Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A single-cell eQTL atlas of the developing human brain

Darren Cameron
Cardiff University https://orcid.org/0000-0001-7520-7741
Claire Tume
Cardiff University
Ngoc-Nga Vinh
Cardiff University
Sara Davies
Cardiff University
J Morgan
University of Leeds
Yusuf Baran
Cardiff University
Hyejung Won
University of North Carolina at Chapel Hill https://orcid.org/0000-0003-3651-0566
Antonio Pardinas
Cardiff University https://orcid.org/0000-0001-6845-7590
Michael O'Donovan
Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, UK https://orcid.org/0000-0001-7073-2379
Nicholas Bray
BrayN3@cardiff.ac.uk
Cardiff University https://orcid.org/0000-0002-4357-574X

Article

Keywords:

Posted Date: March 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9011993/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Abstract

Genetic influences on gene expression in the developing brain are likely to impact a variety of human traits, including susceptibility to neuropsychiatric disorders. However, to date, expression quantitative trait loci (eQTL) studies of the prenatal human brain have been based on 'bulk' tissue, where gene expression measures are averaged across all constituent cell types. Here, we performed single nuclei RNA sequencing (snRNA-Seq) and genome-wide genotyping on cerebral cortex from 134 unrelated samples from the second trimester of gestation to provide the first single nuclei eQTL atlas of the prenatal human brain. We identify 3121 unique eGenes across 7 major cell types and 12 of their subtypes, combining these data with large-scale genome-wide association study (GWAS) data for neuropsychiatric disorders to implicate particular cell populations of the developing human brain in the etiology of these conditions and to identify specific gene expression differences within them that credibly confer susceptibility. Our study significantly advances knowledge of the cellular origins of neuropsychiatric disorders and provides a resource for interpreting GWAS findings for other conditions with a potential neurodevelopmental component.

Main

Common genetic variants associated with complex human traits typically reside in non-coding regions of the genome, where they may impact cell function through effects on gene regulation[1,2]. Variants that are associated with altered gene expression can be mapped on a genomic scale (as expression quantitative trait loci, or eQTL) by combining genome-wide genotyping with transcriptome profiling in modestly sized cohorts (typically, 100 - 1000 individuals) [3]. These eQTL data can then be integrated with the results of genome-wide association studies (GWAS) to identify gene expression changes that potentially mediate associations between complex traits and non-coding genetic variation [4-6].

To date, most eQTL mapping studies have been based on transcriptomic data from 'bulk' tissues (e.g. an individual organ or anatomical region), where gene expression measures are averaged across all constituent cell types. However, advances in single cell / nuclei RNA sequencing have now made it possible to map eQTL in a cell-specific manner [7-10]. These single cell eQTL studies have confirmed that genetic effects on gene expression are often cell type-specific and provide greater power to detect eQTL, particularly when they are active only in rarer cell types. Moreover, when leveraged to identify potential disease mechanisms, single cell eQTL data have the considerable advantage of informing which cell types these are operating in [7-10].

There is now strong evidence that genetic influences on gene expression in the prenatal human brain impact susceptibility to a range of neuropsychiatric conditions [11-13]. In particular, mid-gestation has been proposed as a critical window during which genes associated with neuropsychiatric disorders reach peak expression [14]. However, to date, single cell eQTL studies of the human brain have been limited to adult tissue [8,9,15]. In the present study, we performed genome-wide genotyping and single nuclei RNA

sequencing (snRNA-Seq) on cerebral cortex from 134 samples from the second trimester of gestation to provide the first single cell eQTL atlas of the prenatal human brain. We integrate our eQTL data with large-scale GWAS data for neuropsychiatric disorders to identify gene expression differences within specific cell types of the developing brain that credibly confer risk to these conditions.

Results

Single nuclei RNA sequencing of the prenatal human cerebral cortex

We performed snRNA-Seq on 134 independent human brain samples aged 12-21 post-conception weeks, retaining data from 679,738 nuclei, and a median of 5135 nuclei per sample, after strict quality control (Methods) (Fig. 1a). Nuclei were initially clustered and annotated as 7 major cell types of the mid-gestation human cerebral cortex according to their gene expression profile: upper layer glutamatergic neurons (Glu-UL), deep layer glutamatergic neurons (Glu-DL), GABAergic neurons, neural progenitor cells (NPCs), oligodendrocyte precursor cells (OPCs), endothelial cells / pericytes (Endo-Peri) and microglia (Fig. 1b). Cell annotations were based on expression of canonical markers (Fig. 1c) and other differentially expressed genes (Supplementary Table 1). Nuclei from all 7 major cell types were represented in all 134 samples (Fig. 1d). The 4 most numerous cell populations – Glu-UL, Glu-DL, GABAergic neurons and NPCs – were then individually subclustered to yield 16 subpopulations (Supplementary Fig. 1), of which 12 were represented by nuclei from at least 133 samples and therefore also carried forward for eQTL analysis.

Mapping cis-eQTL in cell populations of the prenatal human cerebral cortex

Genomic DNA from each sample was directly genotyped for approximately 654,000 single nucleotide polymorphisms (SNPs), followed by genotype imputation. After strict quality control, we retained 5,794,957 SNPs with a minor allele frequency > 0.05 for eQTL analyses. We derived expression measures for between 16,091 and 17,369 genes in each broad cell type. Expression measures were quantile normalized and adjusted for sex, age (post-conception weeks), 4 genotype principal components and between 10 and 50 gene expression principal components; the latter empirically-determined to maximise eQTL discovery in each cell population (Supplementary Fig. 2), as performed by others [8,9]. To identify *cis*-eQTL in each cell type, we tested association between each gene's expression and all SNPs within 500kb either side of its transcription start site (TSS) by linear regression of allele dosage against normalized gene expression measures using TensorQTL[17]. We identified high confidence eQTL (false discovery rate [FDR] < 0.05) [18] for 3121 unique genes ('eGenes') across the 7 broad cell types and 12 analysed subpopulations. The number of eGenes identified in each cell population is shown in Fig 2A, ranging from 62 in microglia to 1,985 in Glu-UL. Consistent with previous snRNA-Seq based eQTL studies [8,9,15], we found a strong correlation between the total number of nuclei assayed for each cell population and the number of eGenes detected (Fig. 2B). Accordingly, the majority of eGenes were identified at the broad cell type level, although additional eGenes detected in subpopulations of these cells suggest eQTL operating in a more cell-specific manner (Fig. 2C). Detected eQTL were concentrated

in proximity to the TSS of their target eGene in all cell populations (Fig. 2D) and enriched in open chromatin regions identified in the corresponding cell populations of the prenatal human cerebral cortex by single nuclei ATAC-Seq [19] (Fig. 2E). We list our top eQTL for the eGenes discovered in each cell population in Supplementary Table 2, indicating those SNPs that can be further functionally prioritised through their location in open chromatin regions identified in the developing cerebral cortex [19].

Sharing of cis-eQTL across prenatal and adult neural cell populations

While high confidence (FDR < 0.05) eQTL were frequently detected in only one prenatal cell population (Fig. 3A, 3B), estimates of π_1 [18] for the proportion of FDR < 0.05 eQTL identified in one cell population that have non-null effects in another cell population indicate considerable sharing of eQTL across prenatal cell types (Fig. 3C). However, π_1 estimates of the extent to which the FDR < 0.05 eQTL that we identified in developing neuron populations were shared with neuronal populations of the adult human cerebral cortex [9] did not exceed 0.73 (Fig. 3D), consistent with a proportion of eQTL operating specifically during neuronal development. Figures 3D – 3F show the correlation of effect sizes for eQTL SNP-eGene pairs identified in developing neuronal populations (at FDR 5%) with the same SNP-gene pairs in their closest corresponding neuronal populations from the adult cerebral cortex [9]. SNPs that are associated with distinct effects on eGene expression in prenatal neuronal populations include rs9818993, where notably larger effects on *ABI3BP* expression were observed in GLU-DL compared with adult glutamatergic neurons, and rs77889556, where alleles were found to be associated with opposite effects on *ABCC8* expression in prenatal and adult neuronal populations.

Enrichment of common variant genetic liability to neuropsychiatric disorders in cellular fine-mapped eQTL of the prenatal brain

Hormozdiari and colleagues [20] showed that tissue-specific fine-mapped eQTL are strongly enriched for disease SNP heritability. We reasoned that fine-mapped eQTL identified in individual cell populations of the prenatal brain might therefore serve as useful annotations for assessing the relevance of each cell type to common variant genetic risk for neuropsychiatric disorders. To this end, we performed Bayesian fine-mapping [21] for each eGene (FDR < 0.05) in each cell population, estimating posterior inclusion probabilities (PIPs) for SNPs within a 500kb window of the TSS and deriving 95% credible sets (the minimal sets of SNPs with cumulative PIP ³95% of containing the causal variant(s). Following Hormozdiari et al, we then created a set of continuous annotations in which each SNP was weighted by its PIP of being a causal eQTL in the corresponding cell population. Finally, we applied stratified linkage disequilibrium score regression (S-LDSR) [22] to test enrichment of SNP heritability for neuropsychiatric conditions within these fine-mapped eQTL annotations for each cell type.

At the broad cell type level (Figure 4a), we observed significant enrichment of schizophrenia common variant heritability [23] within fine-mapped eQTL in Glu-UL (P = 0.00031), Glu-DL (P = 0.0013) and GABAergic neurons (P = 0.0037). Common variant liability to schizophrenia was also significantly enriched at the Bonferroni-corrected P-value threshold within fine-mapped eQTL in two subpopulations

of upper layer glutamatergic neurons (Fig. 4B): Glu-UL-1 (P = 0.0018), which prominently expresses *MEF2C* and *DLG2*, and Glu-UL-2 (P = 0.0019), which markedly expresses *PTPRN2* and *MACF1*. Common variant liability for bipolar disorder [24] was most significantly enriched in fine-mapped eQTL in GABAergic neurons (P = 0.0099), while that for major depressive disorder [25] was most significantly enriched in fine-mapped eQTL in GABAergic neurons (P = 0.0089) and NPCs (P = 0.0089), although none of these observations reached the Bonferroni-corrected P-value threshold (Fig. 4A).

