Title: Integrative Functional Genomic Analysis of Human Brain Development and Neuropsychiatric Risks

Short title:

Integrative genomics of human brain development

One sentence summary:

Integrative analysis of human brain development reveals a cup-shaped inter-regional transcriptomic variation and a convergence of neuropsychiatric risk into distinct gene co-expression modules and cell types.

Authors: Mingfeng Li[1+], Gabriel Santpere[1+], Yuka Imamura Kawasawa[1,2+], Oleg V. Evgrafov[3+], Forrest O. Gulden[1+], Sirisha Pochareddy[1+], Susan M. Sunkin[4+], Zhen Li[1+], Yurae Shin[1,5+], Ying Zhu[1], Andre M.M. Sousa[1], Donna M. Werling[8], Robert R. Kitchen[6,7], Hyo Jung Kang[1,9], Mihovil Pletikos[1,10], Jinmyung Choi[1], Sydney Muchnik[1], Xuming Xu[1], Daifeng Wang[11], Belen Lorente-Galdos[1], Shuang Liu[1,6], Paola Giusti-Rodríguez[12], Hyejung Won[12,13], Christiaan A de Leeuw[14], Antonio F. Pardiñas[15], BrainSpan Consortium[++], PsychENCODE Consortium[++], PsychENCODE Developmental Subgroup[++], Ming Hu[16], Fulai Jin[17], Yun Li[18], Michael J. Owen[15], Michael C. O'Donovan[15], James T.R. Walters[15], Danielle Posthuma[14], Patt Levitt[19], Daniel R. Weinberger[20], Joel E. Kleinman[21], Daniel H. Geschwind[21,22,23], Michael J. Hawrylycz[4], Matthew W. State[8], Stephan J. Sanders[8], Patrick F. Sullivan[11], Mark B. Gerstein[6]*, Ed S. Lein[4]*, James A. Knowles[3]*, Nenad Sestan[1,7, 24,25,26]*

Affiliations:

[1]Department of Neuroscience and Kavli Institute for Neuroscience, Yale School of Medicine, New Haven, CT, USA

[2]Departments of Pharmacology and Biochemistry and Molecular Biology, Institute for Personalized Medicine, Pennsylvania State University College of Medicine, Hershey, PA, USA

[3]Department of Cell Biology, SUNY Downstate Medical Center, Brooklyn NY, USA

[4]Allen Institute for Brain Science, Seattle, WA, USA

[5]National Research Foundation of Korea, Daejeon, South Korea

[6]Program in Computational Biology and Bioinformatics, Departments of Molecular Biophysics and Biochemistry and Computer Science, Yale University, New Haven, CT, USA

[7]Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA

[8]Department of Psychiatry, University of California, San Francisco, San Francisco, CA, USA

[9]Department of Life Science, Chung-Ang University, Seoul, Korea

[10]Department of Anatomy & Neurobiology, Boston University School of Medicine, MA, USA

[11]Department of Biomedical Informatics Stony Brook University, NY, USA

[12]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA

[13]UNC Neuroscience Center, University of North Carolina, Chapel Hill, NC 27599, USA

[14]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, VU University, Amsterdam, The Netherlands

[15]MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, UK

[16] Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH, USA

[17]Department of Genetics and Genome Science, Case Western Reserve University, Cleveland, OH, USA

[18]Department of Genetics and Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA

[19]Department of Pediatrics, Institute for the Developing Mind Keck School of Medicine of USC & Children's Hospital Los Angeles, Los Angeles, CA, USA

[20]Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, MD, USA

[21]Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

[22]Center for Autism Research and Treatment, Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

[23]Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

[24]Department of Genetics, Yale School of Medicine, New Haven, CT, USA

[25] Department of Comparative Medicine, Program in Integrative Cell Signaling and Neurobiology of Metabolism, Yale School of Medicine, New Haven, CT, USA

[26]Program in Cellular Neuroscience, Neurodegeneration, and Repair and Yale Child Study Center, Yale School of Medicine, New Haven, CT, USA

[+] These authors contributed equally to this work

[++] The consortium authors are listed at the end of the paper.

* Co-corresponding authors

Abstract

To broaden our understanding of human neurodevelopment, we profiled transcriptomic and epigenomic landscapes across brain regions and/or cell types for the entire span of prenatal and postnatal development. Integrative analysis revealed temporal, regional, sex, and cell type-specific dynamics. We observed a global transcriptomic cup-shaped pattern, characterized by a late-fetal transition associated with sharply decreased regional differences and changes in cellular composition and maturation, followed by reversal in childhood-adolescence, and accompanied by epigenomic reorganizations. Analysis of gene co-expression modules revealed relationships with epigenomic regulation and neurodevelopmental processes. Genes with genetic associations to brain-based traits and neuropsychiatric disorders (including *MEF2C*, *SATB2, SOX5*, *TCF4,* and *TSHZ3*) converged in a small number of modules and distinct cell types, revealing insights into neurodevelopment and the genomic basis of neuropsychiatric risks.

Introduction

The development of the human central nervous system is an intricate process that unfolds over several decades, during which time numerous distinct cell types are generated and assembled into functionally distinct circuits and regions (*1-4*). These basic components of the brain are neither born mature nor are they static throughout their lifetimes; over the course of development, they undergo a variety of molecular and morphological changes. As a consequence, the characteristics of a given cell, circuit, or brain region described at a given time offer only a snapshot of that unit.

The processes guiding the development of the nervous system are reliant on the diversity and precise spatiotemporal regulation of the transcriptome (*1-4*). There is increasingly persuasive evidence that dysregulation of the transcriptional, regulatory, and epigenetic processes underlying the spatial architecture and temporal progression of human neurodevelopment can have dire consequences for brain function or strongly impact the risk of neuropsychiatric disorders (*5-7*). Indeed, many of the regulatory and epigenomic features governing the transcriptome of the developing human nervous system may be specific to particular developmental contexts in humans or closely related primate species. As such, it is difficult to identify or fully study human functional genomic elements using most common model organisms or cell culture systems (*8*). Assaying human cells and postmortem tissues solves some of these problems, but challenges, including the availability and quality of developmental tissue, limit the scale of such analyses. Consequently, despite ongoing efforts, our understanding of different facets of the transcriptional, regulatory, and epigenetic architecture of the human nervous system, particularly during early developmental periods, remains highly incomplete (*8-21*).

To begin rectifying this deficiency, the NIH-funded PsychENCODE (www.psychencode.org) and BrainSpan Consortia (www.brainspan.org) sought to generate and

analyze multi-dimensional genomics data from the developing and adult human brain in healthy and disease states.

Study design and data generation

Here, we describe the generation and integrated analysis of multiple genomic data modalities including transcriptomic profile, DNA methylation status, histone modifications, CTCF binding sites, and genotype generated from bulk tissue (1,230 samples from 48 brains) or at the single cell/nucleus level (18,288 cells/nuclei from 12 brains) from 60 de-identified post-mortem brains obtained from clinically and histopathologically unremarkable donors of both sexes and multiple ancestries. Subject ages ranged from 5 postconceptional weeks (PCW) to 64 postnatal years (PY) (Fig. 1 and tables S1 to S6). Genotyping of DNA extracted from brain with a HumanOmni2.5-8 BeadChip confirmed subject ancestry and revealed no obvious genomic abnormalities (*22*).

For transcriptome analysis, tissue-level mRNA-Seq was performed on a total of 607 histologically verified, high-quality tissue samples from 16 anatomical brain regions (11 areas of the neocortex [NCX], hippocampus [HIP], amygdala [AMY], striatum [STR], mediodorsal nucleus of thalamus [MD], and cerebellar cortex [CBC]) involved in higher-order cognition and behavior (Fig. 2A, (*22*)). These regions were systematically dissected from 41 brains ranging in age from 8 PCW to 40 PY (18 females and 23 males; postmortem interval (PMI) = 12.9 ± 10.4 hours; tissue pH = 6.5 ± 0.3; RNA integrity number = 8.8 ± 1) (Fig. 1 and table S1). Due to the limited amounts of prenatal samples, small RNA sequencing (smRNA-Seq) was performed on 16 regions of 22 postnatal brains with 278 samples passing quality control measures (Fig. 1 and table S2). These tissue-level RNA-Seq analyses were complemented by single-cell RNA-Seq (scRNA-Seq) data generated from 1,195 cells collected from embryonic fronto-parietal neocortical wall

and mid-fetal fronto-parietal neocortical plate and adjacent subplate zone of an independent set of 9 brains ranging in age from 5 to 20 PCW (Fig. 1 and table S3) and single-nuclei RNA-Seq data (snRNA-Seq) generated from 17, 093 nuclei from the dorsolateral prefrontal cortex (DFC, aka DLPFC) of three adult brains (Fig. 1 and table S4). For epigenome analyses, DNA cytosine methylation was profiled with the Infinium HumanMethylation450 BeadChip in 269 postnatal samples covering the same 16 brain regions analyzed by RNA-Seq (Fig. 1 and table S5). Additional epigenomic data was generated with ChIP-Seq for H3K4me3, H3K27me3, and H3K27ac histone marks and the epigenetic regulatory protein CTCF, which together identify a large fraction of active enhancers, promoters, repressors, and insulators. These data were generated from DFC and CBC of a subset of samples from mid-fetal, infant, and adult brains (Fig. 1 and table S6). Stringent quality control measures (figs. S1 to S8) were applied to all datasets before in-depth analyses. We also validated some results by applying independent approaches (figs. S9, S10, S18). Finally, to enable more powerful comparisons, we grouped specimens into 9 time windows (W1-W9) on the basis of major neurodevelopmental milestones and unsupervised transcriptome-based temporal arrangement of constituent specimens (Fig. 1A and tables S1 to S6).

Global spatio-temporal dynamics

We found that most protein coding genes were temporally (67.8%) or spatially (54.5%) differentially expressed (*22*) between at least two time windows or regions, respectively, with the majority of spatially differentially expressed genes (95.8%) also temporally differentially expressed. To gain a broad understanding of this transcriptomic variation, we analyzed the level of similarity between individual samples in the mRNA-Seq dataset using multidimensional scaling applied to both gene and isoform transcript-level analyses (Fig. 2B and figs. S11, S12). In both

analyses, we found a clear divide between samples from embryonic through late mid-fetal development (W1-4) and samples from late infancy through adulthood (W6-9), with samples from the late fetal period through early infancy (W5) generally spanning this divide. To determine the relationship between these three groups, we performed unsupervised hierarchical clustering analysis and found that all samples from W5, including the late fetal samples, were more similar to early postnatal samples than to late mid-fetal samples (fig. S13). Analysis of large-scale, intra-regional changes in the transcriptome across time also suggest a major transition that begins prior to birth: the transcriptomes of major brain regions and neocortical areas correlated well across both embryonic and early to mid-fetal (W1-4) and later postnatal (W6-9) development, but displayed a sharp decrease in correlation across late fetal development and early infancy (W5) (Fig. 2C and fig. S14). This transition was also apparent at the inter-regional level. Pairwise comparisons of gene expression across all 16 brain regions found a reduction in the number of genes showing differential regional expression during W5 relative to all other windows (fig. S15). Taken together, our observation of high variation during embryonic and early to mid-fetal ages followed by a decrease across late fetal ages and the subsequent resumption of higher levels of inter- and intra-regional variation during late childhood and adolescence, revealed a cup-shaped, or hourglass-like, pattern of transcriptomic development (Fig. 2D).

To further explore how regional transcriptomic profiles change with age, we applied the adjustment for confounding principal component analysis algorithm (AC-PCA) (*23*) which adjusts for inter-individual variations. Within any given developmental window, AC-PCA exhibited a clear separation of brain regions, but the average dissimilarity between transcription profiles of brain regions declined from W1 to W5 and then increased with age after W5 (Figs. 2E, 2F and fig. S16). Implying a relationship between transcriptional signatures and developmental origin, we

found that dorsal pallium-derived structures of the cerebrum (i.e., NCX, HIP, and AMY) as well as STR became increasingly similar across prenatal development, while CBC and MD remained most distinct across all time windows. To confirm these observations and to evaluate the contribution of each brain region to the regional variation described by AC-PCA, we quantified the mean distance in the first two principal components across brain regions, excluding from the AC-PCA one region at a time. Due to the relative transcriptomic uniqueness of the CBC, its exclusion unmasked a qualitatively distinct and pronounced cup-shaped pattern with a transition beginning prior to birth and spanning the late fetal period and early infancy (Fig. 2F). CBC was again the most distinct region of the brain following multidimensional scaling analysis for expressed mature miRNAs, a small RNA species enriched within our small RNA-Seq dataset, and the dominant contributor to miRNA expression variance (fig. S17).

The global late fetal transition and overall cup-shaped developmental dynamics we observed were also apparent when this analysis was repeated for the 11 neocortical areas included in this study (Fig. 3A and fig. S16). We observed greater dissimilarity across areas at early fetal ages (Fig. 3A), with prefrontal areas (medial prefrontal cortex [MFC], orbital prefrontal cortex [OFC], DFC, and ventrolateral prefrontal cortex [VFC]) being the most distinct. In addition, reflecting the spatial and functional topography of the neocortex, both rostro-caudal and dorsal-ventral axes were evident in the transcriptome during fetal development. Areal differences were also seen at later ages, with functional considerations likely taking precedence over topographical arrangements. For example, VFC clustered closely with primary motor (M1C) and somatosensory (S1C) cortex, likely reflecting functional relationships with orofacial regions of the motor and somatosensory perisylvian cortex (fig. S16). Across the entirety of human brain development, the transcriptomic similarity between cortical regions also showed a pronounced decrease centered on

the late fetal and early infancy samples of W5 (i.e., perinatal window), again reminiscent of a cup-shaped pattern (Figs. 3A, 3B and fig. S16).

Similar to gene expression, global measures of alternative splicing, such as the ratio between reads including or excluding exons (i.e., the percent spliced in index, PSI), were higher during prenatal than postnatal ages (fig. S18 and table S7). So too was the gene expression of 68 RNA-binding proteins selected due to their involvement in RNA splicing and their analysis in adulthood by the Genotype-Tissue Expression (GTEx) project (*24*). Hierarchical clustering of expression data for these proteins also revealed a late fetal transition (fig. S19). Coincident with these observations, we found that genes exhibiting the highest inter-regional variation in expression in any given window (see (*22*)) exhibited a higher PSI during that window than iteratively chosen control groups of genes (fig. S18). Taken together, these analyses suggest broad phenomena in the developing human brain, including a late fetal transition in intra- and inter-regional transcriptomic variation, may be amplified by alternative splicing.

Cellular heterogeneity and developmental dynamics

The high inter-areal variation observed during embryonic and early-to-mid fetal development (Fig. 3B) coincides with a crucial period in neural development and the suspected etiology of psychiatric diseases *(4)*. To help understand the temporal dynamics underlying this variation in gene expression, we analyzed our scRNA-Seq data from embryonic front-parietal neocortical wall and mid-fetal fronto-parietal neocortical plate and adjacent subplate zone, alongside our snRNA-Seq data from adult human neocortex and other independent datasets from overlapping developmental time points (*12, 25, 26*). To do so, we first applied a clustering and classification algorithm (*27, 28*) to the prenatal scRNA-Seq data following an initial division of the dataset based on age of the

donor brain (i.e. embryonic or fetal), obtaining 24 transcriptomically distinct cell clusters. Reflecting the rapid developmental change occurring across embryonic and fetal development and the relative homogeneity of cell type composition as compared to adult ages, as well as the specific distribution of samples in our dataset, a number of these clusters were comprised of cells from only a single donor brain, and vice versa. Suggesting this resulted from spatiotemporal changes across brain development rather than artifactual changes related to data processing, we confirmed broad classifications of individual cells and general relationships between cell clusters and donor brains using an alternative clustering algorithm (fig. S21). Differential expression analysis and measurements of expression specificity recovered well-known gene markers of distinct types of neuronal and non-neuronal progenitor and postmitotic cell types (figs. S20, S22 and table S8), as well as closely related groups of cell types (i.e., markers enriched in all prenatal excitatory neuron clusters) (fig. S22).

We complemented these data with snRNA-Seq from adult human DFC (fig. S20), where we identified 29 transcriptomically distinct cell clusters representing various populations of glutamatergic excitatory (ExN) projection neurons, GABAergic interneurons (InN), oligodendrocyte progenitor cells, oligodendrocytes, astrocytes, microglia, endothelial cells, and mural cells (i.e., pericytes and vascular smooth muscle cells) (fig. S21). Alignment of our prenatal data with adult snRNA-Seq data revealed hierarchical relationships and similarities between major cell classes, reflecting their developmental origins and functional properties (fig. S23). Notably, putative embryonic and fetal excitatory neurons clustered near, but did not wholly overlap with their adult counterparts. We also observed transient transcriptomic entities, such as fetal cells in the oligodendrocyte lineage that clustered separately from their adult counterparts. Similarly, nascent excitatory neurons generally did not cluster with progenitor cells nor with fetal or adult

excitatory neurons, indicating their maturationally distinct status. Confirming the validity of our prenatal scRNA-Seq and adult snRNA-Seq data, alignment of our prenatal data with cells from a previously published dataset (*9*) consisting of mid-fetal and adult human neocortical cells yielded similar relationships between prenatal and adult cell types (fig. S23). Comparison of neuronal transcriptomes from our prenatal single cells with both our adult single nucleus data and independently generated adult single nucleus data (*27*) also confirmed key differences between embryonic, mid-fetal, and adult populations. We observed limited transcriptional diversity in embryonic and mid-fetal excitatory and inhibitory neuron populations in the neocortex as compared to the adult counterparts. The clusters identified in our prenatal dataset did not express specific combinations of marker genes described for the adult excitatory (fig. S24) and inhibitory (fig. S25) neurons. For example, the embryonic and mid-fetal neocortical excitatory neurons expressed combinations of genes known to be selectively enriched in different layers in adult human or mouse neocortex (*29-31*), as previously shown in the prenatal human and mouse neocortex (*12, 31*). Notably, genes enriched in adult excitatory projection neuron subtypes located in layer (L) 5 and L6, such as *BCL11B* (*CTIP2*) and *FEZF2* (*FEZL*, *ZFP312,* or *ZNF312*), were co-expressed with L2-4 intracerebral excitatory projection neuron markers, such as *CUX2,* in certain embryonic and mid-fetal excitatory cell types (figs. S24, S26). We also observed temporal changes in the co-expression patterns of cell type-specific marker genes in other cell types. For example, single cell data from mid-fetal NCX revealed frequent co-expression of *RELN*, a marker for L1 Cajal-Retzius neurons (*32*), and *PCP4* (75.9% of 133 *PCP4*-expressing cells; RPKM >=1), a marker previously shown to be expressed by deep layer excitatory neurons (*33*). In contrast, analysis of snRNA-Seq suggested only sporadic co-expression of these genes (10.8% of 6084 *PCP4*-expressing cells, UMI >=1) in the adult human DFC. Subsequent immunohistochemistry

on independent specimens confirmed the robust co-expression of these genes in L1 of the prenatal cortex, but not in L1 of the adult cortex or in other cortical layers (fig. S26). These data imply that the molecular identities of many neuronal cell types are not fully resolved before the end of mid-fetal development and are likely malleable during early postmitotic differentiation.

Next, we utilized our single cell/nucleus datasets to deconvolve bulk tissue mRNA-Seq samples and estimate temporal changes in the relative proportions of major cell types in the neocortex. The combined analysis revealed the cellular architecture of distinct neocortical areas and their variations across development. We observed temporal changes in cellular composition and maturational states, including the most dramatic changes during a late fetal transition (Figs. 3C, 3D, 3E). For example, transcriptomic signatures for fetal excitatory neurons and fetal interneurons were generally inversely correlated with progenitor cell signatures during embryonic and early fetal development, but fetal neuron signatures nonetheless decreased across mid-to-late fetal development despite a concomitant reduction in the progenitor cell signature, an observation that was likely affected by our dissection strategy (Fig. 3C, (*22*)). Similarly, signatures for adult excitatory neurons increased rapidly across the late fetal period and early infancy, coincident with the decrease in signatures of fetal excitatory neurons and interneurons (Fig. 3C). As expected, the molecular signatures for early born, deep layer excitatory neurons preceded those for late born, upper layer excitatory neurons (fig. S27). Transcriptomic signatures for prenatal oligodendrocytes and prenatal astrocytes also began to emerge during mid-fetal periods and increased rapidly across the late fetal transition and early infancy (Fig. 3C). Demonstrating the robustness of these observations, independent deconvolution using two alternate fetal single cell datasets (*12, 26*) yielded similar results (figs. S27, S30).

Given the increase in adult cell type signatures during W5, we next reasoned that the observed decrease in inter-regional transcriptomic divergence during late fetal periods and infancy may reflect a synchronized transition from fetal to more mature features of neural cells. Consequently, we analyzed the variance in cell type specific signatures across neocortical areas, which varies in accordance with their relative proportion, and found that the maximum cell type inter-areal variation through time recapitulated the developmental cup-shaped pattern (Fig. 3D), with large variation in the proportion of neural progenitor cells and fetal excitatory neurons (figs. S28, S29). Beginning during early postnatal periods, we observed increased proportions and variance in the signatures of astrocytes and, by adulthood, mature excitatory neurons (Fig. 3E). These observed temporal differences in the magnitudes and variances of the relative proportions of certain cell types and the global heterogeneity of the cell-type composition at each window likely at least partially explain the observed pattern of inter-areal differences across development. Gene Ontology (GO) enrichment analysis using the top variant genes in each window, with all genes expressed in each window as background, provided further support for these changes in cell composition across areas and time. Commensurate with the changes we observed in discrete cell populations, biological processes including neurogenesis in early developmental windows (W3-4), myelination in the perinatal window (W5), and sensory/ion activity calcium-related biological processes in later postnatal windows (W7-9), among others, exhibited regional variation in the global brain transcriptome (fig. S31 and table S9). Similar patterns of inter-regional variation involving discrete cell types were also observed in the macaque neocortical transcriptome (*34*), indicating that these are conserved and consistent features of prenatal primate neocortex.

Other lines of evidence also suggested pronounced and qualitatively distinct regional differences in myelination, synaptic function, and neuronal activity. For example, although we

observed differences in the expression of genes associated with these processes (*10*) across the neocortex (fig. S31 and table S9), *TempShift*, a Gaussian-based model that allows the quantification of temporal shifts in the trajectories of groups of genes represented by their first principal components (*34*), indicated that of these processes only genes associated with myelination displayed such a shift (Fig. 4A). Conversely, perhaps reflecting functional or areal diversity in cell subtypes, we observed no similar temporal shift in the expression of genes associated with synaptogenesis or neuronal activity, confirming these results through reference to published post-translational analyses of myelinated fiber density (*35*) and synaptic density (*36*) conducted across multiple neocortical areas (Fig. 4B). Crucially, although genes associated with these processes were expressed across the late fetal transition (Fig. 4C), of the processes analyzed, only myelination contributed to the increased inter-areal differences we observed during this period (Fig. 4D). Suggesting these differences are a conserved feature of primate development, we also observed similar areal differences in the transcriptional signatures of oligodendrocytes in the macaque neocortex.

Overall, these observations indicate that higher levels of divergence during early prenatal and later postnatal development reflect regional variations in cell type composition, likely arising from topographical variation in progenitor populations and neuron development during prenatal ages and cell type and functional diversification during later postnatal ages.

