Distributed genetic effects of the corpus callosum subregions suggest links to neuropsychiatric disorders and related traits

Megan L. Campbell[1,2] ⓘ, Shareefa Dalvie[3] ⓘ, Alexey Shadrin[4],
Dennis van der Meer[4,5] ⓘ, Kevin O'Connell[4], Oleksander Frei[4], Ole A. Andreassen[4] ⓘ,
Dan J. Stein[1] and 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, 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 and [6]Centre of Research and Education in Forensic Psychiatry (SIFER), Oslo University Hospital, Oslo, Norway

Abstract

Background: The corpus callosum (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 remain largely unclear. We investigated the genetic architectures of the volume of the CC and its subregions and the genetic overlap with psychiatric disorders. *Methods:* We applied multivariate genome-wide association study (GWAS) to genetic and T1-weighted magnetic resonance imaging (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 Single nucleotide polymorphism (SNP)-based heritability of total CC volume and volumes of its subregions as well as their genetic correlations with relevant psychiatric traits. *Results:* We identified 70 independent loci with distributed effects across the five subregions of the CC ($p < 5 \times 10^{-8}$). Additionally, we 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). We found significant genetic correlations of total CC volume with bipolar disorder (BD, $r_g = -0.09$, SE = 0.03; $p = 5.9 \times 10^{-3}$) and drinks consumed per week ($r_g = -0.09$, SE = 0.02; $p = 4.8 \times 10^{-4}$), and volume of the mid-anterior subregion with BD ($r_g = -0.12$, SE = 0.02; $p = 2.5 \times 10^{-4}$), major depressive disorder (MDD) ($r_g = -0.12$, SE = 0.04; $p = 3.6 \times 10^{-3}$), drinks consumed per week ($r_g = -0.13$, SE = 0.04; $p = 1.8 \times 10^{-3}$) and cannabis use ($r_g = -0.09$, SE = 0.03; $p = 8.4 \times 10^{-3}$). *Conclusions:* Our results demonstrate that the CC has a polygenic architecture implicating multiple genes and show that CC subregion volumes are heritable. We found that distinct genetic factors are involved in the development of anterior and posterior subregions, consistent with their divergent functional specialisation. Significant genetic correlation between volumes of the CC and BD, drinks per week, MDD and cannabis consumption subregion volumes with psychiatric traits is noteworthy and deserving of further investigation.

Significant outcomes

We discovered novel loci not previously associated with the corpus callosum and provide evidence for unique genetic aetiology of the subregions. Additionally, we provide the first large-scale genetic evidence of heritability of the subregions and provide evidence to suggest overlap with psychiatric disorders.

Limitations

This study consisted only of individuals from the UK Biobank, and future studies should focus on including diverse populations and those with a lower mean age.

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 magnetic resonance imaging (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 et al., 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 et al., 2021; Bahrami 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, we 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.

Methods

Processing of imaging and genetic data

We accessed raw, T1-weighted MRI data from 35,592 genotyped, Caucasian individuals of British ancestry (age range: 45–81 years; mean age: 64.3 years, SD: 7.5; 52.3% female) from the UK Biobank (Bycroft et al., 2018) (accession code 27412) for our main analysis, and 5302 individuals of non-Caucasian ancestry for our generalisation 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 et al., 2012). Both methods divide the CC into five subregions, and the differences are illustrated and described in more detail in the supplementary material (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, we chose to focus on this method and use the approach proposed by Hofer and Frahm (2006) to validate our findings. To account for possible confounders, neuroCombat was used to pre-residualise for age, age squared, sex, scanning site, Euler score (a proxy of image quality), total intracranial volume 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 genome-wide association study (GWAS). The pre-residualisation 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: Single nucleotide polymorphisms (SNPs) with an imputation score of less than 0.5, a minor allele frequency of less than 0.005, a missingness rate of more than 10% and those that failed the Hardy–Weinberg test of equilibrium ($p < 1 \times 10^{-9}$) were removed. A total of 35,593 individuals remained in the European cohort after QC and an additional 5286 in the non-European cohort.