Identification of gene expression changes associated with genetic risk for neuropsychiatric disorders in individual cell populations of the prenatal brain

A primary motivation for eQTL mapping is to help prioritize credibly causal genes at GWAS loci. We performed cell-specific Summary-data-based Mendelian Randomization (SMR) [4] and causal transcriptome-wide association studies (cTWAS) [6] to identify gene expression differences in cell populations of the prenatal brain associated with risk alleles for neuropsychiatric disorders.

Across cell types, SMR identified 38 genes that potentially mediate genetic associations with schizophrenia through their expression in the prenatal brain (SMR $P_{\text{Bonferroni}} < 0.05$; heterogeneity in dependent instruments [HEIDI] $P > 0.01$; Supplementary Table 3). These included 10 genes located within a common inversion polymorphism on chromosome 17 [28] (*LRRC37A*, *LRRC37A2*, *KANSL1*, *ARL17B* and 6 long non-coding RNA genes), a known locus for differential *cis*-regulation in the prenatal human brain [11], which were significant in SMR tests of eQTL identified across multiple cell types. Other genes at genome-wide significant loci for schizophrenia [23] that were implicated by SMR / HEIDI include *ANKRD45* (risk variation associated with decreased expression in developing glutamatergic and GABAergic neurons), *ARL3* (risk variation associated with decreased expression in upper layer glutamatergic neurons), *AS3MT* (risk variation associated with increased expression in NPCs and a subtype of developing upper layer glutamatergic neuron), *ZKSCAN8* (risk variation associated with decreased expression in deep layer glutamatergic neurons) and *SNX19* (risk variation associated with increased expression in multiple cell types). Common risk alleles for bipolar disorder [24] were also associated with increased expression of *AS3MT* in NPCs, along with decreased expression of a novel gene (ENSG00000285942) overlapping *LRRC57* in NPCs (Supplementary Table 3). SMR / HEIDI implicated expression of 13 genes in major depressive disorder [25] (Supplementary Table 3), including multiple genes within the chromosome 17 inversion in several cell types, increased expression of *ENTREP2* in a subpopulation of developing GABAergic neuron and increased expression of *FAM120AOS* in NPCs. Common risk variants for obsessive-compulsive disorder [27] (OCD) were associated with altered expression of a long non-coding RNA (ENSG00000286100) in multiple types of developing neuron and a gene predicted to encode a novel protein (ENSG00000267127) in a subpopulation of NPCs (Supplementary Table 3).

We performed causal TWAS (cTWAS) to identify gene expression differences associated with genetic risk for neuropsychiatric traits that might be more accurately captured through multiple SNPs at each

gene locus. For each cell type, SNP-weight predictors of *cis*-regulated gene expression were generated for genes with evidence of *cis*-heritable expression ($P < 0.01$) using the FUSION pipeline [5]. We then used these predictors to impute gene expression into large-scale GWAS of neuropsychiatric disorders. We applied the cTWAS framework [6] to control potential genetic confounders and identify only those TWAS associations ($P_{\text{Bonferroni}} < 0.05$) with a >0.8 posterior inclusion probability (PIP) of reflecting a direct non-zero effect on the trait.

Across cell types, cTWAS implicated altered *cis*-regulation of 41 genes in susceptibility to schizophrenia (Supplementary Table 4). These included schizophrenia-associated expression differences identified by SMR, such as decreased expression of *ANKRD45* in subpopulations of glutamatergic (PIP = 0.95) and GABAergic (PIP = 0.89) neurons, increased expression of *AS3MT* in (a subpopulation of) NPCs (PIP = 0.96), decreased expression of *ZKSCAN8* in deep layer glutamatergic neurons (PIP = 1) and increased expression of *SNX19* in deep layer glutamatergic neurons (PIP = 0.91). Schizophrenia-associated gene expression differences uniquely identified by cTWAS included increased expression of *BBS12* in a subpopulation of deep layer glutamatergic neurons (PIP = 0.98), increased expression of *INO80E* in deep layer glutamatergic neurons (PIP = 0.81) and increased expression of *ST3GAL3* in a subpopulation of GABAergic neurons (PIP = 0.98). cTWAS provided evidence for altered expression of 17 genes in susceptibility to bipolar disorder (Supplementary Table 4). These included potential risk mechanisms shared with schizophrenia (e.g. decreased expression of *ARL3* in subtypes of glutamatergic and GABAergic neuron [PIP > 0.95], increased expression of *ST3GAL3* in a subtype of GABAergic neuron [PIP = 0.92] and increased expression of *INO80E* in deep layer glutamatergic neurons [PIP = 0.91]), as well as distinct associations such as increased expression of *GIGYF1* in a subtype of upper layer glutamatergic neuron (PIP = 1) and increased expression of *SNX16* in a subtype of GABAergic neuron (PIP = 0.95). cTWAS prioritised the expression of 19 genes in the etiology of major depressive disorder (Supplementary Table 4). These included increased expression of *NEGR1* in NPCs (PIP = 1), increased expression of *WDR6* in a subpopulation of upper layer glutamatergic neuron (PIP = 1) and decreased expression of *DISP2* in a subtype of NPC (PIP = 0.99). In common with schizophrenia and bipolar disorder, ADHD risk alleles were found to be associated with increased expression of *ST3GAL3 in* GABAergic neurons (PIP = 0.98), while, like schizophrenia, those for OCD were found to be associated with reduced *ZKSCAN8* expression in deep layer glutamatergic neurons (PIP = 0.98) (Supplementary Table 4).

Discussion

Elucidating the molecular mechanisms by which common non-coding genetic variation influences human traits requires not only identification of the genes that are impacted and an understanding of their functional alteration, but also knowledge of which cell types these regulatory effects are operating in. In this study, we have produced the first single nuclei eQTL dataset of the prenatal human brain, leveraging these data to identify gene expression differences in individual developing cell populations that credibly mediate genetic risk for neuropsychiatric disorders.

We identified more than twice as many eGenes in this study (3121 at FDR < 0.05) than we did previously using bulk RNA sequencing data from a human prenatal brain collection of similar size (N = 120; 1329 eGenes at FDR < 0.05) [11], highlighting a major advantage of snRNA-Seq data for eQTL mapping. However, given sample size is known to be an important determinant of eQTL discovery [12,30], further gains can be expected in future single cell eQTL analyses of the human prenatal brain where this is expanded. As in other single nuclei eQTL mapping studies based on pseudobulk analyses [7-10], we observed a strong correlation between the number of nuclei in each cell population and the number of eGenes detected within them, likely reflecting greater precision of gene expression measurement when more nuclei are assayed. Future studies are therefore likely to benefit from sequencing more nuclei / cells per individual, as well as greater sequencing depth. A limitation of this and other currently available single nuclei / cell eQTL datasets is that gene expression measures do not distinguish between alternative mRNA transcripts. The incorporation of long-read RNA sequencing into future single nuclei- / cell- based eQTL studies should help resolve the individual gene transcripts impacted by identified eQTL in each cell population.

Given that many common genetic variants associated with complex human traits will exert their effects through altered gene expression, eQTL identified within individual cell populations constitute a logical functional annotation for identifying cell types that are relevant to their etiology. We find significant enrichment of common variant liability to schizophrenia [25] within fine-mapped eQTL identified in both glutamatergic and GABAergic neuronal populations of the developing cerebral cortex (with most significant enrichment in upper layer glutamatergic neurons), consistent with previous analyses based on cell-specific gene expression[30] and genomic open chromatin regions[19] in this tissue. In contrast to our findings for schizophrenia, common variant liability to bipolar disorder and major depressive disorder was most significantly enriched in fine-mapped eQTL identified in GABAergic neurons of the developing cerebral cortex, although neither observation reached the Bonferroni-corrected *P*-value threshold. We also note the caveat that greater enrichment of genetic liability to these disorders in particular cell types does not preclude certain risk mechanisms operating in other cell populations; indeed, we provide evidence that altered gene expression in a variety of prenatal cell types confers susceptibility to these conditions (discussed below).

Our SMR and cTWAS analyses implicate altered expression of specific genes within individual prenatal brain cell populations in the etiology of neuropsychiatric disorders. These include genetic effects on gene expression that have previously been highlighted through gene expression analyses in (bulk) prenatal human brain, as well as novel candidate gene risk mechanisms. The former includes increased expression of *AS3MT* in association with genetic risk for schizophrenia, an effect that has been reported as pronounced in prenatal (compared to adult) brain [31] and restricted to a particular *AS3MT* RNA isoform (*AS3MT*^{d2d3}) [32]. In the present study, we find that effects on *AS3MT* expression associated with schizophrenia (and bipolar disorder) risk variation are strongest in NPCs, with effects also observed in a population of developing upper layer glutamatergic neurons in our SMR analyses. Recently, Kim et al [33] reported that over-expression of *AS3MT*^{d2d3} in mouse cortical neural stem cells led to altered neuronal

differentiation as well as neuroanatomical and behavioural phenotypes associated with schizophrenia, providing a plausible neurobiological mechanism by which elevated *AS3MT* expression in human cortical NPCs confers risk for the disorder. Another gene that has previously been implicated in schizophrenia is *INO80E*, which we find to have elevated expression in deep layer glutamatergic neurons in association with genetic risk variants for schizophrenia and bipolar disorder. *INO80E* is one of 120 genes prioritized by Trubetskoy et al [23] in common variant susceptibility to schizophrenia and resides within a microduplication on chromosome 16p11.2 that is associated with a greatly increased risk of the disorder [34]. Recently, Hu and colleagues [35] reported that overexpression of *INO80E* in postnatal frontal cortex in mice resulted in impaired sensorimotor gating and altered dendritic spine density. We have previously reported association between schizophrenia risk variation and increased expression of a specific *INO80E* transcript in (bulk) prenatal brain tissue [36]; here, we advance understanding of this potential risk mechanism by defining a specific cell type of the developing cerebral cortex that it operates in.