Spatio-temporal and multimodal integration

We next sought to assess temporal variation in epigenetic signatures and their relationships to gene expression, development, and biological processes. Global DNA methylation profiling revealed that the majority of CpG loci were either hyper (37.5%; Beta value [β]>=0.8) or hypo (31.8%;

ß<=0.2) methylated in at least one sample (fig. S32), but only approximately 10% of the tested methylation sites were progressively hyper- or hypo-methylated through prenatal windows, postnatal windows, or both. Similarly, a majority of methylation sites also exhibited regional variation, with 64% of tested sites differentially methylated between at least two brain regions at postnatal ages. Additionally, 16% of tested sites were differentially methylated between at least two neocortical areas. Conversely, a majority of putative promoters (66%) and a substantial proportion of putative enhancers (43%) were not differentially enriched between DFC and CBC at either fetal or adult ages. However, a greater proportion of putative enhancers (H3K27ac-enriched regions not overlapping H3K4me3-enriched regions or proximal to a transcription start site; TSS) were regionally (15%), temporally (17%), or spatiotemporally (24%) enriched than putative promoters (8%, 14%, and 12%, respectively). These differences, which suggest a greater role for enhancers relative to promoters in contributing to differential spatiotemporal gene expression, were selectively validated using quantitative droplet digital PCR (fig. S10). We next explored correlations between methylation, histone modifications, and gene expression (figs. S32 to S34). In the adult, we found that TSSs that were more highly methylated were associated with genes that were expressed at low levels at the corresponding age, and vice versa. These relationships were not strongly indicated for methylation at other locations in the gene body (fig. S32). The presence of CBC-enriched H3K4me3 and H3K27ac marks in the adult human brain also correlated strongly with increased gene expression in CBC relative to DFC (fig. S33), and vice versa. Similarly, putative fetal-active and adult-active enhancers were associated with higher fetal or adult gene expression, respectively.

In addition to epigenetic effects on gene expression, we observed discrete relationships between specific enhancers, methylation sites, and cell type specific signatures. For example,

enhancers identified during the fetal period were enriched for methylation sites that were progressively more methylated across postnatal ages (post-up) while adult-active enhancers were enriched for methylation sites that were progressively less methylated across postnatal ages (post-down) (Fig. 5A and fig. S35**,** (*22*)). Both post-up and post-down sites were themselves depleted at TSSs and enriched for sites undermethylated in neurons (neuron undermethylated sites, or NUM sites) and undermethylated in non-neurons (non-NUM sites) (fig. S35). They were also enriched for fetal and adult enhancers, respectively (Fig. 5B). Post-up sites were also enriched in both neuron and glia-enriched-genes, while post-down sites were enriched only in glial genes (Fig. 5B). Further suggesting a relationship between enhancer activity, methylation, and cell type, genes associated with fetal-active enhancers, as well as those associated with differentially methylated regions (DMRs) composed of post-up sites (*22*), were enriched for Gene Ontology terms related to early events in neural development, such as neurogenesis, cell differentiation, and synaptic transmission, but generally not for processes occurring later in development (Fig. 5B and fig. S35). In contrast, genes near adult-active enhancers and post-down DMRs exhibited enrichment for postnatal or adult processes including myelination and axon ensheathment (Fig. 5B and fig. S35). Taken together, these data demonstrate relationships between gene expression and epigenetic modifications including methylation status and putative regulatory elements, as well as signatures of specific cell types and developmental programs.

We next sought further evidence that cellular dynamics contributed to the late fetal transition through the analysis of cell type- and spatiotemporal-specific patterns of gene expression and epigenetic regulation. We curated 73 gene co-expression modules resulting from Weighted Gene Correlation Network Analysis (WGCNA) according to spatial relationships between brain regions and the temporal relationships of gene expression in the neocortex across the late fetal

transition (Fig. S36 and tables S10, S11). We found 44 modules that showed expression differences among regions in the brain (spatial), 40 modules that showed expression differences between prenatal and postnatal neocortical areas (temporal), 16 modules that were neither spatially nor temporally dynamic, and 27 modules that exhibited both spatial and temporal differences (Fig. 5C). A significantly greater than expected number of these spatiotemporally dynamic modules (including Modules 2, 10, 32, and 37) exhibited their greatest change in neocortical expression from W2 through W5 ($P<0.0118$, Hypergeometric test) (Fig. 5C, fig. S37 and table S12). Genes whose expression was enriched in excitatory neurons, genes associated with putative fetal-active enhancers, and/or genes associated with NUM sites, a selection of characteristics we refer to collectively as Neuronal (N)-Type associations, were also enriched in spatiotemporal dynamic modules ($P<0.0029$, Hypergeometric test) (Fig. 5C, fig. S37 and table S12). Conversely, genes associated with adult-active enhancers, methylation sites hypomethylated in non-NUM sites, and glial genes (Glial or "G"-Type modules/associations in Fig. 5C, fig. S37 and table S12) were enriched among the 13 modules where temporal ($P<0.0002$, Hypergeometric test), but not spatial, specificity was observed. These observations indicate increased spatial diversity of neuronal cell types relative to glial cell populations.

Analyses by sex revealed that modules enriched for the 783 genes exhibiting sex-differential expression (sex-DEX, fig. S37) in at least two consecutive windows in at least one brain region were enriched among modules with no spatial or temporal differential expression in the neocortex ($P<0.0029$, Hypergeometric test) (Fig. 5C) and depleted among spatiotemporal modules ($P<0.0021$, Hypergeometric test). There were four modules exhibiting temporal expression differences in the neocortex that were also enriched for sex-biased genes, as well as glial and other cell type-enriched markers, but this did not represent a significant enrichment in

sex-DEX enriched modules among temporal modules (P<0.132, Hypergeometric test). In addition, no module comprised of autosomal genes exhibited persistent male- or female-dimorphism across both prenatal development and later postnatal ages such as adolescence or adulthood (Figs. S38, S39); in cases where an autosomal module was sex-DEX throughout development, the sex exhibiting higher expression reversed between early and late postnatal development (fig. S39). This observation was upheld when multiple thresholds were used for the identification of sexual dimorphism (fig. S40). Similarly, we identified no autosomal genes that exhibited sexual dimorphism throughout development in all brain regions or neocortical areas (figs. S38, S39).

Cellular and temporal convergence of neuropsychiatric disease risks

Loci implicated in several neuropsychiatric disorders have been identified through genome-wide association studies (GWAS) and are enriched in putative noncoding regulatory elements (*29*-*31*). We sought to determine whether the proportion of phenotypic variance explained by common SNPs in large neuropsychiatric GWAS (i.e., SNP heritability) was enriched in the *cis*-regulatory elements we identified at W1, W4, W5 and W9 in DFC and CBC. Towards this end, we collected GWAS data concerning neuropsychiatric disorders or personality traits including schizophrenia (SCZ) from CLOZUK (*37*), Alzheimer's disease (AD) from IGAP (*38*), Parkinson's disease (PD; (*39*)), autism spectrum disorder (ASD; (*40*)), attention deficit hyperactivity disorder (ADHD) from iPSYCH (*41*), major depression disorder (MDD; (*42*)), bipolar disorder (BD; (*43*)), IQ (*44*), and neuroticism (*45*), as well as non-neural traits such as height from GIANT (*46*), inflammatory bowel disease (IBD; (*47*)), total cholesterol levels (*48*), and an endophenotype associated with diabetes (HbA1C; (*49*)). Using partitioned LD score regression (pLDSC) analysis, we found that SNP heritability in SCZ, intelligence quotient (IQ), and neuroticism were exclusively enriched in

DFC-specific, but not CBC-specific, regulatory elements as identified by peak regions of H3K27ac activity. In contrast, SNP-heritability in AD or PD rendered no significant associations, and the analysis on ASD, ADHD, BD, and MDD was only nominally enriched or not enriched in putative region-specific fetal enhancers (Fig. 6 and fig. S41, (*22*)). Non-neural traits (such as height and diabetes) were also not enriched in either DFC- or CBC-specific regulatory elements, but were instead enriched in regulatory elements active in the two brain regions (fig. S41), indicating a general enrichment of many of our tested GWASes in H3K27ac regions when considering a set of more ubiquitous regulatory regions.

After aggregating GWAS SNPs and identifying candidate associated regions on the basis of their P-values and linkage disequilibrium patterns in individuals of northwest European ancestry (*50*), we next leveraged partially overlapping Hi-C datasets, derived from mid-fetal and adult neocortex and processed by two independent research groups (*51-53*), as well as H3K27ac activity in the brain, to develop two lists of genes putatively associated with those GWAS-associated regions. To do so, we initially populated both lists of disease-associated genes by identifying TSSs overlapping H3K27ac peaks that themselves overlapped a GWAS significant region, as well as genes directly affected by GWAS significant variants within the LD-region, as predicted by EnsmbleV78. We next expanded these lists of disease-associated genes by identifying TSSs that interact with H3K27ac peaks overlapping GWAS significant regions, excluding interactions that did not overlap with at least one H3K27ac peak at each end or where peak-to-peak interactions were not concordant in time and brain region. In the first, less stringent list (List 1), a single interaction from either of the two Hi-C datasets was sufficient to associate a gene to a GWAS locus (table S13). For the second, more stringent list (List 2), we excluded those genes whose only

association to a GWAS locus was via Hi-C interactions identified in only one of the two Hi-C datasets (table S14).

We next sought to determine the cell types enriched for the expression of the high-stringency genes implicated in neuropsychiatric disorders or brain-based traits, using our prenatal scRNA-Seq and adult snRNA-Seq datasets, and matching prenatal and adult datasets generated from the macaque (*34*). We found numerous cell types enriched for disease-associated loci in both human and macaque (fig. S42). For example, neocortical excitatory neurons were enriched for the expression of genes we associated with IQ in both the fetal and adult human as well as the fetal and adult macaque. However, we found no other excitatory neuron populations in the macaque amygdala, striatum, hippocampus, thalamus, or cerebellum enriched for genes associated with IQ. Similarly, neural progenitors in the prenatal macaque amygdala, but not progenitors in the prenatal macaque hippocampus, thalamus, neocortex, or striatum, were enriched for the expression of genes associated with major depressive disorder, a finding especially intriguing given the variable or potentially increased size of some amygdalar nuclei in MDD patients (*54, 55*). Similarly confirmatory was the enrichment of schizophrenia risk genes in cortical excitatory neurons (*56*), with enrichment also observed in embryonic/fetal progenitor cells and adult cortical interneurons.

Analysis of gene co-expression modules found that genes in the more stringent early onset disease (ADHD, SCZ, and MD) risk lists converged on 7 of 73 co-expression modules while adult onset disease (AD and PD) risk gene lists converged on 5 partially overlapping modules (fig. S37 and table S12). 8 of these 10 total disease-associated modules (Fig. 7A) exhibited spatiotemporal or temporal specificity, and all modules exhibited their greatest spatiotemporal change either during W2 or W5 (fig. S37). A significant number of modules associated with adult onset disorders were enriched for signatures of glial gene expression ($P<0.0266$, Hypergeometric test, table S12),

and of particular interest were modules ME3 and ME7, which in addition to glial signatures were enriched for non-NUM sites, adult-active enhancers, sex-DEX genes, and AD-associated risk genes (Fig. 7A).

Another module of interest was ME37, a module of 145 genes enriched for NUM sites and fetal enhancers, and whose expression was enriched specifically in neurons as opposed to neural progenitors or glia. ME37 was also exceptional for its disease association, as it was enriched for genes associated with SCZ, IQ, and neuroticism but not for non-neurological characteristics such as height or a diabetes-related trait (Fig. 7A). Complementary module-based association analysis with Multi-marker Analysis of GenoMic Annotation (MAGMA), testing for an enrichment in association to disease specifically around genes in any given module, confirmed enrichment for SCZ, IQ, and neuroticism in ME37 (MAGMA-P values<0.01; FDR for all traits and modules <0.3) (table S11). At the gene level, multiple genes in ME37 identified using our less stringent criteria for interaction were associated with up to four or more different traits and disorders, including *MEF2C*, *ZNF184*, *TCF4,* and *SATB2,* all genes critical for neurodevelopment and/or implicated in neurodevelopmental disorders (*57-65*) (Figs. 7B, 7C). We also found that ME37 was specifically enriched in clusters of excitatory neurons in the fetal and adult neocortex (Fig. 7D), and further analysis of adult excitatory neuron populations identified in this study and an independent database of adult single cell data (*27*) suggested this enrichment was selective for deep layer neocortical neurons (fig. S43).

As the ASD GWAS resulted in only 13 significant genes, eight of which were non-protein coding, and because *de novo* germline mutations are known to significantly contribute to ASD risk (*66*), we next developed two non-overlapping lists of neurodevelopmental disorders (NDD; ASD, intellectual disability (ID), and developmental delay (DD)). The first list was comprised of 65

high-confidence ASD risk genes (hcASD) associated with *de novo* mutations (*66*). The second list included all ASD genes documented in the SFARI database (www.gene.sfari.org) under categories "Syndromic" or with scores from 1 to 4 as well as an independent list of genes associated with DD (*67*), with genes overlapping the hcASD list removed. We found that these genes were also significantly enriched in ME37 (FDR <0.0001, Fisher's exact test), and commensurate with the cell type enrichment found in ME37, the expression of genes in both of these lists was also enriched in several clusters of fetal and adult excitatory neurons identified in our single cell dataset (Fig. 7D). Medium spiny neurons in the striatum, a population that has also been previously linked to ASD (*68*), were also enriched for the expression of ASD risk genes in the prenatal macaque (Fig. 7D).

We finally studied the overlap between WGCNA modules and modules significantly enriched in differentially expressed genes in postmortem brains from patients of SCZ, BD, and ASD (*69*). Interestingly, we found little overlap between modules enriched in genes exhibiting postmortem differences in expression between SCZ, BD, or ASD, as compared to neurotypical controls, and modules enriched in GWAS-risk genes for these same disorders. Emphasizing the necessity of studying neurotypical brain development, these observations may suggest a decoupling between the primary genetic causes of some neurological or psychiatric disorders and second order effects manifesting as changes in gene expression months or years following disease onset.

Discussion

In this study, we have presented a comprehensive dataset and a multi-platform functional genomic analysis of the developing and adult human brain. The presence of these multiple data modalities

in a unified resource, and largely from the same tissue samples, allows the integration of information spanning prenatal and postnatal human brain development. Resource description and access are available at development.psychencode.org and brainspan.org.

Although transcriptomic differences between distinct brain regions remain across time, they are developmentally specified and exhibit an overall cup-shaped pattern centered on a late fetal transition following a period of high intra- and inter-regional variation during embryonic and early/mid-fetal development. Multiple analyses of distinct transcriptomic features all confirm this transition begins well prior to birth. Our complementary transcriptomic study of the developing rhesus macaque brain (*34*) also revealed a similar global developmental pattern, with a first transition beginning prior to birth, indicating that this is a conserved feature of catarrhine primate neurodevelopment and not due to an artifact of difficult to acquire samples from late fetal and early postnatal development. Such a phenomenon is consistent with previously observed differences in transcriptomic and methylomic profiles of mid-fetal and postnatal human neocortex (*17-20*), and coincident with processes involved in region-specific cell type generation, differentiation, and maturation (*2*). Crucially, this transition is strikingly distinct from previously reported phylogenetic hourglass-like patterns that occur during the embryonic organogenetic period in several invertebrate and vertebrate species (*70, 71*). Moreover, the developmental (ontogenetic) cup-shaped pattern we observe coincides with an "evolutionary" (phylogenetic) cup-shaped pattern, where developmental periods exhibiting high levels of inter-regional differences (for example, early to mid-fetal periods) also exhibit less conservation in gene expression patterns between human and macaque (*34*).

Among the processes becoming prominent during the late fetal period are astrogliogenesis, synaptogenesis, dendritogenesis, and neuronal activity. In contrast to a previous report of robust

areal differences in the progression of synaptogenesis during the same time period in humans (*36*), this and an accompanying study (*34*) found that genes associated with these processes exhibit largely synchronous expression trajectories across the developing neocortex in both humans and macaque. However, myelination, which sharply increases during late fetal development, peaks after birth, and extends through childhood and adolescence (*72*), is temporally asynchronous. This asynchronicity in oligodendrocyte development and myelination is not apparent at the level of OPCs, which suggests that the maturation of OPCs into myelinating oligodendrocytes is a process with a variable onset and pace across areas. Similar observations were made in macaque (*34*), indicating that this may be another conserved catarrhine feature.

Transcriptomic variation may reflect several unique cellular and maturational reorganizational events. For example, as first described by Brodmann (*73*), an ontogenetic six-layered Grundtypus foreshadows the adult neocortex and transiently transforms the entirety of the neocortical plate beginning in the late fetal period, or in our window 5. Furthermore, consistent with the extensive changes we observed in the cerebellar transcriptome during late fetal development and early postnatal ages, cerebellar granule cells, a cell type that represents approximately 2/3 of all neurons in the brain, are also generated predominately during this period (*74*). The late fetal transition may therefore follow an inflection point after which developmental and spatiotemporal transcriptomic variations are transiently consolidated in advance of the emergence of cellular and functional differences between adult brain regions.

The mid-fetal period of high intra- and inter-regional divergence that immediately precedes the late fetal transition also coincides with a key developmental period previously associated with the etiology of ASD and SCZ (*63, 65, 75*). Consequently, understanding the developmental and evolutionary history of this period may be essential for understanding neuropsychiatric disease.

Integrating our multiple data modalities with gene co-expression modules allowed us to organize and characterize the whole brain developmental transcriptome and identify modules with dynamic spatiotemporal trajectories, many of them showing a sharp late-fetal transition, and enrichment in specific cell types, epigenetic activity, and disease-associated genes. Of particular interest is ME37, a module displaying the greatest rate of change in the neocortex within the late-fetal transition and in which putative risk genes for ASD, NDD, SCZ, IQ, and neuroticism converged. Several of the genes in ME37 were implicated by our study in multiple disorders and traits, and have been linked previously to neurodevelopment and human disease. For example, *MEF2C* controls activity-dependent expression of neuronal genes, including those linked to synapse function and ASD (*61, 63*), and *Mef2c* mutant mice display numerous behaviors reminiscent of ASD, ID, and SCZ (*58*). Similarly, *TCF4* regulates key neurodevelopmental processes, such as neurogenesis and synaptic plasticity, DNA methylation, and memory function processes (*62, 64*). Moreover, mutations in both *MEF2C* and *TCF4* result in intellectual disability in humans (*57, 59, 60*). Numerous other genes in this module are similarly involved in neurodevelopment, have been implicated in human brain disease, and are highly plausible disease risk genes and potentially therapeutic candidates. For example, *NR4A2*, another transcription factor in ME37 that we linked to neuroticism and IQ, has been linked to ASD and SCZ, among other disorders. Our study also links the transcription factor *TSHZ3* to neuroticism and IQ, and previous efforts have linked murine *Tshz3* to ASD and the fetal development of cortical excitatory projection neurons (*76*), a cell type and developmental period also implicated in ASD (*63, 65*). Other genes in ME37, such as *SATB2*, *FEZF2*, *SOX5*, and *TBR1*, play critical roles in the development of cortical excitatory projection neurons and are mutated in NDDs (*29-31, 65, 77, 78*). Similarly, the population of genes included in ME37, as well as genes linked to ASD and NDD, also exhibit regional and cell type specific

convergence in neocortical excitatory neurons. Moreover, the identification of ME37 and the overlap of genes in this module with those implicated in ASD and NDD illustrates how disease association signals from common variants unveiled by GWAS for any given neuropsychiatric disorder can identify genes that have also been associated with the etiology of a different disease through the study of *de novo* mutations in patient populations (*76*). While not every gene in ME37 is likely to contribute to neuropsychiatric disease etiology, the coincident enrichment within this module of genes associated with multiple disorders or neurological traits, along with the multitude of genes in this module that are associated directly, suggests that neuropsychiatric disease might be considered through a broader lens encompassing additional aspects of brain dysfunction.

Interestingly, there is little overlap between the risk-gene associated modules we identified and modules enriched in genes that are differentially expressed in postmortem brains of SCZ, ASD, and BD, as compared to controls (*69*). This comparison may help discriminate gene networks that are primary causes from those that are secondary or reactive in these neuropsychiatric disorders while emphasizing the importance of studying disease in the context of neurotypical development.

Taken together, these observations demonstrate the utility of this resource to perform integrated analysis for the understanding of brain development and function, and for the rapid interpretation of findings from neuropsychiatric genomics.

Materials and Methods

A full description of the materials and methods is provided in the supplementary text. Briefly, we precisely dissected multiple brain regions (hippocampus, striatum, amygdala, cerebellum, thalamus, and 11 neocortical areas) in more than 60 postmortem human brains ranging in age from

5 PCW to 64 PY. We then applied bulk tissue RNA-Seq, single cell and nucleus RNA-Seq, small RNA-Seq, DNA methylation assay, or ChIP-Seq to generate multimodal datasets, often from the same brain. After applying stringent quality control checks and independent analysis of each data set, we performed integrated analyses to gain insights into human brain development, function, and disease.

