

Whole-genome and RNA sequencing reveal variation and transcriptomic coordination in the developing human prefrontal cortex

Donna M. Werling^{1,*}, Sirisha Pochareddy^{2,*}, Jinmyung Choi^{2,*}, Joon-Yong An^{1,3,4,*}, Brooke Sheppard¹, Minshi Peng⁵, Zhen Li^{2,6}, Claudia Dastmalchi¹, Gabriel Santpere², Andre M. M. Sousa², Andrew T. N. Tebbenkamp², Navjot Kaur², Forrest O. Gulden², Michael S. Breen^{7,8,9}, Lindsay Liang¹, Michael C. Gilson¹, Xuefang Zhao^{10,11,12}, Shan Dong¹, Lambertus Klei¹³, A. Ercument Cicek^{14,15}, Joseph D. Buxbaum^{7,8,9,16}, Homa Adle-Biassette¹⁷, Jean-Leon Thomas^{18,19}, Kimberly A. Aldinger²⁰, Diana R. O'Day²¹, Ian A. Glass²¹, Noah A. Zaitlen²², Michael E. Talkowski^{10,11,12,23}, Kathryn Roeder^{5,15}, Matthew W. State^{1,24}, Bernie Devlin¹³, Stephan J. Sanders^{1,24,†}, and Nenad Sestan^{2,25,26,27,28,†}

¹Department of Psychiatry, UCSF Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA 94158, USA.

²Department of Neuroscience and Kavli Institute for Neuroscience, Yale School of Medicine, New Haven, CT 06510, USA.

³Department of Integrated Biomedical and Life Science, Korea University, Seoul, Republic of Korea.

⁴School of Biosystem and Biomedical Science, College of Health Science, Korea University, Seoul, Republic of Korea.

⁵Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA.

⁶Department of Neurosciences, University of California, San Diego, San Diego, CA 92093, USA.

⁷Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

⁸Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

⁹Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

¹⁰Center for Genomic Medicine and Department of Neurology, Massachusetts General Hospital, Boston, MA 02114, USA.

¹¹Department of Neurology, Harvard Medical School, Boston, MA 02115, USA.

¹²Program in Medical and Population Genetics and Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA 02142, USA.

¹³Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA.

¹⁴Department of Computer Engineering, Bilkent University, Ankara, 06800, Turkey.

¹⁵Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA.

¹⁶Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

¹⁷Department of Pathology, Lariboisière Hospital, APHP and Paris Diderot University, Sorbonne Paris Cité, Paris, France.

¹⁸Department of Neurology, Yale University School of Medicine, New Haven, CT 06511, USA.

¹⁹UMRS1127, Sorbonne Université, Institut du Cerveau et de la Moelle épinière, 7, Boulevard de l'Hôpital, 75013 Paris, France.

²⁰Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, WA 98101, USA.

²¹Department of Pediatrics, the University of Washington, Seattle, WA 98105, USA.

²²Department of Medicine, University of California, San Francisco, San Francisco, CA 94158, USA.

²³Departments of Pathology and Psychiatry, Massachusetts General Hospital, Boston, MA 02114, USA.

²⁴Institute for Human Genetics, University of California San Francisco, San Francisco, CA 94158, USA.

²⁵Department of Psychiatry, Yale University School of Medicine, New Haven, CT 06520, USA.

²⁶Department of Genetics, Yale University School of Medicine, New Haven, CT 06520, USA.

²⁷Department of Comparative Medicine, Program in Integrative Cell Signaling and Neurobiology of Metabolism, Yale School of Medicine, New Haven, CT 06510, USA.

²⁸Program in Cellular Neuroscience, Neurodegeneration, and Repair and Yale Child Study Center, Yale School of Medicine, New Haven, CT 06510, USA.

*These authors contributed equally to this work.

†Please address correspondence to: stephan.sanders@ucsf.edu (S. J. S.), nenad.sestan@yale.edu (N. S.)

Summary

Variation in gene expression underlies neurotypical development, while genomic variants contribute to neuropsychiatric disorders. BrainVar is a unique resource of paired whole-genome sequencing and bulk-tissue RNA-sequencing from the human dorsolateral prefrontal cortex of 176 neurotypical individuals across prenatal and postnatal development, providing the opportunity to assay genomic and transcriptomic variation in tandem. Leveraging this resource, we identified rare premature stop codons with commensurate reduced and allele-specific expression of corresponding genes, and common variants that alter gene expression (expression quantitative trait loci, eQTLs). Categorizing eQTLs by prenatal and postnatal effect, genes affected by temporally-specific eQTLs, compared to constitutive eQTLs, are enriched for haploinsufficiency, protein-protein interactions, and neuropsychiatric disorder risk loci. Expression levels of over 12,000 genes rise or fall in a concerted late-fetal transition, with the transitional genes enriched for cell type specific genes and neuropsychiatric disorder loci, underscoring the importance of cataloguing developmental trajectories in understanding cortical physiology and pathology.

Keywords

Gene expression, eQTL, human brain development, dorsolateral prefrontal cortex, whole-genome sequencing, transcriptome, cell type-specific, neuropsychiatric disorders

Highlights

- Whole-genome and RNA-sequencing across human prefrontal cortex development in BrainVar
- Gene-specific developmental trajectories characterize the late-fetal transition
- Identification of constitutive, prenatal-specific, postnatal-specific, and rare eQTLs
- Integrated analysis reveals genetic and developmental influences on CNS traits and disorders

Introduction

The human nervous system develops slowly over several decades, starting during embryogenesis and extending postnatally through infancy, childhood, adolescence and young adulthood (Keshavan et al., 2014; Shaw et al., 2010; Silbereis et al., 2016; Tau and Peterson, 2010). Over this time, myriads of functionally distinct cell types, circuits, and regions are formed (Hu et al., 2014; Lui et al., 2011; Silbereis et al., 2016). These aspects of neurodevelopment are highly context-dependent, since neural cells are born in an immature state and must differentiate, migrate, and establish circuits to produce regionally distinct brain structures. In doing so, the cells undergo a variety of molecular and morphological changes. Consequently, the characteristics of a given cell and brain region at a given time offer only a snapshot of organogenesis and brain function, necessitating consistent profiling across development.

The molecular and cellular processes underlying development of the nervous system rely on the diversity of transcripts and their precise spatiotemporal regulation (Bae and Walsh, 2013; Silbereis et al., 2016). Functional genomic analyses of the developing human brain have revealed highly dynamic spatiotemporal patterns of gene expression and epigenetic changes during prenatal and early postnatal development (Kang et al., 2011; Li et al., 2018). In contrast, these metrics are comparatively stable over several decades of adulthood (Colantuoni et al., 2011; Jaffe et al., 2018; Kang et al., 2011; Li et al., 2018; Pletikos et al., 2014). Disruptions of these developmentally dynamic regulatory processes are likely to be important contributors to multiple neurodevelopmental and neuropsychiatric disorders (Birnbaum and Weinberger, 2017; Breen et al., 2016; Geschwind and Flint, 2015; McCarroll and Hyman, 2013; Rosti et al., 2014; Sestan and State, 2018; Turner and Eichler, 2018). In keeping with this expectation, analyses of spatiotemporal expression patterns have implicated mid-fetal brain development as a vulnerable process, and the prefrontal cortex as a vulnerable region, for both autism spectrum disorder (ASD) and schizophrenia (SCZ) risk genes (Chang et al., 2015b; Gulsuner et al., 2013; Li et al., 2018; Network and Pathway Analysis Subgroup of the Psychiatric Genomics Consortium, 2015; Parikshak et al., 2013; Satterstrom et al., 2018; Willsey et al., 2013; Xu et al., 2014). More generally, atypical trajectories of brain maturation likely affecting pathophysiological manifestations have been described in ASD, SCZ, and other neuropsychiatric disorders (Birnbaum and Weinberger, 2017; Courchesne et al., 2007; Ecker et al., 2015; Insel, 2010; Keshavan et al., 2014; Shaw et al., 2010; Tang and Gur, 2018). Given that neuropsychiatric disorders have discrete ages of onset and patterns of progression and may arise due to genetic or environmental insults at various times and places during the life of an individual, there is a clear need to examine gene expression and neuropsychiatric risk across the span of human brain development.

In addition to spatiotemporal variation, genetic sequence variants are also associated with variation in gene expression levels, which can thereby contribute to differences in brain structure, function, and behavior between individuals (Elliott et al., 2018). Several laboratories and consortia have systematically identified such expression quantitative trait loci (eQTLs) through paired analysis of single nucleotide polymorphism (SNP) genotyping and gene expression data in numerous tissues, including the brain (BrainSeq: A. Human Brain Genomics Consortium, 2015; Dobbyn et al., 2018; Fromer et al., 2016; Gibbs et al., 2010; Heinzen et al., 2008; Jaffe et al., 2018; Liu et al., 2010; Myers et al., 2007; Psych et al., 2015; The GTEx Consortium, 2015; Wang et al., 2018). However, there are relatively few such studies of the developing human brain

(Colantuoni et al., 2011; Jaffe et al., 2018; Kang et al., 2011; O'Brien et al., 2018) and most of these, as a result of limitations forced by then-contemporary technologies or limited tissue availability, were unable to assess the full scope of genetic variants and/or gene expression dynamics across the whole of brain development (Colantuoni et al., 2011; Jaffe et al., 2018; Li et al., 2018; O'Brien et al., 2018). Thus, developmentally regulated and rare variant eQTLs are sparsely represented in the current catalog of human brain eQTLs, highlighting the need for additional resources. Such eQTL catalogues offer the potential to gain insight into the functional consequences of the hundreds of coding and noncoding genetic loci that have been associated with neuropsychiatric disorders, including developmental delay (DD), ASD, SCZ, major depressive disorder (MDD), and Alzheimer's disease (AD) (Deciphering Developmental Disorders Study, 2017; Grove et al., 2017; Kosmicki et al., 2016; Sanders et al., 2015; Sanders et al., 2017; Satterstrom et al., 2018; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). Many of these variants are presumed to act by altering gene expression, including noncoding SNPs that alter expression regulation, copy number variants (CNVs) that delete one copy of a gene, and premature stop codon variants that are predicted to induce nonsense-mediated decay (NMD) of transcripts (Nagy and Maquat, 1998; Rivas et al., 2015).

To help fill this gap, we have generated BrainVar, a unique resource of whole-genome sequencing (WGS) paired with bulk tissue RNA sequencing (RNA-seq) analyses of 176 samples from the human dorsolateral prefrontal cortex (DLPFC) across the span of its development, starting from 6 postconceptional weeks (PCW) to young adulthood (20 years). We focused our analyses on DLPFC due to its importance in higher order cognition (Silbereis et al., 2016) and the co-expression of many ASD and SCZ risk genes observed in this region during mid-fetal development (Gulsuner et al., 2013; Li et al., 2018; Network and Pathway Analysis Subgroup of the Psychiatric Genomics Consortium, 2015; Parikshak et al., 2013; Willsey et al., 2013). The single-base resolution captured by WGS across almost the entire genome allowed us to go beyond common SNPs captured by genotyping arrays used in previous eQTL studies of the human brain and detect other important classes of genetic variants, including those which span multiple basepairs (insertion/deletions (indels), copy number variants (CNVs)) or those that are rare or even unique to a person's genome. Combined with high-depth paired-end RNA-seq in the same tissue samples, this approach allows hitherto unattainable analyses of the developmental transcriptome and the impact of diverse classes of genetic variants identified by WGS.

Our analysis discovers gene-specific developmental trajectories, identifying the genes that underlie the late-fetal transition, a dramatic shift in gene expression between mid-fetal development and infancy (Li et al., 2018). We characterize distinct groupings of genes as defined by their developmental expression trajectories and their co-expression patterns, and then leverage WGS to identify common variants implicated in cis-regulation of gene expression across development or during specific stages, and rare protein-truncating variants with corresponding changes in allele-specific and absolute gene expression. Finally, we evaluate relationships between these variants and genes associated with genome-wide significant loci in ten central nervous system (CNS) traits and disorders to identify unique temporal, cell type, and eQTL signatures.

Results

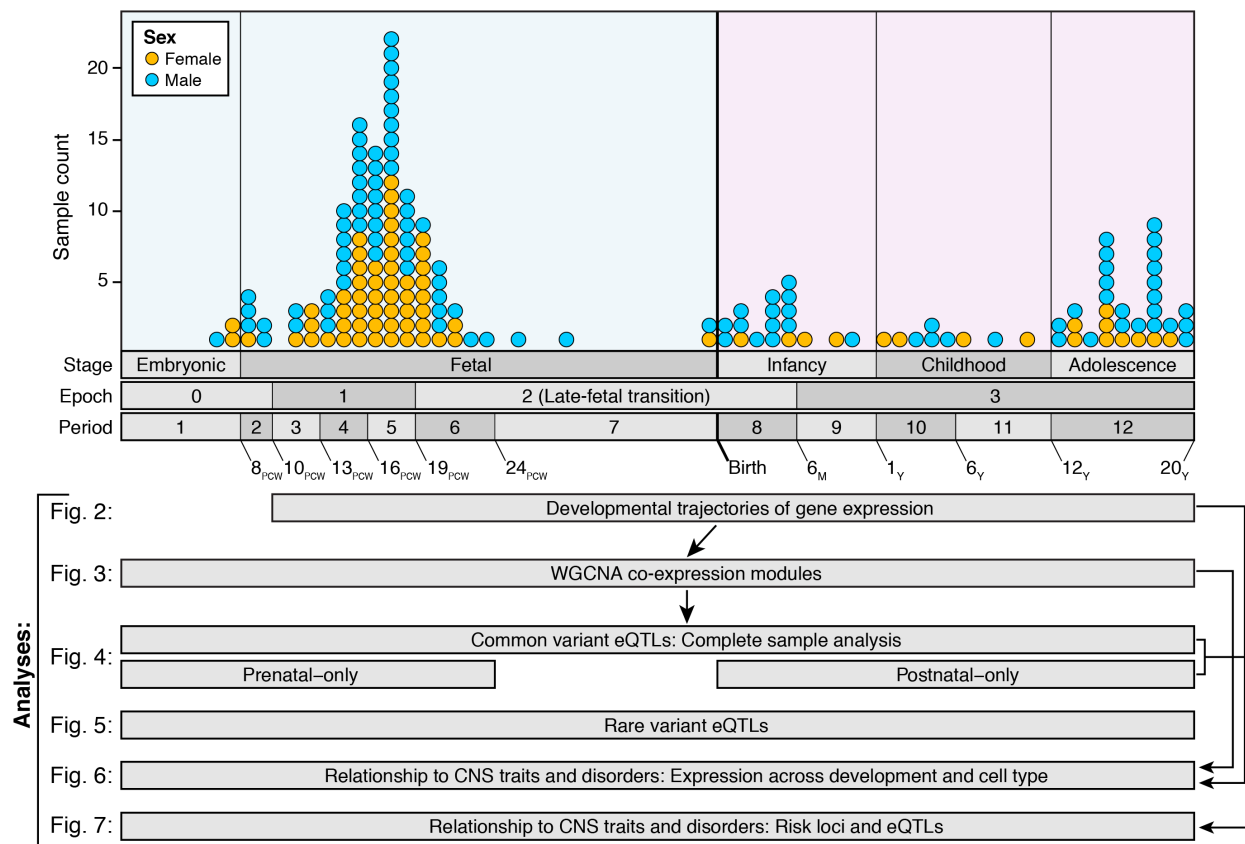


Figure 1: Overview of the dataset and the analysis. 176 samples from the dorsolateral prefrontal cortex (DLPFC) of the developing human brain were processed to generate RNA-seq gene expression data and WGS data (top). The distribution of the samples is shown by sex (color) and developmental stage (x-axis). Periods were defined previously (Kang et al., 2011) and epochs are defined as a superset of periods, based on principal component analysis of these RNA-seq data. The analyses conducted using these data are shown below, and the width of each box corresponds to the age of the samples included in each analysis.