Genome-wide association analyses: multivariate and univariate

The pre-residualised 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 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 5210 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 was

conducted as an additional replication step. Significant SNPs were defined using a significant threshold of $p < 9.6 \times 10^{-9}$ (Bonferroni, accounting for five 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 250 kb. 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 was conducted, including total CC volume as a covariate. An overview of the GWAS is provided in the supplementary material (Supplementary Material, Figure 2).

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).We 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 significant 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, we used Open Target Genetics which integrates functional and biological data from multiple sources to highlight functionally implicated genes and conduct phenome-wide association studies and tests for association with previous GWAS (Mountjoy *et al.*, 2021).

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: ADHD (Stergiakouli *et al.*, 2012), ASD (PGC ASD, 2017), BD (Mullins *et al.*, 2021), MDD (Wray *et al.*, 2018), 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).

Results

Genome-wide association analyses

We found 70 unique genetic loci with distributed associations across the CC, using MOSTest. The polygenic architecture of the CC is shown in Fig. 1A, illustrating the results of the multivariate analysis. The univariate statistics indicate subregion-specific genetic signals for each of the CC subregions (Fig. 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 (Fig. 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 (Supplementary Material, Table 1). Figure 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 the supplementary material (Supplementary Material, Figure 3).

Replication of the multivariate results was conducted using MOSTest-PVS in the 5286 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, $p < 1.51 \times 10^{-3}$) (Supplementary Material, Table 2). The results for the multivariate analysis of the Hofer and Frahm segmentation can be found in the supplementary material (Supplementary Material, Figure 4).

Independent, lead SNPs 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 five 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 (Supplementary Material, Table 2).

Functional annotation

Significant gene sets for the GWAS of each subregion and total CC are listed in the supplementary material (Supplementary Material, Table 3). Gene sets identified for the anterior, mid-anterior and central subregion overlapped significantly with previous GWAS of BD, 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 organisation (van der Meer, Rokicki, *et al.*, 2020). Manhattan plots from the gene-based association tests, conducted through FUMA, can be found in the supplementary material (Supplementary Material, Figure 4). PheWAS, conducted with Open Target, implicated kidney and liver function (Supplementary Material, Table 4).
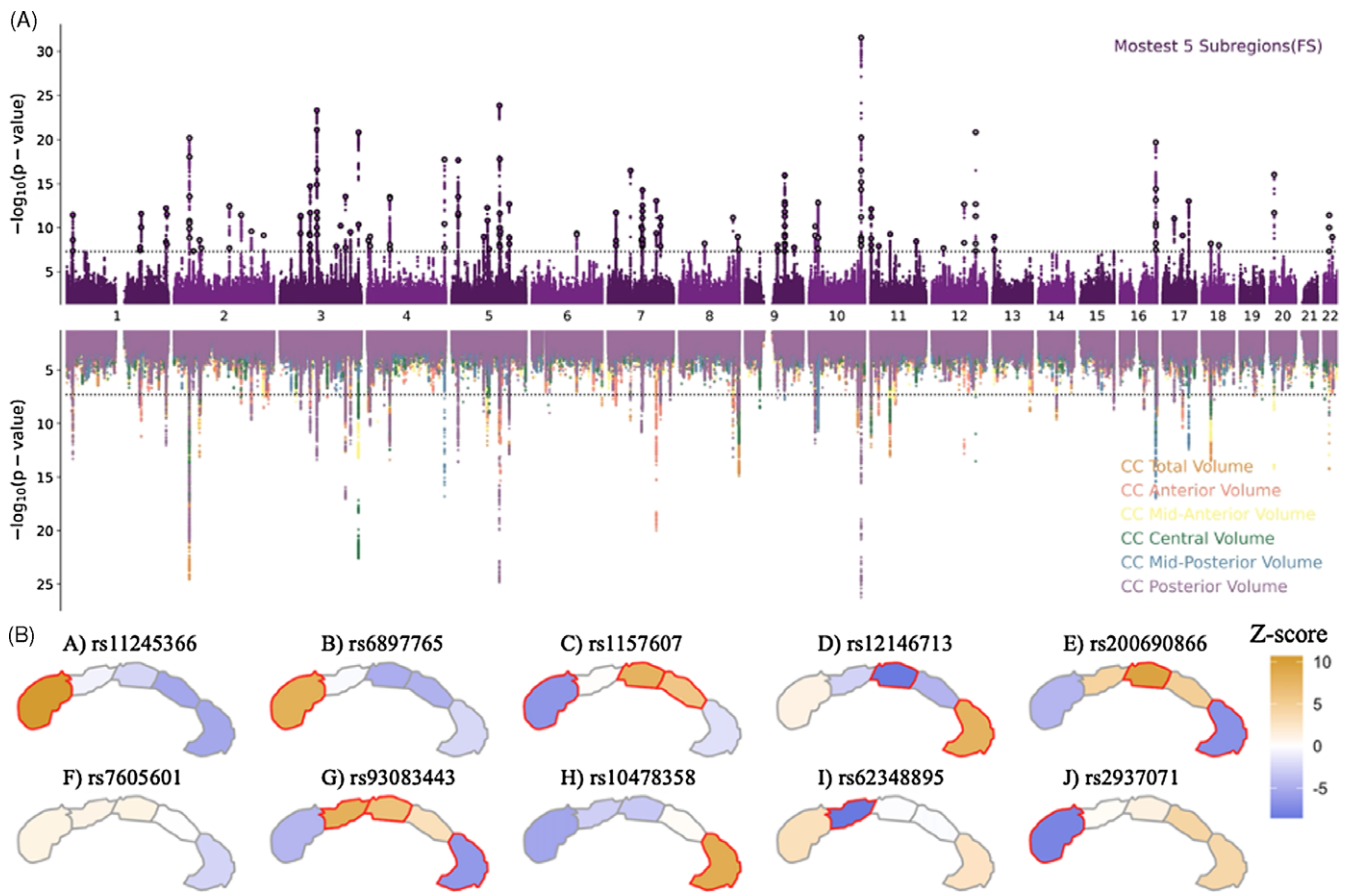
Figure 1. Miami plot of the $-\log_{10}(p)$ statistics from the multivariate (top) and univariate (bottom) GWAS of total CC volume. (A) Top: The multivariate analysis was conducted using MOSTest and illustrate the distributive genetic architecture of the CC. Bottom: Individual univariate GWAS were conducted for total CC volume and each subregion, indicating both a shared and unique genetic aetiology of the CC subregions. (B) Top SNPs from the multivariate GWAS show subregion-specific effects across the CC. The top 10 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.