CONSORTIUM AUTHORS

The BrainSpan Consortium

(team members are listed in alphabetical order of the last name after PI)

Steering committee

James A. Knowles (Co-chair), Ed S. Lein (Co-chair), Nenad Sestan (Co-chair), Mark B. Gerstein, Michael J. Hawrylycz, Pat Levitt

Tissue procurement and processing group

Yale University: Nenad Sestan (Principal Investigator), Anita Huttner, Zeljka Krsnik, Mihovil Pletikos, Goran Sedmak, André M. M. Sousa

Lieber Institute for Brain Development: Joel E. Kleinman (Principal Investigator), Tom M. Hyde, Richard E. Straub, Daniel R. Weinberger

Newcastle University: Susan Lindsay (Principal Investigator), Steven N. Lisgo

Allen Institute for Brain Science: Ed S. Lein (Principal Investigator), Amy Bernard, Nick Dee, Song-Lin Ding, Zack L. Riley, Elaine H. Shen, Susan M. Sunkin

Transcriptome data production group

Yale University: Nenad Sestan (Principal Investigator), Matthew B. Johnson, Hyo Jung Kang, Yuka Imamura Kawasawa, Richard P. Lifton, Shrikant Mane, Christopher E. Mason, James P. Noonan

University of Southern California: James A. Knowles (Principal Investigator), Sandra Abramowicz, Oleg V. Evgrafov, Reaghan Gittin, Teresa Gomez, Hee Jae Yoon

Genome and epigenome data production group

University of Southern California: James A. Knowles (Principal Investigator), Maxine M. Chen, Zemin Deng, Oleg V. Evgrafov, Jessica Stoll, Hee Jae Yoon

Yale University Nenad Sestan (Principal Investigator), Hyo Jung Kang, Yuka Imamura Kawasawa, Yurae Shin

Transcriptome, epigenome and genome data analysis group

Yale University Nenad Sestan (Principal Investigator), Mark B. Gerstein (Principal Investigator), Feng Cheng, Lukas Habegger, Arif Harmanci, Robert R. Kitchen, Jing Leng, Mingfeng Li, Shuang Liu, Joel Rozowsky, Andrea Sboner, Daifeng Wang, Xuming Xu, Ying Zhu; Zhixiang Lin, Hongyu Zhao

University of Southern California: James A. Knowles (Principal Investigator), Ting Chen, Yang-Ho Chen, Ann Chervenak, Andrew Clark, Ewa Deelman, Oleg V. Evgrafov, Jennifer Herstein, Rajiv Mayani, Gaurang Mehta, Tade Souaiaia

University of California, Los Angeles: Daniel H. Geschwind (Principal Investigator), Neelroop Parikshak

Michigan State University: Mark Reimers (Principal Investigator), Paul Manser

High throughput histological gene expression group

Allen Institute for Brain Science: Ed S. Lein (Principal Investigator), Kaylynn Aiona, James Arnold, Crissa Bennet, Amy Bernard, Darren Bertagnolli, Angie L. Guillozet-Bongaarts, Kristina Brouner, Stephanie Butler, Shiella Caldejon, Anita Carey, Christine Cuhaciyan, Rachel A. Dalley, Suvro Datta, Nick Dee, Tsega Desta, Song-Lin Ding, Amanda Ebbert, Michael Fisher, Erich Fulfs, Garrett Gee, Jeff Goldy, Lindsey Gourley, Ben W. Gregor, John G. Hohmann, Robert E. Howard, Jayson M. Jochim, Allan R. Jones, Chang-Kyu Lee, Tracy A. Lemon, Naveed Mastan, Jeremy A. Miller, John Morris, Nerick F. Mosqueda, Nhan Kiet Ngo, Julie Nyhus, Aaron Oldre, Eric Olson, Jody Parente, Patrick D. Parker, Sheana E. Parry, Julie Pendergraft, John W. Phillips, Melissa Reding, Zack L. Riley, Katie Roll, David.Sandman, Melanie Sarreal, Amanda Sekijima, Sheila Shapouri, Nadiya V. Shapovalova, Elaine H. Shen, Cliff R. Slaughterbeck, Kimberly A. Smith, Michael Smith, Susan M. Sunkin, Derric Williams, Paul Wohnoutka

Public resource group

Allen Institute for Brain Science: Ed S. Lein (Principal Investigator), Michael J. Hawrylycz (Principal Investigator), Chinh Dang, Tim A. Dolbeare, David Feng, Tim Fliss, Guangyu Gu, Chihchau L. Kuan, Chris Lau, Felix Lee, Lydia Ng, Nathan Sjoquist, Elaine H. Shen, Aaron Szafer, Andy J. Sodt, Susan M. Sunkin

The PsychENCODE Consortium

(members are listed in alphabetical order)

Executive committee

Alexej Abyzov (Mayo Clinic Rochester), Schahram Akbarian (Icahn School of Medicine at Mount Sinai), Gregory E. Crawford (Duke University), Stella Dracheva (Mount Sinai), Peggy J. Farnham (University of Southern California), Mark Gerstein (Yale University), Daniel H. Geschwind (University of California, Los Angeles), Fernando S. Goes (Johns Hopkins University), Thomas M. Hyde (Lieber Institute for Brain Development), Andrew E. Jaffe (Lieber Institute for Brain Development), James A. Knowles (SUNY Downstate Medical Center), Chunyu Liu (SUNY Upstate Medical University), Angus C. Nairn (Yale University), Junmin Peng (St. Jude Children's Hospital), Dalila Pinto (Icahn School of Medicine at Mount Sinai), Panos Roussos (Mount Sinai), Stephan J. Sanders (University of California, San Francisco), Nenad Sestan (Yale University), Matthew W. State (University of California, San Francisco), Patrick F. Sullivan (University of North Carolina - Chapel Hill), Flora M. Vaccarino (Yale University), Sherman Weissman (Yale University), Zhiping Weng (University of Massachusetts Medical School), Kevin P. White (The University of Chicago), Peter Zandi (Johns Hopkins University)

Data generation

Schahram Akbarian (Icahn School of Medicine at Mount Sinai), Majd Alsayed (The University of Chicago), Anahita Amiri (Yale University), Thomas G. Beach (Banner Sun Health Research Institute), Leanne Brown (Icahn School of Medicine at Mount Sinai), Mimi Brown (The University of Chicago), Adrian Camarena (University of Southern California), Becky C. Carlyle (Yale

University), Lijun Cheng (The University of Chicago), Adriana Cherskov (Yale University), Gregory E. Crawford (Duke University), Luis De La Torre Ubieta (UCLA), Diane DelValle (Icahn School of Medicine at Mount Sinai), Olivia Devillers (Icahn School of Medicine at Mount Sinai), Stella Dracheva (Mount Sinai), Elie Flatow (Icahn School of Medicine at Mount Sinai), Nancy Francoeur (Icahn School of Medicine at Mount Sinai), John F. Fullard (Mount Sinai), Michael J. Gandal (University of California, Los Angeles), Tianliuyun Gao (Yale University), Daniel H. Geschwind (University of California, Los Angeles), Gina Giase (SUNY Upstate Medical University), Paola Giusti-Rodriguez (University of North Carolina - Chapel Hill), Fernando S. Goes (Johns Hopkins University), Kay S. Grennan (SUNY Upstate Medical University), Evi Hadjimichael (Icahn School of Medicine at Mount Sinai), Chang-Gyu Hahn (University of Pennsylvania), Vahram Haroutunian (Icahn School of Medicine at Mount Sinai and James J Peters VA Medical Center), Gabriel E. Hoffman (Icahn School of Medicine at Mount Sinai), Thomas M. Hyde (Lieber Institute for Brain Development), Rivka Jacobov (Icahn School of Medicine at Mount Sinai), Andrew E. Jaffe (Lieber Institute for Brain Development), Yan Jiang (Icahn School of Medicine at Mount Sinai), Graham D. Johnson (Duke University), Bibi S. Kassim (Icahn School of Medicine at Mount Sinai), Joel E. Kleiman (Lieber Institute for Brain Development), Alexey Kozlenkov (Mount Sinai), Marija Kundakovic (Icahn School of Medicine at Mount Sinai), David A. Lewis (University of Pittsburgh), Zhen Li (Yale University), Barbara K. Lipska (Human Brain Collection Core, National Institutes of Health, Bethesda, MD), Chunyu Liu (SUNY Upstate Medical University), Jessica Mariani (Yale University), Daniel J. Miller (Yale University), Angus C. Nairn (Yale University), Mingming Niu (St. Jude Children's Hospital), Royce B. Park (Icahn School of Medicine at Mount Sinai), Junmin Peng (St. Jude Children's Hospital), Dalila Pinto (Icahn School of Medicine at Mount Sinai), Sirisha Pochareddy (Yale University), Damon

Polioudakis (University of California, Los Angeles), Amanda J. Price (Lieber Institute for Brain Development), Mohana Ray (The University of Chicago), Timothy E. Reddy (Duke University), Panos Roussos (Mount Sinai), Alexias Safi (Duke University), Shannon Schreiner (University of Southern California), Soraya Scuderi (Yale University), Nenad Sestan (Yale University), Annie W. Shieh (SUNY Upstate Medical University), Joo Heon Shin (Lieber Institute for Brain Development), Mario Skarica (Yale University), Lingyun Song (Duke University), Andre M.M. Sousa (Yale University), Valeria N. Spitsyna (University of Southern California), Patrick F. Sullivan (University of North Carolina - Chapel Hill), Vivek Swarup (University of California, Los Angeles), Anna Szekely (Yale University), Ran Tao (Lieber Institute for Brain Development), Flora M. Vaccarino (Yale University), Ramu Vadukapuram (SUNY Upstate Medical University), Xusheng Wang (St. Jude Children's Hospital), Yongjun Wang (Central South University), Maree J. Webster (Stanley Medical Research Institute), Kevin P. White (The University of Chicago), A Jeremy. Willsey (University of California, San Francisco), Jennifer R. Wiseman (Icahn School of Medicine at Mount Sinai), Heather Witt (University of Southern California), Hyejung Won (University of California, Los Angeles), Gregory A. Wray (Duke University), Mo Yang (Yale University), Peter Zandi (Johns Hopkins University), Elizabeth Zharovsky (Icahn School of Medicine at Mount Sinai)

Data analysis

Alexej Abyzov (Mayo Clinic Rochester), Schahram Akbarian (Icahn School of Medicine at Mount Sinai), Joon-Yong An (University of California, San Francisco), Christoper Armoskus (University of Southern California), Allison E. Ashley-Koch (Duke University), Judson Belmont (Icahn School of Medicine at Mount Sinai), Jaroslav Bendl (Mount Sinai), Tyler Borrman (University of

Massachusetts Medical School), Miguel Brown (The University of Chicago), Tonya Brunetti (The University of Chicago), Julien Bryois (Karolinska Institutet), Emily E. Burke (Lieber Institute for Brain Development), Becky C. Carlyle (Yale University), Chao Chen (Central South University), Adriana Cherskov (Yale University), Jinmyung Choi (Yale University), Declan Clarke (Yale University), Leonardo Collado-Torres (Lieber Institute for Brain Development), Gianfilippo Coppola (Yale University), Gregory E. Crawford (Duke University), Rujia Dai (Central South University), Stella Dracheva (Mount Sinai), Prashant S. Emani (Yale University), Oleg V. Evgrafov (SUNY Downstate Medical Center), Dominic Fitzgerald (The University of Chicago), Michael J. Gandal (University of California, Los Angeles), Tianliuyun Gao (Yale University), Melanie E. Garrett (Duke University), Mark Gerstein (Yale University), Daniel H. Geschwind (University of California, Los Angeles), Kiran Girdhar (Icahn School of Medicine at Mount Sinai), Paola Giusti-Rodriguez (University of North Carolina - Chapel Hill), Fernando S. Goes (Johns Hopkins University), Thomas Goodman (The University of Chicago), Mengting Gu (Yale University), Gamze Gürsoy (Yale University), Evi Hadjimichael (Icahn School of Medicine at Mount Sinai), Mads E. Hauberg (Mount Sinai), Jack Huey (University of Massachusetts Medical School), Thomas M. Hyde (Lieber Institute for Brain Development), Nikolay A. Ivanov (Lieber Institute for Brain Development), Andrew E. Jaffe (Lieber Institute for Brain Development), Yi Jiang (Central South University), Amira Kefi (University of Illinois at Chicago), Yunjung Kim (University of North Carolina - Chapel Hill), Robert R. Kitchen (Yale University), Alexey Kozlenkov (Mount Sinai), Mingfeng Li (Yale University), Zhen Li (Yale University), Chunyu Liu (SUNY Upstate Medical University), Shuang Liu (Yale University), Eugenio Mattei (University of Massachusetts Medical School), Daniel J. Miller (Yale University), Jill Moore (University of Massachusetts Medical School), Angus C. Nairn (Yale University), Fabio C. P. Navarro (Yale

University), Mingming Niu (St. Jude Children's Hospital), Junmin Peng (St. Jude Children's Hospital), Dalila Pinto (Icahn School of Medicine at Mount Sinai), Sirisha Pochareddy (Yale University), Damon Polioudakis (University of California, Los Angeles), Henry Pratt (University of Massachusetts Medical School), Amanda J. Price (Lieber Institute for Brain Development), Michael Purcaro (University of Massachusetts Medical School), Timothy E. Reddy (Duke University), Suhn Kyong. Rhie (University of Southern California), Panos Roussos (Mount Sinai), Tanmoy Roychowdhury (Mayo Clinic Rochester), Stephan J. Sanders (University of California, San Francisco), Gabriel Santpere (Yale University), Soraya Scuderi (Yale University), Nenad Sestan (Yale University), Brooke Sheppard (University of California, San Francisco), Xu Shi (Yale University), Annie W. Shieh (SUNY Upstate Medical University), Mario Skarica (Yale University), Lingyun Song (Duke University), Andre M.M. Sousa (Yale University), Patrick F. Sullivan (University of North Carolina - Chapel Hill), Vivek Swarup (University of California, Los Angeles), Flora M. Vaccarino (Yale University), Harm van Bakel (Icahn School of Medicine at Mount Sinai), Xusheng Wang (St. Jude Children's Hospital), Daifeng Wang (Yale University), Jonathan Warrell (Yale University), Zhiping Weng (University of Massachusetts Medical School), Donna M. Werling (University of California, San Francisco), Kevin P. White (The University of Chicago), A Jeremy. Willsey (University of California, San Francisco), Hyejung Won (University of California, Los Angeles), Feinan Wu (Yale University), Yan Xia (SUNY Upstate Medical University/Central South University), Min Xu (Yale University), Yucheng T. Yang (Yale University), Mo Yang (Yale University), Peter Zandi (Johns Hopkins University), Jing Zhang (Yale University), Ying Zhu (Yale University)

Data coordination

Yooree Chae (Sage Bionetworks), Lara M. Mangravite (Sage Bionetworks), Mette A. Peters (Sage Bionetworks), Zhiping Weng (University of Massachusetts Medical School)

PsychENCODE Developmental Subgroup

(team members are listed in alphabetical order of the last name after PI)

Tissue procurement and processing group:

Yale University: Nenad Sestan (Principal Investigator), Candace Bichsel, Daniel Franjic, Fuchen Liu, Marco Onorati, Mihovil Pletikos, Mario Skarica, Andre M.M. Sousa

University of Washington: Ian A. Glass (Principal Investigator), Kimberly A. Aldinger, Theresa Naluai-Cecchini

Transcriptome data production and validation group:

Yale University: Nenad Sestan (Principal Investigator), Navjot Kaur, Sirisha Pochareddy, Mihovil Pletikos, Mikihito Shibata, Andre Sousa, Andrew Tebbenkamp

Single cell/nucleus data production and validation group:

Yale University: Nenad Sestan (Principal Investigator), Tianliuyun Gao, Zhen Li, Mario Skarica

Epigenome data production and validation group:

Yale University: Nenad Sestan (Principal Investigator), Zhuo Li, Sirisha Pochareddy, Yurae Shin

Data analysis group:

Yale University: Nenad Sestan (Principal Investigator), Mark B. Gerstein (Principal Investigator), Jinmyung Choi, Robert R. Kitchen, Belen Lorente-Galdos, Mingfeng Li, Zhen Li, Shuang Liu, Shaojie Ma, Sydney Muchnik, Gabriel Santpere, Daifeng Wang, Xuming Xu, Ying Zhu

University of California, San Francisco: Stephan Sanders (Principal Investigator), Donna M. Werling

Acknowledgments

We are grateful to the many individuals in our laboratories and at our institutions who have provided support and contributed to these projects. In addition, we would like to thank the National Institute of Mental Health (NIMH), and in particular Andrea Beckel-Mitchener, Michelle Freund, Thomas Lerner, and Geetha Senthil, and for providing invaluable institutional support and guidance for these projects. **Funding:** Data was generated as part of the PsychENCODE Consortium, supported by: U01MH103392, U01MH103365, U01MH103346, U01MH103340, U01MH103339, R21MH109956, R21MH105881, R21MH105853, R21MH103877, R21MH102791, R01MH111721, R01MH110928, R01MH110927, R01MH110926, R01MH110921, R01MH110920, R01MH110905, R01MH109715, R01MH109677, R01MH105898, R01MH105898, R01MH094714, R01MH109901, P50MH106934, 5R24HD000836 and SFARI #307705 awarded to: Schahram Akbarian (Icahn School of Medicine at Mount Sinai), Gregory Crawford (Duke University), Stella Dracheva (Icahn School of Medicine at Mount Sinai), Peggy Farnham (University of Southern California), Mark Gerstein (Yale University), Daniel Geschwind (University of California, Los Angeles), Ian Glass (Washington University), Fernando Goes (Johns Hopkins University), Thomas M. Hyde (Lieber Institute for Brain Development), Andrew Jaffe (Lieber Institute for Brain Development), James A. Knowles (University of Southern California), Chunyu Liu (SUNY Upstate Medical University), Dalila Pinto (Icahn School of Medicine at Mount Sinai), Panos Roussos (Icahn School of Medicine at Mount Sinai), Stephan Sanders (University of California, San Francisco), Nenad Sestan (Yale University), Pamela Sklar (Icahn School of Medicine at Mount Sinai), Matthew State (University of California, San Francisco), Patrick Sullivan (University of North Carolina), Flora Vaccarino (Yale University), Daniel Weinberger (Lieber Institute for Brain Development), Sherman

Weissman (Yale University), Kevin White (University of Chicago), Jeremy Willsey (University of California, San Francisco), and Peter Zandi (Johns Hopkins University). The BrainSpan Project Consortium was supported by grants MH089929, MH090047, and MH089921 from NIMH. Additional support was provided by the Kavli Foundation, the James S. McDonnell Foundation, the Beatriu de Pinós program (BP-DGR 2014; to Belen Lorente-Galdos), and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (5R24HD000836). The human embryonic and some of the fetal material was provided by the Joint MRC / Wellcome Trust (MR/R006237/1) Human Developmental Biology Resource (www.hdbr.org). This work also received funding from The Netherlands Organization for Scientific Research (NWO VICI 435-14-005), Sophia Foundation for Scientific Research. Part of the analyses were carried out on the Genetic Cluster Computer, which is financed by the Netherlands Scientific Organization (NWO: 480-05-003), by the VU University, Amsterdam, the Netherlands, and by the Dutch Brain Foundation, and is hosted by the Dutch National Computing and Networking Services SurfSARA.
Author contributions: Specific contributions to the work are provided in the annotated author list for the consortia at the end of the manuscript. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** scRNA-Seq and snRNA-Seq data has been deposited at www.psychENCODE.org. Tissue RNA-Seq, ChIP-Seq, DNA methylation, and SNP array data has been deposited in dbGAP under accession number phs000755. All processed files and scripts pertaining to this manuscript are available on www. development.psychencode.org

Supplementary materials

Materials and Methods
Figs. S1 to S43
Table Legends S1 to S16
References (79 to 124)

Figure legends

Fig. 1. Overview of the data generated in this study. (**A**) The developmental timespan of the human brain, from embryonic ages (≤ 8 PCW) through fetal development, infancy, childhood, adolescence, and adulthood, with PCW indicated in text and postnatal years indicated in bold text. Below, distribution of samples in this study across broad developmental phases (embryonic to adulthood), age (5 PCW to 64 PY (*19*)), and developmental windows (W1 – W9). Each circle represents a brain and color indicates the sex; red circles (female) and blue circles (male) (**B**) Postmortem human brains sampled for different data modalities in this study as indicated.

Fig. 2. Global transcriptomic architecture of the developing human brain. (**A**) mRNA-Seq dataset includes 11 neocortical areas (NCX) and five additional regions of the brain. (**B**) First two multidimensional scaling components from gene expression showed samples from late fetal ages and early infancy (W5; gray) clustered between samples from exclusively prenatal windows (W1-4; blue) and exclusively postnatal windows (W6-9; red). (**C**) Intra-regional Pearson's correlation analysis found that samples within exclusively prenatal (W1-4) or postnatal (W6-9) windows correlated within but not across those ages. (**D**) Inter-regional transcriptomic differences revealed a developmental cup-shaped pattern in brain development. The inter-regional difference was measured as the upper-quartile of the average absolute difference in gene expression of each area compared to all other areas. (**E**) Principal component analysis adjusted for inter-individual

variation (AC-PCA) for samples from all brain regions at late mid-fetal ages (W4), late fetal ages and early infancy (W5), and early adulthood (W9) showed that inter-regional differences were generally greater during windows 4 and 9 but reduced across W5. (**F**) Pairwise distance across samples using the first two principal components for all regions (left), or excluding one region at a time (right), demonstrated that the reduction of variation we observed is common across multiple brain regions, once the most differentiated transcriptomic profile (the cerebellum) is excluded.

Fig. 3. Dynamics of cellular heterogeneity in the human neocortex. (**A**) AC-PCA conducted on 11 neocortical areas showed decreased inter-areal variation across W5, similar to our observations of inter-regional variation in major brain regions. (**B**) Pairwise distance across samples using the first two principal components identified a late fetal transition in all of the neocortical areas we assessed, similar to what we observed across other brain regions. (**C**) Deconvolution of tissue level data using cell type enriched markers identified through single cell sequencing of primary cells from 5 – 20 PCW postmortem human donor brain as well as from single nuclei sequencing of adult human brain (*27*). (**D**) Maximum inter-areal variance across cell types for each window. (**E**) Neocortical areal variation in the transcriptomic signatures of each major cell type assayed in each developmental window. Due to dissection protocols and rapid brain growth across early fetal development, progenitor cell proportions are nonreliable estimates after W2 (red dashed line). NPC: neural progenitor cells, ExN: excitatory neurons, InN: interneurons, Astro: astroglial lineage, Oligo: oligodendrocytes, Endo: endothelial.

Fig. 4. Timing and temporal variation of genes expression associated with key neurodevelopmental processes. (**A**) Temporal variation, determined by the TempShift algorithm

(*34*), in the expression of genes associated with myelination showed a broad gradient across the neocortex and other brain regions while synaptogenesis showed only a shift between brain regions (but not neocortical areas) and neuronal activity indicated the distinct nature of the cerebellum. Application of the TempShift algorithm to previously published post-translational analyses of myelinated fiber density (*35*) (**B**) and synaptic density (*36*) (**C**) in multiple neocortical areas yielded relationships between areas similar to those observed in the transcriptome. (**D**) Expression of genes associated with assorted biological processes highlights pronounced change during the late fetal period and W5. (**E**) Variation in myelination-associated genes peaks during W5, as evidenced by the standard deviation of the fitted regional mean, driving inter-regional variation during this and neighboring (W4, 6) windows.

Fig. 5. Integration of gene expression and epigenetic regulation with cell types and biological processes. (**A**) Fetal-active enhancers (top left) were generally enriched for sites where methylation progressively increased across postnatal ages and associated with genes whose expression was higher during fetal development than adulthood and whose expression was enriched in neurons as compared to glia. Conversely, adult-active enhancers were enriched for sites exhibiting progressively higher methylation across postnatal ages and depleted for associations with higher fetal gene expression or enriched in neurons. These enhancers were also enriched for gene ontology terms generally involving neurons and glia, respectively. (**B**) Sites where methylation progressively increased across postnatal ages and where methylation progressively decreased across postnatal ages were generally enriched for fetal enhancers and genes whose expression was enriched in neurons, or adult enhancers and genes whose expression was enriched in glia, respectively, as well as related gene ontology terms. (**C**) Modules identified

through Weighted Gene Correlation Network Analysis (WGCNA) were segregated by regulation across brain regions, prenatal and postnatal gene expression in the neocortex, both, or neither. Spatiotemporal modules (right) were enriched for modules that are themselves enriched for genes associated with enhancers active in the fetal DFC, associated with sites undermethylated in NeuN-positive (neuronal) cells, and/or enriched in neurons (N-type associations). Temporal, non-spatial modules (second from left) were enriched for modules that are themselves enriched for genes associated with enhancers active in the adult DFC, associated with sites undermethylated in non-NeuN-positive (non-neuronal) cells, and/or genes enriched in glia (G-type associations). Modules exhibiting no spatial or temporal specificity (left) were enriched for genes exhibiting sex-biased gene expression across neocortical development. Full circles (grey) indicate the proportion of modules in each category of modules exhibiting their greatest rate of change in windows 1 through 9

Fig. 6. Enrichment analysis for GWAS loci among putative regulatory elements. Putative promoters and enhancers (H3K27ac peaks) specific for DFC or CBC in the fetal, infant, or adult were enriched for SNP heritability identified through partitioned LD score regression analysis from Genome Wide Association Studies (GWAS) for autism spectrum disorder (ASD;(*40*)), attention-deficit hyperactive disorder (ADHD; (*41*)), schizophrenia (SCZ; (*37*)), major depressive disorder (MDD; (*42*)), bipolar disorder (BD; (*43*)), Alzheimer's disease (AD; (*38*)), Parkinson's disease (PD; (*39*)), IQ; (*44*), or neuroticism (Neurot; (*45*)) but not for non-neural disorders or traits such as height (HGT; (*46*)) or diabetes (HBA1C; (*49*)). Solid color indicates significance for Bonferroni adjusted P-value and faint color indicates nominal significance at LDscore regression $P<0.05$.