Description of the cohort and data generation

To characterize gene expression across prenatal and postnatal development of human DLPFC and identify genetic variants associated with expression changes, *post mortem* tissue was obtained from 176 de-identified, clinically unremarkable donors (genotypic sex: 104 male, 72 female) without known neuropsychiatric disorders or large-scale genomic abnormalities, ranging between 6 PCW and 20 years of age (Figure 1 and Table S1). In keeping with prior analyses (Kang et al., 2011), we assign these samples to 12 developmental stages, which we group into four developmental epochs (Figures 1 and 2). Gene expression data were generated using RNA-seq from tissue dissected from the frontal cerebral wall from donors aged <10 PCW or the DLPFC (corresponding mainly to Brodmann area 46) from donors at later developmental stages. WGS data (31.5x median coverage) were generated simultaneously from DNA isolated from the same individuals.

Data processing

RNA-seq reads were aligned to GRCh38 using the STAR aligner (Dobin et al., 2013) with genes defined by Gencode v21 (Harrow et al., 2012). Read counts per gene were calculated using HTSeq (Anders et al., 2015) and converted to counts per million (CPM), which were logarithmically scaled to base 2 ($\log_2\text{CPM}$). Of the 60,155 genes assessed, 23,782 were defined as being cortically-expressed, based on $\text{CPM} \geq 1$ in at least 50% of samples of either sex in at least one of the 12 developmental periods. We restricted further analysis to these 23,782 genes, for which median $\log_2\text{CPM}$ ranges from -5.9 to 12.3, with a median of 2.3. Of these, 16,296 (68.5%) encode proteins, while 7,486 (31.5%) are noncoding, including lncRNA (12.6%) and antisense (9.2%) genes (Table S2). On average, noncoding genes were expressed at lower levels than protein-coding genes ($3.1 \log_2\text{CPM}$ vs. $-0.9 \log_2\text{CPM}$; $p < 2 \times 10^{-16}$, Wilcoxon test), especially lncRNAs ($-1.5 \log_2\text{CPM}$). However, not all noncoding genes follow this pattern, with 1,081 noncoding genes expressed above the median for all cortically-expressed genes and the nuclear lncRNA *MALAT1* being the fifth most highly expressed gene.

Although the vast majority of samples were newly ascertained for this study (162/176, 92%), 14 samples were previously analyzed by the BrainSpan Consortium (Li et al., 2018), but resequenced in BrainVar. For these 14 samples with replicate RNA-seq data in BrainSpan DLPFC samples, we compared the correlation of gene expression values with the current data. Despite differences in library preparation, with BrainSpan using poly-A priming compared to TruSeq random priming in BrainVar, the per sample correlations range from 0.76 to 0.91, while the median gene to gene correlation was 0.54 (Figure S1). The first principal component of gene expression in both data sets correlated strongly with developmental age (Figure S1).

WGS data were processed using a pipeline functionally equivalent to major WGS datasets (Regier et al., 2018), using BWA (Li and Durbin, 2009) for alignment to GRCh38, Picard (<https://github.com/broadinstitute/picard/>) to flag duplicate reads, and GATK v3.8 Haplotype Caller (McKenna et al., 2010) to call variants. Ancestry was estimated based on principal component analysis of 118,849 common SNPs and indels ($\geq 5\%$ minor allele frequency), alongside corresponding WGS data for 3,804 individuals of known ancestry (An et al., 2018). The results correlated strongly with self-reported ancestry and identified clusters corresponding to African-American ancestry (42 samples, 24% of cohort), European ancestry (82 sample, 47%), and Asian ancestry (4 samples, 2%), while 48 samples (27%) were outside of these clusters and enriched for individuals who identified as Hispanic, Alaskan native, or mixed ancestry (Table S1). Ancestry groups were differentially represented across developmental periods ($F=3.4$, $df=5$, $p=0.006$, ANOVA), with posthoc analysis showing a greater representation of African American samples in later developmental periods compared with samples of Hispanic, Alaskan native, or mixed ancestry (median age across cohort: one postnatal month vs. 17 PCW respectively; $p_{\text{adj}} = 0.02$, TukeyHSD).

To confirm concordant sample identity between the WGS and RNA-seq samples, we determined genetic sex, based on expression of *XIST* and chromosome Y genes, and sample genotype for 289 high fixation index (F_{ST}) protein-coding SNPs (Table S1). All 176 samples had concordant sex and genotypes between the expression and sequencing data.

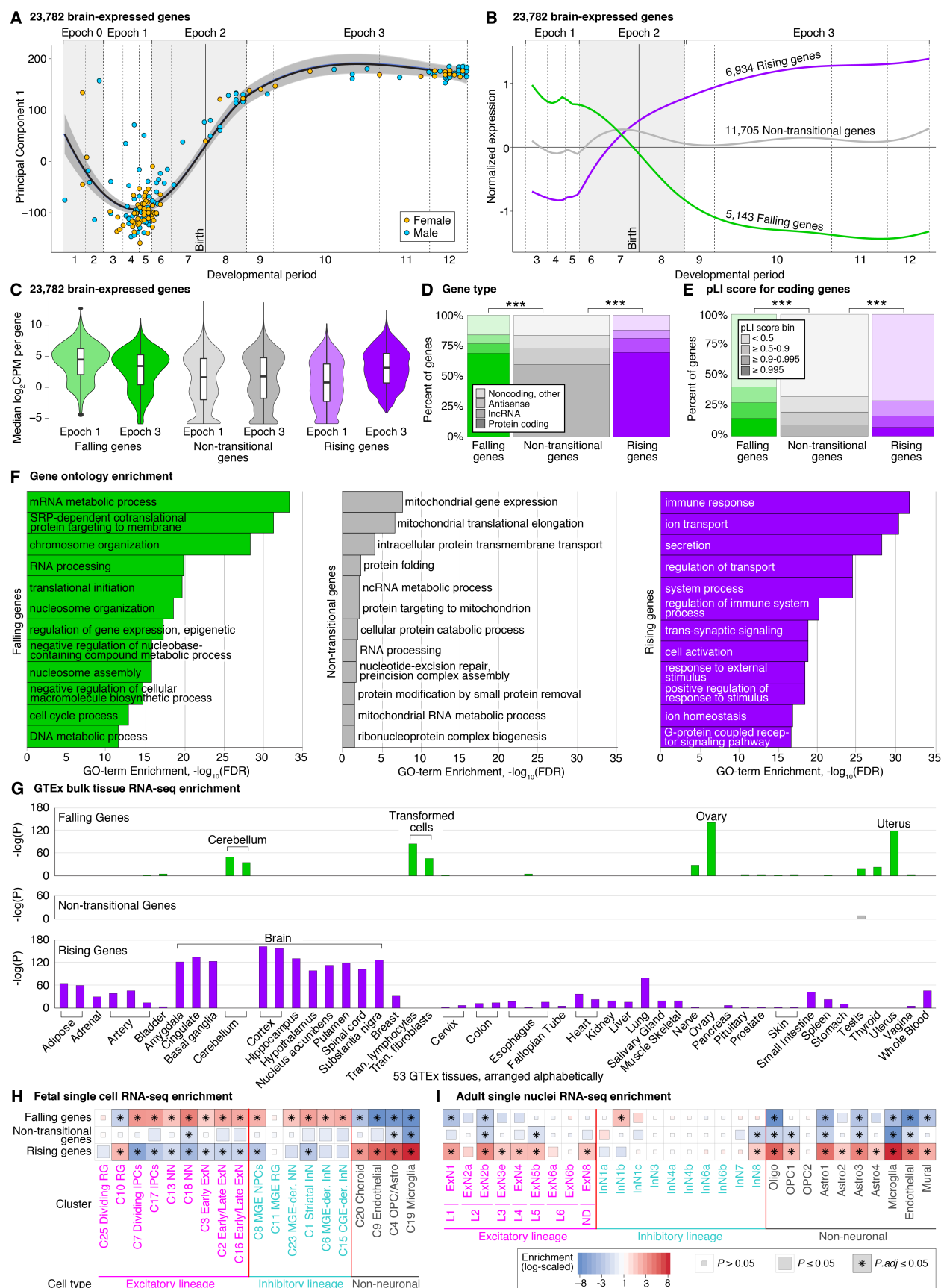


Figure 2. Temporal trajectories of gene expression in the human dorsolateral prefrontal cortex. **A)** Gene expression $\log_2\text{CPM}$ for each sample was used to calculate principal components (Figure S2). The first principal component (PC1) explains 42% of the variance between samples and 81% of variance in PC1 is explained by developmental stage (Figure S2). The changes in PC1 over time were used to define four “epochs” of gene expression. Dotted lines represent the boundaries of the indicated developmental period as defined previously (Kang et al., 2011) **B)** Trajectory analysis was used to identify sets of genes with similar developmental profiles. Three groups of genes were identified and distinguished by their behavior across the late-fetal transition in epoch 2 (Table S2). For each group, the expression over time, normalized by the interquartile range and smoothed using locally weighted smoothing (LOESS) is shown. These three groups are further characterized by plotting: **C)** The median $\log_2\text{CPM}$ across all samples in epoch 1 (E1) and epoch 3 (E3). **D)** The relative proportion of Gencode coding and noncoding genes. **E)** The distribution of probability loss-of-function intolerance (pLI) scores for protein-coding genes (Lek et al., 2016). **F)** The twelve gene ontology biological processes with the lowest FDR values. **G)** Enrichment in the most tissue-specific genes from the 53 tissues with bulk tissue RNA-seq data from GTEx. **H)** Enrichment of the 200 most cell type cluster-specific genes from 19 clusters defined by scRNA-seq data in the developing human cortex (Nowakowski et al., 2017). **I)** Enrichment of the 200 most cell type cluster-specific genes in 29 clusters defined by snRNA-seq data in the adult human DLPFC (Li et al., 2018). Abbreviations: CPM: Counts per million; RG: radial glia; IPC: intermediate progenitor cell; NN: newborn neuron; ExN: excitatory neuron; InN: inhibitory neuron; CGE: caudal ganglionic eminence; MGE: medial ganglionic eminence; OPC: oligodendrocyte precursor cell. Statistical analyses: A: Principal component analysis; B: Longitudinal mixture model with Gaussian noise; C: Wilcoxon two-sided; D: FET; E: gProfiler (Reimand et al., 2011); F: t-test with Benjamini-Hochberg correction; G: FET with Benjamini-Hochberg correction; H: FET, corrected for 57 comparisons; I: FET, corrected for 87 comparisons.

Temporal dynamics of gene expression

Across the 176 samples, the first principal component explains 42% of the variance in gene expression and is highly correlated with developmental age (partial R^2 for Period with PC1 = 0.88; Figure 2A and S2); the same pattern is observed for the 162 brain samples that are new to this analysis. This result is in keeping with prior analyses, highlighting the dramatic shift in gene expression across developmental age (Kang et al., 2011; Li et al., 2018). Similar to earlier reports, we also identified a coordinated shift in gene expression across multiple genes between mid-fetal development and early infancy (periods 6 to 8; epoch 2; Figure 2A), which we will refer to as the “late-fetal transition.” The data suggest a similar coordinated shift in embryonic and early fetal development (periods 1-2; epoch 0; Figure 2A). However, due to the sparse sampling (nine individuals), we did not characterize this “early-fetal transition” in detail. Using trajectory analysis, and excluding the nine epoch 0 individuals, we identified three distinct gene expression trajectories based on the presence or absence of persistent, progressive, and statistically significant expression variance across this late-fetal transition (Figure 2B). The groups are composed of 6,934 “Rising” genes, with expression rising through late-fetal development and infancy leading to higher postnatal expression; 5,143 “Falling” genes, with higher prenatal expression that then falls across late-fetal development leading to lower postnatal expression; and 11,705 “Non-transitional” genes that exhibit no statistically significant change in gene expression between mid-fetal development and childhood, adolescence, and/or adulthood.

Emphasizing the magnitude of this transition, the first principal component of 11,705 genes with a Non-transitional trajectory explains only 18.4% of the variance in gene expression and is weakly correlated with developmental age (partial R^2 for Period with PC1 = 0.3; Figure S2).