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 Fig. 3A, and a full table of heritability estimates for the Hofer and Frahm segmentation can be found in the supplementary material (Supplementary Material, Table 5).

Using LDSC, we found significant genetic correlations between total CC volume, 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, we found significant genetic correlations between volume of the mid-anterior subregion BD ($r_g = -0.12$, SE = 0.02; $p = 2.5 \times 10^{-4}$), MDD ($r_g = -0.12$, 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; $p = 8.4 \times 10^{-3}$) (Fig. 3B). A full

table of these results can be found in the supplementary material (Supplementary Material, Table 6).

Discussion

Several key findings emerged from this work. First, we 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, we note significant genetic overlap of CC and CC subregions with neuropsychiatric disorders, including BD, MDD, ASD, as well as cannabis and alcohol consumption.

We 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, we were able to boost genetic discovery and gain a more comprehensive understanding of the genetic architecture of the CC and its subregions.

Our univariate GWAS indicates that the five subregions making up the CC have both a shared and distinct genetic aetiology, with

(A)

(B)

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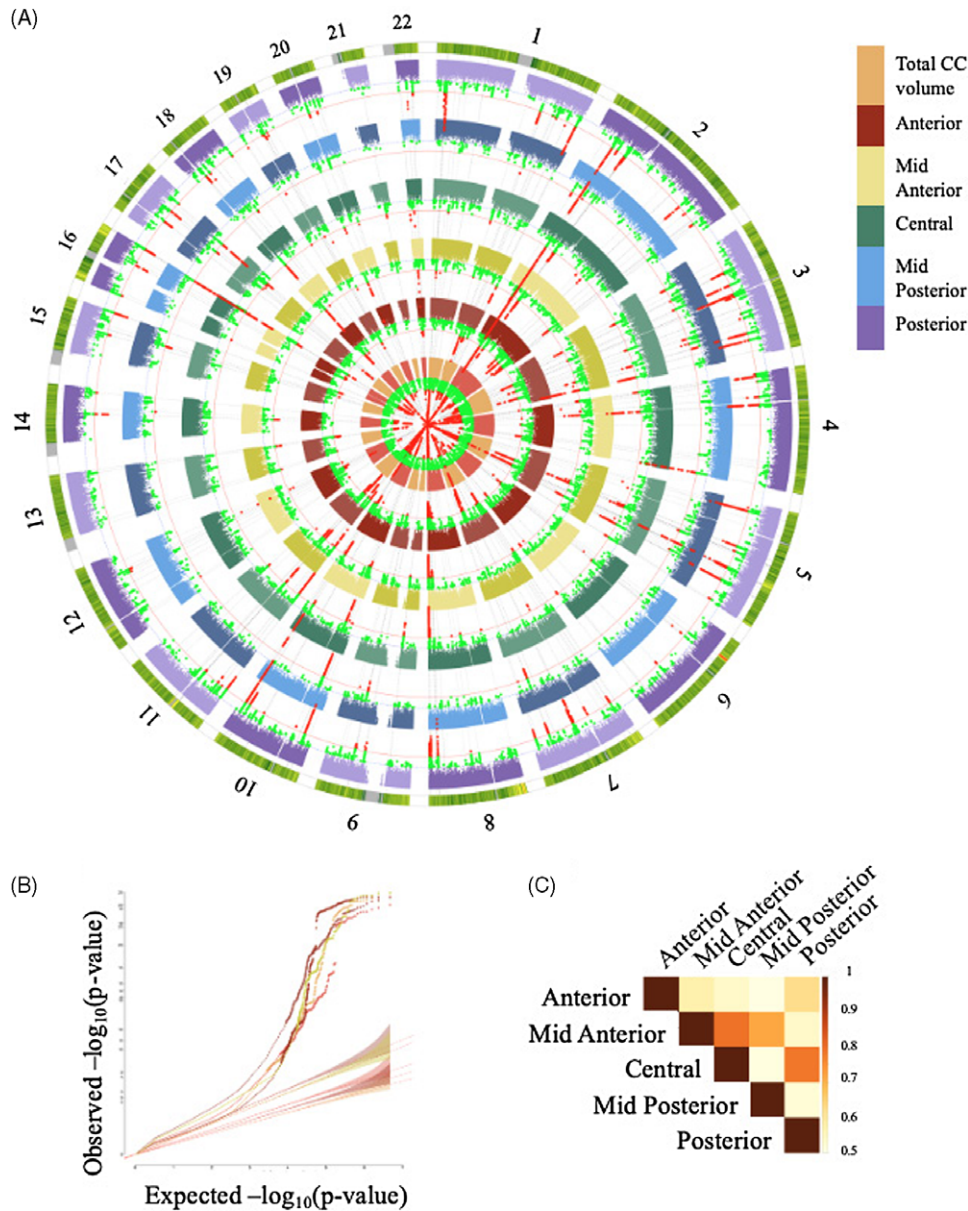
Figure 2. (A) Manhattan plots from univariate GWAS of each CC subregion. (B) Quantile-Quantile plot (QQ plot) of observed versus expected $-\log_{10}(p\text{-values})$ the GWAS of CC subregions. The shaded regions represent the expected null distribution of $-\log_{10}(p\text{-values})$ under the assumption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