Novel credible risk mechanisms for neuropsychiatric disorders identified in this study include decreased expression of *ZKSCAN8* in deep layer glutamatergic neurons in association with risk alleles for schizophrenia and obsessive-compulsive disorder, and decreased expression of *ARL3* in populations of GABAergic and upper layer glutamatergic neurons in association with genetic risk variation for schizophrenia and bipolar disorder. *ZKSCAN8* is yet to be functionally characterised, but encodes a zinc finger protein, and is therefore likely to have a role in gene regulation. *ARL3* is essential for the function of cilia, which are known to have important roles in cortical development, including neuronal migration, differentiation and axon guidance [37]. Other novel findings of our SMR / cTWAS analyses include increased expression of *BBS12* in a subpopulation of deep layer glutamatergic neurons in association with genetic risk variation for schizophrenia, decreased expression of *DISP2* in a subpopulation of NPC in association with genetic risk variants for major depressive disorder, and increased expression of *SNX16* in a subpopulation of GABAergic neuron in association with risk alleles for bipolar disorder.

The results of our analyses should help guide functional follow-up of GWAS findings for neuropsychiatric disorders in appropriate cell types. Although we have focused our analyses on these disorders, we expect our data to also be valuable for translating GWAS findings for other conditions with a potential neurodevelopmental component.

Methods

Samples

Cerebral cortex from an initial collection of 150 karyotypically-normal fetuses derived from elective terminations of pregnancy was obtained from the Human Developmental Biology Resource (HDBR) (http://www.hdbr.org). After quality control (see below), we retained data from 134 samples for eQTL analysis. Ethical approval for the collection and distribution of fetal material for biomedical research was granted to the HDBR by London Fulham research ethics committee (reference 23/LO/0312) and North East - Newcastle & North Tyneside 1 research ethics committee (references: 08/H0906/21+5 and

23/NE/0135). Consent for use of this material for medical research was provided by female donors. Information on gestational age and sex was provided by the HDBR; fetal sex was confirmed in this study through expression of genes on the Y-chromosome in males and heterozygosity for X-chromosome DNA markers and expression of *XIST* in females.

Tissue was obtained frozen, having been first dissected by the HDBR. Each sample was thawed and dounce homogenized in DMEM on ice, followed by nuclei isolation using Nuclei EZ lysis buffer (Sigma-Adrich). Nuclei were passed through a 40mm pluriStrainer (Cambridge Bioscience), followed by a 10mm pluriStrainer, and then pelleted by centrifugation at 300g for 3 mins at 4°C. Nuclei were resuspended in wash buffer (1x PBS with 1% BSA), counted using a Countess automated cell counter (ThermoFisher Scientific) and visualized under a light microscope to assess integrity. Approximately 4 million nuclei from each sample were taken for fixation using Evercode (v2) nuclei fixation reagents (Parse Biosciences) in preparation for snRNA-Seq library generation. Remaining nuclei were used for genomic DNA extraction.

Genotyping

Genomic DNA was extracted from nuclei from each sample using standard phenol-chloroform procedures. DNA was then genotyped for approximately 654,000 single nucleotide polymorphisms (SNPs) using Infinium Global Screening Arrays v3.0 (Illumina). An Illumina HiScan system was used to image the bead chips and GenomeStudio 2011.1 software to analyse the data.

Raw genotypes (aligned to hg19) were converted to PLINK format [38,39]. Pre-imputation genotype quality control (QC) was primarily performed using the GenotypeQCtoHRC pipeline [40]. SNPs were excluded if they exhibited a genotype call rate < 95%, significant deviation from Hardy-Weinberg equilibrium ($P < 1 \times 10^{-6}$) or a minor allele frequency (MAF) < 0.01. Strand-ambiguous SNPs (A/T or G/C) with MAF > 0.4 were also removed. Remaining bi-allelic SNPs were lifted over to the hg38 build using rtracklayer [41], excluding any markers located in conversion-unstable loci [42]. Where possible, variants were aligned to dbSNP release 156 rsIDs and then harmonized to the Trans-Omics for Precision Medicine (TOPMed) reference panel (release 3) [43] using Genotype Harmonizer [44], yielding 523,575 successfully aligned markers. Per-chromosome VCF files were submitted to the TOPMed Imputation Server for phasing (Eagle v2.4) [45] and imputation (Minimac v4.1.6) [46] against the TOPMed-r3 reference panel.

Post-imputation, SNPs were excluded if they had a low imputation accuracy score ($R^2 < 0.8$), a MAF < 0.05, or deviation from Hardy-Weinberg equilibrium at $P < 1 \times 10^{-4}$. The final count of high confidence SNPs across all samples was 5,794,957. Finally, a principal component analysis (PCA) was run on linkage disequilibrium (LD) pruned SNPs using plink (*--indep-pairwise 250 5 0.2*). Inspection of the PC plot indicated that the first four genotype PCs captured the majority of the variance (Supplementary Fig. 3).

Single nuclei RNA sequencing

Single nuclei RNA sequencing libraries were prepared for each sample by combinatorial barcoding using three Evercode WT (v2) Mega kits (Parse Biosciences), which can generate libraries from up to 1 million nuclei from 96 samples per kit. We prepared libraries for the majority of samples (N = 95) by loading up to 10,000 nuclei from each sample on each of two Evercode WT (v2) Mega kits and then combining the sequencing data. Sixteen sequencing sublibraries (each containing library from an estimated 62,500 nuclei from up to 96 samples) were generated from each Evercode WT v2 Mega kit. The concentration and quality of each sublibrary was determined using a Qubit 4 fluorometer (Invitrogen) and 5200 Fragment Analyzer System (Agilent). Sublibraries from each kit were then pooled at equal concentrations and sequenced across 3 X Illumina NovaSeq 6000 S4 flowcells to generate approximately 30,000 reads per nucleus. A median of 12,387 nuclei were sequenced per sample.

Raw sequencing was converted to FASTQ files usingBCL Convert (Dragen v.4.3.6; Illumina) and reads aligned to the hg38 genome reference panel (Ensembl release 113) using the split-pipe software suite (Parse, version 1.3.1). Following read alignment, and the generation of unique molecular identifier (UMI) count matrices, downstream processing of the transcriptomic data was primarily conducted using Scanpy (v1.10.4) [16]. To ensure only high-quality single nuclei data were retained for downstream analyses, a stringent quality control procedure was followed. Nuclei were excluded if they expressed fewer than 1000 or more than 5000 genes, if they had fewer than 500 UMI counts, if more than 5% of their transcriptomic content aligned to mitochondrial or ribosomal genes, or if they were classified as doublets (Scrublet, v0.2.3) [47]. Genes were excluded if detected in fewer than 10 nuclei per sample. *MALAT1* and all mitochondrial genes listed in the Human MitoCarta 3.0 inventory [48] were also excluded. Retained nuclei had a median of 3022 UMI counts and 1742 detected genes per nucleus.

Sample exclusion

Of the 150 samples for which both DNA genotype and snRNA-Seq data were available, 16 were excluded prior to eQTL analysis. One sample was excluded based on deviation in autosomal heterozygosity in the genotype data (+4 standard deviations from the mean), suggestive of sample contamination. Two samples were identified as monozygotic twins / duplicate samples from the genotype data (kinship coefficient ≈ 0.996), and therefore one sample from this pair was excluded. Three samples were excluded for having a low number of sequenced nuclei, and two samples for missing age information. Three samples were excluded due to ambiguous sex classification in the snRNA-Seq data based on expression of *XIST* and genes on the Y-chromosome. Six samples were excluded because genotypes for expressed SNPs in snRNA-Seq data did not match genotypes from genomic DNA for those samples. The latter was determined by generating per-nucleus allele counts (sequencing pileups) at each expressed SNP using cellSNP-lite (v1.2.2) [49] with a minimum sequencing read depth of 20 reads per SNP. Sample assignments were subsequently inferred using Vireo (v0.5.9) [50], which probabilistically matches the inferred SNP profiles for each nucleus to pre-imputed sample genotypes (MAF ≥ 0.01; harmonised to hg38) via Bayesian decomposition of read counts, producing a predicted sample ID per nucleus.

Demographics of final collection

After quality control, data from 134 unrelated samples remained and were carried forward for downstream analyses. Of these 134 samples, 76 were confirmed as male and 58 female, with a median age of 16 PCW (range: 12 – 21 PCW). Sample genetic ancestry was inferred from genotypes in genomic DNA using ancestry-informative markers derived from the Allen Ancient DNA Resource reference panel [51] followed by linear discriminant analysis, as described previously [52]. The majority of samples (N = 118; 88%) were classified as being of European ancestry. Of the remainder, 9 were classified as being of African American / Afro-Caribbean, 2 of sub-Saharan African, 2 of Latino, 2 of Central / South Asian and 1 of East Asian, ancestry.

Cell clustering and annotation

We defined cell populations at two levels of granularity using Scanpy (v1.10.4)[16] : 7 broad 'level 1' cell types and 12 'level 2' cell types that were subclusters of level 1 cell clusters comprising more than 100,000 nuclei.

For the level 1 annotation, raw UMI counts were normalised to counts per 10,000 and log-transformed. Principal component analysis (PCA) was performed on the 6000 most highly variable genes, retaining the top 50 principal components. Batch effects across samples were corrected using Harmony integration [53] and a k-nearest neighbor graph constructed on the Harmony-corrected embedding. Leiden clustering at a resolution of 1.0 yielded 31 clusters, which were manually identified according to differentially expressed genes. Nuclei from cell clusters of ambiguous identity (N = 3) or which highly expressed markers of non-cortical cell types (N = 5) were removed. Retained nuclei (N = 679,738) were reclustered at a resolution of 0.2, which produced stable, biologically coherent clusters with representation from all samples. Clusters were visualized using Uniform Manifold Approximation and Projection (UMAP) embeddings and annotated as 7 major cell types of the mid-gestation human cerebral cortex: upper layer glutamatergic neurons (Glu-UL), deep layer glutamatergic neurons (Glu-DL), neural progenitor cells (NPC), GABAergic neurons (GABA), endocytes / pericytes (Endo-Peri), oligodendrocyte precursor cells (OPC) and microglia (MG) based on the results of differential gene expression analysis using Scanpy[16] (Supplementary Table 1) and UMAP visualisation of canonical markers (*SATB2* and *EPHA3* for Glu-UL; *TLE4* for Glu-DL; *GLI2* and *PAX6* for NPC; *GAD1* for GABA; *FN1* for Endo-Per; *PDGFRA* for OPC; and *C3* for MG).