Characteristics of transitional and non-transitional genes

To further characterize the nature of the late-fetal transition, we assessed the characteristics of the genes within the Rising, Falling, and Non-transitional groups. On average, the Falling genes were expressed at a higher level than both the Rising and Non-transitional groups throughout development, even in epoch 3, when their expression was lowest (3.4 vs. 3.2 median $\log_2$ CPM for Falling vs. Rising genes in epoch 3; $p=0.003$, Wilcoxon test; Figure 2C). Falling genes had a higher proportion of coding genes than the Non-transitional genes (73.9 vs. 63.3%, $p<1\times 10^{-40}$, chi-squared; Figure 2D) and were also enriched for genes with high probability Loss-of-function Intolerant scores ($pLI \geq 0.995$; 10.1% of Falling genes, 1.6-fold expectation; Figure 2E), which identify genes with fewer protein truncating variants than expected in exome sequencing of ~60,000 population samples (Lek et al., 2016), suggesting haploinsufficiency resulting in selective pressure.

Gene ontology analysis of the Falling genes demonstrated enrichment for biological processes relating to development and anabolism, including cell division, gene expression regulation, RNA synthesis and processing, and protein translation (Figure 2F). Comparison with genes enriched in specific adult tissues from the Genotype-Tissue Expression (GTEx) project dataset (The GTEx Consortium, 2015) showed enrichment for several non-cortical tissues (Figure 2G), driven by genes with ontology terms relating to RNA transcription in the cerebellum, ovaries, and uterus and cell division in transformed cells (Table S2). Compared to H3K27ac DLPFC peaks from BrainSpan (Li et al., 2018), Falling genes were enriched in genes associated with fetal-biased H3K27ac peaks (Figure S2). To further understand how these temporal changes relate to cell differentiation and cell type, we assessed the overlap of our bulk tissue expression with data from single-cell RNA-seq (scRNA-seq) in prenatal human forebrain (Nowakowski et al., 2017) and single-nucleus RNA-seq (snRNA-seq) in postnatal DLPFC (Li et al., 2018). Using cell clusters previously defined in each of these studies, we applied a Tau metric to define the 200 protein-coding genes most enriched in each cell type cluster. The Falling genes are enriched in multiple neuronal cell type clusters of the excitatory and inhibitory lineages, mainly from the prenatal data (Figure 2H).

The Rising genes had median postnatal expression values that were lower than Falling genes (3.2 vs. 3.4 median $\log_2$ CPM in epoch 3), though a similar proportion of protein-coding genes (73.1 vs. 73.9%). The Rising genes were enriched for adult-biased H3K27ac peaks (Figure S2) but depleted for genes with high pLI scores (5.1% $pLI \geq 0.995$; 0.80-fold expectation), suggesting greater tolerance to variation in expression. Gene ontology terms enriched in the Rising genes related to immune processes and metabolic processes, such as transportation, secretion, and ion homeostasis, with the genes highly enriched in all adult brain tissues in GTEx except the cerebellum. In the single cell data, the Rising genes are enriched in multiple non-neuronal cell types, including microglia, oligodendrocyte precursor cells (OPC), oligodendrocytes, and astrocytes in both the pre- and postnatal data, and multiple classes of excitatory neurons, mainly in the postnatal data (Figure 2I).

The Non-transitional genes had the lowest levels of expression, the lowest proportion of protein-coding genes, and few genes with high pLI score (5.4%, 0.86-fold expectation). They were not enriched for either fetal-biased or adult-biased H3K27ac peaks (Figure S2), mapped to gene ontology terms relating to mitochondria and protein folding and transport and were expressed ubiquitously across GTEx tissues. None of the cell type clusters in pre- or postnatal scRNA-seq data were enriched for Non-transitional genes and, compared to Transitional genes (Falling and Rising), the Non-transitional genes were less likely to be enriched in a cell type cluster (14.3% of Transitional genes vs. 8.3% of Non-transitional genes; odds ratio (OR) = 1.8 (95%CI: 1.7-2.0); $p \leq 2 \times 10^{-16}$ FET).

Co-expression modules in the developing human cortex

To further define the relationships between the 23,782 cortically-expressed genes, we applied weighted gene co-expression network analysis (WGCNA) to define 19 consensus modules that included 10,459 genes (Figures 3A and S3). As expected, genes within each module shared functional roles (Figure 3B), temporal trajectories of gene expression (Figure 3C and 3D), and regulatory transcription factors (Figure S3). In contrast, genes that were not assigned to modules were more likely to be noncoding (OR = 2.5; 95%CI: 2.3-2.6; $p < 2 \times 10^{-16}$, FET), have lower expression (0.74 vs. 3.20 median $\log_2$ CPM; $p < 2 \times 10^{-16}$, Wilcoxon test), and belong to the Non-transitional group (OR = 1.7; 95%CI: 1.6-1.8; $p < 2 \times 10^{-16}$, FET). Module preservation analysis using BrainSpan data (Li et al., 2018) identifies similar co-expression patterns in independent DLPFC samples and across all neocortical samples, but not in non-neocortical regions (Figure 3E). Comparison of the modules to the prenatal and postnatal scRNA-seq data showed enrichment for specific cell type clusters for ten of the modules, with patterns similar to those observed for the temporal trajectories (Figure 3F and 3G).

In keeping with our trajectory analysis (Figure 2), multidimensional scaling analysis of the module eigengenes demonstrated that developmental age accounted for 44.7% and 36.1% of the variance in the first two dimensions. Plotting the relationship of 19 modules between these two dimensions, and considering the developmental trajectories of the genes in each module, we identified five groups of related modules (Figure 3A and 3B). Group 1 modules (M1 Black, M2 Royal-blue, M3 Green-yellow, M4 Yellow) are closely related to the Falling genes (Figure 3D) and have similar characteristics, including enrichment for gene ontology terms related to cell division, protein translation and targeting, and histone modifications, along with enrichment in developing neurons of the excitatory and inhibitory lineages. Of note, the M4 Yellow module is enriched for numerous ontology terms related to neuronal development, contains genes specific to neuronal stem cells (e.g. *NCAM1/ncam*, *PROM1/cd133*), and is highly enriched for genes related to maturing excitatory and inhibitory neurons, while the M2 Royal-blue module, which captures cell cycle, is enriched in neuroprogenitor cells, radial glia, and intermediate progenitor cells. Group 2 modules (M5 Pink, M6 Tan, M7 Salmon) are enriched for factors relating to RNA processing (Figure 3C) and contain targets of numerous transcription factors and microRNAs (Figure S3). Gene expression of group 2 modules increases between periods 10-11 (1-12 years post birth), suggesting a key role for RNA processing during childhood cortical development.

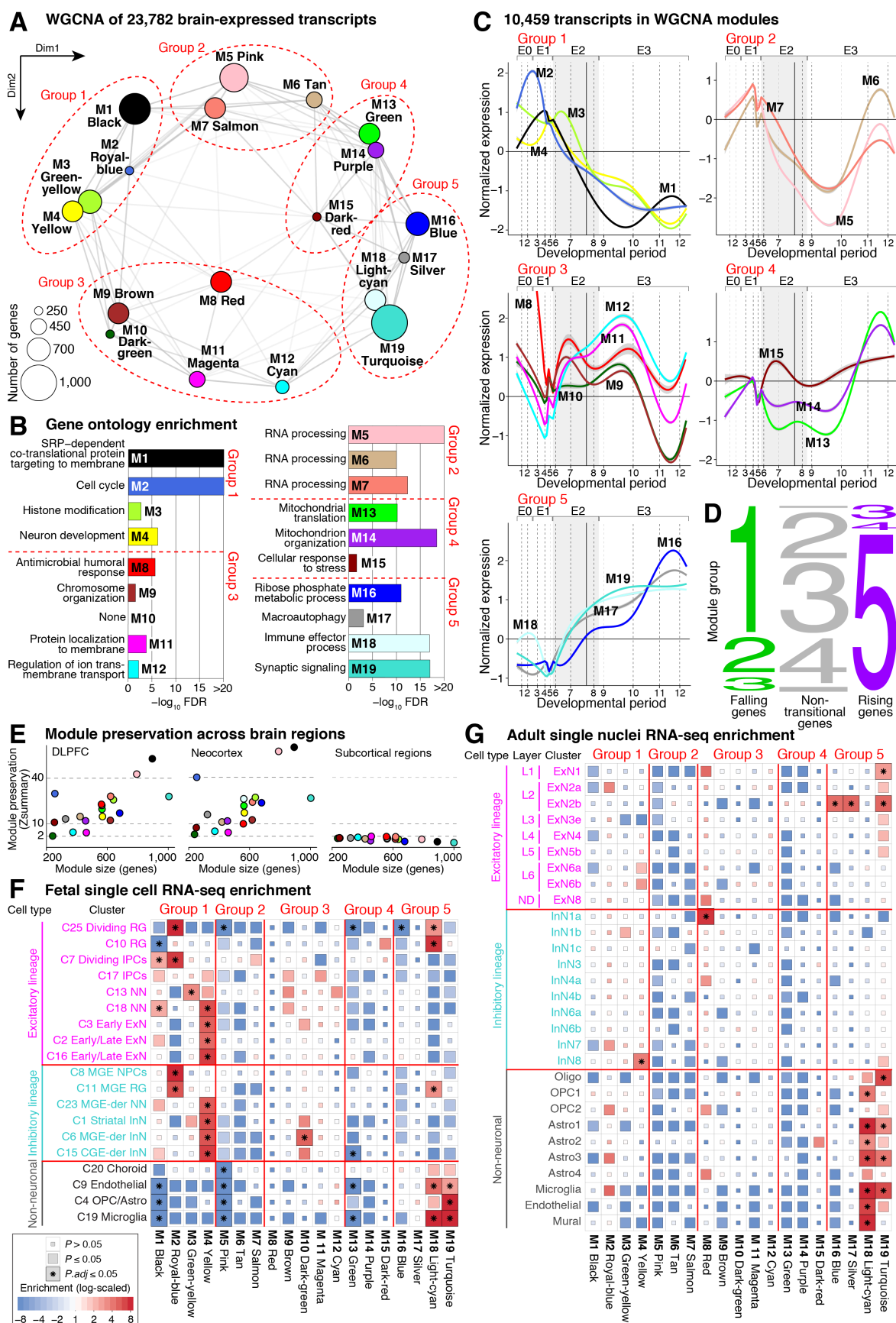


Figure 3. Co-expression modules in the developing human cortex. **A)** WGCNA identified 19 modules comprised of 10,459 genes, shown as colored nodes, from patterns of co-expression for 23,782 expressed genes. The nodes are plotted based on the first two dimensions from multidimensional scaling. The weight of the connecting lines represents the degree of correlation between module eigengenes. **B)** Normalized expression is shown for the 19 modules arranged in five groups based on proximity in 'A' and similar temporal trajectories. **C)** Gene ontology enrichment analysis for each module, showing only biological processes with the lowest FDR; M10 Dark-green, was not enriched for any process after correcting for multiple comparisons. **D)** The relationship between the five groups of co-expression modules (from 'A') and the genes with Falling, Rising, or Non-transitional temporal trajectories (Figure 2). The number represents the WGCNA group, with the height proportional to the number of genes in the group and the width proportional to the number of genes in each temporal trajectory group. Detailed relationships between modules and temporal trajectories are shown in Figure S3. **E)** Module preservation in independent BrainSpan samples (Li et al., 2018) from the same brain region (left), other cortical regions (middle), and five subcortical regions (right). **F)** Enrichment between the 19 modules and the 200 genes most specific to 19 cell type clusters defined by scRNA-seq data in the developing human cortex (Nowakowski et al., 2017). **G)** Enrichment between the 19 modules and the 200 genes most specific to 29 cell type clusters defined by snRNA-seq data in the adult human DLPFC (Li et al., 2018). Abbreviations: FDR: False discovery rate; SRP: Signal Recognition Particle. Statistical analysis: A: WGCNA with the consensus module detection from 100 random resampling; B: FET, corrected for gProfiler gene ontology pathways (10 to 2000 term size); C: gene expression normalized by interquartile range; E: FET, corrected for 361 comparisons; F: FET, corrected for 551 comparisons. G: FET, corrected for 19 comparisons.

Group 3 modules (M8 Red, M9 Brown, M10 Dark-green, M11 Magenta, M12 Cyan) overlap with the Non-transitional gene set, with expression peaking between periods 6-9 (19 PCW to 1-year post birth) before returning to mid-fetal levels. They are enriched for terms related to protein localization, ion transport, synapses, and receptors. The M8 Red module, which is enriched for ontology terms relating to cell fate and morphogenesis, is highly enriched for noncoding genes and has an expression peak in early fetal development, hinting at an early-fetal transition in epoch 0 analogous to the epoch 2 late-fetal transition. Several genes associated with regional patterning in non-cortical tissues, including hindbrain (e.g. *UNCX*, *CCDC140*) and hypothalamus (e.g. *DMBX1*, *SOX14*), are expressed at high levels in this module. Group 4 modules (M13 Green, M14 Purple, M15 Dark-red) are relatively static during development before rising steeply between periods 10 and 11 (1-12 years post birth), mirroring the group 2 modules. They are enriched for multiple ontology terms relating to mitochondria and cellular respiration, possibly reflecting the metabolic activity necessary to support neuronal activity during this postnatal stage (Ashrafi and Ryan, 2017). Finally, the group 5 modules (M16 Blue, M17 Silver, M18 Light-cyan, M19 Turquoise) are closely related to the Rising genes (Figure 2) with expression rising throughout development. Both the M18 Light-cyan and M19 Turquoise modules are strongly enriched in glial and other non-neuronal cell clusters, accordingly, both modules are enriched for ontology terms related to immune responses. The M19 Turquoise module is also enriched in excitatory neurons on the postnatal cortex and ontology terms relating to synaptic signaling and neurotransmitter transport. The other two modules in group 5, M16 Blue and M17 Silver, are enriched for terms relating to numerous metabolic processes, including glycolysis, respiration, and protein catabolism, along with neuronal death and autophagy.