the loci associated with whole CC volume only partially overlapping with those of the individual CC subregions. We 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 over-lapped 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 heritabilities 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 *et al.*, 2021; Bahrami *et al.*, 2022; Ou *et al.*, 2023). Our 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 *et al.* (2014), who found the anterior and posterior subregions of the CC to have the highest heritability estimates. Our 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
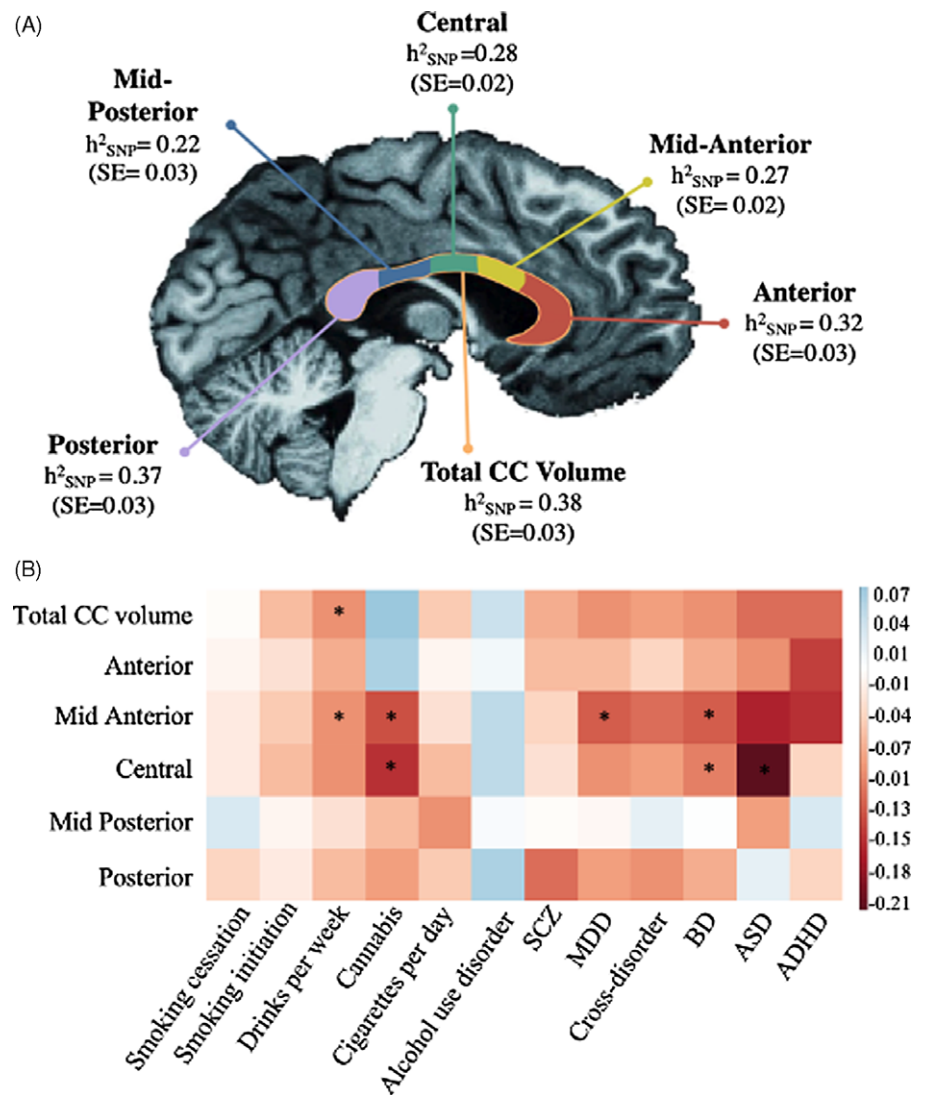
Figure 3. (A) SNP-based heritability estimates of total CC volume and volume of the subregions. LDSC was used to estimate the SNP-based heritability of total CC volume and the five subregions. A schematic of the CC and the five subregions, extracted using FreeSurfer, based on the Witelson segmentation. The CC and subregions have been overlaid onto an anatomical reference image. (B) Heat map of genetic correlations between CC subregions and relevant psychiatric and substance use traits each square represents the r-squared values, and significant correlations are denoted by *

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 we 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).

We 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). Our genetic correlation results also echoed findings from previous imaging studies. We 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 with healthy controls (Arnone, *et al.*, 2006; Gifuni *et al.*, 2017; Prunas *et al.*, 2018). Volumetric MRI studies have consistently demonstrated reduced 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). Our 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). Our genetic correlation results are in line with these findings: there was a significant correlation between ASD and the central subregion.