Fig. 7. Convergence of risk for brain-based traits and disorders on discrete co-expression modules and cell types. (**A**) Genes associated with disease risk (right; light yellow indicates neuropsychiatric disorder or brain-based trait; dark yellow indicates adult-onset disorder) were identified by integrating GWAS, Hi-C, and H3K27ac data and converged on 10 WGCNA modules. Many of these modules exhibited dynamic expression across time; bold rectangle in left of panel A indicates the window with greatest rate of change. Many were also enriched for gene expression associated with distinct cell types (orange), putative active enhancers (green), and/or sites undermethylated in NeuN-positive (NUM) or NeuN-negative cells (blue; non-NUM). (**B**) Schematic highlighting genes in ME37 that were implicated by our study in multiple neurodevelopmental disorders (ADHD, SCZ, MDD, or BD) and neurological traits (IQ or NEUROT), as well as neurodevelopmental disorder risk genes (NDD) including two independent lists of high-confidence risk genes associated with ASD through de novo mutations or copy number variants (dark blue; (*66*)) as well as ASD risk genes identified from the SFARI dataset (light blue; www.gene.sfari.org) or for developmental delay (*67*); genes implicated in only a single disorder or trait are not shown in this panel. (**C**) Network representation of module 37 showing connectivity between genes based on Pearson correlation. Genes linked to neurodevelopmental disorders or neurological characteristics in our study are indicated using either dark blue-shaded hexagons, indicating they are associated with the NDD list and/or List 2, or light blue-shaded hexagons, indicating genes only present List 1. The size of a given hexagon (or circle, indicating no association in this study) is proportional to the degree of each gene under a minimum correlation value of 0.7. (**D**) Enrichment for genes in ME37 or two lists of ASD risk genes among the fetal and adult cell types we identified from human neocortex and multiple regions of the macaque (*34*)

brain. ExN: excitatory projection neurons, InN: inhibitory interneurons, Astro: astroglial lineage, Oligo: oligodendrocytes, Endo: endothelial. MSN: medium spiny neurons, NasN: nascent neurons, GraN: granule neurons, PurkN: Purkinje neurons, IPC: intermediate progenitor cells, OPC: oligodendrocyte progenitor cells.

Supplementary Materials for

Integrative Functional Genomic Analysis of Human Brain Development and Neuropsychiatric Risk

Mingfeng Li[1+], Gabriel Santpere[1+], Yuka Imamura Kawasawa[1,2+], Oleg V. Evgrafov[3+], Forrest O. Gulden[1+], Sirisha Pochareddy[1+], Susan M. Sunkin[4+], Zhen Li[1+], Yurae Shin[1,5+], Ying Zhu[1], Andre M.M. Sousa[1], Donna M. Werling[8], Robert R. Kitchen[6,7], Hyo Jung Kang[1,9], Mihovil Pletikos[1,10], Jinmyung Choi[1], Sydney Muchnik[1], Xuming Xu[1], Daifeng Wang[11], Belen Lorente-Galdos[1], Shuang Liu[1,6], Paola Giusti-Rodríguez[12], Hyejung Won[12,13], Christiaan A de Leeuw[14], Antonio F. Pardiñas[15], BrainSpan Consortium[++], PsychENCODE Consortium[++], PsychENCODE Developmental Subgroup[++], Ming Hu[16], Fulai Jin[17], Yun Li[18], Michael J. Owen[15], Michael C. O'Donovan[15], James T.R. Walters[15], Danielle Posthuma[14], Patt Levitt[19], Daniel R. Weinberger[20], Joel E. Kleinman[21], Daniel H. Geschwind[21,22,23], Michael J. Hawrylycz[4], Matthew W. State[8], Stephan J. Sanders[8], Patrick F. Sullivan[11], Mark B. Gerstein[6]*, Ed S. Lein[4]*, James A. Knowles[3]*, Nenad Sestan[1,7, 24,25]*

[+] These authors contributed equally to this work

[++] The consortium authors are listed at the end of the paper.

* Co-corresponding authors: mark.gerstein@yale.edu, edl@alleninstitute.org,

james.knowles@downstate.edu, nenad.sestan@yale.edu

This PDF file includes:

Materials and Methods

Supplementary Text

Figs. S1 to S43

Table Legends S1 to S16

References (79 to 124)

Other Supplementary Materials for this manuscript include the following:

Tables S1 to S16 as in a single excel file

Materials and Methods

In this Supplementary Information, we provide further information regarding the study design, materials and methods, and additional data. The materials and methods section provide detailed description of the collection, dissection methods, and quality control assessments of postmortem human brain tissue used in this study. We provide technical descriptions of data generation and analyses.

1. Tissue procurement

This study was conducted using postmortem human brain specimens from tissue collections at the Department of Neurobiology at Yale University School of Medicine and the Clinical Brain Disorders Branch of the National Institute of Mental Health. Additional specimens were procured from the Human Fetal Tissue Repository at the Albert Einstein College of Medicine (AECOM), the Brain and Tissue Bank for Developmental Disorders at the University of Maryland, the Birth Defects Research Laboratory at the University of Washington, Advanced Bioscience Resources Inc. and the MRC-Wellcome Trust Human Developmental Biology Resource at the Institute of Human Genetics, University of Newcastle, UK. Tissue was collected after obtaining parental or next of kin consent and with approval by the institutional review boards at the Yale University School of Medicine, the National Institutes of Health, and at each institution from which tissue specimens were obtained. Tissue was handled in accordance with ethical guidelines and regulations for the research use of human brain tissue set forth by the NIH (http://bioethics.od.nih.gov/humantissue.html) and the WMA Declaration of Helsinki (http://www.wma.net/en/30publications/10policies/b3/index.html).

All available non-identifying information was recorded for each specimen. 60 postmortem brain specimens ranging in age from 5 post-conception weeks (PCW) to 64 postnatal years (PY) (Fig. 1 and table S1) were included in this study. Fetal age was extrapolated based on the date of the mother's last menstruation, characteristics of the fetus noted upon ultrasonography scanning, foot length of the fetus, and visual inspection. The postmortem interval (PMI) was defined as hours between time of death and time when tissue samples were frozen.

2. Brain regions and neocortical areas definition

Brain development is a highly dynamic process during which each region undergoes distinct organizational and maturational changes. Thus, we created a structural ontology that contains brain structures (e.g., NCX areas, HIP, AMY, STR, MD, CBC) that are well defined throughout most of time periods, and several transient structures (e.g., MGE, LGE, CGE, URL). In total, 10 regions for 5 - 9 PCW (window 1) specimens, and up to 16 regions for all ages after 9 PCW (windows 2-9). Regions and areas were as described in Kang *et al* but are noted here again for comprehensive description of methods (*19*). Below we describe this ontology and anatomical definition of sampled brain regions and NCX areas based on histological verification.

2.1. Cerebrum

2.1.1. Neocortex (NCX)

Samples collected from 8 – 9 PCW specimens contained the entire thickness of the cerebral wall. Samples collected from 12 – 22 PCW specimens contained the marginal zone, cortical plate, and part of the underlying subplate. Samples from 35 PCW – 40 PY specimens were dissected such that the entire gray matter (layer 1-6) and part of the underlying subplate (4- 12 M) or white matter (1 – 40 PY) were collected. Nissl staining of the neighboring thin block was used to histologically verify the identity of the dissected area and to microscopically evaluate tissue. Neocortical cytoarchitecture of each sample was compared to a real cytoarchitectonic maps to distinguish Brodmann areas (BA) (*73*). Samples with incorrect cytoarchitecture or abnormal microscopical appearance were excluded from the study. Neocortical areas (see below) were grouped according to the lobes from which they were sampled.

2.1.1.1. Frontal cortex

For 8 – 9 PCW specimens, the sampled area corresponded to different parts (orbital (OFC), dorsolateral (DFC, aka DLPFC), ventrolateral (VFC), and medial (MFC) of the anterior part of telencephalic vesicle (cerebral wall) corresponding to prospective FC. In addition, paracentral region corresponding approximately to the prospective motor and parietal somatosensory (M1C/S1C) cerebral wall was dissected as one sample (MSC).

For 12 – 22 PCW specimens, prior to the appearance of all gyri and sulci, multiple areas of the FC were sampled as follows:

- Orbital prefrontal cortex (OFC) was sampled from the middle part of the orbital surface of the cerebral hemisphere, immediately next to the prospective gyrus rectus.
- Dorsolateral prefrontal cortex (DFC) was sampled from the middle third of the dorsolateral surface of anterior third of the cerebral hemisphere.
- Ventrolateral prefrontal cortex (VFC) was sampled from the posterior part of the frontal operculum, above the lateral sulcus and prospective insula.
- Medial prefrontal cortex (MFC) was sampled from the perigenual and subgenual region of the medial surface.
- Primary motor cortex (M1C), prior to the appearance of the central sulcus, was sampled from the anterior third of the middle third of the cerebral hemisphere, medial third and upper part of the lower third of the dorsolateral surface. The striatum at the septal level was used as the landmark between the anterior and middle one third of the dorsolateral cortical surface. In some cases, M1C and S1C areas were sampled as single area and termed it motor-somatosensory cortex (M1C/S1C) due to the lack of clear anatomical and histological boundaries between immature M1C and S1C. After the appearance of the central sulcus M1C was sampled in front of the central sulcus from the middle and upper part of the lower third of the dorsolateral surface of the hemisphere.

<u>For 35 PCW – 40 PY (64 PY for DFC) specimens</u>, sampled areas were as follows:

- OFC was sampled from the anterolateral two thirds of the orbital gyri. OFC corresponds approximately to BA 11.

- DFC was sampled from approximate border between the anterior and middle third of the medial frontal gyrus. DFC corresponds approximately to BA 9 and 46.
- VFC was sampled from the posterior third of the inferior frontal gyrus, corresponding to the opercular and triangular part of the inferior frontal gyrus. VFC corresponds approximately to BA 44 and 45.
- MFC was sampled from perigenual and subgenual parts of the anterior cingulate gyrus and the anteromedial part of the superior frontal gyrus. MFC corresponds approximately to BA 24, 32 and 33.
- M1C was sampled from the ventrolateral part of the precentral gyrus, corresponding most closely to the orofacial region of M1C. M1C corresponds to BA 4.

2.1.1.2. Parietal cortex

For 8 – 9 PCW specimens, the sampled areas included the paracentral region corresponding approximately to the prospective motor and parietal somatosensory (M1C/S1C) cerebral wall, and the posterior half of the dorsal middle third of the cerebral wall corresponding approximately to the prospective inferior parietal cortex (IPC).

For 12 – 22 PCW specimens, prior to the appearance of gyri and sulci, multiple areas of the PC were sampled as follows:

- Primary somatosensory cortex (S1C), prior to the appearance of the central sulcus, was sampled immediately caudal to the M1C (see M1C description above). After the appearance of the

central sulcus, S1C was sampled behind the central sulcus from the middle and upper part of the lower third of the dorsolateral surface of the cerebral hemisphere adjacent to the M1C area.

- Posterior inferior parietal cortex (IPC) was sampled from the lower posterior part of the dorsolateral surface of the middle third of the cerebral hemisphere adjacent to the end of the lateral sulcus.

For 35 PCW – 40 PY specimens, sampled areas were as follows:

- S1C was sampled from the ventrolateral part of the postcentral gyrus adjacent to the M1C area. S1C corresponds to BA 1–3.
- IPC was sampled from the posterior half of the supramarginal gyrus. IPC corresponds approximately to BA 40.

2.1.1.3. Temporal cortex

For 8 – 9 PCW specimens, the sampled areas included the posterior two thirds of TC corresponding approximately to the prospective auditory and superior temporal cortex (A1C/STC) cerebral wall, and the anterior third corresponding approximately to the prospective inferior temporal cortex (ITC).

For 12 – 22 PCW specimens, prior to the appearance of gyri and sulci, multiple areas of the TC were sampled as follows:

- Primary auditory cortex (A1C) was sampled from the upper part of the temporal bank of the lateral sulcus.

• Posterior superior temporal cortex (STC) was sampled from the upper part of the superior third of the temporal lobe adjacent to the lateral sulcus and A1C area.

• Inferior temporal cortex (ITC) was sampled from the lower part of the inferior third of the temporal lobe adjacent to the temporal lobe pole.

For 35 PCW – 40 PY specimens, sampled areas were as follows:

• A1C was sampled from the planum temporale and the transverse temporal gyri. A1C corresponds to BA 41.

• STC was sampled from the posterior third of the superior temporal gyrus. STC corresponds approximately to BA 22.

• ITC was sampled from the anterior third of the inferior temporal gyrus. ITC corresponds approximately to BA 20.

2.1.1.4. Occipital cortex

For 8 – 9 PCW specimens, sampled tissue corresponded to the posterior (occipital) part of the cerebral wall.

For 12 – 22 PCW specimens, prior to the appearance of gyri and sulci, sampled tissue corresponded to prospective primary visual cortex (V1C). Prior to the appearance of the calcarine fissure, V1C was sampled from the posterior third of the medial wall of the prospective occipital lobe. After appearance of the calcarine fissure, V1C was sampled as described below.

For 35 PCW – 40 PY specimens, V1C was sampled from the area surrounding the calcarine fissure. Only samples in which the stria of Gennari could be recognized were included. V1C corresponds to BA 17. Small pieces of the neighbouring BA 18 could have been occasionally present in the sample, but the majority of the sample corresponded to BA 17.

2.1.2. Hippocampus (HIP)

For 8 – 9 PCW specimens, HIP was sampled from the hippocampal anlage, located on the ventromedial side of the cerebral hemisphere.

For 12 PCW – 40 PY, HIP was sampled from the middle third of the retrocommissural hippocampal formation, located on the medial side of the temporal lobe. Sampled areas always contained dentate gyrus and the Ammon's horn. Samples dissected from the frozen tissue may contain small quantities of the neighboring choroid plexus.

2.1.3. Amygdala (AMY)

For 12 PCW – 40 PY specimens, at the aim was to dissect the whole AMY. Potentially very small quantities of surrounding white matter and other surrounding structures in the basal telencephalon were included in samples.

2.1.4. Striatum (STR)

For 8 – 9 PCW specimens, the Medial ganglionic eminence (MGE), Lateral ganglionic eminence (LGE) and caudal ganglionic eminence (CGE) were sampled separately. Small quantities of surrounding tissue may be included in the samples.

For 12 PCW – 40 PY specimens, striatum (STR) was sampled as follows: the anterior part of striatum containing the head of the caudate nucleus and the putamen were dissected, separated by the internal capsule and ventrally connected to the nucleus accumbens. Small quantities of surrounding white matter are included in the samples.

2.2. Thalamus (MD)

For 8 – 9 PCW specimens, the sampled region corresponds to the dorsal part of the thalamic anlage (DTH).

For 12 PCW – 40 PY specimens, the whole mediodorsal nucleus of the thalamus (MD) was sampled from the dorsal and medial thalamus. Small quantities of surrounding thalamic nuclei could be present in the samples.

2.3. Cerebellar cortex (CBC)

For 8 – 9 PCW specimens, region corresponding to the upper (rostral) rhombic lip and adjacent tissue located above the upper rhomboid fossa were dissected.

For 12 PCW – 40 PY specimens, CBC was sampled from the lateral part of the posterior lobe. The sampled area contained all three layers of cerebellar cortex and underlying white matter but not the deep cerebellar nuclei. CBC approximately corresponds to the lateral pontocerebellum.

3. Tissue dissection

Depending on the condition and period of the procured specimens, four different dissection methods were used. Photos and/or video were used to document dissections using digital cameras. Regions of interest were matched between different specimens, ages, and hemispheres of each brain. Specific dissection protocol depended upon the period of the specimen and the method by which it was preserved. For all brain specimens procured at Yale University School of Medicine and the Human Fetal Tissue Repository at AECOM, brain regions and NCX areas of interest were collected from fresh tissue. For all other specimens, regions/areas were collected from frozen tissue slabs or whole specimens stored at -80 ºC. To ensure consistency between specimens, all dissections were performed by the same person. Small samples of fresh or frozen CBC were used to measure tissue pH.

3.1. Tissue dissection methods

Different dissection procedures were used for each specimen, depending upon the period of the brain (see below). Our pilot experiments indicated that the quality of RNA and DNA was largely unaffected by variation between the dissection methods used.

3.1.1. Regional sampling from fresh brain specimens

Brains were chilled on ice for 15–30 minutes prior to sectioning. Brains were placed ventral side up onto a chilled aluminium plate (1 cm thick) on ice. The brainstem and cerebellum were removed from the cerebrum by making a transverse cut at the junction between the diencephalon and midbrain. Next, the cerebrum was divided into left and right hemispheres by cutting along the midline using a Tissue-Tek Accu-Edge trimming blade, 260 mm. The cerebellum was separated from the brainstem by cutting directly posterior to the brainstem, along the cerebellar peduncles.

The regions of interest were dissected using a scalpel blade and immediately frozen in liquid nitrogen. Dissected samples were either immediately processed for RNA extraction or stored at -80 °C for later RNA extraction. The remaining brain tissue was cut to obtain 1 cm (specimens from 16 PCW) or 0.5 cm (12 – 13 PCW specimens) thick serial, coronal sections. The tissue slabs were snap frozen in isopentane (J.T. Baker)/dry ice at -30 to -40 ºC and stored at -80 ºC.

3.1.2. Regional sampling from frozen brain specimens

All previously frozen 12 PCW – 40 PY specimens and tissue slabs were microscopically inspected and the desired region was demarcated, then dissected using a dental drill (AnyXing, 300D) and a Lindemann Bone Cutter H162A.11.016 or diamond disk saw (Dental Burs USA; r=11 mm) on a 1 cm thick aluminium plate over dry ice. Dissected tissue samples were stored at -80 ºC prior to further processing.

3.1.3. Regional sampling from specimens processed in RNAlater ICE

8 – 9 PCW frozen specimens were sectioned coronally at approximately 2 mm, beginning at the frontal pole, using a dental diamond disk saw. For gradual thawing, tissue slabs were transferred from -80 °C storage to overnight storage in RNAlater ICE (Ambion) at -20 °C. Tissue slabs were visually inspected for gross anatomical neuropathological abnormalities. Next, regions of interest were sampled under a dissection microscope at 4 °C and stored in Buffer RLT Plus from the RNeasy Plus Mini Kit (Qiagen) at 4 °C. RNA was immediately extracted.

3.2. Histological verification of tissue sampling

To verify that the region or NCX area of interest is properly and consistently sampled, we also collected small tissue blocks, from both frozen and fresh brain specimens, adjacent to the tissue sample dissected for the RNA extraction. We have done this for the majority of M1C, S1C, IPC, A1C and V1C samples, which in our experience were hard to match across different specimens but can be histologically verified using Nissl method in postnatal specimens due to cytoarchitectonic differences. This method was also occasionally used for other regions or NCX areas. These tissue blocks were then fixed in 4% paraformaldehyde for 48 h, sectioned at 50 μm thickness using a vibratome, and Nissl stained to verify the identities of dissected adjacent tissue.

3.3. Dissection scoring

We developed a scoring system to evaluate the precision of how well the sampled region/area was represented at the same position of corresponding samples of the same period.

Score description

- 1 or 2: The region/area of interest was absent (1) or largely absent (2) and thus not collected.
- 3: The region/area of interest was not complete but was of suitable quality to collect.
- 4: The region/area of interest was largely intact but was not histologically verified or could not be collected at precisely the same position from which the corresponding contralateral sample was collected.
- 5: The region/area of interest was fully intact, verified by gross inspection or Nissl staining (NCX areas), and collected at precisely the same position as corresponding samples of the same period.

3.4. Tissue pulverization

To ensure proper representation of the region of interest, frozen tissue samples were pulverized in liquid nitrogen using a ceramic mortar and pestle (Fisher Scientific, cat# 12-961C and 12-961-5C). Pulverized samples were transferred to chilled wide-mouth cryogenic vials (Nalgene, cat# 03-337-7B) and stored at -80 ºC until used for RNA extraction.

4. RNA sample preparation and sequencing

4.1. RNA extraction

RNA was extracted using RNeasy Plus Mini Kit (Qiagen) for mRNA and mirVana kit (Ambion) for small RNA. Either approximately 30 mg of pulverized tissue (12 PCW – 40 PY specimens) or entire amount of dissected brain piece (8 – 9 PCW, smaller than 30 mg) was processed. Tissue was pulverized with liquid nitrogen in a chilled mortar and pestle and transferred to a chilled safe-lock microcentrifuge tube (Eppendorf). Per tissue mass, equal mass of chilled stainless steel beads (Next Advance, cat# SSB14B) along with two volumes of lysis buffer were added. Tissue was homogenized for 1 min in Bullet Blender (Next Advance) at speed 6 and incubated at 37°C for 5 min. Lysis buffer up to 0.6 ml was again added, tissue homogenized for 1 min and incubated at 37°C for 1 min. Extraction was further carried out according to manufacturer's protocol. Genomic DNA was removed by a proprietary column provided in RNeasy Plus Mini Kit (Qiagen) or by DNase treatment using TURBO DNA-free Kit (Ambion/ Life technologies). 260:A280 ratio and RNA Integrity Number (RIN) were determined for each sample with NanoDrop (Thermo Scientific) and Agilent 2100 Bioanalyzer system, respectively.

4.2. mRNA-Sequencing library preparation

The mRNA-Sequencing (mRNA-Seq) Sample preparation Kit (Illumina) was used to prepare cDNA libraries per manufacturer instructions with some modifications. Briefly, polyA RNA was purified from 1 to 5 μg of total RNA using Oligo (dT) beads. Quaint-IT RiboGreen RNA Assay Kit (Invitrogen) was used to quantitate purified mRNA with the NanoDrop 3300. Following mRNA quantitation, 2.5 μl spike-in master mixes, containing five different types of RNA molecules at varying amounts (2.5×10^{-7} to 2.5×10^{-14} mol), were added per 100 ng of mRNA (*79*). Spike-in RNAs were synthesized by the External RNA Control Consortium (ERCC) by *in vitro* transcription of de novo DNA sequences or DNA derived from *B. subtilis* or the deep-sea vent microbe *M. jannaschii* and were a generous gift of Dr. Mark Salit at The National Institute of Standards and Technology (NIST). Each sample was tagged by adding two spike-in RNAs unique to the region from which the sample was taken. Besides, three common spike-in RNAs with gradient concentrations were added to each sample, aiming at the assessment of sequencing quality (table S10). Spike-in sequences are available at http://archive.gersteinlab.org/proj/brainseq/spike_in/spike_in.fa. The mixture of mRNA and spike-in RNAs was subjected to fragmentation, reverse transcription, end repair, 3' end adenylation, and adapter ligation to generate libraries of short cDNA molecules, followed by PCR amplification. The PCR enriched product was assessed for its size distribution and concentration using Bioanalyzer DNA 1000 Kit.

4.3. smallRNA-Sequencing library preparation

The TruSeq Small RNA Sample Kit (Illumina) was used to prepare libraries for small RNA-Sequencing (smRNA-Seq) per manufacturer instructions. Briefly, 1 μg of total RNA was ligated with 3'- and 5'- adapters, followed by reverse transcription and PCR amplification. Unique index

sequence tags were introduced during PCR to enable multiplexed sequencing. Each library was assessed for the presence of desired micro RNA population and quantity by Bioanalyzer High Sensitivity DNA Kit. 21 samples (16 samples from one brain and 5 technical replicates) were pooled and size selected for fragments between 145-160 bp, including the ligated 5' and 3' adapters, by gel excision. The extracted product was assessed for its size distribution and concentration using Bioanalyzer DNA 1000 Kit.

4.4. Sequencing

The mRNA-Sequencing was carried out on Illumina Genome Analyzer IIx (GAIIx), loading one sample per lane. Each flow cell was run for 76 cycles using a single-read recipe (v4 sequencing kits) according to the manufacturer's instructions.

The small RNA-Sequencing was done on Illumina HiSeq 2000 by loading 21 samples per lane. The HiSeq flow cell was run for 50 cycles using a single-read recipe (v2 sequencing kit) according to the manufacturer's instructions.

A total of 607 samples, 16 brain regions from 41 post-mortem individuals, which passed stringent tissue quality measures, including postmortem interval (PMI) = 12.9 ± 10.4 (hours); tissue pH = 6.5 ± 0.3; and RNA integrity number (RIN) = 8.8 ± 1 (table S1) were subjected to profiling by mRNA- and smRNA-Seq. For mRNA-Seq, 16 billion uniquely mapped reads (26.5 ± 8.6 million per sample) were generated (table S1), accounting for ~20,000X coverage of human transcriptome. For smRNA-Seq, 1.9 billion reads (6.9 ± 2.7 million per sample) were generated (table S2).