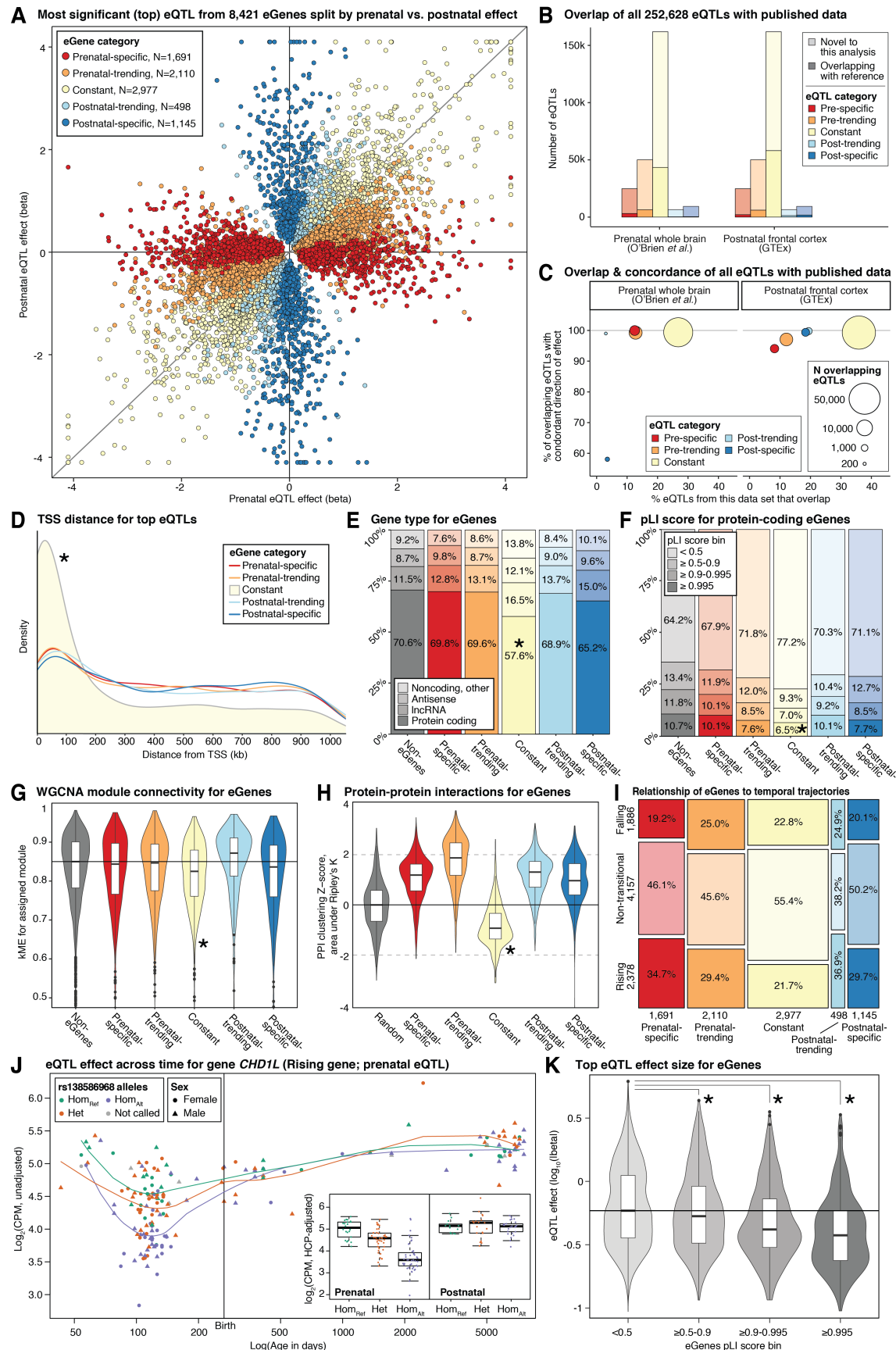


Figure 4. Common variant cis-eQTLs. **A)** Prenatal (x-axis) and postnatal (y-axis) effects for the eQTLs with the smallest p-value for 8,421 eGenes (points). The eQTLs are split into five categories based on temporal specificity using effect size and statistical thresholds; categories are represented by color. **B)** The total number of eQTLs in each temporal category is represented by the height of the bar. The number of eQTLs that match previously identified significant eQTLs in the prenatal whole brain (O'Brien et al., 2018) (left) and postnatal frontal cortex (The GTEx Consortium et al., 2017) (right) is shown by the darker shade. **C)** The percentage overlapping eQTLs between datasets (from 'B') is shown on the x-axis against the percentage of these overlapping eQTLs with concordant direction of effect in both datasets. **D)** Density plot of the distance of top eQTLs per eGene from the TSS by eGene temporal category. Characteristics of non-eGenes and eGenes in each temporal category are shown by plotting the: **E)** proportion of coding and noncoding genes, **F)** proportion of genes with pLI scores in different bins, **G)** centrality to an assigned WGCNA module (kME values, black line is the non-eGene median), and **H)** BioGRID protein-protein interactions (permuted Z-scores from Ripley's K-net function (Cornish and Markowetz, 2014), black line is the non-eGene median). **I)** Mosaic plot of the proportion of genes in each temporal trajectory with eGenes split by temporal category. **J)** Expression data binned by genotype for the top Prenatal-specific eQTL for CHD1L, a gene with a Rising trajectory. Main panel: Gene expression by sample age across development. Lines are loess-smoothed trajectories for expression in samples with each of three genotypes for rs138586968. Inset: Boxplots for prenatal (left) and postnatal (right) samples with each of three rs138586968 genotypes. **K)** Distribution of eQTL effect size for eGenes binned by pLI scores, the black line is the median of the pLI < 0.5 group. Abbreviations: eQTL: expression quantitative trait locus; eGene: gene regulated by an eQTL; TSS: transcription start site; pLI: probability loss-of-function intolerant. Statistical analysis: D, G, H: two-sided Wilcoxon rank sum test for Constant versus other eGenes; E, F: two-sided FET for Constant versus other eGenes; K: two-sided Wilcoxon rank sum test for three top pLI bins versus genes with pLI < 0.5.

Common genetic variants regulating gene expression

We next leveraged WGS data to identify 6,573,196 high-quality SNPs and indels, observed at an allele frequency of at least 5% in both our prenatal (periods 1-6, N=112) and postnatal (periods 8-12, N=60) samples. The four samples from period 7 were excluded due to the sparse representation over this dynamic developmental phase (Figure 2). This common variant set included only variants located outside of low-complexity regions (Li, 2014) that passed variant quality score recalibration, with allelic balance between 0.22-0.78 (SNPs; 0.2-0.8 for indels), Hardy Weinberg equilibrium p-value $\geq 1 \times 10^{-12}$, call rate $\geq 90\%$ across all samples, and that reached variant and genotype thresholds established and validated in a previous analysis of family-based WGS data (Werling et al., 2018).

Linear regression was used to identify common variants within 1Mb of a gene associated with altered levels of gene expression, i.e. cis-eQTLs. Developmental period, sex, and the first five principal components for ancestry were included as covariates. Subsequently, we applied two approaches to control for technical covariates: Hidden Covariates with Prior (HCP) and surrogate variable analysis (SVA). To distinguish temporally specific eQTLs, we performed cis-eQTL analysis for three sets of samples: all 176 samples (complete sample), 112 samples in periods 1-6 (prenatal), and 60 samples in periods 8-12 (postnatal). Within each analysis, results

for all tested variant-gene pairs were corrected for multiple testing using the Benjamini-Hochberg procedure.

Across these three analyses, we identified 252,629 unique eQTLs ($FDR \leq 0.05$) associated with the expression of 8,421 unique genes (eGenes) (Table S4). While HCP-adjusted results were used for our primary analysis, results for 167,665 unique eQTLs identified using SVA are also provided (Table S4). The rate of eGene discovery from each analysis aligns with expectation given sample size (Figure S4).

We observed no significant difference in the effect size between SNVs and indels ($p = 0.55$, two-sided Wilcoxon test) or between insertions and deletions ($p = 0.87$, two-sided Wilcoxon test; Figure S4). Considering functional annotations, we identified greater overlap with H3K27ac peaks derived from the human brain (Li et al., 2018; Reilly et al., 2015) for eQTLs than non-eQTL variants (Figure S4). Analysis with GREGOR (Schmidt et al., 2015) using 18 human frontal lobe derived chromatin states (Roadmap Epigenomics Consortium et al., 2015) showed enrichment of eQTLs near transcription start sites and genic enhancers (Figure S4).

Temporal specificity of eQTLs

Given the dramatic changes in gene expression across development, we leveraged the inclusion of both prenatal and postnatal samples to assess whether eQTLs also varied across development. To delineate the temporal specificity of eQTLs, we categorized eQTLs and their associated eGenes according to the magnitude and significance of their effects in our prenatal, postnatal, and complete analyses. We defined 161,293 (63.8% of 252,629 eQTLs) “Constant” eQTLs as those observed in all three analyses ($FDR \leq 0.05$ in the complete sample, uncorrected $p \leq 0.05$ in both the prenatal and postnatal samples) with the same direction of effect. The 19,142 (7.6%) “Prenatal-specific” eQTLs were observed in the prenatal data ($FDR \leq 0.05$), but not the postnatal data ($p > 0.05$) and had a significant difference in effect size between these two cohorts (FDR-corrected Z-test of regression coefficients $p \leq 0.05$). Analogous criteria were used to define 7,973 (3.2%) “Postnatal-specific” eQTLs. For the remaining eQTLs, the absolute value of the effect size in the prenatal and postnatal data sets was used to define 55,659 (22.0%) “Prenatal-trending” eQTLs and 7,931 (3.1%) “Postnatal-trending” eQTLs (Figure 4A).

To assess concordance with comparable datasets, we assessed FDR-significant eQTLs that overlapped between our data and prenatal whole brain (O'Brien et al., 2018) or postnatal frontal cortex (The GTEx Consortium et al., 2017). The percentage of overlapping eQTLs was highest for the Constant group (26.8% with prenatal brain and 35.9% with postnatal frontal cortex; Figure 4B), while 73.2% and 64.1% (118,543 and 103,868 constant eQTLs), respectively, were novel to this analysis. Of the Constant eQTLs that did overlap, the direction of effect was concordant for 99.4% and 99.3% of eQTLs respectively, despite differing cohorts, library preparation methods, genome-build, the use of WGS, and analysis pipelines. As expected, temporally specific eQTLs showed higher overlap and concordance in samples of an equivalent developmental age (e.g. Prenatal-specific eQTLs in prenatal whole brain) compared to samples of a different developmental age (e.g. Postnatal-specific eQTLs in prenatal whole brain; Figure 4C). A comparison of the prenatal whole brain and postnatal cortex reference data sets to each other demonstrated comparable or a lower percentage of overlapping eQTLs (29.7% of prenatal brain eQTLs in postnatal frontal cortex; 13.9% of postnatal frontal cortex eQTLs in prenatal brain)

than we attain for Constant eQTLs, with only 93.9% concordance in the direction of effect (Figure S4).

Characteristics of genes influenced by eQTLs

More than one eQTL was identified for 5,538 of the 8,421 eGenes (65.8%), often with similar direction and magnitude of effect due to linkage disequilibrium. We therefore identified the top eQTL for each eGene, defined as the lowest FDR-significant p-value in any of the three sample sets (Table S4). By allowing the eGene to inherit the temporal classification of this top eQTL, we categorized 1,691 eGenes as Prenatal-specific, 2,110 as Prenatal-trending, 2,977 as Constant, 498 as Postnatal-trending, and 1,145 as Postnatal-specific (Figure 4A). The temporal-specific category of the top eQTL matched the majority of eQTLs in the gene for 7,425 (88.2%) of eGenes (Fig S4), with a further 688 (8.2%) of genes shifting by no more than one category (e.g. Prenatal-specific to Prenatal-trending, Postnatal-trending to Constant). While 63.8% of all eQTLs are Constant, only 3,804 (45.2%) eGenes have any Constant eQTLs. This difference is because many Constant eQTLs have high rates of linkage disequilibrium and regulate the same gene.

eQTLs are often observed in close proximity to the Transcription Start Site (TSS) of the associated gene (Fromer et al., 2016; The GTEx Consortium et al., 2017). This pattern was evident for the top eQTLs of the 2,977 Constant eGenes with a median distance of 92,223bp and both statistical significance and effect size of the eQTL increasing with proximity to the TSS ($p = 6.6 \times 10^{-66}$ for $-\log_{10}(P)$, $p = 1.0 \times 10^{-12}$ for $|\beta|$, linear regression). In contrast, the median distance for the top eQTLs of the 5,444 eGenes with a degree of temporal specificity was 402,591bp ($p = 1.3 \times 10^{-148}$ two-sided Wilcoxon rank sum test compared to Constant eGenes, Figure 4D) with statistical significance, but not the effect size, of the eQTL increasing with proximity to the TSS ($p = 2.8 \times 10^{-41}$ for $-\log_{10}(P)$, $p = 0.25$ for $|\beta|$, linear regressions for distance from TSS).

This distinction between Constant eGenes and those with a degree of temporal specificity was also observed across multiple other metrics. Compared to other eGenes, Constant eGenes include fewer protein-coding genes (OR = 0.62 (95%CI: 0.57-0.68), $p = 6.5 \times 10^{-24}$, two-sided Fisher's exact test (FET); Figure 4E). Of the protein-coding genes, Constant eGenes also include fewer genes with high pLI scores than other eGenes (≥ 0.995 ; OR = 0.74 (95%CI: 0.58-0.93), $p = 0.01$, two-sided FET; Figure 4F). Considering the WGCNA results, Constant eGenes are less likely to be assigned to a module (OR = 0.66 (95%CI: 0.61-0.73), $p = 3.0 \times 10^{-18}$, two-sided FET) and, for those that are assigned a module, show weaker co-expression with the module eigengene (0.83 vs. 0.85 median kME, $p = 2.5 \times 10^{-9}$, two-sided Wilcoxon rank sum test; Figure 4G). Constant eGenes also show weaker connectivity to PPI networks (median Z-score -0.914 for Constant vs. 1.294 for other eGenes, $p = \leq 2 \times 10^{-16}$, two-sided Wilcoxon rank sum test; Figure 4H). Across these metrics, eGenes with a degree of temporal specificity tend to be more similar to non-eGenes than the Constant eGenes.