Previous work has also shown that alcohol consumption affects CC morphometry, irrespective of an alcohol use disorder diagnosis (de Souza *et al.*, 2019). Here, we 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, our 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 our 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, we provide preliminary evidence for a shared genetic aetiology 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, our study was limited to individuals residing in the United Kingdom and consisted largely of individuals of European ancestry. We did, however, replicate our 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 generalisability. Third, the comparison of our results with the literature is challenging due to differences in approach to subregion divisions. Although we implemented two segmentation protocols with consistent results, a robust CC segmentation is needed to improve replicability of research in this field.

In conclusion, our 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 specialisation. The pheWAS findings emphasise the importance of liver and kidney function in CC volume. We provide the first SNP-based evidence that CC and subregion volumes are heritable. Genetic correlations of CC and subregions volumes with specific psychiatric traits are noteworthy and deserv further investigation.

References

Abramovic L, Boks MPM, Vreeker A, *et al.* (2018) White matter disruptions in patients with bipolar disorder. *European Neuropsychopharmacology* **28**, 743–751. DOI: 10.1016/j.euroneuro.2018.01.001.

Arnone D, Abou-Saleh MT and Barrick TR (2006) Diffusion tensor imaging of the corpus callosum in addiction. *Neuropsychobiology* **54**(2), 107–113.

Bahrami S, Nordengen K, Shadrin AA, *et al.* (2022) Distributed genetic architecture across the hippocampal formation implies common neuropathology across brain disorders. *Nature Communications* **13**(1), 3436. DOI: 10.1038/s41467-022-31086-w.

Barton AR, Sherman MA, Mukamel RE and Loh P-R (2021) Whole-exome imputation within UK Biobank powers rare coding variant association and fine-mapping analyses. *Nature Genetics* **53**(8), 1260–1269. DOI: 10.1038/s41588-021-00892-1.

Bearden CE, van Erp TGM, Dutton RA, *et al.* (2011) Mapping corpus callosum morphology in twin pairs discordant for bipolar disorder. *Cerebral Cortex* **21**(10), 2415–2424. DOI: 10.1093/cercor/bhr030.

Bulik-Sullivan B, Finucane HK, Anttila V, *et al.* (2015) An atlas of genetic correlations across human diseases and traits. *Nature Genetics* **47**(11), 1236–1241.

Bulik-Sullivan BK, Loh P-R, Finucane HK, *et al.* (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**(3), 291–295.

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

Campbell M, Dalvie S, Shadrin A, *et al.* (2022) The genetic architecture of the corpus callosum and its subregions. *European Neuropsychopharmacology* **63**, e142. DOI: 10.1016/j.euroneuro.2022.07.263.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM and Lee JJ (2015) Second-Generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, s13742-015.

Chen S-J, Wu B-S, Ge Y-J, *et al.* (2023) The genetic architecture of the corpus callosum and its genetic overlap with common neuropsychiatric diseases. *Journal of Affective Disorders* **335**, 418–430. DOI: 10.1016/j.jad.2023.05.002.

Connaughton M, Whelan R, O'Hanlon E and McGrath J (2022) White matter microstructure in children and adolescents with ADHD. *NeuroImage: Clinical* **33**, 102957. DOI: 10.1016/j.nicl.2022.102957.

Cover GS, Herrera WG, Bento MP, Appenzeller S and Rittner L (2018) Computational methods for corpus callosum segmentation on MRI: a systematic literature review. *Computer Methods and Programs in Biomedicine* **154**, 25–35. DOI: 10.1016/j.cmpb.2017.10.025.

de Souza RSM, Rosa Jr. M, Escobar TDC, Gasparetto EL and Nakamura-Palacios EM (2019) Anterior to midposterior corpus callosum subregions are volumetrically reduced in male alcoholics but only the anterior segment is associated to alcohol use. *Frontiers in Psychiatry* **10**, 196. DOI: 10.3389/fpsyt.2019.00196.

Fortin J-P, Cullen N, Sheline YI, *et al.* (2018) Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage* **167**, 104–120. DOI: 10.1016/j.neuroimage.2017.11.024.