5. DNA methylation

5.1. DNA extraction

Genomic DNA was isolated using the DNeasy Blood and Tissue kit (Qiagen). In detail, approximately 25 mg of brain tissue from each brain region was lysed in Buffer ATL supplemented with proteinase K at 56˚C for 3 to 4 hours. If necessary, tissue was homogenized with a pellet pestle motor homogenizer (Kontes) before adding Proteinase K. Genomic DNA was purified from the lysate by passing it through the column provided in the kit. The DNA was eluted with 200ul of AE buffer. The concentration of the DNA was measured using the PicoGreen dsDNA assay kit (Invitrogen) on a NanoDrop (Thermo Scientific). DNA quality was assessed by visualization on 2% agarose gels.

5.2. Bisulfite conversion

Bisulfite treatment of the DNA using Zymo Research EZ DNA Methylation kit (D5001) was carried out at the USC Epigenome Center. Briefly, 500 ng of DNA was bisulfite treated by incubating in the dark at 50˚C for 12-16 hours with CT Conversion Reagent provided in the kit. Each sample was purified on the column provided in the kit. C to T converted DNA was eluted with 10 μl of M-Elution Buffer in the kit.

5.3. Illumina Infinium HumanMethylation450 BeadChip

Further steps of DNA amplification, fragmentation, hybridization, single base extension and signal detection were performed at the USC core facility, according to standard Illumina protocols. The signals were imported into Illumina's GenomeStudio software. Two samples were discarded at

this stage because of detectable drop-outs (an indicator suggestive of poor quality) as measured by standard Illumina metrics.

6. ChIP and sequencing

6.1. Chromatin isolation

DFC and CBC were used for ChIP-Seq. Due to tissue limitations, tissues from the right and left hemispheres, when available, were pooled. Dissected tissue was pulverized in liquid nitrogen using a ceramic mortar and pestle and cross-linked with formaldehyde at a final concentration of 1% for 15 minutes at room temperature. Cross-linking was quenched by adding glycine to a final concentration of 125 mM and tissue was washed twice in ice-cold 1X Dulbecco's PBS (Sigma-Aldrich, P5368). Tissue was then dounced in ice-cold 1X-PBS and centrifuged to obtain a pellet of single cells. To isolate nuclei, the cell pellet was first incubated in Lysis buffer 1 (50 mM HEPES-KOH, pH 7.5; 140 mM sodium chloride; 1 mM EDTA, pH 8; 10% glycerol; 0.5% NP-40; 0.25% TritonX-100; 1X protease inhibitor cocktail) on a rotator at 4 C, followed by incubation in Lysis buffer 2 (200 mM sodium chloride; 1 mM EDTA, pH 8; 0.5 mM EGTA, pH 8; 10 mM Tris-HCl, pH 8; 1X protease inhibitor cocktail) at RT, with a centrifugation step after each incubation. Nuclei in second pellet were lysed by suspending in Lysis buffer 3 (1 mM EDTA, pH 8; 0.5 mM EGTA, pH 8; 10 mM Tris-HCl, pH 8; 20% Sarkosyl; 1X protease inhibitor cocktail). Chromatin was sonicated to sizes between 200 bp and 500 bp with Misonix 4000 (S-4000).

6.2. Immunoprecipitation

Antibodies validated and used in ENCODE studies (*80*) were used in this study: Millipore - H3K4me3 (07-743), and Abcam – H3K27ac (ab4729). For each ChIP reaction, 50ul of washed

Dyna Protein G beads (Invitrogen, 10004D) were incubated with 5 µl of the respective antibody for 12 h at 4 ˚C. Beads were washed thrice with ice-cold BSA solution (5 mg/ ml BSA in 1X PBS), once with 1X TE, and then suspended in 50 µl of ice-cold BSA solution. 50 µg of the sheared chromatin was added to beads and the mixture was incubated for at least 16 h at 4 ˚C. Dynabeads were washed at room temperature eight times with RIPA buffer (50 mM HEPES, pH 8; 1 mM EDTA, pH 8; 0.7% deoxycholate; 1% NP-40; 0.5 M lithium chloride; 1X protease inhibitor cocktail) and once with 1X TE. Immunoprecipitated chromatin was eluted in 200 µl of Elution buffer (20% SDS in 1X TE) at 65˚C for 15 minutes, with occasional vortexing. The eluted protein-chromatin complex was reverse cross-linked by overnight incubation at 65˚C. ChIPped and input DNA were treated with RNaseA (Sigma-Aldrich, R4875), Proteinase K (Denville Scientific Inc, CB3210-7) and purified with a Qiagen purification kit (Qiagen, 28104).

6.3. ChIP-Seq library preparation and sequencing

For adult samples, 5 ng of the purified immunoprecipitated DNA was used to prepare sequencing libraries for the Illumina sequencing platform. Sequencing libraries were generated using the ChIP-Seq DNA sample prep kit (Illumina, IP-102-1001) according to the manufacturer's protocol with some modifications, as follows. DNA fragments were end-repaired and adenylated, as per manufacturer's instructions. In order to reduce the free adaptor dimers that can be preferentially amplified during PCR, a 1:40 dilution of the adaptor was used instead of the recommended 1:10 dilution. The recommended gel purification step was performed after DNA fragments were enriched with PCR amplification. DNA fragments between 250bp and 350bp were size selected by gel electrophoresis and band excision. During gel extraction, the gel was melted at room

temperature, instead of 50˚C. For multiplexing, index primers from the Multiplexing Sample Preparation Oligonucleotide Kit (Illumina, PE-400-1001) were used.

For fetal and infant samples, 5 ng of the purified immunoprecipitated DNA was used to prepare sequencing libraries using the TruSeq ChIP Library Prep Kit (Illumina, IP-202-1012) according to the manufacturer's protocol with the following modification. The recommended gel purification step was performed after DNA fragments were enriched with PCR amplification.

Sequencing of the libraries was carried out at the Yale Center for Genome Analysis using either an Illumina Genome Analyzer IIx (GAIIx) or an Illumina HiSeq 2000 (adult samples) and an Illumina HiSeq 2500 (fetal/ infant samples).

6.4. Quantitative PCR followed by chromatin immunoprecipation (ChIP-qPCR)

Independent ChIP was carried out as described above. Selected regions were tested for enrichment by droplet digital PCR (QX100™ Droplet Digital™ PCR System, Bio-Rad). The primer-probe pairs used are listed in Supplementary table 16. Each reaction included 500 pg of immunoprecipitated DNA or input DNA, 500 nM each primer, 250 nM of the probe and 1X ddPCR Master mix (Bio-Rad). Droplets were generated and DNA was PCR amplified as follows: 95ºC/10 minutes, (95 ºC/30 s, 60 ºC/ 1 min) x 40 cycles, 98 ºC/ 2 minutes. The PCR plate was read in a droplet reader (Bio-Rad) and the absolute values obtained by QuantaSoft analysis software. Enrichment was calculated as percentage of input (%IP). Three independent PCR reactions were carried out for each target region and the error bars indicate the standard error of the mean.

7. Single cell RNA-Seq

7.1. Tissue dissociation

Brains from donors from 5 PCW to 64 PY were dissected as described in section 2 in accordance with samples ages. Fresh human brain tissue was transferred from hibernate into a tube with 37 ºC pre-warmed standard Papain protease and DNAse (PPD) solution. The mix was pipetted with a wide-bore pipette tip for 10 times, and then the tube was incubated in 37 ºC water bath for 30 minutes. The mix was gently pipetted for 10-20 times every 10 minutes during incubation. After incubation, the cell suspension was re-pipetted again until no visible clots could be seen. Finally, the well-dissociated cell suspension was centrifuged at 300g for 3 minutes and washed with DPBS.

7.2. Single cell capture and cDNA synthesis

Single cells were captured with Fluidigm C1 Single-Cell mRNA Seq IFC, 10–17 μm using the Fluidigm C1 system according to the manufacturer's instructions. Immediately before cell capture, viability was assessed by trypan blue staining. In all experiments, the cell viability was close to 100%. A concentration of 1,000–3,000 cells/μL was used for cell loading. In total, 1,512 single cells were captured from 5, 6, 8, 16, 19 and 20 PCW postmortem fetal brains (table S3). After capture, C1 chips were examined visually and the number of cells at each capture site was recorded manually. Cells captured by C1 were subsequently processed through lysis, reverse transcription and PCR amplification to generate single-cell cDNA libraries using the Smarter Ultra Low Input RNA kit for Fluidigm (Clontech). The cDNAs from all capture sites were harvested the next morning in about 3 μL C1 harvest reagent and were transferred to a 96-cell PCR plate in 10 μL C1 DNA dilution reagent.

7.3. Single cell RNA-Seq library preparation

cDNA concentration was quantified with Quant-iT™ PicoGreen ® (Invitrogen™, P7589). cDNA samples were diluted to 200 ng/μL with C1 harvest reagent and libraries were prepared and indexed using Nextera XT Library Prep Kit (Illumina), according to the manufacturer's protocol. All sequencing libraries were assessed for quality by Agilent Bioanalyzer, using high sensitivity dsDNA assay.

Library of single cells pooled from each C1 IFC was denatured using the Illumina protocol. The denatured libraries were diluted to 6 pM, followed by cluster generation on a single-end HiSeq flow cell using an Illumina cBOT, according to the manufacturer's instructions. The HiSeq flow cell was run for 100 cycles using a single-read recipe according to the manufacturer's instructions.

8. Single nucleus RNA-Seq

Dorsolateral prefrontal cortex (DFC) was precisely dissected from the adult human brains and finely pulverized to powder in liquid nitrogen with mortar and pestle (Coorstek: #60316, #60317). All buffers were ice-cold and all reagents were molecular biology grade. 50 -100 mg of pulverized tissue was added into 5 ml of the ice-cold lysis buffer: 320 mM sucrose (Sigma #S0389), 5 mM CaCl2 (Sigma #21115), 3 mM Mg(Ace)2 (Sigma #63052), 10mM Tris-HCl (pH 8) (AmericanBio #AB14043), protease inhibitors w/o EDTA (Roche #11836170001), 0.1 mM EDTA (AmericanBio #AB00502), RNAse inhibitor (80U/ml) (Roche #03335402001), 1mM DTT (Sigma #43186), 0.1% Triton X-100 (v/v) (Sigma #T8787). DTT, RNAse inhibitor, protease inhibitors, Triton X-100 were added immediately before use. The suspension was transferred to Dounce tissue grinder

(15ml volume, Wheaton # 357544; autoclaved, RNAse free, ice-cold) and homogenized with loose and tight pestles, 30 cycles each, with constant pressure and without introduction of air. The homogenate was strained through 40 um tube top cell strainer (Corning #352340) which was pre-wetted with 1ml of the isolation buffer (1800 mM sucrose, 3 mM Mg(Ace)2, 10mM Tris-HCl (pH 8), protease inhibitors w/o EDTA, RNAse inhibitor, and 1mM DTT). Additional 9 ml of the isolation buffer was added to wash the strainer. Final 15 ml of solution was mixed by inverting the tube ten times and carefully pipetted onto the 5 ml isolation buffer cushion in each of the two ultracentrifuge tubes (Beckman Coulter #344059), without disrupting the phases. The samples were centrifuged at 30,000 x g, for 60 min at 4 ˚C in the Beckman SW41-Ti rotor of the Beckman L7-65 ultracentrifuge. After ultracentrifugation, the supernatant was carefully and completely removed and 100 ul of the resuspension buffer (250 mM sucrose, 25 mM KCl (Sigma #60142), 5mM MgCl2 (Sigma #1028) 20mM Tris-HCl (pH 7.5) (AmericanBio #AB14043; Sigma #T2413), protease inhibitors w/o EDTA, RNAse inhibitor (80U/ml), 1mM DTT) was added dropwise on the pellet in each tube and incubated on ice for 15 minutes. Pellets were gently dissolved by pipetting 30 times with 1ml pipette tip, pooled and filtered through 40 um tube top cell strainer. Finally, nuclei were counted on hemocytometer and diluted to a concentration of 1 million/ml with sample-run buffer: 0.1% BSA (Gemini Bio-Products #700-106P), RNAse inhibitor (80U/ml), 1mM DTT in DPBS (Gibco #14190).

The nuclei suspension was further processed at Yale Center for Genome Analysis (YCGA) core facility. Single nuclei were captured and barcoded with 10X Genomics chromium platform (10x Genomics), using Chromium Single Cell 3’ Library and Gel Bead Kit v2, (10x Genomics #PN-120237) and Chromium Single Cell A Chip Kit (10x Genomics #PN-120236) and following

manufacturer's protocol (https://support.10xgenomics.com/permalink/3vzDu3zQjY0o2AqkkkI4CC, CG00052_RevE). A maximum of 10,000 nuclei were targeted for recovery. One modification we made in the protocol was to increase the total number of cDNA amplification cycles to 14 cycles from the recommended 8 cycles for targeted recovery of 10000 cells, to generate sufficient mass of cDNA for the library construction from nuclear RNA. The final Single Cell 3' Libraries were paired-end sequenced with single indexing on the HiSeq 4000 platform (Illumina). To avoid lane bias, multiple uniquely indexed samples were mixed and distributed over several lanes.

9. Genotyping

For genotyping analysis, genomic DNA was extracted from approximately 25 mg of brain tissue, usually collected from the CBC using the DNeasy Blood and Tissue kit (Qiagen). In detail, brain tissue from each brain was lysed in Buffer ATL supplemented with proteinase K at 56˚C for 3 to 4 hours. If necessary, tissue was homogenized with a pellet pestle motor homogenizer (Kontes) before adding Proteinase K. Genomic DNA was purified from the lysate by passing it through the column provided in the kit. The DNA was eluted with 200ul of AE buffer. The concentration of the DNA was measured using the PicoGreen dsDNA assay kit (Invitrogen) on a NanoDrop (Thermo Scientific). DNA integrity was confirmed by agarose gel electrophoresis. Illumina Omni-2.5 million SNP arrays were used for genotyping analysis. DNA samples were processed at the Yale Center for Genome Analysis according to the Infinium HD Assay Super, Automated Protocol for Human Omni 2.5-Quad Bead Chip (Illumina). All SNP-arrays were scanned using the Illumina iScan system.

10. mRNA-Seq data analyses

mRNA-Seq data processing was based on the RSEQtools framework, version 0.5 (*81*) and the workflow is shown in fig. S1. Details of each processing step are given below.

10.1. Sequence alignment

mRNA-Seq reads were aligned to the human genome assembly hg38/GRCh38 excluding minor haplotypes, random and unknown sequences. Hence, only the 22 autosomes, X, Y, and M chromosomes were considered for the alignment. FASTA files were downloaded from the UCSC genome browser (*82*). Additionally, the sequences of spike-in RNAs were included in the reference. The FASTA file of the spike-in RNAs is available at http://archive.gersteinlab.org/proj/brainseq/spike_in/spike_in.fa.

Alignment of the reads was performed by STAR (version 2.4.0e) (*83*). Command line"--runMode genomeGenerate" was used to build the sequence index for both human genome and spike-in sequences. To improve the mapping quality of splice junction reads, human gene annotation retrieved from the GENCODE project (version 21) was additionally provided (*84*). Command line "--sjdbOverhang 74" was used to construct splice junction library.

As an example, the alignment of Sample A is given below

```
--runMode alignReads --readFilesIn A.fq --outFileNamePrefix A.out --genomeDir hg38ANDSpikein --runThreadN 8 --outSAMattributes All --outSAMtype BAM SortedByCoordinate --limitBAMsortRAM 62000000000 --quantMode TranscriptomeSAM --outFilterMismatchNoverLmax 0.1 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --outSAMunmapped Within --outFilterType BySJout --alignMatesGapMax 500 --outFilterMultimapNmax 50 --alignEndsType Local --outSAMstrandField intronMotif --outFilterIntronMotifs RemoveNoncanonical
```

where,"hg38ANDSpikein" is the indexed human genome and spike-in sequences. More details about the parameters are available at https://github.com/alexdobin/STAR/releases. Some default parameters were changed to obtain high confidence alignment. Options "--outSAMstrandField intronMotif" and "--outFilterIntronMotifs RemoveNoncanonical" were used to make the STAR alignment compatible with Cufflinks in the downstream splicing analyses.

Finally, the entire mRNA-Seq information were wrapped in the BAM format alignments.

10.2. Alignment visualization and measuring expression

After the reads were mapped to the reference sequences, the alignment was visualized in the UCSC genome browser and the expression level of genes, exons, and spike-in RNAs were measured in the commonly used units of RPKM (reads per kilobase of exon model per million mapped reads) (*85*). SAMtools (*86*), RSEQtools (*81*) and UCSC applications (*82*) were used to perform this task.

First, the BAM format alignment was converted into SAM format alignment by using the "view" function in SAMtools, and then the "sam2mrf" function in RSEQtools was used to convert the SAM format to Mapped Read Format (MRF).

To visualize sequence alignments in UCSC genome browser, the MRF format was converted to wig format using “mrf2bedGraph” function in RSEQtools. Then, the “wigToBigWig” applications in UCSC was used to convert wig format to bigwig format. The bigwig format alignments will display the mapped reads in the genome browser as a graph.

Thereafter, analyses were carried out using only uniquely mapped reads because of the uncertainty of the reads mapped to multiple sites. In addition, mitochondrial reads were excluded due to their large variability across different individuals. After filtering, RPKM values were computed using "*mrfQuantifier*" function in RSEQtools. This program required an annotation set, which includes elements whose expression level is to be measured. Due to the presence of multiple transcripts for each gene in the human genome, assignment of reads to specific transcript is not straightforward. Hence, a composite model of a gene, which is a union of all exonic nucleotides across all of its transcripts, was defined. "*mergeTranscript*" function in RSEQtools was used to generate the composite gene model from GENCODE (version 21). From the composite model of genes, the composite model exons were also extracted. The *mrfQuantifier* calculated an expression value for each annotation entry by counting reads from all nucleotides that overlap that annotation entry. Subsequently, this value was normalized per million mapped nucleotides and the length of the annotation item per kb (RPKM).

In summary, the steps to compute the expression levels for sample A:

Convert BAM format alignment to SAM format alignment:
$ samtools view A.bam > A.sam
Convert SAM format to MRF:
$sam2mrf < A.sam > A.mrf
Build composite model of gene:
$mergeTranscripts knownIsoforms.txt transcript.interval compositeModel > geneComposite.interval
where "knownIsoforms.txt" determines which transcript isoforms belong together; "*transcript.interval*" is the interval format annotation of transcript isoforms; "*geneComposite.interval*" is the interval format annotation of gene composite model, from which the interval format annotation of exon composite model is extracted;
Calculate expression of any assigned element:
$mrfQuantifier annotation.interval singleOverlap < A.mrf > A.expression
where "*annotation.interval*" should be the interval format annotations for gene composite model, exon composite model and spike-in RNAs; "*singleOverlap*" means reads that overlap with multiple annotated features were excluded.

STAR, SAMtools format conversion, and measuring expression by RSEQtools were driven by in-house Perl scripts and were run at the Yale High Performance Computing clusters.

Moreover, we used HTSeq (*87*) to generate read count for every gene annotated in GENCODE (version 21) such that these raw read counts were used as the primary input for DESeq2 (*88*) .

10.3. Normalization and batch correction

Gene expression values were further processed to remove unwanted technical variates. Firstly, since the mitochondrial reads were excluded due to their large variability across different individuals, the mitochondrial genes were excluded from further analyses. The in-house analysis found their expression values strongly correlated to PMI values (Spearman correlation coefficient = 0.62). Secondly, conditional quantile normalization was performed using CQN Bioconductor

(*89*) to correct global distortion and to exclude GC-content bias. Finally, ComBat (*90*) was used to remove batch effects due to processing at two different sequencing centers, i.e., Yale and USC.

10.4. Quality control assessments

Several quality control measures were implemented throughout sample preparation and data analysis steps. Samples that failed to pass the quality control measures were resequenced, when possible, or removed from analyses. A work flow of all data analysis quality control steps is shown in the figs. S1 and S2.

10.4.1. Spike-in RNAs and sequencing error rate

Multiple spike-in RNAs were used to tag different samples (table S15). A pair of spike-in RNAs unique to the region from which the sample was taken was used to identify any possible sample swaps from library preparation to sample loading on the sequencer. In addition, three common spike-in RNAs were also added. All five spike-in RNAs were used to calculate sequence error rate. Since the sequence of the spike-in RNAs was known, mismatches between the sequenced reads and the reference spike-in RNAs was used to estimate sequencing error rate of Genome Analyzer. The dependence between the percentage of mismatches and the sequencing cycles were plotted and mostly low sequencing error rates (median values <5%) were found. As expected, a higher error rate was observed towards last sequencing cycles (fig. S2)

10.4.2. Reads from mitochondrial DNA, rRNA and tRNA

For every sample, the number of reads or percentage of reads that mapped to the 25 chromosomes and different RNA categories was calculated (fig. S2 and table S1). Notably, the proportion of

reads mapped to X and Y chromosomes were concordant with the gender of the individuals. To assess the genome-wide uniformity, the percentages of genome and exons covered by reads were counted for each chromosome. Additionally, we counted the percentage of reads mapped to protein-coding genes, non-coding RNAs, intronic regions and intergenic regions. Reads mapped to non-coding RNAs were further categorized to pseudogene, lncRNA, miRNA, snRNA/snoRNA and others. Most importantly, we assessed the number of reads aligned to mitochondrial, rRNA and tRNA to evaluate the level of RNA quality, because abnormally high percentage of aligned reads could be a sign of low quality of sample or failure depletion of ribosomal RNA.

10.4.3. Reads distribution across gene length

All samples were analyzed to detect any bias in read distribution along the body of each annotated gene. The composite exon-model gene was split into 100 equal segments, from 5'-end to 3'-end and RPKM of each segment was calculated. Then a ratio of the RPKM of each segment relative to the median RPKM value of the gene was taken. The median ratios of all genes in all samples (represented as a flat black line in the fig. S2) around zero indicated uniform read distribution along genes. There was only a slight trend of 5' underrepresentation and 3' overrepresentation.

10.4.4. Sequencing depth simulation

In our previous work (*20*), we performed the sequencing depth simulation to investigate the dependence between total sequenced reads and expression variation when measuring the abundance of gene and exon. The simulation suggested ten million of uniquely mapped reads could efficiently quantify the abundance of gene and exon, even higher sequencing depth could gain less expression variation. We adopted this work to the quality control pipeline.

10.4.5. Gene expression distribution

The fractions of protein-coding and non-coding genes expressed over a range of abundance levels (RPKM ≥0.1, ≥0.5, ≥1, ≥5) were counted for NCX areas and subcortical regions (HIP, AMY, STR, MD, CBC) (fig. S2 and table S1).

10.5. Clustering analyses

10.5.1. AC-PCA: PCA adjusting for individual variation

Strong individual effects were observed in mRNA-Seq data. Given the small differences observed between neocortical areas in the standard PCA, the first several principal components likely represent individual variation rather than inter-areal variation. To adjust for individual variation, AC-PCA was implemented. Within a time window, suppose there are K individuals. For every individual, samples from B regions were taken. For every sample, the expression levels of G genes were measured. Let X_i denote the $B \times G$ matrix for individual i. Let w denote a $G \times 1$ vector, representing the loading for the principal components. AC-PCA maximized the following objective function:

$$\sum_{i=1}^{K} w^T X_i^T X_i w - \lambda \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} w^T \left(X_i - X_j\right)^T \left(X_i - X_j\right) w,$$

where λ is a tuning parameter. To encourage the coordinates for the same brain region across individuals to be similar, a penalty term was added in the objective function of standard PCA. The

spatial pattern of the brain regions is quite robust to the choice of λ, and, $\lambda = 1$ was chosen. To gain robustness, the rank of RPKM across samples was used instead of RPKM itself. Furthermore, genes that have median RPKM less than 2 across all samples were filtered out in the analysis. Full implementation details for AC-PCA can be found in Lin et al. 2016 (*23*).