Temporally specific eQTLs could be a consequence of higher expression yielding better power to detect variation between samples at one developmental stage; this explanation would predict that Prenatal-specific eGenes would be enriched for Falling genes and Postnatal-specific eGenes would be enriched for Rising genes. We do not observe this effect (Figure 4I); instead, the Prenatal-specific eGenes are enriched for Rising genes (OR = 1.5 (95%CI: 1.3-1.6), $p =$

1.1×10^{-10} , two-sided FET) and depleted for Falling genes (OR = 0.79 (95%CI: 0.69-0.9), $p = 4.2 \times 10^{-4}$, two-sided FET), while the distribution of Postnatal-specific eGenes across trajectories does not differ significantly from expectation. The Rising gene *CHD1L* and Prenatal-specific eQTL rs138586968 provides a representative example of this pattern (Figure 4J).

Comparing eGenes to WGCNA modules (Figure S4), we find that Constant eGenes are enriched for four of the five Group 3 modules, with significant enrichment for the M9 Brown module (OR = 2.1 (95%CI: 1.6-2.8), $p = 1.46 \times 10^{-6}$, two-sided FET); Constant eGenes are also significantly depleted from the M19 Turquoise module (OR = 0.38 (95%CI: 0.27-0.53), $p = 1.1 \times 10^{-8}$, two-sided FET) and the M8 Red module (OR = 0.47 (95%CI: 0.32-0.69), $p = 2.83 \times 10^{-3}$, two-sided FET). Postnatal-specific eGenes are enriched for the M8 Red module (OR = 2.5 (95%CI: 1.8-3.5), $p = 1.77 \times 10^{-5}$, two-sided FET; Figure S4), which is characterized by high expression in epoch 0 and low expression thereafter (Figure 3; Table S2), and Prenatal-specific eGenes are depleted for the M3 Green-yellow module (OR = 0.51 (95%CI: 0.33-0.76), $p = 0.04$, two-sided FET), which has high prenatal and lower postnatal expression.

The lack of enrichment for Prenatal-specific eGenes in the Group 1 modules and the Falling genes or Postnatal-specific eGenes in the Group 5 modules and the Rising genes raises the possibility that natural selection constrains the variability of gene expression at developmental stages where the genes perform critical functions, necessitating high expression. We reasoned that selective pressure would similarly constrain the variability in gene expression of eQTLs near dosage sensitive genes. Accordingly, we observed an inverse relationship between eQTL effect size and pLI score (0.59 vs. 0.38 median $|\beta|$ for genes with $pLI < 0.5$ vs. $pLI \geq 0.995$; $p = 4.0 \times 10^{-36}$; two-sided Wilcoxon rank sum test; Figure 4K).

Allele specific expression and eQTLs for rare variants

Along with identifying common variant eQTLs, WGS data raises the possibility of detecting rare variants that alter gene expression. However, at allele frequencies below 5%, rare variants would be present in eight or fewer samples in our cohort of 176, limiting the power for detection in a genome-wide analysis. We therefore focused on rare protein truncating variants (PTVs) with a premature stop codon and exonic deletions in cortically expressed protein-coding genes predicted to halve gene expression. To identify the rarest variants, which have the potential to mediate the highest effects, we defined rare for PTVs as never observed in 15,708 individuals with WGS in the Genome Aggregation Database (GnomAD; (Lek et al., 2016) or in 3,804 unrelated samples with WGS data from the Simons Simplex Collection (SSC) and, for deletions, never observed in 9,097 samples from SSC families (An et al., 2018; Werling et al., 2018).

We expect a PTV to induce nonsense-mediate decay (NMD) of the mRNA, resulting in allele specific expression, i.e. the absence of reads containing the variant in RNA-seq data (Nagy and Maquat, 1998; Rivas et al., 2015). We identified 88 rare PTVs in Gencode-defined protein-coding genes with at least five RNA-seq reads overlaying the variant and classified them into three categories: 22 “Escaping NMD” variants in the last coding exon or 50bp upstream of the last splice site (Nagy and Maquat, 1998); 11 “Possible NMD” variants in exons that are not expressed in all isoforms or in single-exon genes; and “Probable NMD” for the remaining 55. For the Probable NMD group, 45 of the 55 PTVs (81.8%) had fewer than 50% of reads that contained the variant ($p = 1 \times 10^{-6}$, binomial exact test; Figure 5A), with a mean of 35.3% (95% CI: 30.4 to

40.2%), which was lower than the mean of 48.8% (95% CI: 44.2 to 53.5%; $p=0.001$, Wilcoxon test) in the Escaping NMD group, in which 11 of 22 PTVs (50.0%) had fewer than 50% variant reads ($p=0.58$, binomial exact test). The Possible NMD group is similar to the Probable NMD group (10 out of 11 had below 50% variant reads, $p=0.006$, binomial exact test; mean of 31.3%, 95% CI: 27.9 to 34.6%).

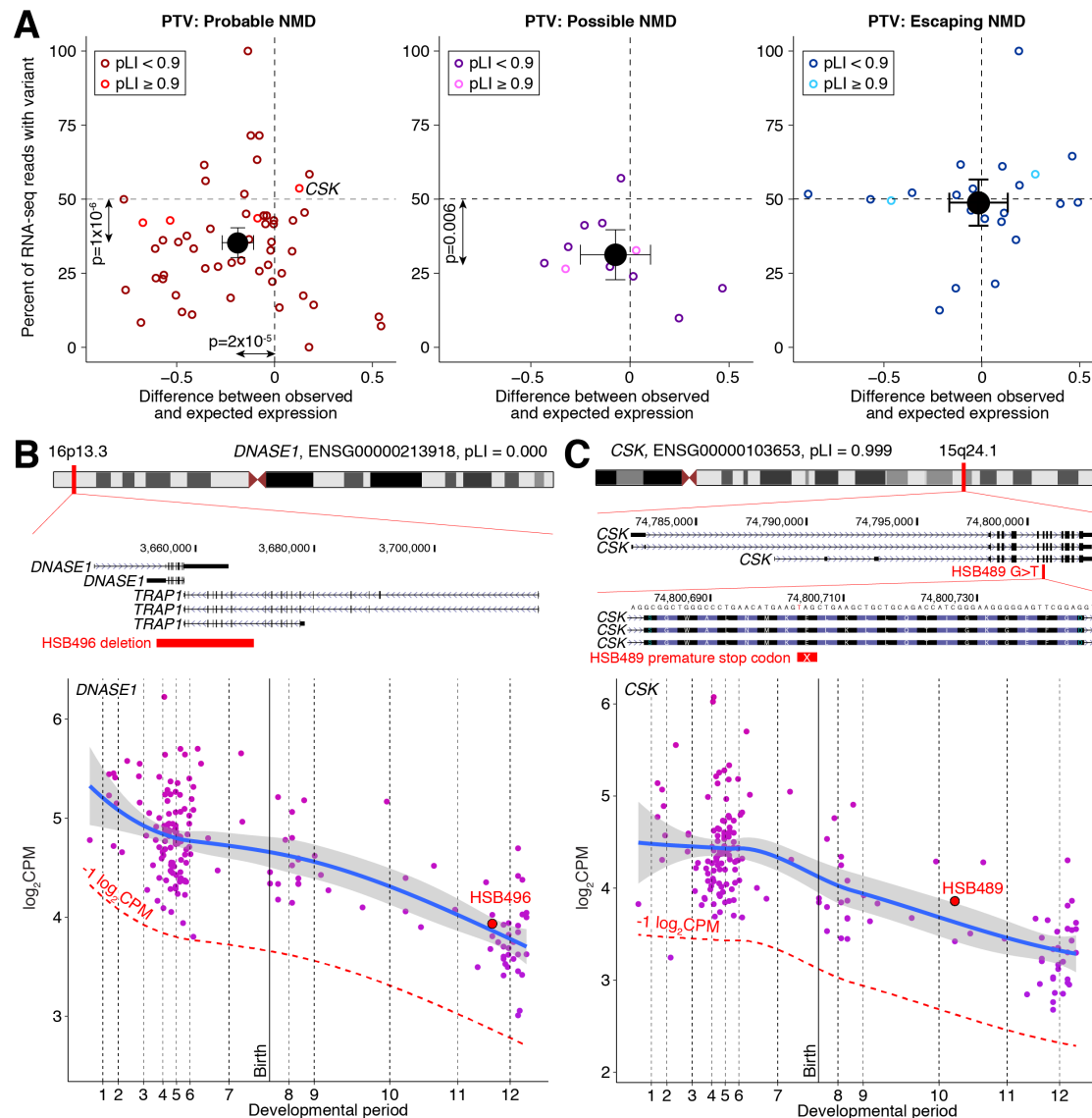


Figure 5. Rare variant eQTLs. **A)** For 55 rare PTVs (red points) predicted to undergo NMD (Probable NMD) with at least 5 RNA-seq reads at the site of the variant, the percentage of RNA-seq reads containing the variant (y-axis) is plotted against the difference between the observed and imputed log₂CPM in the individual with the PTV (x-axis). The black point shows the centroid and 95% confidence intervals across all 55 PTVs. This plot is repeated for 12 “Possible NMD” PTVs in single exon genes or with the PTV-containing exon varying between isoforms (middle) and 22 “Escaping NMD” PTVs in the last exon or 50bp upstream of the last splice site (right) (Nagy and Maquat, 1998; Rivas et al., 2015). **B)** A deletion of all protein-coding exons in the

Deoxyribonuclease 1 (DNASE1) gene and the first seven protein-coding exons of TNF receptor associated protein 1 (TRAP1) in sample HSB496 has no appreciable effect on the expression level of DNASE1 (shown) or TRAP1 (Figure S5). C) A variant causing a premature stop codon in a constitutive exon in the C-Terminal Src Kinase (CSK) gene in sample HSB489 has no appreciable effect on expression level or allele specific expression (53.7% variant reads), despite being predicted to cause NMD. Abbreviations: PTV: Protein truncating variant; NMD: Nonsense-mediated decay; pLI: probability loss-of-function intolerant. Statistical analysis: A: binomial exact test.

Alongside allele specific expression, we also expect NMD to induce a 50% reduction in gene expression, equivalent to $-1 \log_2\text{CPM}$. To evaluate this expectation, we assessed the difference between observed and expected expression, with the later estimated by imputation accounting for variation between samples and across developmental periods. In the Probable NMD group, expression was lower than expectation for 43 of the 55 PTVs ($p=2 \times 10^{-5}$, binomial exact test), with a mean of $-0.19 \log_2\text{CPM}$ (95%CI: -0.11 to -0.27), equivalent to a 12.1% (95%CI: 7.2 to 16.9%) reduction in expression. This was also lower than the mean of $-0.02 \log_2\text{CPM}$ observed in the Escaping NMD group ($p=0.02$, Wilcoxon test). For genes with lower median $\log_2\text{CPMs}$, and consequently fewer reads from which to estimate expression, we observed greater variability in gene expression (Figure S5). Of note, none of the Probable NMD PTVs resulted in the expected magnitude of decreased expression, with the largest difference being $-0.77 \log_2\text{CPM}$, equivalent to a 41% reduction. In contrast, we observed numerous examples of variants predicted to decrease expression that did not, including a heterozygous deletion of all protein-coding exons in the *DNASE1* gene ($p\text{LI} = 0.000$; Figure 5B, S5) and a premature stop codon in the haploinsufficient *CSK* gene ($p\text{LI} = 0.999$; Figure 5C).

Temporal trajectories in human traits and disorders

To leverage this dataset to provide insight into the biology underlying CNS traits and disorders, we assessed the enrichment of gene lists based on developmental trajectory, module, and cell type for genes associated with human traits and disorders. We selected ten datasets that applied genome-wide statistical thresholds to identify multiple associated variants. For DD, ASD, and epilepsy with developmental delay (EPDD) we used gene lists derived from rare and *de novo* variants identified with exome sequencing (Deciphering Developmental Disorders Study, 2017; Heyne et al., 2018; Satterstrom et al., 2018). For educational attainment (EA), attention deficit/hyperactivity disorder (ADHD), SCZ, MDD, multiple sclerosis (MS), Parkinson's Disease (PD), and AD we used genes within 10kb of common variants detected with genotyping arrays in genome-wide association studies (GWAS) (Chang et al., 2017; Demontis et al., 2019; International Multiple Sclerosis Genetics Consortium et al., 2013; Lambert et al., 2013; Lee et al., 2018; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Wray et al., 2018). For our analyses, we excluded genes within the Major Histocompatibility Complex on chromosome 6, due to the complicated nature of this region (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014).

To assess temporal relationships for CNS traits and disorders we plotted the normalized gene expression across development, ordering the disorders by the mean period of maximal expression for each gene (Figure 6A), and assessed the enrichment for the transitional gene groups (Figure 6B). In general, early-onset traits and disorders, including DD, EA, and ASD,

have higher prenatal than postnatal expression and are enriched for Falling genes. In contrast, later-onset disorders, including PD and AD, have higher postnatal expression and are enriched for Rising genes. EPDD would be expected to group with DD if there were a simple relationship between maximal gene expression and the onset of a disorder, however it is ranked last on this list, after PD and AD. We note that, while GWAS loci are genome-wide significant, the genes associated with these loci are likely to contain false positives, which may obscure clearer temporal relationships. Furthermore, critical periods of neurological vulnerability may precede symptom onset.