For the level 2 annotation, level 1 clusters comprising more than 100,000 nuclei (Glu-UL, Glu-DL, GABA, NPC), were subset into independent Scanpy data objects. Raw counts were restored, and the preprocessing pipeline described for the level 1 clusters repeated per subset. At this level, Leiden clustering was then applied at cell population-specific resolutions (0.2 for Glu-UL, Glu-DL, and NPC; 0.1 for GABA) to yield interpretable subclusters. To ensure sufficient statistical power for downstream eQTL mapping, level 2 subclusters were required to have nuclei represented from ≥130 samples; subclusters failing this coverage threshold were excluded from further analysis. This criterion resulted in the exclusion of 3 NPC subclusters and one GABA inhibitory neuron subcluster, with a total of 12 subclusters retained for downstream analyses.

Pseudobulk expression matrices were generated for each of the level 1 and level 2 cell populations by summing raw UMI counts across all nuclei within each cell population for each sample, thereby aggregating the gene expression signal to reduce sparsity and enhance eQTL detection.

Expression quantitative trait loci mapping

Cis-expression quantitative trait loci (*cis*-eQTL) discovery was performed independently for each cell population using normalized pseudobulk gene expression and matched genotype data. For each cell population, pseudobulk read counts were normalized for library size and composition bias using trimmed mean of M-values (TMM) normalisation[54], followed by $\log_2$(CPM + 1) transformation and quantile normalization. Genes were annotated to Ensembl identifiers using BioMart-derived lookup tables (v2.60.1) [55], and TSS were used to construct BED-formatted expression files (chromosome, TSS, TSS+1, gene ID, sample expression). Only autosomal genes were retained. To account for latent sources of expression variation, PCA was performed on the normalised pseudobulk expression data for each cell type, and *cis*-eQTL mapping was evaluated using covariate models including between 10 and 50 gene expression principal components (PCs), with the optimal number of gene expression PCs selected per population based on eQTL discovery performance (Supplementary Fig. 2). All models additionally included sex, gestational age (post-conception weeks), and the top four genotype PCs to control for population structure.

Cis-eQTL mapping was performed using TensorQTL (v1.0.8) [17] in a two-step procedure. In the nominal pass, linear regression was performed for every SNP–gene pair within a ±500 kb window of each gene's TSS, testing the additive effect of ALT allele dosage on normalized gene expression while adjusting for covariates. This produced nominal association *P*-values for all tested SNP–gene pairs. In the permutation pass, each SNP–gene pair in the gene *cis*-window was re-tested 10,000 times, shuffling the gene expression values across samples while preserving the original genotype and covariate structure. For each permutation and gene, the minimum nominal *P*-value across all *cis*-variants was recorded and used to fit a gene-specific Beta distribution approximating the null distribution of the best *P*-value, thereby correcting for the number of tested variants and local linkage disequilibrium. The observed minimum nominal *P*-value per gene was then converted to an empirical gene-level *P*-value using the fitted Beta cumulative distribution function. Finally, empirical *P*-values for each gene were corrected using the method of Storey and Tibshirani [18] to control the genome-wide false discovery rate (FDR). Genes with eQTL q-values < 0.05 were considered significant eGenes, and the SNP with the lowest nominal *P*-value reported as the top *cis*-eQTL.

Sharing of eQTL between cell populations

For eQTL sharing across cell types, we used π_1 as a metric for replication. For each pair of cell types (reference and query), SNP–gene pairs were merged by Ensembl gene ID and SNP ID. We then extracted the nominal *P*-values from the query dataset corresponding to the significant (FDR < 0.05) SNP–gene pairs discovered in the reference dataset. These *P*-values were submitted to the qvalue R package

(v.2.36.0) [56] which estimates π_0 (the proportion of true null hypotheses) using the Storey and Tibshirani method. [18] The π_1 statistic was computed as $\pi_1 = 1 - \pi_0$, representing the estimated proportion of true positive (non-null) associations among the tested nominal *P*-values in the query dataset. The π_1 values were presented in a heatmap illustrating replication strength across all level-1 cell type pairs.

We applied the same π_1 framework to evaluate sharing of high confidence *cis*-eQTL signals identified in our developing neuronal populations within a mature (adult) neuronal context. Specifically, we compared significant *cis*-eQTLs (FDR < 0.05) from the 3 'level 1' prenatal neuronal populations (Glu-UL, Glu-DL and GABA neurons) against nominal *cis*-eQTL association statistics from corresponding adult neuronal populations (excitatory neurons, Exc; Inhibitory neurons, Inh) generated from the adult human cerebral cortex by Fujita et al[9]. To assess the consistency of effect sizes between eQTL detected in developing neuron populations and the same SNPs in corresponding adult neuron populations, we computed pairwise correlations of beta coefficients for overlapping SNP–gene pairs. For each prenatal neuron population, significant *cis*-eQTLs were merged with the full nominal beta values from the corresponding adult neuronal population (i.e. Exc for prenatal Glu-UL and Glu-DL neurons, and Inh for prenatal GABA neurons). The resulting paired betas were used to calculate Spearman's rank correlation coefficients (ρ) evaluating the concordance in effect magnitude and direction between prenatal and adult neuronal contexts.

Testing enrichment of eQTL in open chromatin regions mapped in the prenatal cerebral cortex

To test whether the top eQTL identified in each cell population are collectively over-represented in predicted regulatory genomic elements mapped in corresponding cell populations of the developing human brain, we assessed their enrichment within open chromatin regions (OCRs) identified by single-nucleus ATAC-seq (snATAC-seq) on human prenatal cerebral cortex [19]. For each level 1 and level 2 cell type, genomic coordinates (hg38) for the top eQTL for each significant (FDR < 0.05) eGene were retrieved using biomaRt (Ensembl SNP mart) and converted into separate Genomic Ranges (v1.56.2) objects [57]. Variants mapping to non-standard chromosomes were excluded. Enrichment was evaluated by quantifying the overlap between the genomic coordinates of top eQTL in each cell type with those of OCRs from the corresponding cell type in the snATAC-seq data (e.g., Glu-UL eQTL tested against upper-layer excitatory neuron OCRs [ulEN]; NPC eQTL tested radial glia OCRs [RG]). Significance was determined using block bootstrapping as implemented in nullranges (v1.10.0) [58] with 1,000 iterations and 500 kb block lengths. This approach generates null OCR sets by sampling genomic blocks with replacement while preserving within-chromosome structure. Unlike simple random shuffling, which can overestimate significance by disrupting spatial clustering and local density of accessible regions, block bootstrapping maintains these genomic correlations, yielding a more accurate and conservative null distribution. Fold enrichment was calculated as the ratio of observed overlaps to the mean from bootstrapped sets, with empirical *P*-values derived from the bootstrap distribution. *P*-values were then Bonferroni corrected for the number of cell type comparisons (24, including eQTL mapped in 4 GABA cell populations against OCRs mapped in GABAergic neurons from either the medial or caudal ganglionic

eminences)[19]. Open chromatin regions identified in the prenatal cerebral cortex[19] were also used to functionally annotate individual top eQTL within each cell type (Supplementary Table 2).

Testing enrichment of SNP heritability for neuropsychiatric disorders in eQTL mapped in prenatal brain cell populations

To assess whether *cis*-eQTL in developing brain cell populations capture a disproportionate share of SNP heritability for neuropsychiatric traits, we first generated a maximum credible set posterior probability (MaxCPP) annotation for eQTL in each cell type, following a framework described previously [12,20]. This involves using a Bayesian fine-mapping method, Sum of Single Effects (SuSiE; v0.11.92) [21], to define the 95% credible set of SNPs containing the causal eQTL for each identified eGene (FDR < 0.05) in each cell population and the posterior inclusion probability (PIP) of each SNP being the causal eQTL. Genotype dosages were extracted from the imputed VCF using bcftools (v1.9) [59] and gene expression and covariate matrices used during eQTL discovery were harmonized for SuSiE compatibility using custom scripts. Then, for each eGene in each level 1 and level 2 cell type, SNPs within a ±500 kb *cis*-window of the TSS were extracted from the genotype dosage matrix and normalized expression values and genotypes were residualized against the covariate matrix to remove confounding. SuSiE was used to estimate PIPs with L = 10 to accommodate multiple independent signals per locus [21]. High-purity 95% credible sets (min $r^2 \geq 0.5$) were retained, and overlapping sets were filtered to prevent redundancy. From these credible sets, a continuous MaxCPP metric was computed per variant as the maximum PIP across all eGenes in which it appeared, with non-credible-set variants assigned MaxCPP = 0. Variants were harmonized to the 1000 genomes hg38 reference panel and per-chromosome MaxCPP files were used as annotation files in stratified linkage disequilibrium score regression (S-LDSR; v1.0.1)[22] to test for significant enrichment of SNP heritability for neuropsychiatric traits in fine-mapped eQTL in each cell population.

Genome-wide association study (GWAS) data for 5 neuropsychiatric traits (schizophrenia [23], bipolar disorder [24], major depressive disorder [25], attention deficit hyperactivity disorder [26], and obsessive-compulsive disorder [27]) were downloaded from figshare repositories via the Psychiatric Genomics Consortium website (https://pgc.unc.edu/for-researchers/download-results) (Supplementary Table 5), lifted over to genome build hg38, and prepared for S-LDSR using the *munge_sumstats.py* script [22]. For each level 1 and level 2 cell type, per-chromosome MaxCPP annotation files were aligned to the European-ancestry 1000 Genomes Phase 3 reference panel (hg38 build) and restricted to HapMap3 SNPs (as recommended [22]) for stable LD estimation. Cell-type-specific LD scores were computed separately for each annotation file using the same reference panel. S-LDSR was then performed jointly on the baseline-LD model plus the cell-type MaxCPP annotation, allowing conditional estimation of the unique per-SNP heritability contribution of the MaxCPP annotation (standardised enrichment τ^*) while accounting for overlap with baseline categories (v1.2; encompassing 53 functional categories, including coding, UTR, promoter, intron, and chromatin marks) and genome-wide LD structure. Significance was assessed with a one-tailed *P*-value derived from S-LDSR Z-scores, testing for positive enrichment of the

MaxCPP coefficient. Enrichment was deemed significant after Bonferroni correction applied separately within each level: $P < 0.05 / 7 \approx 7.14 \times 10^{-3}$ for the 7 level 1 cell populations, and $P < 0.05 / 12 \approx 4.16 \times 10^{-3}$ for the 12 level 2 cell populations.