The median gene expression levels (log2 RPKM) were calculated across samples and genes with median expression greater than 2 were selected. This step retained 10,952 genes. To gain robustness, the genes across samples were ranked and the rank used for dimension reduction instead of the RPKM itself. The pairwise Euclidean distances between brain regions/neocortical areas were calculated for each brain based on the first two PCs. The average pairwise distance was calculated for each brain and plotted by time window (Fig. 2 and fig. S16).

10.5.2. Multidimensional scaling (MDS)

The multidimensional scaling (MDS) was analyzed by using normalized RPKM values of all expressed genes in the autosomes from 607 samples. The Euclidean distance between any pair of samples was calculated to construct a distance matrix with dimension 607 x 607. The R function *cmdscale* was used to calculate two-dimensional MDS. The returned two vectors were used as coordinates x and y for each of the 607 samples to make a scatter plot in a 2-dimensional plane (Fig. 2 and fig. S11). To survey the effects from confounders, we performed linear regression analysis between top 10 multidimensional scaling components and 10 potential confounders: age, region, sequence depth, pH, RIN, sex, ethnicity, brain hemisphere, sample dissection score, and sequencing processing site. The squared correlation coefficient was utilized to measure the effect extent originated from confounders (fig. S11).

10.5.3. Hierarchical analysis

The hierarchical analysis was implemented by using normalized RPKM values of all expressed genes in the autosomes from 607 samples. The 1- Pearson correlation between any pair of samples was calculated to construct a distance matrix with dimension 607 x 607. The R function *hclust* with average or UPGMA model was used to perform hierarchical analysis. The returned clustering tree was plot as circular dendrogram by using the R function *as.phylo* (Fig. 2 and fig. S13). To survey the effects from confounders, we matched the 10 potential confounders (i.e. age, region, sequence depth, pH, RIN, sex, ethnicity, brain hemisphere, sample dissection score, and sequencing processing site) to each of 607 dendrogram branches (fig. S13).

10.6. Identification of temporally/spatially differentially expressed genes

Differentially expressed (DEX) genes were computed for every pair of 16 brain regions and 9 time windows (1-9). That is, the comparisons were separately performed in two different dimensions, of which the spatial dimension was the pairwise region comparison in each time window and the temporal dimension was the pairwise windows comparison for each region.

The DESeq2 package was utilized to perform the differential expression analyses (*88*). The reads count per gene served as the input for DESeq2. The GC content correction from CQN package was incorporated to DESeq2. Besides, the two sequencing sites (Yale and USC) were set as covariates to reduce the batch effects. Since we collected more than one sample in either condition, they were treated as biological replicates to improve the reliability of DEX genes, because DESeq2 was more reliable at comparing groups with replicates. When performing the comparison, DESeq2

firstly gets the mean expression level as a joint estimate for both groups, and then calculates the difference as well as the p-value for the statistical significance of this change. Statistical tests for differential expression were based on a model using the negative binomial distribution. The reported statistical significances were corrected for multiple testing using the Benjamini-Hochberg procedure (*91*) with a false discovery rate less than 0.01. In addition, to be called DEX genes we required RPKM > 1, reads count > 10 in at least one condition, and fold change > 2. DEX in multiple regions and across time is plotted in fig. S15.

10.7. Identification of genes differentially expressed by sex

Sex-differential expression analyses were run for samples in windows represented by both sexes (windows 2-9; excludes outlier individual HSB155, by PCA and inter-sample correlation). We filtered to expressed genes, defined as genes with a read count ≥10 in at least one sample from each of ≥2 individuals, leaving a total of 29,875 (49.7%) transcripts for analysis. We adjusted $\log_2$-transformed counts for processing site using ComBat (*90*), and for sample-level quantitative variables (RIN, PMI, and sequencing depth) using linear regression, with missing RIN and PMI values imputed to the median of non-missing values. By the nature of the sample collection methods, the male and female samples are not closely matched for developmental age; however, age is a strong contributor to the variance in gene expression and strongly correlated (r = 0.82) with the first principal component from all 29,875 expressed transcripts. To account for the differences in developmental stage between male and females we used linear interpolation ('approx' function in R) to estimate the expression of each transcript at a given developmental window for males and females. This interpolation was performed separately for the male and female samples within each of 16 brain regions for all ages from 13 PCW to 40 years. Post-

interpolated counts were estimated at 30 designated developmental time points within this age range. We then applied conditional quantile normalization (CQN) (*89*) to the interpolated counts to correct for transcript-level variables GC-content and gene length.

Differential expression analysis was assessed using DESeq2 (*88*) on the interpolated counts and normalization factors from CQN within each of the 16 brain regions and 8 developmental stages for a total of 128 tests. Instead of selecting arbitrary p-value and/or fold difference (FD) thresholds for defining sex-DEX genes, we opted to identify the top 100 genes expressed more highly in males (male-DEX), and the top 100 genes expressed more highly in females (female-DEX), from each region-stage, according to p-value and fold difference direction.

Based on the logic that a truly sex-DEX gene would show a degree of sexual dimorphism in neighboring developmental windows (represented by different individuals) we sought to estimate the probability of observing a gene in the top 100 for each sex in neighboring windows by chance in the same brain region. By randomizing the assignment of sex (17 random females and 21 random males, as in the true data) to each of the samples we ran 100 permutations of the experiment, including the interpolation, differential expression, and selection of 100 top "male"-DEX and "female"-DEX genes. By determining the proportion of genes that appeared in N consecutive developmental windows across all regions, time periods, and permutations we estimated the false discovery rate (q-value) for a given value of N. We report and characterize all genes with q-values ≤ 0.05, which represents a threshold of two or more neighboring time periods.

10.8. Gene co-expression network analyses and module characterization

Weighted gene co-expression network analysis (WGCNA) (*92*) was performed using R package WGCNA to investigate the spatio-temporal dynamics of gene expression (fig. S31 and table S9). The normalized and batch corrected gene RPKM values were first log2-transformed. Next, *pickSoftThreshold* function was used to analyze the network topology with 25 as soft-threshold power. In order to do automatic network construction and module detection, *blockwiseModules* function was used with several customized parameters, i.e., mergeCutHeight = 0.20 and networkType = signed. Modules with fewer than 10 genes were merged to their closest larger module. For each module, WGCNA generated an eigengene to represent modular features. To check the reliability of detected modules, custom R scripts were used to calculate the correlation between a gene and modular eigengene. Genes were re-assigned to another module if the gene had larger correlation coefficient with that module's eigengene than with its own. Overall, only a small number of genes were re-assigned to other modules. For those changed modules, the *moduleEigengenes* function was used to re-calculate the eigengene. To illustrate the modular feature, the smoothed fitted trajectories of the modular eigengenes were plotted as function of developmental windows using *lowess* function in R. To analyze the gene ontology item enrichment for genes in each module, two online resources were utilized, i.e., DAVID (*93*) and Gene Ontology Consortium (*94*).

10.9. Alternative splicing analyses

10.9.1. Exon percent spliced in (PSI)

For this analysis, to reduce unwanted influences due to inadequate sequencing depth, we solely chose these subjects with over 20 million uniquely mapped reads, resulting in 479 samples from

41 donor brains (table S1). The efficiency of one exon splicing into transcript was indicated by the percent spliced in (PSI) index. The PSI score summarized the possibility of one specific splicing event. For instance, a PSI score of 1 (or 100%) meant constitutive exons that were included in every gene transcript, whereas PSI score less than 1 (or 100%) denoted this exon was excluded from some gene transcripts. In short words, exon PSI could be the identifier of gene alternative splicing event. For one specific exon, the PSI score was calculated using the normalized counts as the following formula:

$$PSI = IR / (IR + ER)$$

where IR is the number of reads supporting exon inclusion and ER corresponds to the number of reads supporting exon exclusion. In application to genome-wide calculation for hundred thousand exons, the BAM format base-alignment, BED format junction-alignment, and GTF format exon-annotation were incorporated to implement our pipeline that was adopted from (*95*) and was embedded in our PERL scripts to automatically analyze hundreds of samples. In brief, the exon annotations were firstly processed to generate unique exonic signature using *dexseq_prepare_annotation.py* function from DEXSeq (*96*), and subsequently the base- and junction- alignments were assembled to exonic entries to compute IR and ER using *coverageBed* and *intersectBed* functions from bedtools (*97*). To illustrate the spatio-temporal dynamics, the smoothed fitted trajectories with confidential intervals were plotted as function of developmental windows using *ggplot2* package in R (fig. S18).

10.9.2. Multidimensional scaling (MDS) of alternative splicing

The MDS analysis of alternative splicing was done by plotting exon PSI. The Euclidean distance between any pair of samples was calculated to construct distance matrix with dimensions 607 x 607. The R function *cmdscale* was used to calculate two-dimensional MDS. The returned two vectors were used as coordinates x and y for each of the 607 samples to make a scatter plot in a 2-dimensional plane (fig. S12). To survey the effects from confounders, we performed linear regression analysis between top 10 multidimensional scaling components and 10 potential confounders: age, region, sequence depth, pH, RIN, sex, ethnicity, brain hemisphere, sample dissection score, and sequencing processing site. The squared correlation coefficient was utilized to measure the effect extent originated from confounders (fig. S12).

10.9.3. Identification of top variable genes and Gene Ontology analysis

We calculated the squared coefficient of variation (CV2) for each gene in each individual across all neocortical areas. The CV2 values for all individuals in the same window were averaged. For each window, and for protein-coding genes with an average RPKM>1 in the window, we fitted a model of CV2 to the inverse of the mean to model the dependence of CV2 on the mean of expression. Genes with observed CV2 values above 3 standard deviation from the predicted value were selected as top variable genes. For top variable genes in each window we performed a Gene Ontology enrichment analysis using the R package GOstats and using genes expressed at RPKM>1 in the corresponding window as background. P-values were corrected by false discovery rate using the Benjamini-Hochberg method.

10.9.4. Identification of genes with differential alternative splicing

We first split the samples to prenatal and postnatal groups to identify genes with statistically significant enrichment of alternative splicing in either stage and then compared between stages.

W5 was excluded from both prenatal and postnatal groups. The comparison was implemented for DFC, HIP, AMY, STR, MD, and CBC independently. To optimize the comparison and mostly reduce the unwanted technical variance, we did filtering based on following criteria (1) exon inclusion or exclusion with at least 2 supporting reads, (2) exon PSI larger than 0.05 and less than 0.95, (3) gene with more than 2 exons. After identifying the genes and exons, we performed t-test for each gene by comparing the exon PSI of all its exons between prenatal and postnatal stage. The genes with one-side Student test p-values less than 0.01 were chosen as differentially alternative spliced. In consequence, 1,710 genes were identified, of which some genes were recurrent in different brain regions or in different developmental stages (table S7).

We also computed exon PSI for the highly variable neocortical genes (as described in the Identification of top variable genes and Gene Ontology analysis), and subsequently compared them with those of the background genes in each developmental window. To perform a fair comparison, the background genes, equal to the number of variable genes, were randomly extracted from non-variable genes with expression values larger than 1 RPKM but excluded from the highly variables genes. To avoid sampling bias, we repeated sampling 100 times and used the mean value to represent the exon PSI for the background genes. The Student t-test was calculated to measure the statistical significance of the global difference between highly variable neocortical genes and background genes in term of alternative splicing.

10.9.5. RNA-binding protein (RBP) analysis

A number of RNA-binding proteins (RBPs) are involved in regulating alternative splicing. Assuming gene expression could represent the abundance of its translated protein, the regulation

intensity accompanied with RBP is essentially proportional to the expression of RBP gene. Therefore, profiling the expression of RBP gene could potentially approach the exploration of gene alternative splicing. To do that, we firstly profiled the expression of RBP genes involved in alternative splicing regulation (fig. S19) (*98*). Secondly, we compared the profiles of RBP gene expression to those of exon PSI. By using quantitative correlation analyses, it is shown the strong positive correlation for NCX, HIP, AMY, STR, and MD (Median correlation coefficients > 0.5), except for less correlation for CBC (Median correlation coefficient ~ 0). Notably, the CBC is occupied by relatively higher expression in postnatal stages, which is consistent with GTEx's report but expands their finding from adult to a wider postnatal stage.

10.9.6. Validation of alternative splicing by exon specific RT-PCR

An aliquot of total RNA that was previously extracted from each brain region was used for secondary validation with RT-PCR. One µg of total RNA was used for cDNA synthesis using SuperScript III First-strand synthesis Supermix (Invitrogen). cDNA was subsequently diluted with nuclease-free water to 1 ng/µl. 3 µl of cDNA solution was used for the PCR with Advantage 2 DNA Polymerase (Clontech, Mountain View, CA) under the following conditions: activation at 95ºC for 1 minute, followed by 30 cycles at 95°C for 30 seconds and 68ºC for 1 minute. Transcript or gene-specific high-melting temperature primers were designed using NCBI/Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) (table S16). Expressed sequence information was obtained from GenBank (NCBI) and our mRNA-Seq data set. Exon specific RT-PCR for *GRIN1* was performed using the same cDNA template PCR products were run on Agilent Bioanalyzer for quantification of each band that is specific to either inclusion or exclusion of an alternative exon.

11. Small RNA-Seq data analyses

For this analysis, the schematic of the workflow is shown in fig. S3. All samples (table S2) passed quality control using FastQC (*99*), considering a wide range of quality metrics including per-Sequence and per-base quality, per-Sequence and per-base GC content, per base N content, sequence duplication levels, and overrepresented sequences. Sequence reads were clipped to remove the Illumina small RNA 3' adapter (TGGAATTCTCGGGTGCCAAGG). Clipped reads with length less than 18 bases were excluded in the downstream analyses. The length distribution analysis of the clipped reads showed an enrichment for microRNAs (miRNAs, ~22 bps in length) (fig. S17). Hence, the clipped reads were collapsed into the multi-fasta format that contains only unique sequences and associated counts for each read were used as input into the pipeline constructed around the miRDeep2 miRNA analysis software described in (*100*).

The trimmed reads were searched against human mature and precursor/hairpin databases in miRBase (release 21) (*101*). Reads not mapped to known targets were searched against non-human mature miRNAs in miRBase, as well the entries in Rfam database (*102*). Still remaining reads were mapped to the human genome (hg38) to identify read clusters corresponding to potential novel miRNAs. All mapping procedures were performed using Bowtie (version 1.2.1.1).

The multidimensional scaling analysis was implemented for expressed mature miRNAs with count per million (CPM) greater thanor equal to 5 in at least 5 samples. The Euclidean distance between any pair of samples were calculated and the R function *cmdscale* was used to calculate two-dimensional MDS. The returned two vectors were used as coordinates x and y for each sample to

make a scatter plot in a 2-dimensional plane (fig. S17). To survey the effects from confounders, we performed linear regression analysis between top 10 multidimensional scaling components and 9 potential confounders: age, region, sequence depth, pH, RIN, sex, ethnicity, brain hemisphere, sample dissection score. The squared correlation coefficient was utilized to measure the effect extent originated from confounders (fig. S17).

12. Single cell RNA-Seq data analyses

12.1. Quality assessment and data analyses

Illumina CASAVA was used to purify the low-quality and non-identified reads, and FastQC (*99*) was used to report the fundamental quality parameters (e.g., raw reads number, sequencing base quality score distribution and GC content). To avoid adapter contamination and higher error rates in reads boundary, 12 nucleotides in 5'-end and 13 nucleotides in 3'-end were trimmed, leaving 75 nucleotide long reads for sequence alignment. Similar to data analysis conducted for bulk mRNA-Seq, the alignment of the reads was performed by STAR (version 2.4.0e) (*83*). Command line"--runMode genomeGenerate" was used to build the sequence index for both human genome and spike-in sequences. To improve the mapping quality of splice junction reads, human gene annotation retrieved from the GENCODE project (version 21) (*84*) was additionally provided. Command line "--sjdbOverhang 74" was used to construct splice junction library. In term of gene expression measurement, RSEQTools (*81*) and SAMtools (*86*) were used to calculate the gene RPKM values, and featureCounts (v1.5.0) (*103*) was used to calculate the gene reads count values for each annotation entry.

In addition to FastQC, a series of quality control measures were implemented (fig. S4). Firstly, we excluded some artificial single cells, because under the microscope the wells they occupied were observed to be empty or to be filled with more than one cell in Fluidigm C1. Secondly, uniquely mapping reads per cell were counted and only cells with at least 100,000 unique exonic reads were used. Next, the reads distribution across different chromosomes was checked. Cells with more than 20% unique reads mapped to mitochondrial chromosome were considered low quality and removed. Besides, the reads coverage uniformity is considered as an indicator for RNA degradation. We checked the reads coverage uniformity to choose those cells with fold change of average reads coverage within gene less than 3. To ensure the enough interrogation of human transcriptome, cells with less than 1000 genes having RPKM > 1 were excluded. In total, 1198 out of the 1,512 captured single cells passed all QC filters (table S3).

12.2. Clustering and classification

We filtered genes and cells previous to clustering. We kept 25,663 RNA transcripts which were detected with at least 25 read counts cumulatively in all single cells that passed QC and log2 RPKM>1 in at least 5 single cells. We used iteratively the method described in Lake et al. (*27*) for clustering and classification which uses unsupervised hierarchical clustering coupled to random forest classification. We applied the method in a first prospective round and excluded cells derived from brain HSB433 that appeared suspiciously clustered together showing only markers for interneurons. We re-ran the analysis on the remaining 814 cells, independently on embryo and fetal cells, obtaining 24 clusters (fig. S20). We labelled obtained clusters on the basis of well-known cell-type markers. Log transformed RPKMs of selected marker genes were plotted using R package heatmap.3 (*104*). We selected those genes with FC>10 between any two nodes in the

tree of clusters to perform a T-distributed stochastic neighbor embedding plot (*105*), which was merely used to present data in two dimensional coordinates. Perplexity was set to 20. To analyze robustness to data processing we re-clustered cells using the Seurat package. We regressed out the per-cell number of genes and clustered based on the first 30 principal components on the top variable genes. The resolution parameter to find clusters from the same PCs was set to 1.5.

12.3. Marker genes selection for SC clusters

To identify marker genes for each individual cell type we performed differential expression (DE) analyses using Seurat function FindAllMarkers (*106*). In brief, we took one group of cells and compared it with the rest of the cells, using a binomial model. For any given comparison, we only considered genes that were expressed by at least 50% of cells in either population. Genes that exhibit p-value under 0.01 were considered statistically significant. Heatmaps were generated with R function heatmap.3 (*104*) (fig. S22).

12.4. Alignment of single-cell datasets

We used the method described in (*107*) and implemented it in Seurat package. When aligning prenatal single cells to adult single nuclei, we used the union of the top 3000 most variable genes. Since a large number of nuclei were captured in the adult brains, the alignment was performed by randomly sampling 100 nuclei from each cell type as representative. In addition, we aligned prenatal cells to adult cells from Darmanis et al. (*9*) using the union of the top 2000 most variable genes. After alignment of the datasets, a tSNE plot was produced to visualize the cell type correspondence between datasets on the aligned canonical correlation dimensions (fig. S23).

12.5. Cell type deconvolution

The integrated analysis of bulk mRNA-Seq and scRNAseq provides an opportunity to survey the cell type composition of tissue samples. To achieve this goal, the first step is to get the cell type signature. Conceptually, the cell type signature should be a combination of a series of gene markers specific to the cell type. Theoretically, the number of gene markers is not a big contributor, but the level of specificity strongly affects the performance of deconvolution. Considering our tissue samples span from early embryonic period to adulthood, we used adult scRNA-Seq data set from (*27*), in addition to our prenatal scRNA-Seq data set. To balance the number of cells, we randomly chose 100 cells in each cell type. For a few cell types with less than 100 cells, all cells were utilized. In total, nine cell types - neural progenitor cells (NPC), excitatory neurons (ExN – prenatal and adult): interneuron0s (InN– prenatal and adult), astrocytic lineage (Astro), oligodendrocyte lineage (oligo) endothelial cells (Endo) were involved in this analysis. Using *glmLRT* function from edgeR packages (*108*), we performed differential expression analysis to find genes specific to certain cell type beyond the statistical significance. For each cell type, the top significantly enriched genes with at least 2-fold change compared to any other cell type were chosen to construct cell type signature. Notably, the qualified cell type signature matrix should ensure each gene has the significantly enriched expression values in only one cell type. Next, we used CIBERSORT R script to perform cell type deconvolution (*109*). The expression profiles of tissue mRNAseq sample and the expression signatures of nine cell types were simultaneously input to the pipeline, calculating the cell composition of complex brain tissue. These analyses were applied to 11 neocortex areas (fig. S29). The results were converted to the percentages of each cell type and were presented as function of developmental windows, leading to the illustration of temporal dynamics of cell

composition. For replication of the deconvolution analysis using two published fetal datasets (*12, 26*) from the prefrontal cortex we used the same method and threshold criteria.

To determine when neuronal layers appear, we divided excitatory neurons derived from adult single cell dataset (*27*) into deep-layer (i.e., L5 and L6) and upper-layer (i.e., L1, 2/3, 4), and performed our deconvolution analysis. This analysis was replicated using three different fetal datasets (ours, (*12, 26*)) in conjunction with Lake et al 2016 (*27*) adult dataset (fig. S27).

13. Single nuclei RNA-Seq data analyses

13.1. Processing of single nuclei RNA-Seq

We employed the commercial software *cellranger* to implement the preliminary data processing for single nuclei RNA sequencing generated from 10x Genomics Chromium 3' Solution (10x Chromium) platform. In brief, the *cellranger mkfastq* converts binary base call (BCL) files to FASTQ files and simultaneously decodes the multiplexed samples. *cellranger count* took the FASTQ files to perform sequencing alignment against reference genome, quality filtering of low quality reads, cell barcode counting and filtering, and unique molecular identifier (UMI) counting for each annotated gene. We used the default parameters, except for customizing the number of excepted cell number to 10,000 and constructing the "pre-mRNA" reference for single nuclei RNA-Seq. We utilized GRCh38/hg38 reference genome and GENCODE v21 annotation for human.

13.2. Classification of cell type and subtype

The sparse expression matrix generated from *cellranger* analysis pipeline serves as an input to the Seurat software (*110*) for the classification of cell type and subtype. To optimize cell classification analysis and reduce unwanted technical variance, we first did quality filtering, normalization, and data scaling according to Seurat's guidelines. For example, we excluded a cell with the number of expressed genes less than 300 or more than 7000, or with the number of UMI less than 300 or more than 20000, or the percentage of mitochondria reads more than 5%. The normalization method was "*LogNormalize*" and the scale factor was 10000. In addition, the linear regression was performed by choosing the percentage of mitochondria reads as variable. After purifying the expression matrix, the 100 principle components were computed and their significances were tested by *JackStraw* and visualized by *Elbow* plot, with the result of finally choosing the top 25 principle components for tSNE and clustering analyses.

Next, we used the R scripts *SpecScore.R* from (*111*) to compute the specificity score for each gene in each cell cluster. Since a cell cluster mostly corresponds to the cell type, we computed the correlation matrix to study the relationship between cell clusters by using the gene specificity score. We manually checked the hierarchical cluster and the top ranked genes in each cell cluster to determine which cell clusters could be assigned to what cell type. After determining the cell type, we re-computed the specificity score for each gene to choose the top ranked gene as the cell type signature markers for other downstream analyses (table S8). Under each cell type, the cell clusters were assigned to different cell subtypes. Lastly, we recruited the well-known gene markers to verify the assignment of cell types and subtypes. Notably, cells with inconsistent or no assignment were grouped to "Unassigned" category.

14. <u>DNA methylation data analyses</u>

14.1. Data normalization

The raw data from the Illumina 450K arrays were normalized by Fresco (https://github.com/paulmanser/fresco). Briefly, a group of loci from housekeeping genes whose CpG sites were known to be consistently highly methylated or unmethylated in all cell types within brain tissue were identified (*112*). Then a three-dimensional local regression surface to the deviations of each array from the average was fit, as a function of non-biological variables such as average signal intensities and probe CG content. This gave an estimate of the distortion due to technical factors on each array for each combination of our non-biological predictors. This estimated distortion was subtracted from raw signals and these adjusted signals were then used to obtain the β-value, an estimate of the proportion of cells methylated at each locus (*113*). It has to be noted that this process does not use any information about batches, but is nonetheless able to remove batch effects as measured by sample clustering and one-way ANOVAs for batch effect (fig. S6). After normalization, a high correlation was observed between biological replicates (fig. S7).