Francis AN, Mothi SS, Mathew IT, *et al.* (2016) Callosal abnormalities across the psychosis dimension: bipolar schizophrenia network on intermediate phenotypes. *Biological Psychiatry* **80**(8), 627–635. DOI: 10.1016/j.biopsych.2015.12.026.

Franco I, Kogan S, Fisher J, *et al.* (1993) Genitourinary malformations associated with agenesis of the corpus callosum. *The Journal of Urology* **149**(5), 1119–1121. DOI: 10.1016/s0022-5347(17)36313-9.

Frazier TW and Hardan AY (2009) A meta-analysis of the corpus callosum in autism. *Biological Psychiatry* **66**(10), 935–941. DOI: 10.1016/j.biopsych.2009.07.022.

Gifuni AJ, Olié E, Ding Y, *et al.* (2017) Corpus callosum volumes in bipolar disorders and suicidal vulnerability. *Psychiatry Research: Neuroimaging* **262**, 47–54. DOI: 10.1016/j.pscychresns.2017.02.002.

Grasby KL, Jahanshad N, Painter JN, *et al.* (2020) The genetic architecture of the human cerebral cortex. *Science* **367**(6484). DOI: 10.1126/science.aay6690.

Herron TJ, Kang X and Woods DL (2012) Automated measurement of the human corpus callosum using MRI. *Frontiers in Neuroinformatics* **6**, 25. DOI: 10.3389/fninf.2012.00025.

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), 989–994. DOI: 10.1016/j.neuroimage.2006.05.044.

Hopkins WD, Westerhausen Ré, Schapiro S and Sherwood CC (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), e12784. DOI: 10.1111/gbb.12784.

Jansen IE, Savage JE, Watanabe K, *et al.* (2019) Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimers disease risk. *Nature Genetics* **51**(3), 404–413. DOI: 10.1038/s41588-018-0311-9.

Kranzler HR, Zhou H, Kember RL, *et al.* (2019) Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nature Communications* **10**(1), 1499. DOI: 10.1038/s41467-019-09480-8.

Liu J, Li S, Li X, *et al.* (2021) Genome-wide association study followed by trans-ancestry meta-analysis identify 17 new risk loci for schizophrenia. *BMC Medicine* **19**(1), 177. DOI: 10.1186/s12916-021-02039-9.

Liu M, Wu Y, Wu X, *et al.* (2020) White matter microstructure changes and cognitive impairment in the progression of chronic kidney disease. *Frontiers in Neuroscience* **14**. Availabl at https://www.frontiersin.org/articles/10.3389/fnins.2020.559117 (accessed 2 April 2023).

Liu M, Jiang Y, Wedow R, *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), 237–244. DOI: 10.1038/s41588-018-0307-5.

Loughnan RJ, Shadrin AA, Frei O, *et al.* (2022) Generalization of cortical MOSTest genome-wide associations within and across samples. *NeuroImage* **263**, 119632. DOI: 10.1016/j.neuroimage.2022.119632.

Mountjoy E, Schmidt EM, Carmona M, *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), 1527–1533. DOI: 10.1038/s41588-021-00945-5.

Mufford MS, van der Meer D, Kaufmann T, *et al.* (2021) The genetic architecture of amygdala nuclei. DOI: 10.1101/2021.06.30.21258615.

Mullins N, Forstner AJ, O'Connell KS, *et al.* (2021) Genome-wide association study of over 40,000 bipolar disorder cases provides new insights into the underlying biology. *medRxiv*. DOI: 10.1101/2020.09.17.20187054.

Nagel M, Jansen PR, Stringer S, *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), 920–927. DOI: 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**, 91. DOI: 10.3389/fgene.2012.00091.

Ou Y-N, Wu B-S, Ge Y-J, *et al.* (2023) The genetic architecture of human amygdala volumes and their overlap with common brain disorders. *Translational Psychiatry* **13**(1), 90. DOI: 10.1038/s41398-023-02387-5.

Pasman JA, Verweij KJH, Gerring Z, *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), 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**(1), 21. DOI: 10.1186/s13229-017-0137-9.

Pouget JG, Wu Y, Mignot E, *et al.* (2019) Cross-disorder analysis of schizophrenia and 19 immune-mediated diseases identifies shared genetic risk. *Human Molecular Genetics* **28**(20), 3498–3513. DOI: 10.1093/hmg/ddz145.