Summary-data-based Mendelian randomization

To test for single-variant pleiotropic associations between *cis*-eQTL in each cell population and neuropsychiatric traits (i.e. whether the same causal variant influences both gene expression and trait risk), we performed summary-data-based Mendelian randomization (SMR) and multi-SNP heterogeneity testing (HEIDI) using the SMR tool (v1.4.0)[4]. SNP positions and alleles for nominal *P*-value *cis*-eQTL in each cell type were extracted from the European-ancestry 1000 Genomes Phase 3 hg38 reference panel, and allele frequencies obtained from the corresponding merged frequency file. Gene metadata (chromosome, position, strand, Ensembl ID) were incorporated, and SNP alleles harmonized against the reference panel (flipping where necessary); SNPs with mismatches, missing frequencies, or other inconsistencies were discarded. A cell-type-specific query file was generated containing harmonized SNP-gene pairs, which was then converted to binary BESD format. GWAS summary statistics for schizophrenia [23], bipolar disorder [24], major depressive disorder [25], attention deficit hyperactivity disorder[26] and obsessive-compulsive disorder [27] (Supplementary Table 5) were formatted into *.ma* files, ensuring A1 was the effect allele and allele frequencies corresponded to A1. SMR was run for each cell type and trait pairing using the merged hg38 reference panel as the LD source. To increase power while controlling instrument strength, analyses included all *cis*-eQTLs with nominal $P < 5 \times 10^{-8}$. The HEIDI heterogeneity test was performed concurrently to distinguish associations that are likely to result from linkage disequilibrium between distinct causal variants. We highlight gene expression-trait associations with SMR *P*-values below a per-cell-type Bonferroni threshold (0.05 / number of unique genes tested in that cell type) and with HEIDI *P*-values > 0.01, supporting a pleiotropic model rather an association through linkage disequilibrium.

Causal transcriptome-wide association studies

Transcriptome-wide association studies test whether the genetically regulated *cis* component of a gene's expression in a particular tissue or cell type is associated with a complex trait[5]. This is achieved by first training predictive models (SNP weights) for a gene's expression from joint genotype-expression data in an eQTL reference panel and then imputing *cis*-regulated gene expression into large GWAS cohorts using only genotype data. Causal transcriptome-wide association studies (cTWAS)[6] extend traditional TWAS by employing a Bayesian fine-mapping framework that jointly models imputed gene expression and direct SNP effects within LD blocks, thereby distinguishing likely causal genes from those spuriously associated due to linkage disequilibrium, co-regulation with nearby genes, or horizontal pleiotropy. This approach is reported to substantially reduce false-positive gene-trait associations common in classical TWAS while improving prioritisation of putatively causal genes, even in loci with multiple causal variants or genes [6].

To generate SNP weights for genes expressed in each level 1 and level 2 cell population, quantile-normalized gene expression values were used directly as input. Genotypes were restricted to SNPs present in the European-ancestry 1000 Genomes Phase 3 reference panel (hg38 build), and *cis*-windows were defined as ±500 kb around the TSS of each gene. LD matrices and variant information files were generated from European-ancestry 1000 Genomes Phase 3 hg38 genotypes and partitioned into LDetect defined EUR blocks [60]. Following cTWAS recommendations [6], SNP weights were computed using L1-penalised regression (LASSO) models in FUSION software [5], as their sparsity and robustness improve expression imputation stability and aid fine-mapping in the presence of correlated predictors. Covariates (sex, post-conception weeks, 4 genotype PCs, and cell-type specific expression PCs) were supplied directly to the weight computation step via a transposed covariate file. Genes were retained only if *cis*-heritability was significant at $P \leq 0.01$ and performance was validated by 5-fold cross-validation. We tested associations between prenatal brain cell gene expression and schizophrenia [23], bipolar disorder [24], major depressive disorder [25], attention deficit hyperactivity disorder [26] and obsessive-compulsive disorder [27], using the same GWAS data as used for SMR (Supplementary Table 5). GWAS summary statistics were harmonized (allele flipping where necessary) and provided as signed Z-scores. cTWAS was run separately for each prenatal brain cell type using default parameters [6]. Posterior inclusion probabilities (PIPs) were estimated for each gene and SNP. Nominal *P*-values were derived from the gene-level Z-scores, and then Bonferroni corrected based on the number of genes tested in the cell type (ranging from 373 in MG to 1680 in Glu-UL). Expression of genes with SuSiE PIP > 0.8 within credible sets were highlighted as high-confidence fine-mapped causal candidates in the respective cell type.

Declarations

Acknowledgements

This work was supported by a Medical Research Council / UK Research and Innovation project grant (MR/Y003756/1) to D.C., M.C.O'D. and N.J.B.

C.E.T. is supported by a GW4 BIOMED Medical Research Council DTP (Grant No. MR/W006308/1), awarded to the Universities of Bath, Bristol, Cardiff, and Exeter from the Medical Research Council/ UKRI.

Y.B. is supported by a PhD scholarship awarded from the Ministry of National Education of the Republic of Türkiye under Law No. 1416.

References

1. Maurano, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337, 1190–1195 (2012).
2. Gallagher, M. D., & Chen-Plotkin, A. S. The Post-GWAS Era: From Association to Function. *American Journal of Human Genetics* 102, 717–730 (2018).

3. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369, 1318–1330 (2020).
4. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature Genetics* 48, 481–487 (2016).
5. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics* 48, 245–252 (2016).
6. Zhao, S. et al. Adjusting for genetic confounders in transcriptome-wide association studies improves discovery of risk genes of complex traits. *Nature Genetics* 56, 336–347 (2024).
7. Yazar, S. et al. Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science* 376, eabf3041 (2022).
8. Bryois, J. et al. Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders. *Nature Neuroscience* 25, 1104–1112 (2022).
9. Fujita, M. et al. Cell subtype-specific effects of genetic variation in the Alzheimer's disease brain. *Nature Genetics* 56, 605–614 (2024).
10. Ray, A. et al. Single-cell transcriptome-wide Mendelian randomization and colocalization analyses uncover cell-specific mechanisms in atherosclerotic cardiovascular disease. *American Journal of Human Genetics* 112, 1597–1609 (2025).
11. O'Brien, H. E. et al. Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. *Genome Biology* 19, 194 (2018).
12. Wen, C. et al. Cross-ancestry atlas of gene, isoform, and splicing regulation in the developing human brain. *Science* 384, eadh0829 (2024).
13. Toste CC, O'Donovan MC, Bray NJ. Mapping microRNA expression quantitative trait loci in the prenatal human brain implicates miR-1908-5p expression in bipolar disorder and other brain-related traits. *Hum Mol Genet* 32, 2941-2949 (2023).
14. Sey, N. Y. A. et al. A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nature Neuroscience* 23, 583–593 (2020).
15. Emani, P. S. et al. Single-cell genomics and regulatory networks for 388 human brains. *Science* 384, eadi5199 (2024).
16. Wolf, F. A., Angerer, P., & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology* 19, 15 (2018).
17. Taylor-Weiner, A. et al. Scaling computational genomics to millions of individuals with GPUs. *Genome Biology* 20, 228 (2019).
18. Storey, J. D., & Tibshirani, R. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America 100*, 9440–9445 (2003).
19. Ziffra, R. S. et al, Single-cell epigenomics reveals mechanisms of human cortical development. *Nature* 598, 205–213 (2021).

20. Hormozdiari, F. et al. Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. *Nature Genetics* 50, 1041–1047 (2018).
21. Wang, G., Sarkar, A., Carbonetto, P., & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society. Series B, Statistical methodology* 82, 1273–1300 (2020).
22. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nature Genetics* 50, 621–629 (2018).
23. Trubetskoy, V. et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* 604, 502–508 (2022).
24. O'Connell, K. S. Genomics yields biological and phenotypic insights into bipolar disorder. *Nature* 639, 968–975 (2025).
25. Adams, M. J. et al. Trans-ancestry genome-wide study of depression identifies 697 associations implicating cell types and pharmacotherapies. *Cell* 188, 640–652.e9. (2025).
26. Demontis, D. et al. Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. *Nature Genetics* 55, 198–208 (2023).
27. Strom, N. I. et al. Genome-wide analyses identify 30 loci associated with obsessive-compulsive disorder. *Nature Genetics* 57, 1389–1401 (2025).
28. Stefansson H. et al. A common inversion under selection in Europeans. *Nature Genetics* 37, 129–137 (2005).
29. de Klein, N. et al. Brain expression quantitative trait locus and network analyses reveal downstream effects and putative drivers for brain-related diseases. *Nature Genetics* 55, 377–388 (2023).
30. Cameron, D. et al. Single-Nuclei RNA Sequencing of 5 Regions of the Human Prenatal Brain Implicates Developing Neuron Populations in Genetic Risk for Schizophrenia. *Biological Psychiatry 93*, 157–166 (2023).
31. Duarte, R. R. R. et al. Genome-wide significant schizophrenia risk variation on chromosome 10q24 is associated with altered cis-regulation of BORCS7, AS3MT, and NT5C2 in the human brain. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics* 171, 806–814 (2016).
32. Li, M. et al. A human-specific AS3MT isoform and BORCS7 are molecular risk factors in the 10q24.32 schizophrenia-associated locus. *Nature Medicine* 22, 649–656 (2016).
33. Kim, S. et al. Perturbed cell fate decision by schizophrenia-associated $AS3MT^{d2d3}$ isoform during corticogenesis. *Science Advances* 11, eadp8271 (2025).
34. Marshall, C. R. et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nature Genetics* 49, 27–35 (2017).
35. Hu, B. et al. The INO80E at 16p11.2 locus increases risk of schizophrenia in humans and induces schizophrenia-like phenotypes in mice. *EBioMedicine* 114, 105645 (2025).
36. Hall, L. S. et al. Cis-effects on gene expression in the human prenatal brain associated with genetic risk for neuropsychiatric disorders. *Molecular Psychiatry* 26, 2082–2088 (2021).