14.2. Identification of cell type specific methylation sites

Cell type specific methylation probes were defined as those CpG sites that exhibit 30% methylation level differences (a delta (β) $\geq$0.3) between the neuronal and non-neuronal cell samples from (*114*). A total of 27,349 sites were identified.

14.3. Comparison of DNA methylation trajectories in fetal versus postnatal development.

For each one of the 461,594 methylation sites (MS) analyzed, DNA methylation values were fitted to a multiple linear model with the log2 of the post-conception days and sex. Analysis was

performed with data from all neocortical regions, independently testing prenatal and postnatal samples to overcome any possible uncorrected batch effect and the fact that prenatal samples came from an independent external dataset (*115*). The probes were then classified into three categories – those showing an increasing methylation trajectory, decreasing, or non-changing through age. Age-correlated categories required a Bonferroni corrected p-value <0.05 and absolute difference of beta (delta(β)) values between the first and the last fitted points predicted by the model to be greater than 0.1. Non-changing trajectories required nominal p-values <0.05 and delta(β) <0.1. All possible combinations of these three categories for pre- and post-natal periods were then constructed to obtain the nine subcategories.

MS in each category were tested for enrichment of 4 different types of elements:

1) Genomic elements related to genes. For defining genomic elements, the GENCODE (version 21) gene model in hg38 was used to map MS to exons, introns, intergenic regions or promoters (defined as 1Kb regions around the TSS).

2) Putative enhancers, defined as H3K27ac ChIP-Seq peaks in six DFC samples (3 fetal, 3 adult). Peaks overlapping TSS were excluded to avoid assessing promoters. Adult and fetal-specific enhancers were defined as explained in section 15.3.

3) Cell type specific methylation sites in brain from (*114*), as in section 14.2.

4) Genes with cell type specific expression in brain (as described in section 12.3)

All cited categories were compared to sites that do not change through age in pre- or post-natal periods. When testing for enrichment in genes instead of MS, genes related to each category's MS according to GENCODE annotations were used. Enrichment p-values were obtained by means of

a two-tailed Fisher exact test. Odd ratios were also calculated applying Haldane correction when needed. Odds ratios for each category are shown in the rightmost panel in fig. S35.

14.4. Differentially methylated regions

To obtain regions of the genome with significant clustering of MS belonging to each category of trajectories we calculated differentially methylated regions (DMRs) independently in prenatal and postnatal samples by considering the multiple linear model of the DNA methylation values per MS (see above) and by determining series of adjacent correlated p-values (comb-p --seed 0.05 --dist 300) (*116*). Briefly, what comb-p does is to calculate the auto-correlation between p-values, to combine adjacent ones performing a false discovery adjustment and to assign significance to each reported region. Only MS that were annotated in hg38 (461,453 MS) were evaluated. We filtered regions by imposing a corrected p-value lower than 0.05 and at least 3 MS included in the DMR. We then recognized hyper methylated or hypo methylated DMRs taking into account the direction and the amount of beta change: the slope of the linear model of all MS within a DMR should have same sign, positive for hyper and negative for hypo methylated regions, and the average absolute delta values should be higher than 0.1. Comparable DMRs between prenatal and postnatal stages were identified as the intersection of hyper or hypo methylated regions for any of the two developmental periods and imposing that smaller regions still maintain an absolute delta value higher than 0.1. Finally, we called a prenatal hyper or hypo DMR as fixed in postnatal if the region does not overlap with any reported postnatal hyper or hypo DMR and has an average delta for all MS in the region lower than 0.1 in postnatal samples. In an analogous way, we determined postnatal hyper or hypo DMRs that are fixed during pre-natal development. For the DMRs

included in each of the 8 categories resulting of these comparisons, we obtained the list of genes for which MS belonging to these DMRs were annotated (see above).

15. ChIP-Seq analyses

15.1. Peak calling, annotation and normalization

Sequencing reads from each sample (table 6) were mapped to human genome assembly hg38 (NCBI GRCh 38) using Bowtie2 (*117*),

bowtie2 --no-unal -p 4 --fast --end-to-end -q -x hg38 -U fastq -S SAM

MACS 2.0 (*118*) was used to detect peaks for H3K4me3, H3K27ac and CTCF. For these marks, aligned reads from all DFC and CBC samples were pooled and analyzed by MACS (FDR ≤ 0.05) to detect the peaks enriched in immunoprecipitated samples over the input.

macs2 callpeak –t BAMFILE -c INPUT.BAM --name=NAME --format=BAM --gsize=hs --tsize=73 --keep-dup=auto --bw=300 --nomodel -q 0.05 –broad (narrow peaks for CTCF).

Each brain sample was then analyzed individually. If there was at least a 50 % overlap between a peak in an individual brain sample and a peak in the master list, that peak was scored as detected in that brain. A peak was defined as present if it was detected in multiple brains (2 out of 3 or 3 out of 4) in DFC or CBC. Enriched peaks in H3K27me3 were detected by SICER 1.1 (*119*), with parameter settings of window size as 1000bp, gap size as 3000bp, and FDR ≤ 0.01. An approach similar to other marks was applied to construct a master list of H3K27me3 peaks and to score a

present peak after comparing individual brain sample peaks. To compare biological replicates, peaks with FDR ≤ 0.1 were selected and plotted as scatter plots. The function 'kde2d' in R package was used to create a 2D density image for each comparison. A high correlation was observed between replicates (fig. S8).

All peaks were annotated by our in-house Python scripts according to database GENCODEv21. We selected all transcripts in the database with types “protein_coding”, “lincRNA”, “snoRNA”, “snRNA”, “miRNA”, “misc_RNA”, and “rRNA” for our annotation. We matched each peak with transcripts in the database as follows. First, the peaks were matched with extended promoter regions, defined as TSS ± 1kb, in the database. If peaks were within the promoter regions, the genic location of the peaks was assigned as “TSS”. If peaks were outside the promoter regions, we then tested if they were matched with any transcripts. Where peaks matched multiple transcripts, we selected the longest one as the target transcript for further analysis. The genic location of a peak matching a transcript was assigned as “Exon” if that peak overlapped any exon. If the peak did not overlap with an exon it was assigned as an “Intron”. If peaks did not overlap any transcripts we assigned their genic location as “Intergenic” (fig. S33).

All ChIP-Seq peaks in the master lists were normalized prior to any downstream analysis. For each ChIP-Seq sample, signal values for each base pair in target peaks created by MACS were read and an average signal value calculated using the R package Conductor. Total signal value was also calculated from all peaks of each sample. The average signal value for each peak was normalized by the total signal value. These normalized average values were used for all subsequent analyses.

15.2. Differential enrichment of histone modifications and gene expression

Differential enrichment was calculated as $w=(S_{DFC} - S_{CBC}) / (S_{DFC} + S_{CBC} +1)$, where S_{DFC} and S_{CBC} are average values of normalized signals in DFC and CBC, respectively. Based on w values, peaks were defined as DFC-enriched ($w > 0.3$), CBC-enriched ($w < -0.3$) and nonspecific (-0.3 to 0.3) (fig. S34). To compare differential enrichment of histone marks with gene expression, expression RPKM value of the gene that the peak overlaps with was considered. For each of the 3 groups defined above, the RPKM values for all genes were utilized to perform t-test to check the expression difference between the groups.

15.3. Identification and analysis of temporal and spatial enhancers

In order to make possible comparisons among samples (spatial and temporal differences) we homogenized ChIP-Seq H3K27ac and H3K4me3 read libraries by evaluating read qualities using FastQC (*99*) and trimming all reads to 76bp using seqtk (*120*). To get similar cycle effect in all libraries we further clipped reads using a dynamic trimming algorithm with sickle (*121*). We then followed the same mapping and peak calling methodology than above with the exception that we trimmed the first 3bp of each read using the corresponding Bowtie2 parameter.

To construct a master table of H3K27ac-enriched regions we merged all peaks into contiguous genomic segments when peaks overlapped within or between samples at least 1bp. We excluded all peaks that were present in only one sample. We annotated peaks using the minimum distance to a TSS described in GENCODE V21. We also indicated overlap with H3K4me3 promoter enriched mark, to help discriminate enhancers from promoters. Enhancers were defined as H3K27ac-enriched segments at least at 1kb from a TSS and without any overlap with a H3K4me3-

enriched peak. Active promoters were defined as any H3K27ac-enriched region overlapping a region of 1kb around any TSS.

We classified enhancer and promoter peaks as 'fetal' or 'adult', for temporal, and 'DFC' or 'CBC' for spatial, if they were present in at least 2 of 3 brains in one group of samples and in 0 brains of the other group. To avoid the possible effect of sample-specific thresholds in peak calling we tested the difference in signal between the putative fetal or adult-specific peaks (or DFC/CBC-specific). To do so we counted reads in peaks using the R package Subread (*122*). We tested differential enrichment between fetal and adult brains with DESeq2 independently in DFC and CBC, or DFC versus CBC, in fetal and adult, using size factors estimated from all brains together from the counts of all peaks identified in any brain in the dataset. Peaks showing and adjusted P-value<0.01 and abs(log2FC)>=1 were retained in the corresponding category -specific groups.

15.4. Putative enhancers enrichment in categories

For all genes annotated near enhancers identified in DFC we obtained the results from the DESeq2 expression level contrasts between W4 and W9 brains (section 10.6). For fetal and adult enhancers, we then calculated the enrichment in genes showing differential expression between W4 and W9 (Adjusted P<0.01 and log2(FC)>=1) by means of a Fisher's exact test against genes annotated on the rest of DFC identified enhancers that were not fetal or adult specific.

We used the same test to calculate the enrichment in cell type specific genes from (list), in overlapping NUM and GUM CpG sites from (*114*) and in dynamically methylated sites categories identified in section 14.24.

16. Gene Ontology Analysis

We analyzed enrichments in gene ontology (GO) categories for genes annotated in differentially methylated regions (DMRs) for the different categories of methylated sites (see section 14.4) or the nearest gene (TSS) associated to different categories of putative enhancer peaks. We used GOstats R package for GO enrichment in a set of preselected GO categories depicting relevant neurological processes (e.g. neurogenesis or axogenesis).

17. Classification of WGCNA modules and integrative enrichment analyses

The WGCNA modules (section 10.8) were further analyzed to characterize the different spatio-temporal enrichment and consequently to unite them to one general feature. We proposed to categorize the WGCNA modules into spatiotemporally dynamic, only temporally dynamic, only spatially dynamic, and invariant. Using *lm* function from R package, we firstly fitted the linear regression model to survey the relationship between modular eigengene and two explanatory variables, i.e., six brain regions and log10 (development days). To reduce the birth disturb, the whole samples were divided into prenatal samples (W1 to W4) and perinatal/postnatal samples (W5 to W9), and were applied with the same linear regression analysis. Through summarizing the statistical significances inferred from these three linear regression analyses, most of WGCNA modules were assigned to one particular developmental and regional enrichments (table S10). On the other hand, we applied the same analysis to survey the relationship between modular eigengene and two explanatory variables, i.e., 11 NCX areas and log10 (development days), so as to gain the knowledge of developmental and regional enrichments occurred in neocortex. Next, to associate this modular classification to a particular biological function, cellular composition or biological

process, we calculated hypergeometric test for modular enrichment of several gene lists gained from different analyses of this study, including sex-DEX, enhancers, methylation sites, GUMs, NUMs, cell types, disease associated genes (fig. S37 and table S11). Finally, using hypergeometric tests, we observed the significant coincident occurrence within individual modules of genes whose expression was enriched in neurons, genes associated with NUM sites, and genes associated with fetal enhancers, which we refer to as Neuronal or N-type associations. Similarly, we observed the coincident occurrence of genes whose expression was enriched in glia, genes associated with non-NUM sites, and genes associated with adult enhancers, which we deemed Glial or G-type associations. Again using hypergeometric tests, we found that N-type associations were enriched among modules in the spatio-temporal set, G-type associations were enriched among modules in the temporal, non-spatial set, and modules enriched for genes exhibiting sex-different expression were enriched among the non-spatial, non-temporal module set. Multiple testing correction was achieved by Benjamini-Hochberg FDR < 0.05 taken together P-values for sex-DEX, enhancers, methylation sites, GUM and NUMs, at P-Bonferroni < 0.05 for cell types, and FDR < 0.10 for disease associated genes.

For the MAGMA gene-set analysis, an initial gene analysis was performed on the SNP p-values for each of the phenotypes, using the SNP-wise (multi) model, using the European panel of the 1,000 Genomes Phase 3 data was used as reference data. Sample sizes per SNP were included in the analysis for the phenotypes for which they were available. SNPs were annotated to genes based on dbSNP 147 SNP locations and GENCODE V21 gene definitions, mapping a SNP to a gene if it was located in the transcription region of that gene, or within 2 kb upstream or 1 kb downstream of the transcription region. In total, 59,140 had at least one SNP mapped to them.

The gene-set analysis itself was performed at default settings for each phenotype; the average sample size per gene was included as a covariate in the analysis if available for that phenotype. In addition, the GENCODE gene definitions where used to categorize all genes as protein-coding or not. This was included as a covariate in the analysis as well, to correct for possible systematic difference between coding and non-coding genes.

18. Hi-C data integration

We have leveraged Hi-C data from the adult dorsolateral prefrontal cortex (DLPFC, (*52*)) and fetal cortices (*53*) to annotate GWAS risk loci. To compare interaction profiles in the adult and fetal cortex, we combined previously generated Hi-C datasets from two fetal cortical layers to obtain comparable read depths (details provided in the companion paper, (*52*)). After mapping and filtering the reads, we constructed normalized contact matrices at 10kb resolution for loop analyses. Promoter-based loops were identified as previously described (*53*). Briefly, we constructed background interaction profiles from randomly selected length- and GC content-matched regions to promoters (defined as 2kb upstream of transcription start sites based on Gencode v21). Using these background interaction profiles, we fit interaction frequencies into Weibull distribution at each distance for each chromosome using the *fitdistrplus* package in R. Significance of interaction from each promoter was calculated as the probability of observing higher interaction frequencies under the fitted Weibull distribution, and interactions with FDR<0.01 (which corresponds to P-values~$1x10^{-4}$) were selected as significant promoter-based interactions. In total, we detected 561,634 and 688,782 promoter-based interactions in the adult and fetal cortex, respectively.

19. Enrichment of disease and personality trains associated variants in regulatory regions

We obtained files with summary statistics (SS) for a number of neuronal and non-neuronal disorders or traits:

- Autism spectrum disorder (ASD; (*40*))
- Attention-deficit/hyperactivity disorder (ADHD; (*41*))
- Schizophrenia (SCZ). CLOZUK (*37*)
- Major depression disorder (MDD); (*42*))
- Bipolar disorder (BD; (*43*))
- Intelligence quotient (IQ; (*44*))
- Neuroticism (*45*)
- Parkinson's disease (PD; (*39*))
- Alzheimer's disease. IGAP. (*38*)
- Height. GIANT (*46*)
- Total cholesterol levels (*48*)
- Inflammatory bowel disease (IBD; (*47*))
- Glycated haemoglobin (HbA1C; (*49*))

We calculated enrichment in SNP-heritability using H3K27ac (active promoter and active enhancer mark) enriched regions identified in fetal and adult brains in this study. For comparative

purposes and to gain temporal resolution we complemented the dataset with H3K27ac enriched regions identified in three infant brains in our dataset and also embryonic brains from (*21*) study. The additional samples were trimmed, mapped and analyzed following the same methods described above, and a new set of merged regulatory elements were created. We also obtained coordinates for active enhancers identified in other tissues in the Roadmap epigenomics project (http://www.roadmapepigenomics.org). Classification of active enhancers and promoter regions in embryonic/fetal/infant/adult and DFC/CBC groups was done following these criteria.

- DFC: Present in at least 2 samples in DFC brains.
- CBC: Present in at least 2 samples in CBC brains.
- Adult: Present in at least 2 adult samples.
- Fetal: Present in at least 2 fetal samples.
- Infant: Present in at least 2 infant samples.
- Embryonic: Present in at least 2 embryonic samples.
- Adult-DFC: Present in at least 2 adult samples from DFC.
- Adult-CBC: Present in at least 2 adult samples from CBC.
- Fetal-DFC: Present in at least 2 fetal samples from DFC.
- Fetal-CBC: Present in at least 2 fetal samples from CBC.
- Infant-DFC: Present in at least 2 infant samples from DFC.
- Infant-CBC: Present in at least 2 infant samples from CBC.
- Adult-DFC_sp: Present in at least 2 adult samples from DFC and 0 from CBC.
- Adult-CBC_sp: Present in at least 2 adult samples from CBC and 0 from DFC.
- Fetal-DFC_sp: Present in at least 2 fetal samples from DFC and 0 from CBC.
- Fetal-CBC_sp: Present in at least 2 fetal samples from CBC and 0 from DFC.
- Infant-DFC_sp: Present in at least 2 infant samples from DFC and 0 from CBC.

- Infant-CBC_sp: Present in at least 2 infant samples from CBC and 0 from DFC.

We clumped GWAS hits trait-associated regions using PLINK (*123*). We used a P<1e-7 for the index SNP and clumped based on adjacent SNPs in a window of 200 kb with and LD r2>0.8 and a P-value <1e-4. LD values were calculated for a subset of 86 CEU samples from the 1000 Genomes Project (*50*).

To produce a comprehensive list of putatively disease-associated genes, we crossed SNP clumps with the complete ChIP-Seq dataset of H3K27ac regions which includes embryo, fetal, infant and adult samples (161,641 regions in total). We then annotated those ChIP-Seq peaks putatively associated to disease with the following strategy:

1 All peaks were crossed with 10 Kb resolution loops determined by Hi-C from (*51-53*). Together, these Hi-C datasets identify loops occurring in developmental fetal and adult brain. We annotated all interacting genomic windows of 10Kb with the overlapping TSS (+/- 1kb) from GencodeV21. Then, we annotated each peak according to all of its interactions with TSSs.
2 We discarded all peak-to-peak interaction that were not concordant in time and brain region (e.g. a DFC-fetal only peak interacting with a TSS with a CBC or DFC-Adult peak). We also discarded interactions that were only detected at one time period (fetal or adult) and involved peaks only detected in the opposite time period.
3 All peaks that were themselves overlapping a TSS (+/- 1kb) were also associated to the corresponding gene.

4 Peaks with no evidence of Hi-C interactions with a TSS or overlapping a TSS were annotated to the closest open TSS that were concordant in time and brain region. Finally, genes containing significant SNPs (P<1x10-4) in associated clumps that were STOP-gain or missense mutations, or fell in UTRs or splicing regions were also included. We annotated SNPs using Ensembl Biomart V78 (*124*).

5 The method was run independently using Hi-C from (*51*) and from (*52*) and the resulting genes were combined.

6 A high confidence list of genes was produced by excluding those genes that were only supported by the Hi-C loop of only one dataset or was annotated by proximity using only one Hi-C dataset and that was the only supporting evidence.

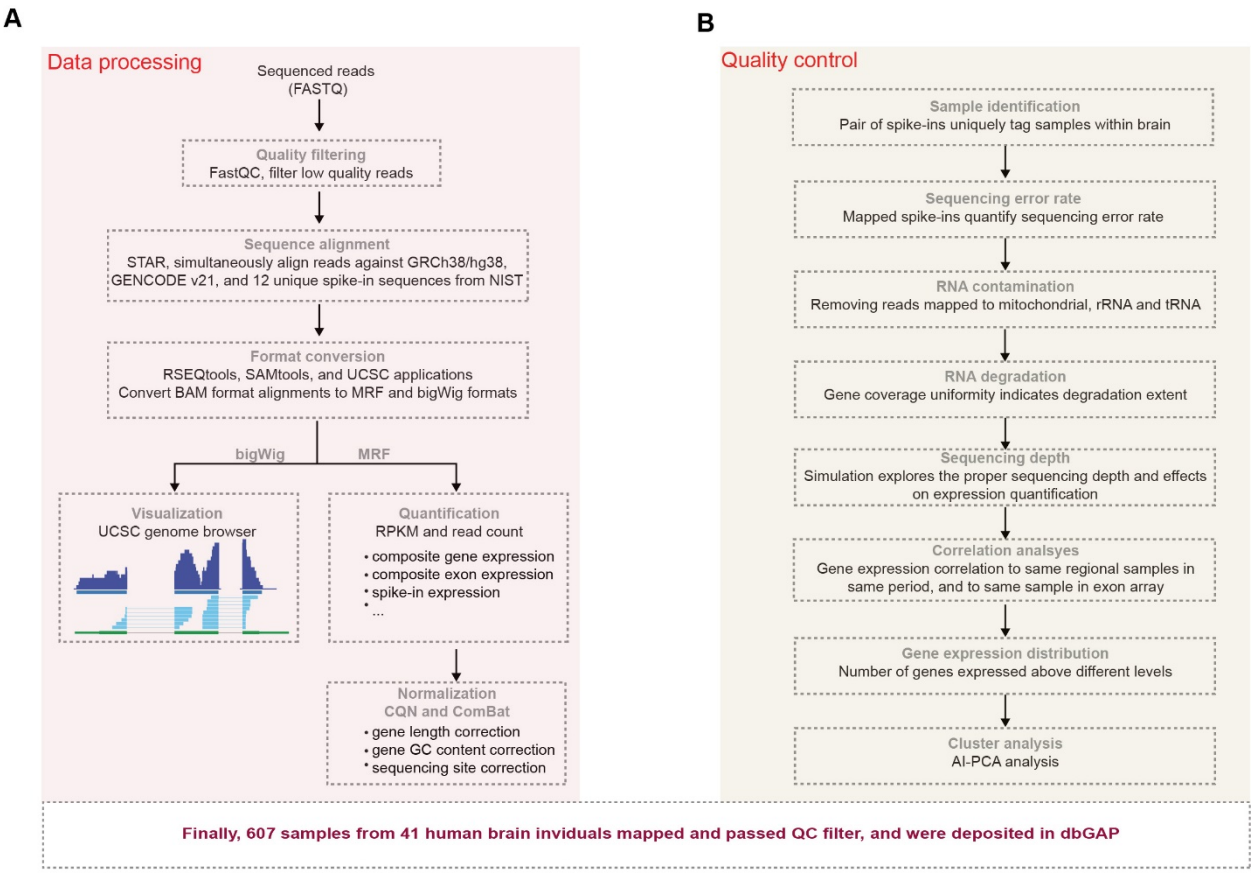


Fig. S1. mRNA-Seq analysis pipeline including data processing and quality control.

(A) Starting with sequenced reads in FASTQ format, we processed data by performing quality filtering, sequence alignment, format conversion, data visualization, expression quantification, data normalization, and batch correction. (**B**) Library preparation and sequencing quality were also subject to a quality control pipeline assessing sequencing error rate, RNA contamination and degradation, sequencing depth, and other metrics (see (*22*)).