Prunas C, Delvecchio G, Perlini C, *et al.* (2018) Diffusion imaging study of the Corpus Callosum in bipolar disorder. *Psychiatry Research: Neuroimaging* **271**, 75–81. DOI: 10.1016/j.pscychresns.2017.11.001.

Rosenbloom M, Sullivan EV and Pfefferbaum A (2003) Using magnetic resonance imaging and diffusion tensor imaging to assess brain damage in alcoholics. *Alcohol Research & Health* **27**, 146–152.

Satizabal CL, Adams HHH, Hibar DP, *et al.* (2019) Genetic architecture of subcortical brain structures in 38,851 individuals. *Nature Genetics* **51**(11), 1624–1636. DOI: 10.1038/s41588-019-0511-y.

Scamvougeras A, Kigar DL, Jones D, Weinberger DR and Witelson SF (2003) Size of the human corpus callosum is genetically determined: an MRI study in mono and dizygotic twins. *Neuroscience Letters* **338**, 91–94. DOI: 10.1016/S0304-3940(02)01333-2.

Stahl EA, Breen G, Forstner AJ, *et al.* (2019) Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature Genetics* **51**(5), 793–803.

Stergiakouli E, Hamshere M, Holmans P, *et al.* (2012) Investigating the contribution of common genetic variants to the risk and pathogenesis of ADHD. *The American Journal of Psychiatry* **169**(2), 186–194. DOI: 10.1176/appi.ajp.2011.11040551.

Thompson PM, Narr KL, Blanton RE and Toga AW (2003) Mapping structural alterations of the corpus callosum during brain development and degeneration.

Turley P, Walters RK, Maghzian O, *et al.* (2018) Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nature Genetics* **50**(2), 229–237. DOI: 10.1038/s41588-017-0009-4.

Valenti M, Pino MC, Mazza M, Panzarino G, Di Paolantonio C and Verrotti A (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), 46–62. DOI: 10.1007/s40489-019-00176-9.

van der Meer D, Frei O, Kaufmann T, *et al.* (2020) Understanding the genetic determinants of the brain with MOSTest. *Nature Communications* **11**(1), 3512. DOI: 10.1038/s41467-020-17368-1.

van der Meer D, Rokicki J, Kaufmann T, *et al.* (2020) Brain scans from 21,297 individuals reveal the genetic architecture of hippocampal subfield volumes. *Molecular Psychiatry* **25**(11), 3053–3065. DOI: 10.1038/s41380-018-0262-7.

Walterfang M, Malhi GS, Wood AG, *et al.* (2009) Corpus callosum size and shape in established bipolar affective disorder. *Australian & New Zealand Journal of Psychiatry* **43**(9), 838–845. DOI: 10.1080/00048670903107534.

Wang P, Meng C, Yuan R, *et al.* (2020) The organization of the human corpus callosum estimated by intrinsic functional connectivity with white-matter functional networks. *Cerebral Cortex* **30**(5), 3313–3324. DOI: 10.1093/cercor/bhz311.

Watanabe K, Taskesen E, van Bochoven A and Posthuma D (2017) Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* **8**(1), 1–11.

Witelson SF (1989) Hand and sex differences in the isthmus and genu of the human corpus callosum: a postmortem morphological study. *Brain* **112**, 799–835. DOI: 10.1093/brain/112.3.799.

Woldehawariat G, Martinez PE, Hauser P, *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), e99980. DOI: 10.1371/journal.pone.0099980.

Wray NR, Ripke S, Mattheisen M, *et al.* (2018) Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics* **50**(5), 668–681. DOI: 10.1038/s41588-018-0090-3.

Zhao B, Li T, Yang Y, *et al.* (2020) Common genetic variation influencing human white matter microstructure. preprint. *Genetics*. DOI: 10.1101/2020.05.23.112409.

Zhao G, Lau WK W, Wang C, *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), 69–79. DOI: 10.1093/schbul/sbab093.

Zhou L, Wang L, Wang M, *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**, 111482. DOI: 10.1016/j.pscychresns.2022.111482.