37. Liu, S., Trupiano, M. X., Simon, J., Guo, J., & Anton, E. S. The essential role of primary cilia in cerebral cortical development and disorders. *Current Topics in Developmental Biology* 142, 99–146 (2021).
38. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* 81, 559–575 (2007).
39. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4, 7 (2015).
40. Lynham, A. J. et al. DRAGON-Data: a platform and protocol for integrating genomic and phenotypic data across large psychiatric cohorts. *BJPsych Open* 9, e32 (2023).
41. Lawrence, M., Gentleman, R., & Carey, V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* 25, 1841–1842 (2009).
42. Ormond, C., Ryan, N. M., Corvin, A., & Heron, E. A. Converting single nucleotide variants between genome builds: from cautionary tale to solution. *Briefings in Bioinformatics 22*, bbab069 (2021).
43. Taliun, D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* 590, 290–299 (2021)
44. Deelen, P. et al. Genotype harmonizer: automatic strand alignment and format conversion for genotype data integration. *BMC Research Notes* 7, 901 (2014).
45. Loh, P. R. et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nature Genetics* 48, 1443–1448 (2016).
46. Fuchsberger, C., Abecasis, G. R., & Hinds, D. A. minimac2: faster genotype imputation. *Bioinformatics* 31*,* 782–784 (2015).
47. Wolock, S. L., Lopez, R., & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Systems* 8, 281–291.e9 (2019).
48. Rath, S. et al. MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic acids research* 49, D1541–D1547 (2021).
49. Huang, X., & Huang, Y. Cellsnp-lite: an efficient tool for genotyping single cells. *Bioinformatics* 37, 4569–4571 (2021).
50. Huang, Y., McCarthy, D. J., & Stegle, O. Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. *Genome Biology* 20, 273 (2019).
51. Mallick, S. et al. The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. *Scientific Data* 11, 182 (2024).
52. Smart, S. E. et al. The Role of SLC39A8.p.(Ala391Thr) in Schizophrenia Symptom Severity and Cognitive Ability: Cross-Sectional Studies of Schizophrenia and the General UK Population. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics* 198, 135–146 (2025).
53. Korsunsky I, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods* 16, 1289-1296 (2019).
54. Robinson, M. D., McCarthy, D. J., & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140 (2010).

55. Durinck, S., Spellman, P. T., Birney, E., & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols* 4, 1184–1191 (2009).
56. Storey JD, Bass AJ, Dabney A, Robinson D. qvalue: Q-value estimation for false discovery rate control. R package version 2.36.0, (http://github.com/jdstorey/qvalue) (2024).
57. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Computational Biology* 9, e1003118 (2013).
58. Mu, W. et al. bootRanges: flexible generation of null sets of genomic ranges for hypothesis testing. *Bioinformatics* 39, btad190 (2023).
59. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993 (2011).
60. Berisa, T., & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* 32, 283–285 (2016).

Figures

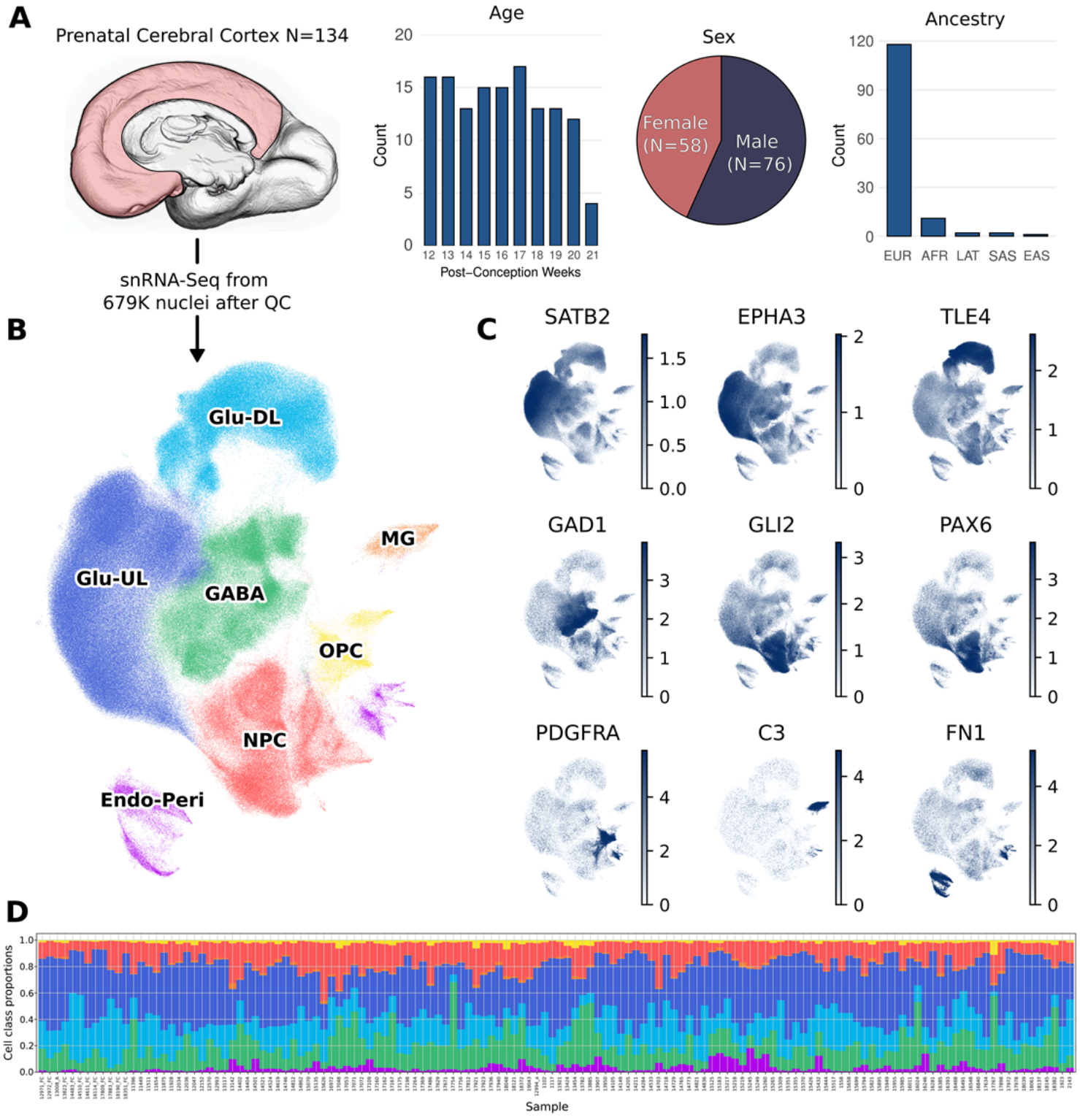


Figure 1

Sample description and primary analysis of single nuclei RNA sequencing data. A) Single nuclei RNA sequencing (snRNA-Seq) was performed on cerebral cortex from 134 unrelated samples from the second trimester of gestation, yielding trancriptional profiles from 679,738 nuclei after strict quality control. The age distribution (post-conception weeks), biological sex and inferred genetic ancestry (see Methods) of the 134 samples are shown. B) Nuclei were clustered based on transcriptional similarity

using a k-nearest neighbor graph constructed in PCA space, followed by Leiden community detection and visualized in two-dimensional space via Uniform Manifold Approximation and Projection (UMAP) in Scanpy [16]. Clusters were annotated as 7 major cell types of the second trimester human cerebral cortex: upper layer glutamatergic neurons (Glu-UL), deep layer glutamatergic neurons (Glu-DL), neural progenitor cells (NPC), GABAergic neurons (GABA), endocytes / pericytes (Endo-Peri), oligodendrocyte precursor cells (OPC) and microglia (MG) based on differential gene expression (Supplementary Table 1). C) UMAP visualisation of canonical markers: *SATB2* and *EPHA3* for Glu-UL; *TLE4* for Glu-DL; *GAD1* for GABA; *GLI2* and *PAX6* for NPC; *PDGFRA* for OPC; *C3* for MG; *FN1* for Endo-Peri. D) Proportion of nuclei from each of the 134 samples annotated to each of the 7 major cell types of the developing cerebral cortex. Color scheme follows that in panel B.