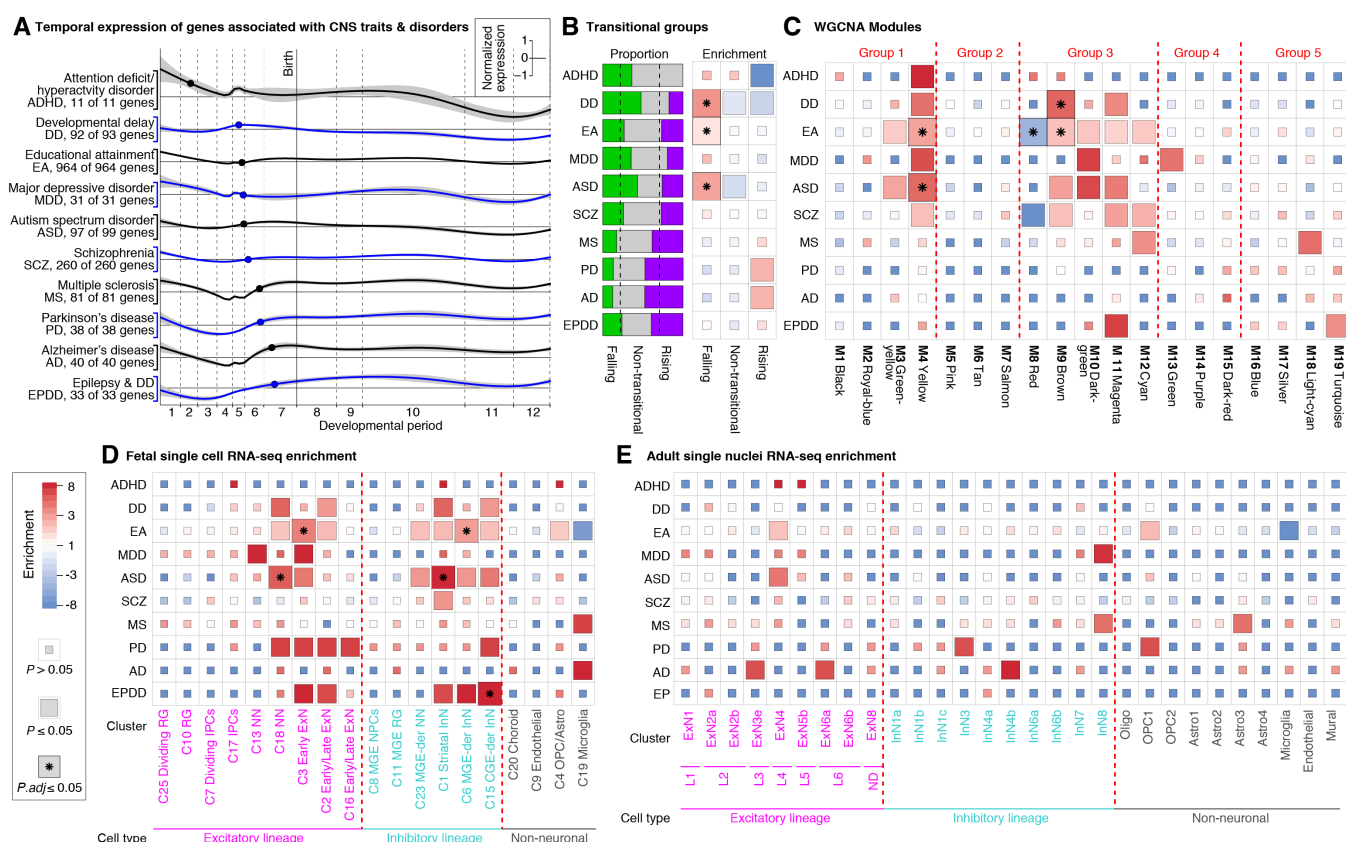


Figure 6. Expression of genes associated with CNS traits and disorders. **A)** Genome-wide significant loci were collated for ten CNS traits and disorders from exome sequencing (DD, ASD, EPDD) or GWAS (ADHD, MDD, EA, SCZ, MS, PD, AD) and genes associated with these loci were identified. The normalized expression of these gene lists is shown for each disorder with alternating colors (black and blue). For each gene the developmental period with the maximum $\log_2$ CPM value was identified and the mean period across the genes in each disorder is indicated with a point. **B)** The relative proportion (left, dashed lines represent expectation) and gene set enrichment (right) for trait/disorder-associated genes to the three trajectory groups (Figure 2). **C)** The analysis in 'B' is repeated for WGCNA modules. **D)** The analysis in 'B' is repeated for genes enriched for cell type clusters from scRNA-seq of the prenatal human brain. **E)** The analysis in 'D' is repeated for snRNA-seq of the adult human DLPFC. Abbreviations: DD: developmental delay; ASD: autism spectrum disorder; EPDD: epilepsy with developmental delay; GWAS: Genome-wide association study; ADHD: Attention deficit/hyperactivity disorder; MDD: major depressive disorder; EA: educational attainment; SCZ: schizophrenia; MS: multiple

sclerosis; PD: Parkinson's disease; AD: Alzheimer's disease. Statistical analysis: B: FET, corrected for 30 comparisons. C: FET, corrected for 190 comparisons. D: FET, corrected for 190 comparisons. E: FET, corrected for 290 comparisons.

Enrichment of CNS traits and disorders in co-expression and single cell gene lists

An alternative approach to identifying convergence across genes within a trait or disorder is to use co-expression (Gandal et al., 2018; Li et al., 2018; Parikshak et al., 2013; Willsey et al., 2013); therefore, we assessed enrichment for disease and trait-associated genes across the 19 WGCNA modules (Figure 3). Two modules are enriched after correcting for multiple comparisons (Figure 6C). The M4 Yellow module is enriched for genes associated with ASD and EA, including *NRXN1*, *TCF4*, and *BCL11A*, with a nominally significant trend in ADHD, DD, MDD, and SCZ. The M9 Brown module is enriched for genes associated with DD and EA, including *CDK13*, *PACS1*, and *EP300*, with a nominally significant trend in ASD and SCZ. Other modules in group 3 (M10 Dark-green, M11 Magenta, M12 Cyan) show nominally significant trends in multiple disorders and these include many of the genes with the strongest evidence for association with ASD and DD, including *ARID1B* in M10, and *ANKRD11*, *SETD5*, and *SYNGAP1* in M11.

Enrichment of cell type-specific genes in disorder-associated genes has been used to implicate cell types in disease pathology (Li et al., 2018; Satterstrom et al., 2018; Willsey et al., 2013; Xu et al., 2014). Using the gene lists described above (Figure 2), we perform a consistent analysis of cell type gene enrichment across the ten CNS traits and disorders (Li et al., 2018; Nowakowski et al., 2017). Five clusters showed significant enrichment, all from the fetal single cell data: C18 Excitatory newborn neurons and C1 Striatal interneurons in ASD, C3 Early excitatory neurons and C6 MGE-derived interneurons in EA, and C15 CGE-derived interneurons in EPDD (Figure 6D and 6E). Considering the nominally significant results too, we observe enrichment for a mixture of excitatory and inhibitory lineages in DD, ASD, EA, EPDD, and PD. For these disorders, the results in postnatal data are less pronounced, but not contradictory. For MS and AD, we observe nominally significant enrichment in prenatal microglia cluster-specific genes, while the postnatal data shows nominally significant enrichment for interneurons and astrocytes in MS and excitatory and inhibitory neurons in AD. Overall, protein-coding genes associated with a trait or disorder were more likely to be enriched in a cell type cluster in fetal single cell data (23.6% of trait or disorder genes vs. 16.1% of non-associated genes; OR=1.6 (95%CI:1.4-1.9); $p=1.5 \times 10^{-10}$ FET).

eQTLs in CNS traits and disorders

We next considered whether the eGenes and eQTLs identified in this dataset overlapped with the genes and variants identified in the four GWAS analyses with the most loci: EA, SCZ, MS, and AD. To assess whether GWAS loci were enriched for eQTLs, as has been observed previously (Fromer et al., 2016; Nicolae et al., 2010), we used a permutation-based method that assessed expectation based on SNPs with similar LD structure, MAF, and gene density. Correcting for multiple comparisons, we find that SCZ, EA, and MS, but not AD, are enriched for eQTLs, with the lowest p-values obtained with Constant eQTLs rather than temporally specific eQTLs (Figure 7A), though we note that these temporal-specific enrichments are driven by small numbers of overlapping loci (Table S7). A complementary analysis, considering if eGenes are enriched for GWAS signal, did not show evidence of enrichment (Figure S6). Given that GWAS

loci are enriched for eQTLs, we assessed whether the specific variants with strongest evidence at GWAS loci for EA and SCZ also had the strongest evidence of acting as eQTLs. This co-localization analysis showed that, while the loci overlapped, we did not see statistical evidence of overlap between the specific variants implicated in GWAS and acting as eQTLs (Figure S6).

To further understand the relationship between eQTLs and variants associated with human traits and disorders, we assessed the characteristics of the genes in four categories: Constant eQTLs, temporally specific eQTLs, GWAS loci (ADHD, EA, MDD, SCZ, MS, PD, AD), and Rare variant loci (DD, ASD, EPDD). The genes in proximity to GWAS loci are more likely to be protein-coding (Figure 7B), have high pLI scores (Figure 7C), and have multiple protein-protein interactions (Figure 7D) to a greater extent than temporally specific eQTLs (Figure 4) and similar to disorder-associated genes identified through rare variants.

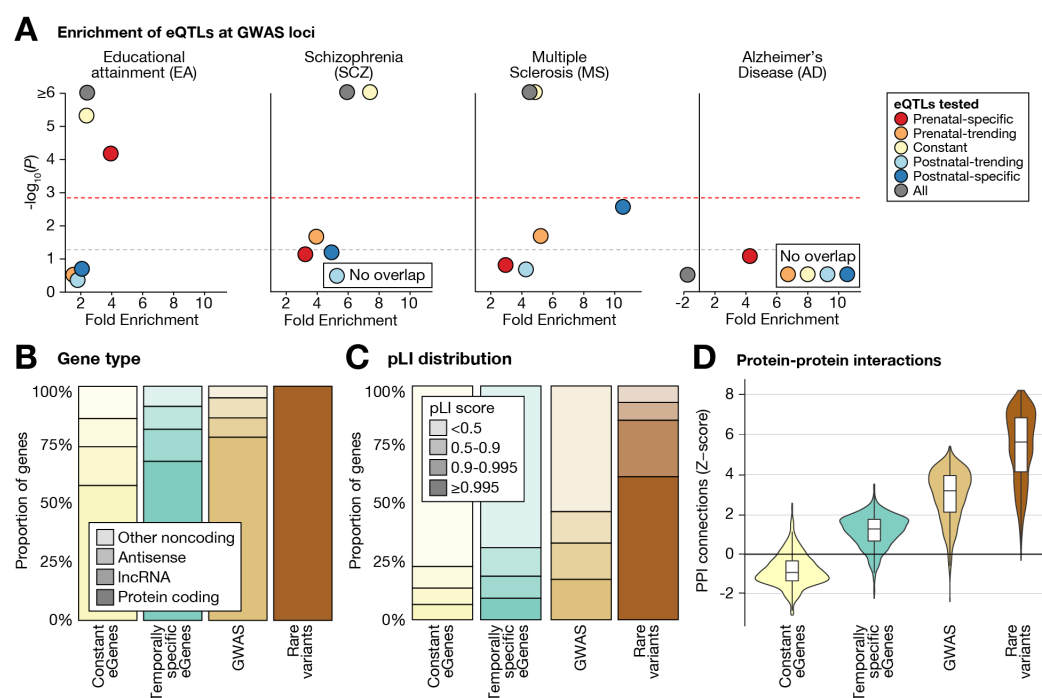


Figure 7. Relationship of eQTLs to genes associated with CNS traits and disorders. A) Enrichment of the eQTLs is shown for the four CNS GWAS datasets with p-value plotted against the degree of enrichment. Gray dashed line indicates nominal significance; red dashed line indicates significance threshold after correcting for 30 multiple comparisons. **B)** The relative proportions of coding and noncoding genes are shown for four gene sets defined by: eQTLs (Constant or Temporally specific, with Prenatal-specific and Postnatal-specific eGenes considered together, Figure 4), the union of the seven GWAS datasets, and the union of the three exome sequencing datasets (Rare variants). **C)** The distribution of pLI scores are shown for protein-coding genes in these four gene sets. **D)** BioGRID protein-protein interaction data was used to assess clustering on a network (Ripley's K-net function), shown as a z-score, for these four gene sets. Abbreviations: GWAS, genome-wide association study; pLI, probability of loss-of-function intolerance. Statistical analysis: A: Permutation procedure comparing observed proportion of GWAS-significant SNPs that were eQTLs to a null distribution matched on LD structure, MAF, gene density and corrected for 30 comparisons.

Discussion

In this manuscript, we describe BrainVar, a unique resource of paired genome (WGS) and transcriptome (bulk tissue RNA-seq) data, derived from 176 human dorsolateral prefrontal cortex samples across prenatal and postnatal development (Figure 1). We leverage these data to identify 23,782 genes expressed during human cortical development, gene lists relating to developmental trajectories and co-expression, common and rare variants that alter gene expression (eQTLs), and how these datasets relate to cell type, derived from scRNA-seq data, and CNS traits and disorders, derived from genomic analyses (exome sequencing and GWAS). In addition to developing a resource with utility for future studies on human development, neurobiology, and neuropsychiatric disorders, we also describe key biological insights, including the nature of the late-fetal transition in gene expression (Figure 2 and 3), the relative independence of variation in gene expression across development compared to gene expression between individuals (Figure 4), the variable penetrance of many haploinsufficient neuropsychiatric disorders (Figure 5), the developmental processes and cell types implicated in CNS traits and disorders (Figure 6), and the distinctive characteristics of genes with Constant eQTLs compared to those with temporally specific eQTLs or associated with CNS traits and disorders (Figure 4 and 7). Across these analyses, we observe complicated relationships between gene function and development, consistent with the impact of natural selection.

Principal component analysis identifies developmental age as the most important factor underlying the variance in gene expression in this dataset. The majority of this temporal variance occurs in two transitional phases (Figure 2). The early-fetal transition is a coordinated decrease in expression of multiple genes in early development (epoch 0; periods 1-2; 6-10 PCW), captured in the lncRNA-enriched M8 Red module (Figure 3). This transition may relate to the establishment of regional identity across the brain since it includes genes associated with identity of non-cortical tissues, however the paucity of samples in this stage limits our interpretation of this shift. The late-fetal transition between mid-fetal development and infancy involves over 12,000 genes, with similar numbers rising and falling (Figure 2). Prior reports in humans (Li et al., 2018) and primates (Zhu et al., 2018) associated this transition with a reduction in intra- and inter-regional variation evident at the levels of bulk tissue and individual cell types. In keeping with these initial reports, our data suggest this transition represents a combination of changes in the relative proportions of various cell types and biological processes within these cells (Figure 2 and 3). Specifically, the late-fetal transition appears to be the co-occurrence of decreasing cell division (M2 Royal-blue), neuronal differentiation (M4 Yellow) and anabolic processes (M1 Black) with increasing gliogenesis (M18 Light-cyan, M19 Turquoise), neuronal maturation and synaptogenesis (M19 Turquoise), and the metabolic demands of supporting neuronal activity (M13 Green, M14 Purple, M16 Blue). Critically, this larger sample set allowed us to better define the inflection point at which the late-fetal transition begins as being within period 5, further distinguishing the late-fetal transition from previously reported organotypic changes (Domazet-Loso and Tautz, 2010; Kalinka et al., 2010; Li et al., 2018).

