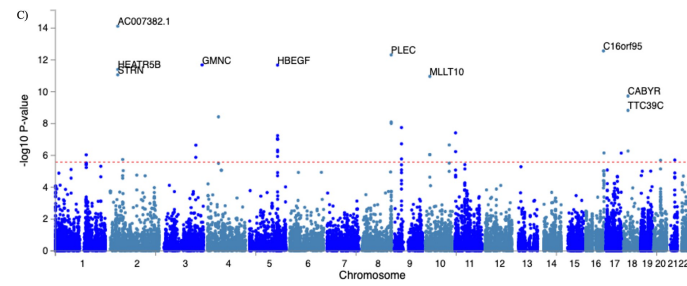
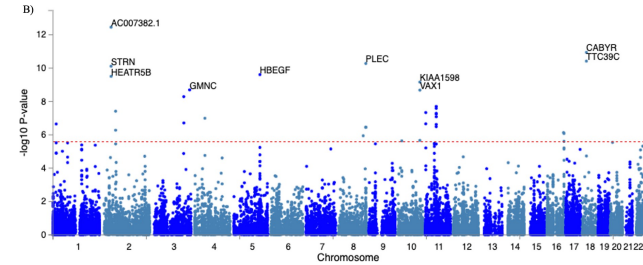
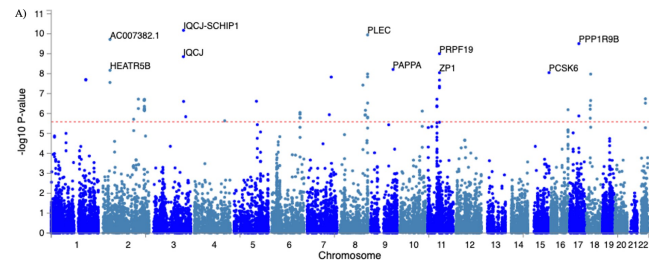
A1: Supplementary Material Figure 2 Overview of the GWAS conducted in this study



A1: Supplementary Material Figure 3 Manhattan plots from the univariate GWAS of each CC subregion, co-varying for total CC volume. A) 42 SNPs were significant for total CC volume, B) 31 for volume of the anterior subregion, C) 22 for volume of the mid anterior subregion D) 26 for volume of the central subregion, E) 7 for volume of the mid posterior subregion F) 49 for volume of the posterior subregion



A1: Supplementary Material Figure 4 Overlap between the FreeSurfer and Hofer and Frahm segmentation, multivariate analyses. A) Miami plot illustrating the $-\log(p\text{-value})$ multivariate GWAS results from the top) FreeSurfer and bottom) Hofer & Frahm segmentation of the CC subregions B) illustrating the significant, independent SNPs in both segmentations and C) the number of overlapping genes from the gene-based association tests. The multivariate analyses were conducted using MOSTest and illustrates the distributive genetic architecture of the CC. SNPs in red are those that are significant, independent loci in both analyses. Nineteen independent loci were identified in the multivariate GWAS of the Hofer and Frahm segmentation; 58% (7 loci) were also significant in the multivariate GWAS of the FreeSurfer segmentation. In the gene-based association tests, 11 of the 13 significant genes overlapped with results based on the FreeSurfer segmentation.



A1: Supplementary Material Figure 5 Manhattan plots of gene-based association tests, conducted through FUMA. A) Anterior, B) Mid anterior, C) Central, D) Mid posterior, E) Posterior

A1 Supplementary Material Table 1 Significant loci from univariate GWAS

- Found in Supplementary Files: A1 SM Table 1

A1 Supplementary Material Table 2 Replication of significant loci from univariate GWAS

- Found in Supplementary Files: A1 SM Table 2

A1 Supplementary Material Table 3 Gene set enrichment analysis results

- Found in Supplementary Files: A1 SM Table 3

A1 Supplementary Material Table 4 PheWAS results from significant loci and overlap with previous GWAS

- Found in Supplementary Files: A1 SM Table 4

A1 Supplementary Material Table 5 SNP-based heritability estimates from two segmentation approaches

Subregion	SNP-based heritability estimates (SE)	
	FreeSurfer Segmentation	Hofer Segmentation
Anterior	0.32 (0.03)	0.29 (0.02)
Mid Anterior	0.27 (0.02)	0.32 (0.03)
Central	0.28 (0.02)	0.28 (0.02)
Mid Posterior	0.22 (0.03)	0.26 (0.02)
Posterior	0.37 (0.03)	0.30 (0.02)
Total CC	0.38 (0.03)	-

SE, Standard Error.

A1 Supplementary Material Table 6 Genetic correlation results between the CC and psychiatric and substance use traits

- Found in Supplementary Files: A1 SM Table 6

Appendix 2: Supplementary Figures and Tables for Chapter 4

A2 Supplementary Material Table 1 Significant genetic loci from univariate GWAS

- Found in Supplementary Files: A2 SM Table 1

A2 Supplementary Material Table 2 Genetic correlation results between CC subregions and psychiatric and substance use phenotypes, across 28 diffusion metrics

- Found in Supplementary Files: A2 SM Table 2

A2 Supplementary Material Table 3 Results from the gene-set enrichment analysis across 28 diffusion metrics

- Found in Supplementary Files: A2 SM Table 3

Appendix 3: Supplementary Figures and Tables for Chapter 5

Appendix 3 Supplementary Material Table 1 Descriptions of the original GWAS

	GWAS	N	Reference	DOI
Psychiatric	PGC ADHD (2022)	225534	Demontis et al., 2023	https://doi.org/10.1038/s41588-023-01350-w
	PGC ASD (2017)	44366	PGC ASD 2017	https://doi.org/10.1186/s13229-017-0137-9
	BD (2021)	150669	Mullins et al 2021	https://doi.org/10.1038/s41588-021-00857-4
	PGC MDD (excl 23and Me & UKB; 2018)	156582	Wray et al 2018	https://doi.org/10.1038/s41588-018-0090-3
	PGC SCZ (2021)	126281	Trubetskoy et al 2022	https://doi.org/10.1038/s41586-022-04434-5
Substance Use	Drinks per week (2019)	941280	Liu et al 2019	https://doi.org/10.1038/s41588-018-0307-5
	Cigarettes per day (2019)	337334	Liu et al 2019	https://doi.org/10.1038/s41588-018-0307-5
	Cannabis (2018)	184765	Pasman et al 2018	https://doi.org/10.1038/s41593-018-0206-1
	Smoking initiation (2019)	1232091	Liu et al 2019	https://doi.org/10.1038/s41588-018-0307-5
	Smoking cessation (2019)	547219	Liu et al 2019	https://doi.org/10.1038/s41588-018-0307-5

Appendix 3 Supplementary Material Figure 1 predicted vs observed conditional quantile-quantile plots and log likelihood plots that visualize the parameter estimation procedure

- Found in Supplementary Files: A3 SM Figure 1

Appendix 3 Supplementary Material Table 2 Results from the conjunctional FDR analysis of the psychiatric disorders

- Found in Supplementary Files: A3 SM Table 2

Appendix 3 Supplementary Material Table 3 Results from the conjunctional analysis of the substance use traits

- Found in Supplementary Files: A3 SM Table 3

Appendix 3 Supplementary Material Table 4 Biologival processes implicated by the significant loci from the conjunctional FDR analyses

- Found in Supplementary Files: A3 SM Table 4