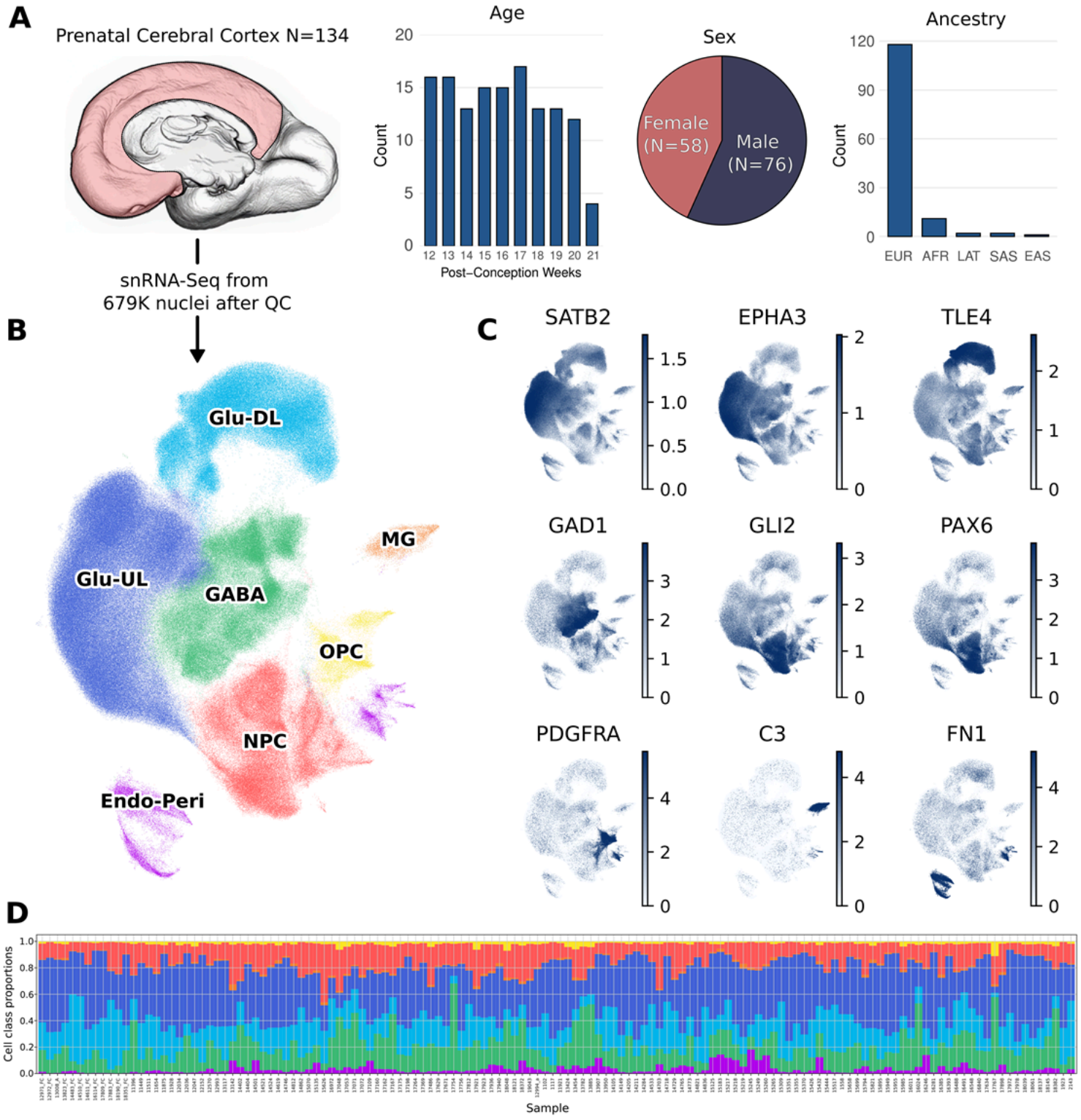


Figure 1

Sample description and primary analysis of single nuclei RNA sequencing data. A) Single nuclei RNA sequencing (snRNA-Seq) was performed on cerebral cortex from 134 unrelated samples from the second trimester of gestation, yielding trancriptional profiles from 679,738 nuclei after strict quality control. The age distribution (post-conception weeks), biological sex and inferred genetic ancestry (see Methods) of the 134 samples are shown. B) Nuclei were clustered based on transcriptional similarity

using a k-nearest neighbor graph constructed in PCA space, followed by Leiden community detection and visualized in two-dimensional space via Uniform Manifold Approximation and Projection (UMAP) in Scanpy 16. Clusters were annotated as 7 major cell types of the second trimester human cerebral cortex: upper layer glutamatergic neurons (Glu-UL), deep layer glutamatergic neurons (Glu-DL), neural progenitor cells (NPC), GABAergic neurons (GABA), endocytes / pericytes (Endo-Peri), oligodendrocyte precursor cells (OPC) and microglia (MG) based on differential gene expression (Supplementary Table 1). C) UMAP visualisation of canonical markers: *SATB2* and *EPHA3* for Glu-UL; *TLE4* for Glu-DL; *GAD1* for GABA; *GLI2* and *PAX6* for NPC; *PDGFRA* for OPC; *C3* for MG; *FN1* for Endo-Peri. D) Proportion of nuclei from each of the 134 samples annotated to each of the 7 major cell types of the developing cerebral cortex. Color scheme follows that in panel B.

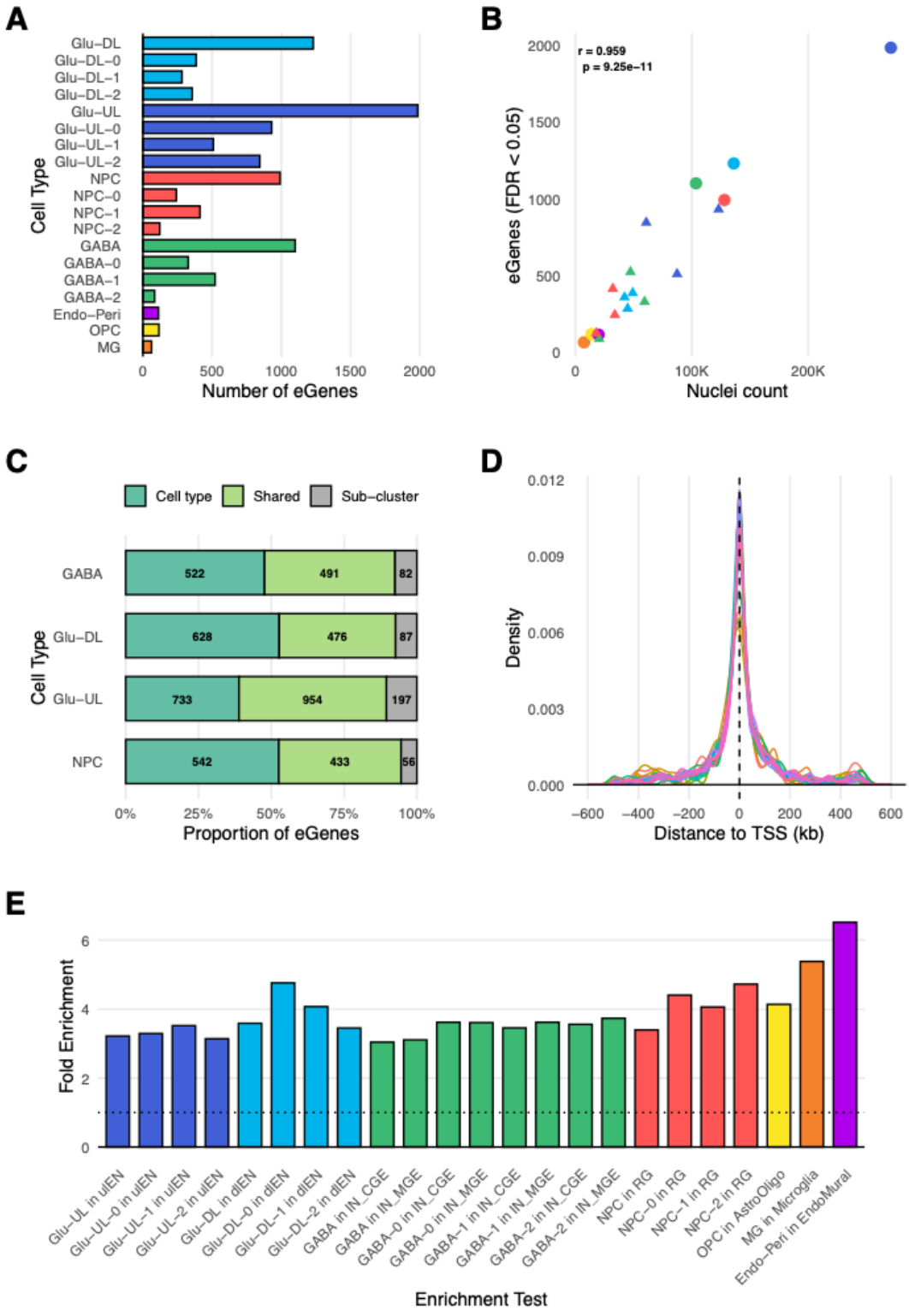


Figure 2

Identification of eQTL in individual cell populations of the prenatal cerebral cortex. A) The number of eGenes (genes with eQTL detected at FDR < 0.05) in each of 19 tested cell populations (7 broad 'level 1' cell types and 12 'level 2' subpopulations). Glu-UL = upper layer glutamatergic neurons; Glu-DL = deep layer glutamatergic neurons; NPC = neural progenitor cells; GABA = GABAergic neurons; Endo-Peri = endocytes / pericytes; OPC = oligodendrocyte precursor cells; MG = microglia. B) The number of eGenes

detected in each cell population plotted against the number of nuclei sequenced for those cell populations. Circles indicate 'level 1' cell types and triangles 'level 2' subpopulations. C) The number of eGenes identified in 4 major level 1 cell types and their level 2 subpopulations, indicating the proportion of eGenes uniquely identified at level 1 and level 2 or shared between them. Shared eGenes are those present in level 1 and at least one level 2 population. D) Density of top eQTL for eGenes in relation to distance from their *cis*-regulated gene's transcription start site (TSS) in the 19 tested cell populations. E) Enrichment of top eQTL for eGenes in open chromatin regions identified in corresponding cell populations of the prenatal human cerebral cortex by single nuclei ATAC-Seq[19]: ulEN = upper layer excitatory neuron; dlEN = deep layer excitatory neuron; IN_CGE = Inhibitory neurons from the caudal ganglionic eminence; IN_MGE = Inhibitory neurons from the medial ganglionic eminence; RG = radial glia; AstroOligo = astrocytes / oligodendrocytes. All enrichments $P < 0.001$.