While previous analyses have identified eQTLs in human brain tissue both postnatally, (Fromer et al., 2016; The GTEx Consortium, 2015) and prenatally (O'Brien et al., 2018), this dataset has two unique characteristics. First, no prior study has assessed the impact of genomic variation in the context of gene expression across the whole of brain development, from embryogenesis through fetal development, infancy, childhood, adolescence, and into young adulthood.

Consequently, we were able to identify temporally specific eQTLs that only modulate expression prenatally or postnatally. Second, by using deep coverage WGS rather than genotyping arrays to detect various types of genomic variation, we were able to assess the impact of both common and rare variants on gene expression.

For common variation, we identify 252,629 eQTLs that act on 8,241 of the 23,782 cortically-expressed genes. The majority are Constant eQTLs, observed in both the prenatal and postnatal samples, while 91,336 eQTLs (36.1%) are temporally specific and act on 5,444 genes (Figure 4A). Considering the FDR-significant variant-gene pairs that overlap with prior eQTL catalogues (O'Brien et al., 2018; The GTEx Consortium et al., 2017) we observe high concordance in both the direction and magnitude of effect (Figure 4C), despite differing in cohorts, methods, and analysis. Concordance was greatest for Constant eQTLs and temporally specific eQTLs from cohorts of equivalent developmental stages (Figure 4C).

Across multiple metrics, we observe dramatic differences between the eGenes with Constant and temporally specific eQTLs (Figure 4 and 7). Compared to other genes expressed in the cortex, the genes affected by Constant eQTLs are more likely to be noncoding, be weakly co-expressed, and have few protein-protein interactions. In contrast, the genes regulated by eQTLs with a degree of temporal specificity are similar to genes for which we did not detect eQTLs. A similar distinction was observed for distance from the TSS, with Constant eQTLs close to the gene (~90kb), while temporally specific eQTLs are further (~420kb). Critically, we find that pLI score, a measure of sensitivity to variation in gene expression, and eQTL effect size are inversely related (Figure 4K). Furthermore, Prenatal-specific eGenes are more common among Rising genes with low prenatal expression. These observations suggest that developmental and evolutionary constraints limit the population frequency or impact of eQTLs on key developmental processes, a hypothesis that might be testable in future studies as additional information concerning spatiotemporal and cell type specificity of enhancers and eQTLs becomes available for a variety of tissues. Under this model, Constant eQTLs with high effect sizes tend to influence genes that tolerate variation in expression (e.g. non-rate-limiting metabolic steps) or are non-critical to brain function, while temporally specific eQTLs tend to influence genes with critical roles that are sensitive to variation in expression, but only to a small degree or at a stage in development when expression variation is tolerated. These temporally specific effects appear to be modulated through distal regulatory elements, perhaps reflecting the influence of transcription factors with temporal and/or cell type specificity.

Relationships between developmental stage, cell type specificity and other confounding factors, such as natural selection, are evident in other aspects of our dataset. For example, although GWAS loci in SCZ, MS, and EA are enriched for eQTLs (Figure 7A), colocalization analysis suggests that different variants mediate the effects in each dataset (Figure S7). A potential explanation may lie in the observation that those characteristics distinguishing temporally specific eQTLs from Constant eQTLs, such as higher pLI scores and interactions with more proteins, are exaggerated at GWAS loci and disorder-associated rare variant loci. Alternatively, neuropsychiatric disorders may be mediated by cell types and subtypes, which our study is not powered to detect. Indeed, relationships between genes associated with neuropsychiatric disorders and specific cell types are evident in the data presented in this manuscript, as we observe the enrichment of genes associated with DD, ASD, and EA in developing neurons of

the excitatory and inhibitory lineages in the prenatal brain and increased associations between late-onset disorders and glial cell types (Figure 6). eQTLs derived from single cell expression data or other brain regions may provide greater clarity. Also suggesting differing specificity or other confounding issues, the magnitude of effects we observed for rare heterozygous PTVs and deletions was less than anticipated and inconsistent between variants, potentially reflecting a dilution of cell type specific signals against a background of inter-individual variation, which may explain variable penetrance, or other effectors including autoregulation, protective alleles, alternate isoforms, or an incomplete model of PTV effects (MacArthur et al., 2012; Rivas et al., 2015). Higher resolution datasets across development, including single cell, additional brain regions, and larger sample sizes, along with complementary analyses in the brains of individuals with neuropsychiatric disorders and rare genetic disorders, are likely to provide additional insights into these relationships.

The combination of genomic and transcriptomic data across development allows us to interrogate human cortical development from a molecular perspective at a higher resolution than before. Understanding patterns of temporal and cell type specificity has already provided insights into the pathology underlying neuropsychiatric disorders and further delineation of these patterns is likely to be critical to a detailed understanding of etiology as a foundation for therapeutic development.

Acknowledgments

Data were generated as part of the PsychENCODE Consortium, supported by: U01MH103339, U01MH103365, U01MH103392, U01MH103340, U01MH103346, R01MH105472, R01MH094714, R01MH105898, R21MH102791, R21MH105881, R21MH103877, and P50MH106934 awarded to: Schahram Akbarian (Icahn School of Medicine at Mount Sinai), Gregory Crawford (Duke), Stella Dracheva (Icahn School of Medicine at Mount Sinai), Peggy Farnham (USC), Mark Gerstein (Yale), Daniel Geschwind (UCLA), Thomas M. Hyde (LIBD), Andrew Jaffe (LIBD), James A. Knowles (USC), Chunyu Liu (UIC), Dalila Pinto (Icahn School of Medicine at Mount Sinai), Nenad Sestan (Yale), Pamela Sklar (Icahn School of Medicine at Mount Sinai), Matthew State (UCSF), Patrick Sullivan (UNC), Flora Vaccarino (Yale), Sherman Weissman (Yale), Kevin White (UChicago) and Peter Zandi (JHU). This work was supported by funding provided by the Autism Science Foundation (DW, JA, SJS); Simons Foundation Autism Research Initiative (SFARI) grants #574598 (SJS) and #402281 (SJS, MWS, BD, KR); National Institute for Mental Health (NIMH) grants R01 MH109901 (SJS, MWS), R01 MH110928 (SJS, MWS), U01 MH103339 (MWS), R01 MH111662 (SJS, MWS), U01 MH105575 (MWS), U01 MH106874 (NS), P50 MH106934 (NS), R01 MH109904 (NS), R01 MH110926 (NS), U01 MH116488 (NS); and National Institute of Child Health and Human Development (NICHD) grant R24 HD000836 (IAG). We thank Thomas Lehner, Anjené Addington, Geetha Senthil, and Alexander Arguello at the NIMH for their support of the PsychENCODE consortium, the Yale Center for Genome Analysis for generating the RNA-seq data, GENEWIZ for generating the WGS data, and Sentieon for the use of a computationally efficient implementation of GATK Haplotype Caller.

Author Contributions

Conceptualization, KR, MWS, BD, SJS, and NS; Methodology, DMW, SP, JC, JA, HAB, JLT, KAA, DRO, IAG, KR, MWS, BD, SJS, and NS; Software, DMW, JA, CD, LL, MCG, and SJS; Validation, DMW, SP, JC, JA, CD, SD, BD, SJS, and NS; Formal Analysis, DMW, SP, JC, JA, BS, MP, GS, MSB, LK, AEC,

BD, and SJS; Investigation, DMW, SP, JC, JA, BS, BD, SJS, and NS; Resources, SP, JC, ZL, CD, AMS, ATT, NK, FOG, LL, MCG, XZ, HAB, JLT, KAA, DRO, IAG, MET, SJS, and NS; Data Curation, DMW, SP, JC, CD, LL, MCG, XZ, SD, and SJS; Statistical analysis, DMW, JC, JA, BS, MP, MSB, BD, and SJS; Writing Original Draft, DMW, SP, JC, JA, BS, BD, SJS, and NS; Writing Review & Editing, DMW, SP, JC, JA, BS, MP, ZL, CD, GS, AMS, ATT, NK, FOG, MSB, LL, MCG, XZ, SD, LK, AEC, JDB, HAB, JLT, KAA, DRO, IAG, NAZ, MET, KR, MWS, BD, SJS, and NS; Visualization, DMW, SP, JC, JA, BS, SJS, and NS; Supervision, JDB, NAZ, MET, KR, MWS, BD, SJS, and NS; Project Administration, DMW, SP, SJS, and NS; Funding Acquisition, KR, MWS, BD, SJS, and NS.

Declaration of Interests

The authors declare no competing interests

References

- An, J.Y., Lin, K., Zhu, L., Werling, D.M., Dong, S., Brand, H., Wang, H.Z., Zhao, X., Schwartz, G.B., Collins, R.L., *et al.* (2018). Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder. *Science* 362.
- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166-169.
- Ashrafi, G., and Ryan, T.A. (2017). Glucose metabolism in nerve terminals. *Curr Opin Neurobiol* 45, 156-161.
- Bae, B.I., and Walsh, C.A. (2013). Neuroscience. What are mini-brains? *Science* 342, 200-201.
- Bioconductor Core Team and Bioconductor Package Maintainer (2016). TxDb.Hsapiens.UCSC.hg38.knownGene: Annotation package for TxDb object(s).
- Birnbaum, R., and Weinberger, D.R. (2017). Genetic insights into the neurodevelopmental origins of schizophrenia. *Nat Rev Neurosci* 18, 727-740.
- BrainSeq: A. Human Brain Genomics Consortium (2015). BrainSeq: Neurogenomics to Drive Novel Target Discovery for Neuropsychiatric Disorders. *Neuron* 88, 1078-1083.
- Breen, G., Li, Q., Roth, B.L., O'Donnell, P., Didriksen, M., Dolmetsch, R., O'Reilly, P.F., Gaspar, H.A., Manji, H., Huebel, C., *et al.* (2016). Translating genome-wide association findings into new therapeutics for psychiatry. *Nat Neurosci* 19, 1392-1396.
- Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015a). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7.
- Chang, D., Nalls, M.A., Hallgrimsdottir, I.B., Hunkapiller, J., van der Brug, M., Cai, F., International Parkinson's Disease Genomics, C., and Me Research, T., Kerchner, G.A., Ayalon, G., *et al.* (2017). A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet* 49, 1511-1516.
- Chang, J., Gilman, S.R., Chiang, A.H., Sanders, S.J., and Vitkup, D. (2015b). Genotype to phenotype relationships in autism spectrum disorders. *Nat Neurosci* 18, 191-198.
- Chou, C.H., Shrestha, S., Yang, C.D., Chang, N.W., Lin, Y.L., Liao, K.W., Huang, W.C., Sun, T.H., Tu, S.J., Lee, W.H., *et al.* (2018). miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res* 46, D296-D302.

- Colantuoni, C., Lipska, B.K., Ye, T., Hyde, T.M., Tao, R., Leek, J.T., Colantuoni, E.A., Elkahoul, A.G., Herman, M.M., Weinberger, D.R., *et al.* (2011). Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* 478, 519-523.
- Cornish, A.J., and Markowetz, F. (2014). SANTA: quantifying the functional content of molecular networks. *PLoS Comput Biol* 10, e1003808.
- Courchesne, E., Pierce, K., Schumann, C.M., Redcay, E., Buckwalter, J.A., Kennedy, D.P., and Morgan, J. (2007). Mapping early brain development in autism. *Neuron* 56, 399-413.
- de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 11, e1004219.
- Deciphering Developmental Disorders Study (2017). Prevalence and architecture of de novo mutations in developmental disorders. *Nature* 542, 433-438.
- Demontis, D., Walters, R.K., Martin, J., Mattheisen, M., Als, T.D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Baekvad-Hansen, M., *et al.* (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* 51, 63-75.
- Dobyn, A., Huckins, L.M., Boocock, J., Sloofman, L.G., Glicksberg, B.S., Giambartolomei, C., Hoffman, G.E., Perumal, T.M., Girdhar, K., Jiang, Y., *et al.* (2018). Landscape of Conditional eQTL in Dorsolateral Prefrontal Cortex and Co-localization with Schizophrenia GWAS. *Am J Hum Genet* 102, 1169-1184.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15-21.
- Domazet-Loso, T., and Tautz, D. (2010). A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* 468, 815-818.
- Ecker, C., Bookheimer, S.Y., and Murphy, D.G. (2015). Neuroimaging in autism spectrum disorder: brain structure and function across the lifespan. *Lancet Neurol* 14, 1121-1134.
- Elliott, L.T., Sharp, K., Alfaro-Almagro, F., Shi, S., Miller, K.L., Douaud, G., Marchini, J., and Smith, S.M. (2018). Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* 562, 210-216.
- Freed, D.N., Aldana, R., Weber, J.A., and Edwards, J.S. (2017). The Sentieon Genomics Tools - A fast and accurate solution to variant calling from next-generation sequence data. *bioRxiv*, 115717.
- Fromer, M., Roussos, P., Sieberts, S.K., Johnson, J.S., Kavanagh, D.H., Perumal, T.M., Ruderfer, D.M., Oh, E.C., Topol, A., Shah, H.R., *et al.* (2016). Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci* 19, 1442-1453.
- Gandal, M.J., Haney, J.R., Parikshak, N.N., Leppa, V., Ramaswami, G., Hartl, C., Schork, A.J., Appadurai, V., Buil, A., Werge, T.M., *et al.* (2018). Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* 359, 693-697.
- Geschwind, D.H., and Flint, J. (2015). Genetics and genomics of psychiatric disease. *Science* 349, 1489-1494.

- Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 10, e1004383.
- Gibbs, J.R., van der Brug, M.P., Hernandez, D.G., Traynor, B.J., Nalls, M.A., Lai, S.L., Arepalli, S., Dillman, A., Rafferty, I.P., Troncoso, J., *et al.* (2010). Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet* 6, e1000952.
- Gogarten, S.M., Bhangale, T., Conomos, M.P., Laurie, C.A., McHugh, C.P., Painter, I., Zheng, X., Crosslin, D.R., Levine, D., Lumley, T., *et al.* (2012). GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. *Bioinformatics* 28, 3329-3331.
- Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R., Won, H., Pallesen, J., Agerbo, E., Andreassen, O.A., Anney, R., *et al.* (2017). Common risk variants identified in autism spectrum disorder. *bioRxiv*, 224774.
- Gulsuner, S., Walsh, T., Watts, A.C., Lee, M.K., Thornton, A.M., Casadei, S., Rippey, C., Shahin, H., Consortium on the Genetics of, S., Group, P.S., *et al.* (2013). Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* 154, 518-529.
- Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., *et al.* (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 22, 1760-1774.
- Heinzen, E.L., Ge, D., Cronin, K.D., Maia, J.M., Shianna, K.V., Gabriel, W.N., Welsh-Bohmer, K.A., Hulet, C.M., Denny, T.N., and Goldstein, D.B. (2008). Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS Biol* 6, e1.
- Heyne, H.O., Singh, T., Stamberger, H., Abou Jamra, R., Caglayan, H., Craiu, D., De Jonghe, P., Guerrini, R., Helbig, K.L., Koeleman, B.P.C., *et al.* (2018). De novo variants in neurodevelopmental disorders with epilepsy. *Nat Genet* 50, 1048-1053.
- Hu, W.F., Chahrour, M.H., and Walsh, C.A. (2014). The diverse genetic landscape of neurodevelopmental disorders. *Annu Rev Genomics Hum Genet* 15, 195-213.
- Insel, T.R. (2010). Rethinking schizophrenia. *Nature* 468, 187-193.
- International Multiple Sclerosis Genetics Consortium, Beecham, A.H., Patsopoulos, N.A., Xifara, D.K., Davis, M.F., Kempainen, A., Cotsapas, C., Shah, T.S., Spencer, C., Booth, D., *et al.* (2013). Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet* 45, 1353-1360.
- Jaffe, A.E., Straub, R.E., Shin, J.H., Tao, R., Gao, Y., Collado-Torres, L., Kam-Thong, T., Xi, H.S., Quan, J., Chen, Q., *et al.* (2018). Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat Neurosci* 21, 1117-1125.
- Jones, B.L., Nagin, D.S., and Roeder, K. (2001). A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories. *Sociological Methods & Research* 29, 374-393.
- Jun, G., Flickinger, M., Hetrick, K.N., Romm, J.M., Doheny, K.F., Abecasis, G.R., Boehnke, M., and Kang, H.M. (2012). Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am J Hum Genet* 91, 839-848.

- Kalinka, A.T., Varga, K.M., Gerrard, D.T., Preibisch, S., Corcoran, D.L., Jarrells, J., Ohler, U., Bergman, C.M., and Tomancak, P. (2010). Gene expression divergence recapitulates the developmental hourglass model. *Nature* 468, 811-814.
- Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M., Pletikos, M., Meyer, K.A., Sedmak, G., *et al.* (2011). Spatio-temporal transcriptome of the human brain. *Nature* 478, 483-489.
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res* 12, 996-1006.
- Keshavan, M.S., Giedd, J., Lau, J.Y., Lewis, D.A., and Paus, T. (2014). Changes in the adolescent brain and the pathophysiology of psychotic disorders. *Lancet Psychiatry* 1, 549-558.
- Kosmicki, J.A., Churchhouse, C.L., Rivas, M.A., and Neale, B.M. (2016). Discovery of rare variants for complex phenotypes. *Hum Genet* 135, 625-634.
- Kryuchkova-Mostacci, N., and Robinson-Rechavi, M. (2017). A benchmark of gene expression tissue-specificity metrics. *Brief Bioinform* 18, 205-214.
- Lachmann, A., Xu, H., Krishnan, J., Berger, S.I., Mazloom, A.R., and Ma'ayan, A. (2010). ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* 26, 2438-2444.
- Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., *et al.* (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 45, 1452-1458.
- Langfelder, P., and Horvath, S. (2007). Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol* 1, 54.
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559.
- Langfelder, P., Luo, R., Oldham, M.C., and Horvath, S. (2011). Is my network module preserved and reproducible? *PLoS Comput Biol* 7, e1001057.
- Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology* 15, R29.
- Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T.A., Bowers, P., Sidorenko, J., Karlsson Linner, R., *et al.* (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* 50, 1112-1121.
- Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28, 882-883.
- Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., *et al.* (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285-291.

- Li, H. (2014). Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* 30, 2843-2851.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754-1760.
- Li, M., Santpere, G., Imamura Kawasawa, Y., Evgrafov, O.V., Gulden, F.O., Pochareddy, S., Sunkin, S.M., Li, Z., Shin, Y., Zhu, Y., *et al.* (2018). Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* 362.
- Liu, C., Cheng, L., Badner, J.A., Zhang, D., Craig, D.W., Redman, M., and Gershon, E.S. (2010). Whole-genome association mapping of gene expression in the human prefrontal cortex. *Mol Psychiatry* 15, 779-784.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550.
- Lui, J.H., Hansen, D.V., and Kriegstein, A.R. (2011). Development and evolution of the human neocortex. *Cell* 146, 18-36.
- MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., *et al.* (2012). A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 335, 823-828.
- Matys, V., Fricke, E., Geffers, R., Gossling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A.E., Kel-Margoulis, O.V., *et al.* (2003). TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* 31, 374-378.
- McCarroll, S.A., and Hyman, S.E. (2013). Progress in the genetics of polygenic brain disorders: significant new challenges for neurobiology. *Neuron* 80, 578-587.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., *et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20, 1297-1303.
- McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biol* 17, 122.
- Mostafavi, S., Battle, A., Zhu, X., Urban, A.E., Levinson, D., Montgomery, S.B., and Koller, D. (2013). Normalizing RNA-sequencing data by modeling hidden covariates with prior knowledge. *PLoS One* 8, e68141.
- Myers, A.J., Gibbs, J.R., Webster, J.A., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., Leung, D., Bryden, L., Nath, P., *et al.* (2007). A survey of genetic human cortical gene expression. *Nat Genet* 39, 1494-1499.
- Nagy, E., and Maquat, L.E. (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci* 23, 198-199.
- Network and Pathway Analysis Subgroup of the Psychiatric Genomics Consortium (2015). Corrigendum: Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat Neurosci* 18, 926.

- Nicolae, D.L., Gamazon, E., Zhang, W., Duan, S., Dolan, M.E., and Cox, N.J. (2010). Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet* 6, e1000888.
- Nowakowski, T.J., Bhaduri, A., Pollen, A.A., Alvarado, B., Mostajo-Radji, M.A., Di Lullo, E., Haeussler, M., Sandoval-Espinosa, C., Liu, S.J., Velmeshev, D., *et al.* (2017). Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* 358, 1318-1323.
- O'Brien, H.E., Hannon, E., Hill, M.J., Toste, C.C., Robertson, M.J., Morgan, J.E., McLaughlin, G., Lewis, C.M., Schalkwyk, L.C., Hall, L.S., *et al.* (2018). Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. *Genome Biol* 19, 194.
- Parikshak, N.N., Luo, R., Zhang, A., Won, H., Lowe, J.K., Chandran, V., Horvath, S., and Geschwind, D.H. (2013). Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* 155, 1008-1021.
- Pedersen, B.S., and Quinlan, A.R. (2017). Who's Who? Detecting and Resolving Sample Anomalies in Human DNA Sequencing Studies with Peddy. *Am J Hum Genet* 100, 406-413.
- Pletikos, M., Sousa, A.M., Sedmak, G., Meyer, K.A., Zhu, Y., Cheng, F., Li, M., Kawasawa, Y.I., and Sestan, N. (2014). Temporal specification and bilaterality of human neocortical topographic gene expression. *Neuron* 81, 321-332.
- Psych, E.C., Akbarian, S., Liu, C., Knowles, J.A., Vaccarino, F.M., Farnham, P.J., Crawford, G.E., Jaffe, A.E., Pinto, D., Dracheva, S., *et al.* (2015). The PsychENCODE project. *Nat Neurosci* 18, 1707-1712.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841-842.
- Regier, A.A., Farjoun, Y., Larson, D.E., Krasheninina, O., Kang, H.M., Howrigan, D.P., Chen, B.J., Kher, M., Banks, E., Ames, D.C., *et al.* (2018). Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. *Nat Commun* 9, 4038.
- Reilly, S.K., Yin, J., Ayoub, A.E., Emera, D., Leng, J., Cotney, J., Sarro, R., Rakic, P., and Noonan, J.P. (2015). Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science* 347, 1155-1159.
- Reimand, J., Arak, T., and Vilo, J. (2011). g:Profiler--a web server for functional interpretation of gene lists (2011 update). *Nucleic Acids Res* 39, W307-315.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43, e47.
- Rivas, M.A., Pirinen, M., Conrad, D.F., Lek, M., Tsang, E.K., Karczewski, K.J., Maller, J.B., Kukurba, K.R., DeLuca, D.S., Fromer, M., *et al.* (2015). Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science* 348, 666-669.

- Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., *et al.* (2015). Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317-330.
- Roeder, K., Lynch, K.G., and Nagin, D.S. (1999). Modeling Uncertainty in Latent Class Membership: A Case Study in Criminology. *Journal of the American Statistical Association* 94, 766-776.
- Rosti, R.O., Sadek, A.A., Vaux, K.K., and Gleeson, J.G. (2014). The genetic landscape of autism spectrum disorders. *Dev Med Child Neurol* 56, 12-18.
- Sanders, S.J., He, X., Willsey, A.J., Ercan-Sencicek, A.G., Samocha, K.E., Cicek, A.E., Murtha, M.T., Bal, V.H., Bishop, S.L., Dong, S., *et al.* (2015). Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* 87, 1215-1233.
- Sanders, S.J., Neale, B.M., Huang, H., Werling, D.M., An, J.Y., Dong, S., Whole Genome Sequencing for Psychiatric, D., Abecasis, G., Arguello, P.A., Blangero, J., *et al.* (2017). Whole genome sequencing in psychiatric disorders: the WGSPD consortium. *Nat Neurosci* 20, 1661-1668.
- Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.-Y., Peng, M., Collins, R.L., Grove, J., Klei, L., *et al.* (2018). Novel genes for autism implicate both excitatory and inhibitory cell lineages in risk. *bioRxiv*, 484113.
- Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511, 421-427.
- Schmidt, E.M., Zhang, J., Zhou, W., Chen, J., Mohlke, K.L., Chen, Y.E., and Willer, C.J. (2015). GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics* 31, 2601-2606.
- Sestan, N., and State, M.W. (2018). Lost in Translation: Traversing the Complex Path from Genomics to Therapeutics in Autism Spectrum Disorder. *Neuron* 100, 406-423.
- Shaw, P., Gogtay, N., and Rapoport, J. (2010). Childhood psychiatric disorders as anomalies in neurodevelopmental trajectories. *Hum Brain Mapp* 31, 917-925.
- Silbereis, J.C., Pochareddy, S., Zhu, Y., Li, M., and Sestan, N. (2016). The Cellular and Molecular Landscapes of the Developing Human Central Nervous System. *Neuron* 89, 248-268.
- Tang, S.X., and Gur, R.E. (2018). Longitudinal perspectives on the psychosis spectrum in 22q11.2 deletion syndrome. *Am J Med Genet A* 176, 2192-2202.
- Tau, G.Z., and Peterson, B.S. (2010). Normal development of brain circuits. *Neuropsychopharmacology* 35, 147-168.
- The GTEx Consortium (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348, 648-660.
- The GTEx Consortium, Laboratory, D.A., Coordinating Center -Analysis Working, G., Statistical Methods groups-Analysis Working, G., Enhancing, G.g., Fund, N.I.H.C., Nih/Nci, Nih/Nhgri, Nih/Nimh, Nih/Nida, *et al.* (2017). Genetic effects on gene expression across human tissues. *Nature* 550, 204-213.

- Turner, T.N., and Eichler, E.E. (2018). The Role of De Novo Noncoding Regulatory Mutations in Neurodevelopmental Disorders. *Trends Neurosci.*
- van Buuren, S., and Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *2011* 45, 67.
- Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F.C.P., Clarke, D., Gu, M., Emani, P., Yang, Y.T., *et al.* (2018). Comprehensive functional genomic resource and integrative model for the human brain. *Science* 362.
- Werling, D.M., Brand, H., An, J.Y., Stone, M.R., Zhu, L., Glessner, J.T., Collins, R.L., Dong, S., Layer, R.M., Markenscoff-Papadimitriou, E., *et al.* (2018). An analytical framework for whole-genome sequence association studies and its implications for autism spectrum disorder. *Nat Genet* 50, 727-736.
- Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., *et al.* (2013). Discovery and refinement of loci associated with lipid levels. *Nat Genet* 45, 1274-1283.
- Willsey, A.J., Sanders, S.J., Li, M., Dong, S., Tebbenkamp, A.T., Muhle, R.A., Reilly, S.K., Lin, L., Fertuzinhos, S., Miller, J.A., *et al.* (2013). Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* 155, 997-1007.
- Winter, A.G., Wildenhain, J., and Tyers, M. (2011). BioGRID REST Service, BiogridPlugin2 and BioGRID WebGraph: new tools for access to interaction data at BioGRID. *Bioinformatics* 27, 1043-1044.
- Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z., *et al.* (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* 46, 1173-1186.
- Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F., *et al.* (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* 50, 668-681.
- Xu, X., Wells, A.B., O'Brien, D.R., Nehorai, A., and Dougherty, J.D. (2014). Cell type-specific expression analysis to identify putative cellular mechanisms for neurogenetic disorders. *J Neurosci* 34, 1420-1431.
- Zhu, Y., Sousa, A.M.M., Gao, T., Skarica, M., Li, M., Santpere, G., Esteller-Cucala, P., Juan, D., Ferrandez-Peral, L., Gulden, F.O., *et al.* (2018). Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* 362.