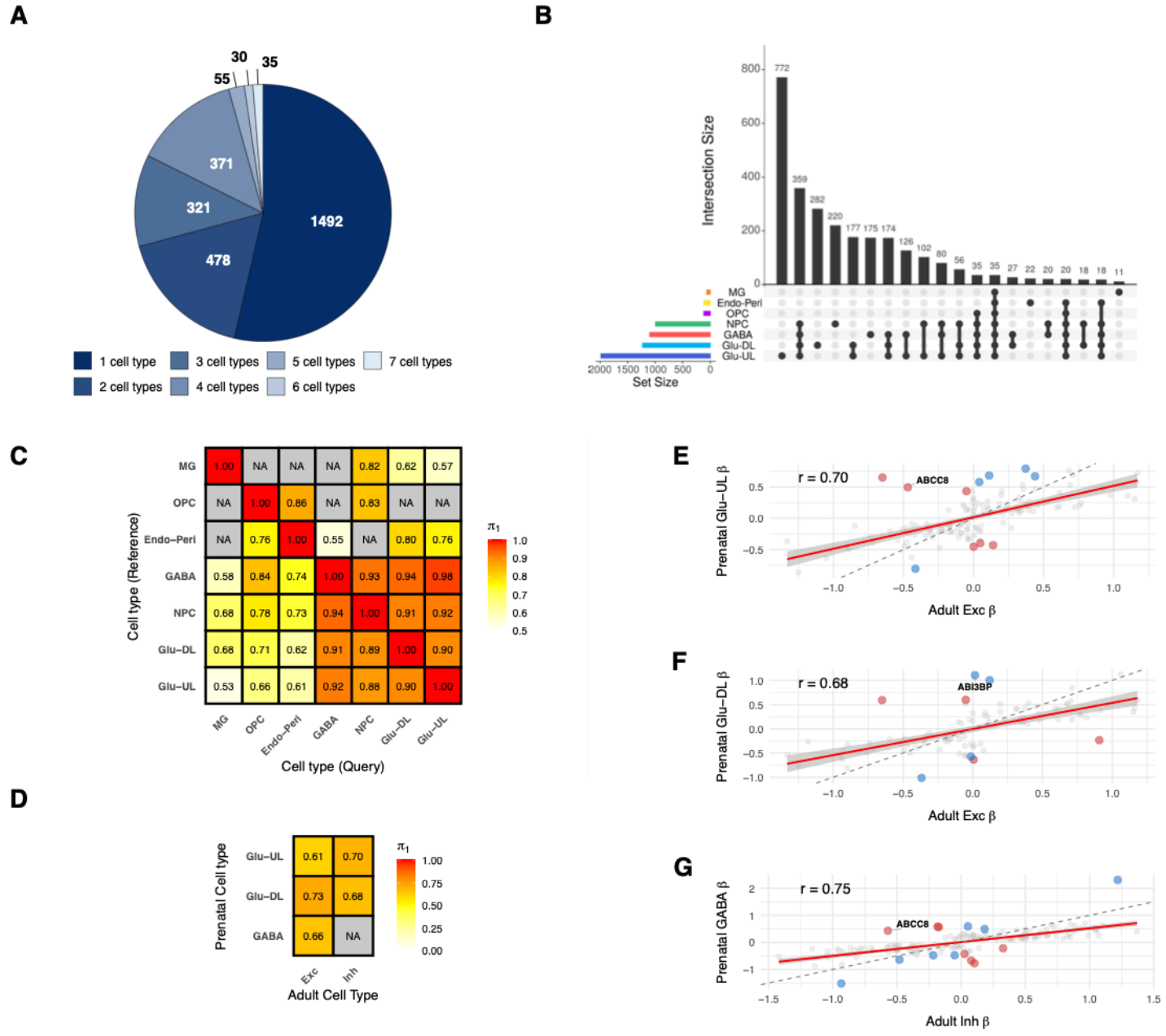


Figure 3

Sharing of identified eQTL between cell populations of the prenatal and adult cerebral cortex. A) The number of eGenes (FDR < 0.05) shared between the 7 broad 'level 1' cell types of the prenatal cerebral cortex. B) The number of eGenes (FDR < 0.05) unique to or shared between specific 'level 1' cell types of the prenatal cerebral cortex (intersections containing fewer than 10 eGenes are not shown). C) Heatmap of estimated π_1[18] for the proportion of FDR < 0.05 eQTL identified in one prenatal cell population (the reference) that have non-null effects in another prenatal cell population (the query). NA denotes comparisons with insufficient overlap of SNP-gene pairs to permit stable estimation of π_1. Glu-UL = upper layer glutamatergic neurons; Glu-DL = deep layer glutamatergic neurons; NPC = neural progenitor cells; GABA = GABAergic neurons; Endo-Peri = endocytes / pericytes; OPC = oligodendrocyte precursor cells; MG = microglia. D) Heatmap of estimated π_1 for the proportion of FDR < 0.05 eQTL identified in developing neuron populations of the prenatal brain that have non-null effects in neuronal populations of the adult cerebral cortex [9]. Exc = glutamatergic neuron of the adult cerebral cortex; Inh = GABAergic neuron of the adult cerebral cortex. NA denotes comparisons with insufficient overlap of SNP-gene pairs to permit stable estimation of π_1. E – G) Correlation of effect sizes (betas) for eQTL SNP-eGene pairs identified in developing neuronal populations (at FDR 5%) with the same SNP-gene pairs in their closest corresponding neuronal populations from the adult cerebral cortex [9]. Blue dots indicate outlier SNP-gene pairs that have a concordant direction of effect, whereas red dots indicate discordant allelic effects on the eGene's expression.

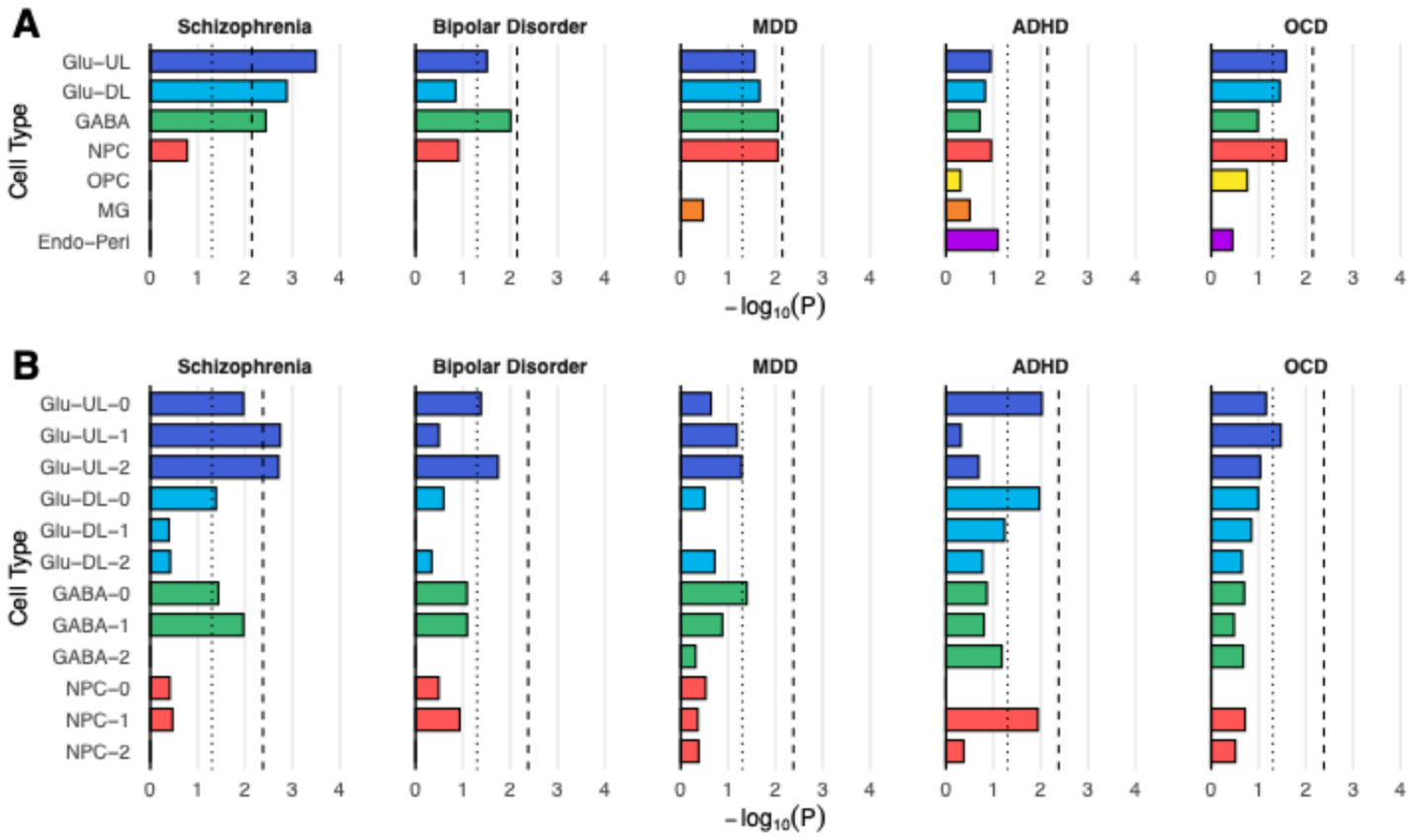


Figure 4

Enrichment of SNP heritability for neuropsychiatric disorders in fine-mapped eQTL operating in cell populations of the prenatal cerebral cortex. A) -$\log_{10}$ *P*-values for enrichment of SNP heritability for neuropsychiatric disorders in fine-mapped eQTL operating in broad 'level 1' cell types. The dotted line indicates the uncorrected $P < 0.05$ threshold; the dashed line indicates the Bonferroni *P*-value threshold for 7 tested cell types ($P = 0.0071$). B) -$\log_{10}$ *P*-values for enrichment of SNP heritability for neuropsychiatric disorders in fine-mapped eQTL operating in subpopulations of Glu-UL, GLU-DL, GABA and NPC. The dotted line indicates the uncorrected $P < 0.05$ threshold; the dashed line indicates the Bonferroni *P*-value threshold for 12 tested cell types ($P = 0.0042$). Glu-UL = upper layer glutamatergic neurons; Glu-DL = deep layer glutamatergic neurons; NPC = neural progenitor cells; GABA = GABAergic neurons; Endo-Peri = endocytes / pericytes; OPC = oligodendrocyte precursor cells; MG = microglia; MDD = major depressive disorder [25]; ADHD = attention deficit hyperactivity disorder [26]; OCD = obsessive-compulsive disorder [27].

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.xlsx
- SupplementaryTable2.xlsx
- SupplementaryTable4.xlsx
- SupplementaryTable3.xlsx
- SupplementaryTable22.xlsx
- SupplementaryTable5.xlsx
- SupplementaryTable3.xlsx
- SupplementaryFigures.docx
- SupplementaryTable3.xlsx