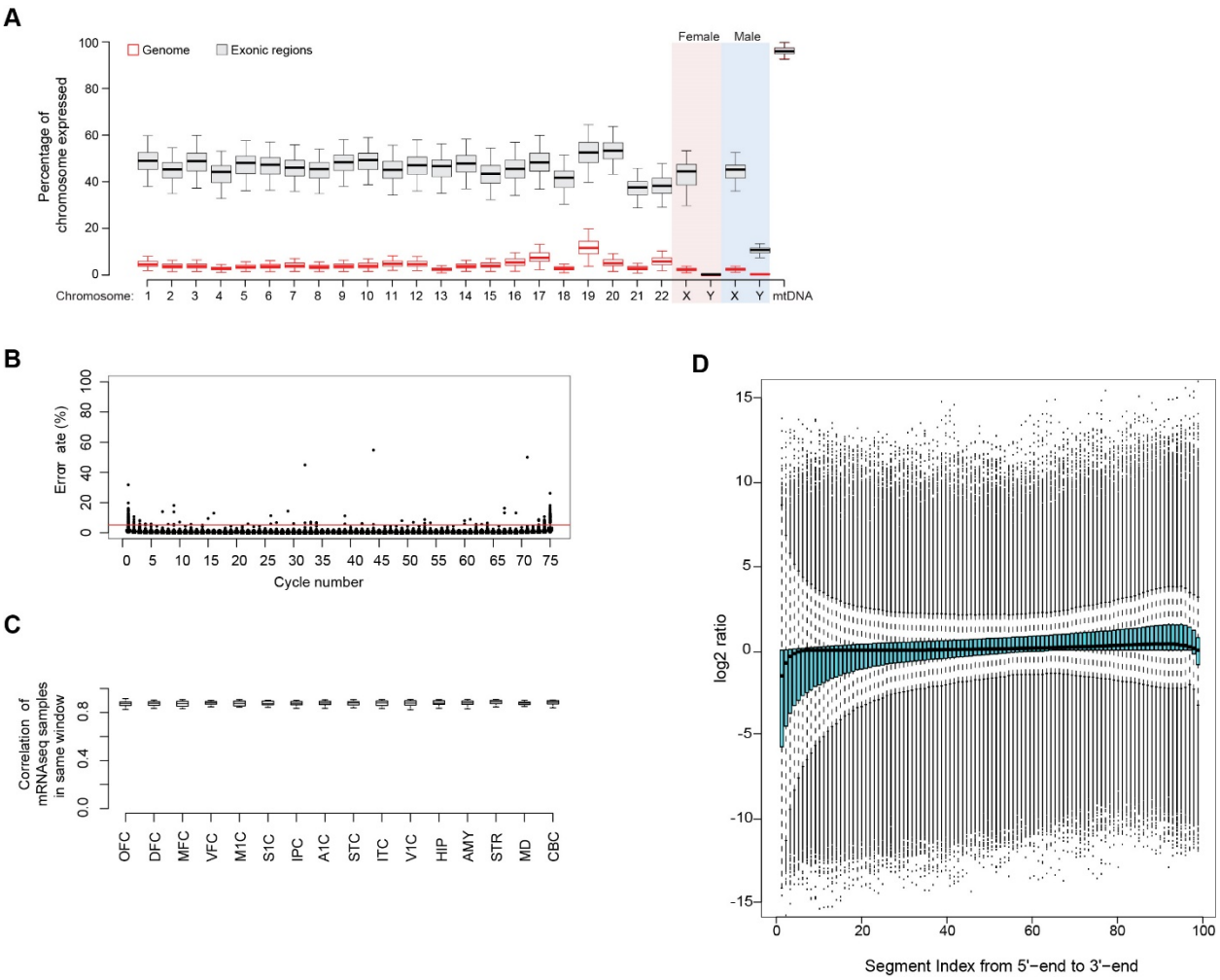


Fig. S2. Quality assessments for mRNA-Seq dataset.

(**A**) Distribution of the ratios of transcribed protein-coding genes per chromosome relative to the genome (red) and the exonic regions (gray). Male and female samples are separated for sex chromosomes. (**B**) Box plots of the sequencing error rate per cycle number. The sequencing error rate was estimated by the ratio of alignment mismatch of the sequenced spike-in reads against the reference spike-in RNAs. The red line represents the 5% limited baseline error rate. All box plots are under the baseline except a small number of outliers at the highest number of cycles (> 70 cycles). (**C**) Distribution showing the correlation between mRNA-Seq samples in the expression of a given gene, in a given neocortical area or brain region, in a given window; mRNA-Seq samples

from the same brain region and same developmental window were treated as biological replicates to implement Spearman correlation analysis. (**D**) Uniformity of read coverage along GENCODE genes. The composite exon-model gene was split into 100 equal segments from the 5'-end to the 3'-end. For each segment, the median RPKM value of one gene was calculated. The distribution of the ratios in all genes is shown by cyan colored box plots and ordered by increasing segment number.

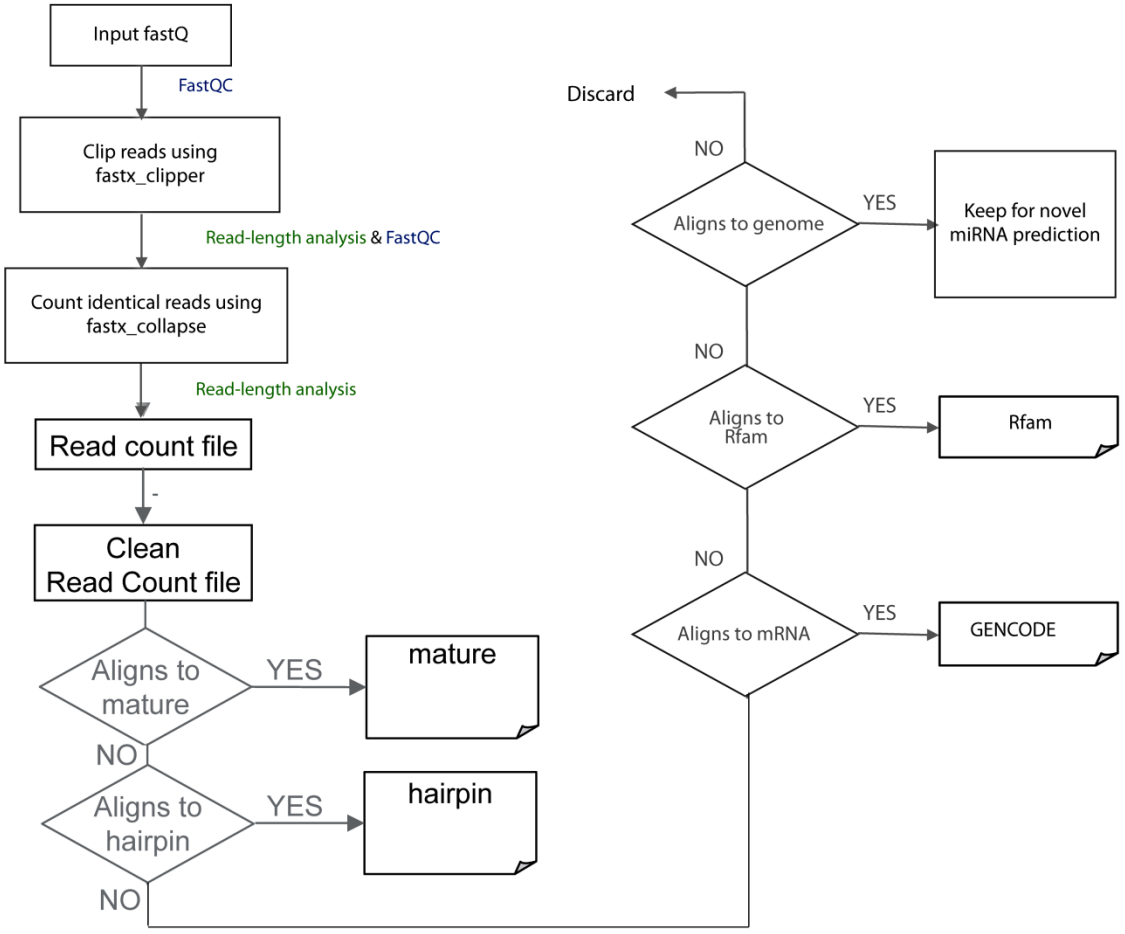


Fig. S3. Schematic workflow used for analyzing small RNA-Seq.

Using a modified version of the miRanalyzer pipeline, raw small RNA-Seq reads were clipped to remove the 3' adapter sequence, collapsed into a canonical set of sequences and associated frequencies, and searched against databases of mature miRNA sequences, hairpin miRNA sequences, GENCODE mRNAs, and entries in the Rfam database. Reads that failed to match any of the previous databases but still could be aligned to the genome were retained to explore potential novel miRNA candidates.

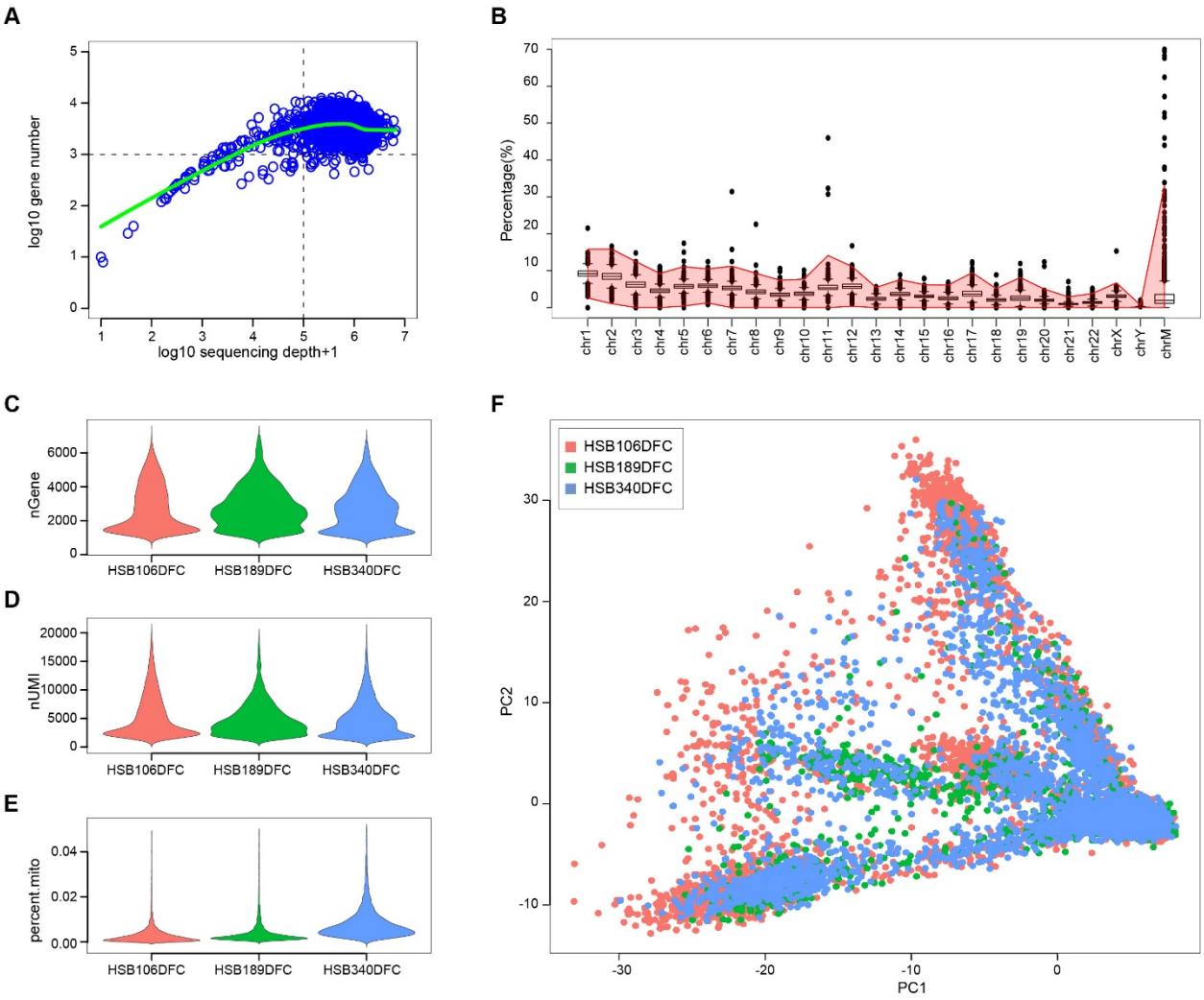


Fig. S4. Quality assessment of single cell and single nuclei RNA-Seq.

(**A**) Dependency between sequencing depth (log10 transformed) and detectable gene number (log10 transformed), as depicted by the green line. The horizontal and vertical dash lines indicate 1,000 genes and 100,000 sequence depth, respectively. (**B**) Percentages of reads from the entire single cell dataset aligned to different chromosomes. The pink shaded area indicates the SEM±3SD. (**C**)(**D**)(**E**) and (**F**) are metrics of the single nuclei RNA-Seq generated from DFC of three adult brains using 10X Genomics chromium platform. (**C**) (**D**) and (**E**) Violin plots show the distribution of number of genes detected, total UMI counts and percent of reads aligning to

mitochondrial genome in each sample. (**F**) PCA plot of the single nuclei from the three brain samples.

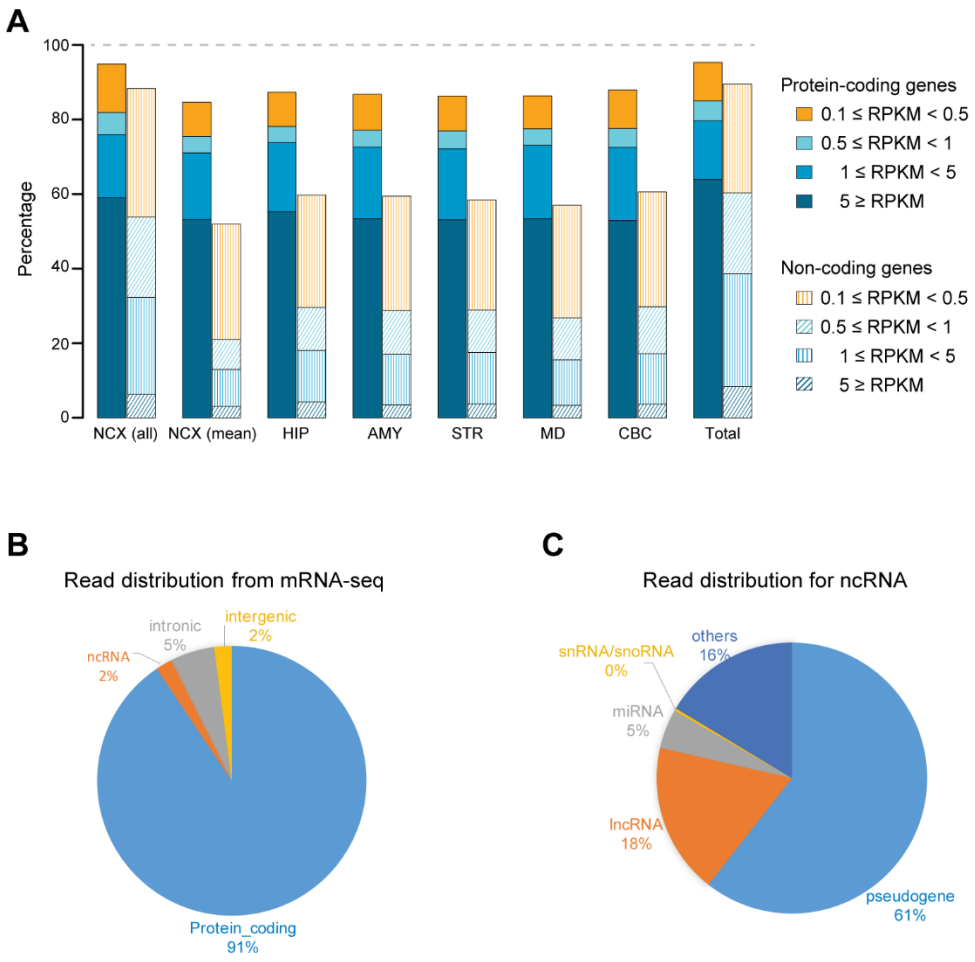


Fig. S5. Overview of gene expression distribution and categories for mRNA-Seq.

(**A**) Bar plots representing the percentage of protein-coding genes (filled bars) and non-coding genes (thatched bars) expressed at varying RPKM levels (RPKM ≥ 0.1, orange; RPKM ≥ 0.5, turquoise; RPKM ≥ 1, blue; RPKM ≥ 5, dark blue) in at least one sample in each region. The 11 NCX areas are either combined (all) or averaged (mean). The last grouped bars (Total) are the combination of all 607 samples from 41 specimens. The genomic attributes of gene annotation were retrieved from GENCODE v21. (**B**) Pie chart representing the percentages of uniquely mapped reads aligned to protein-coding genes (blue), ncRNAs (orange), intronic (grey), and intergenic (yellow) regions. (**C**) Pie chart representing the percentages of ncRNA aligned reads

categorized into pseudogene (blue), lnRNA (orange), miRNA (grey), snRNA/snoRNA (yellow), and others (dark blue).

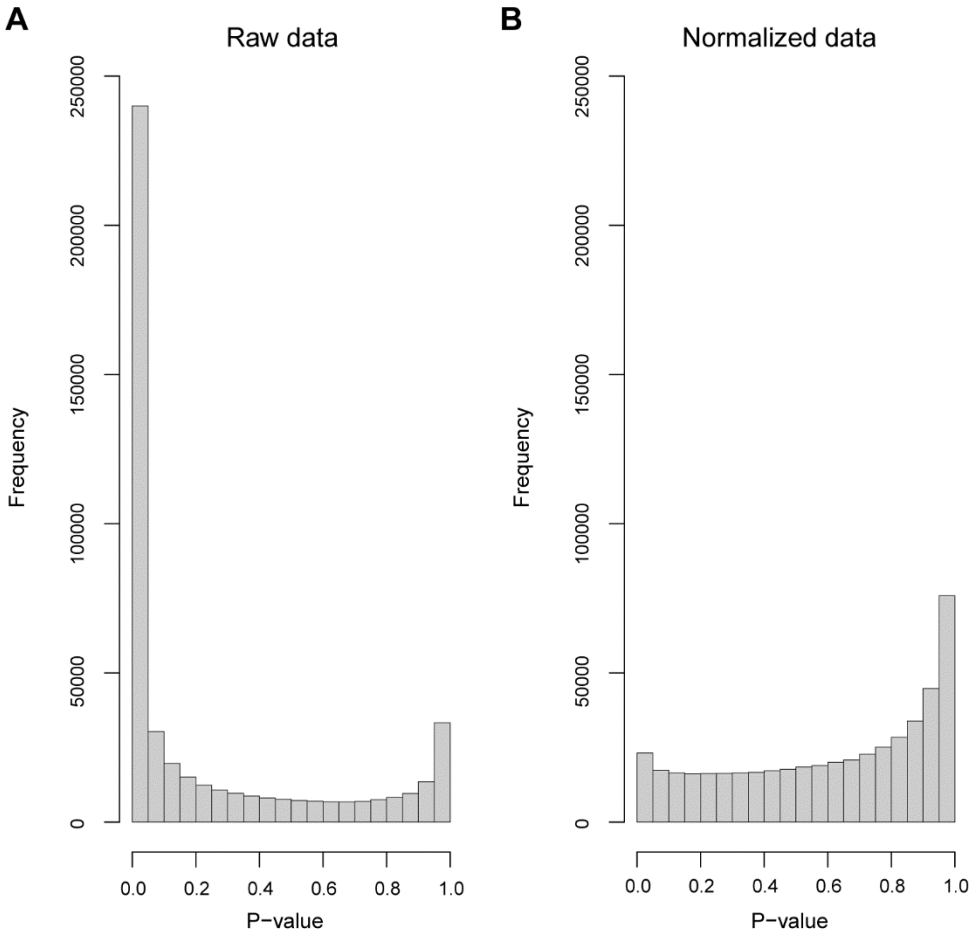


Fig. S6. Effect of DNA methylation data normalization on batch effects.

(A) The distribution of p-values for batch effect over all probes in one-way ANOVA analysis using beta-values after Illumina standard normalization procedure. (**B)** The distribution of p-values for batch effect using beta-values after our normalization procedure showing that batch effect has been removed to a large extent, despite our normalization procedure not using any information about batch identity.

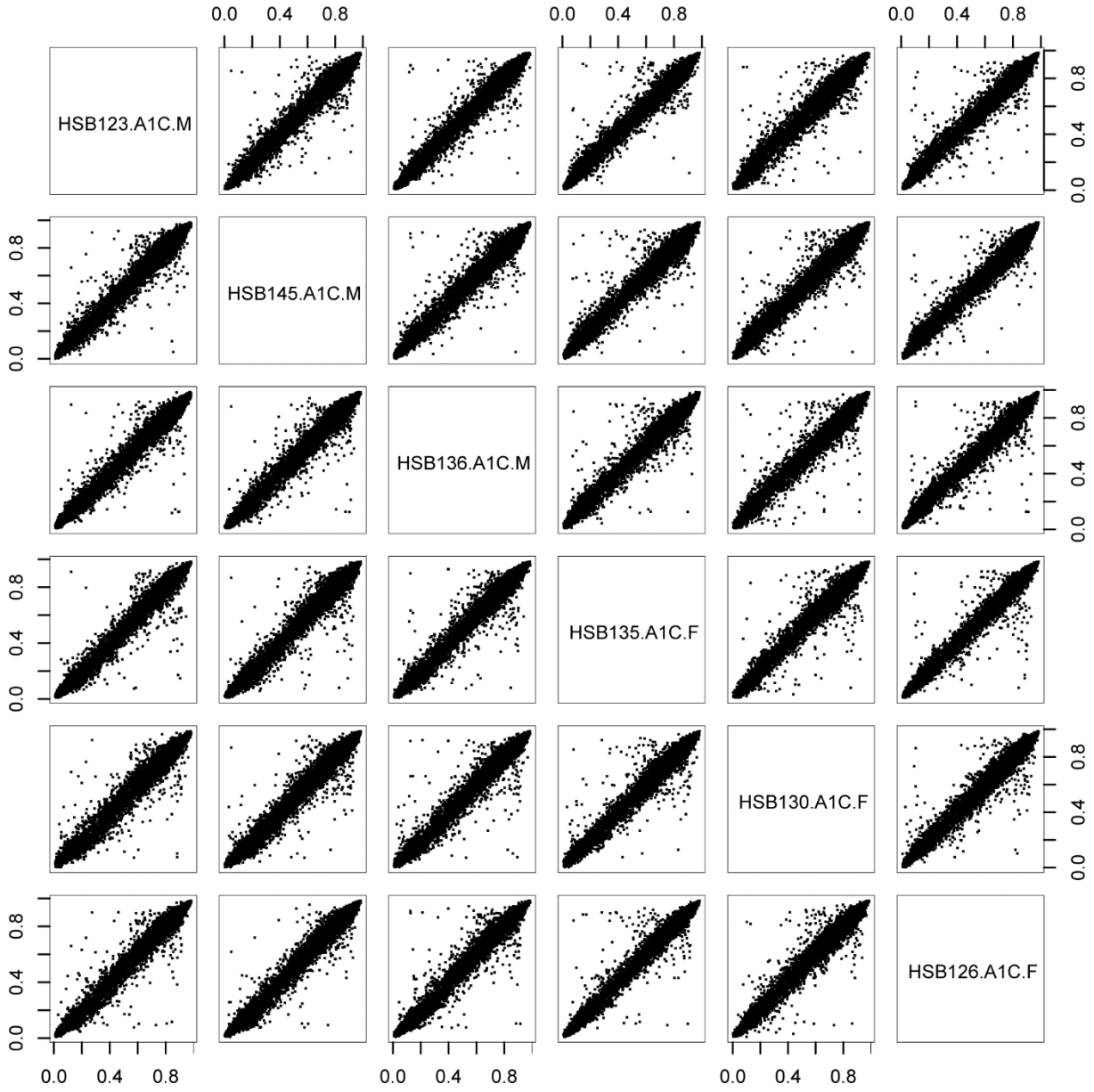


Fig. S7. Correlation between biological replicates for DNA methylation.

Biological replicates for DNA methylation analysis correlated strongly with one another. Examples from A1C samples after the normalization procedure are shown here, with similar results observed with all other regions.

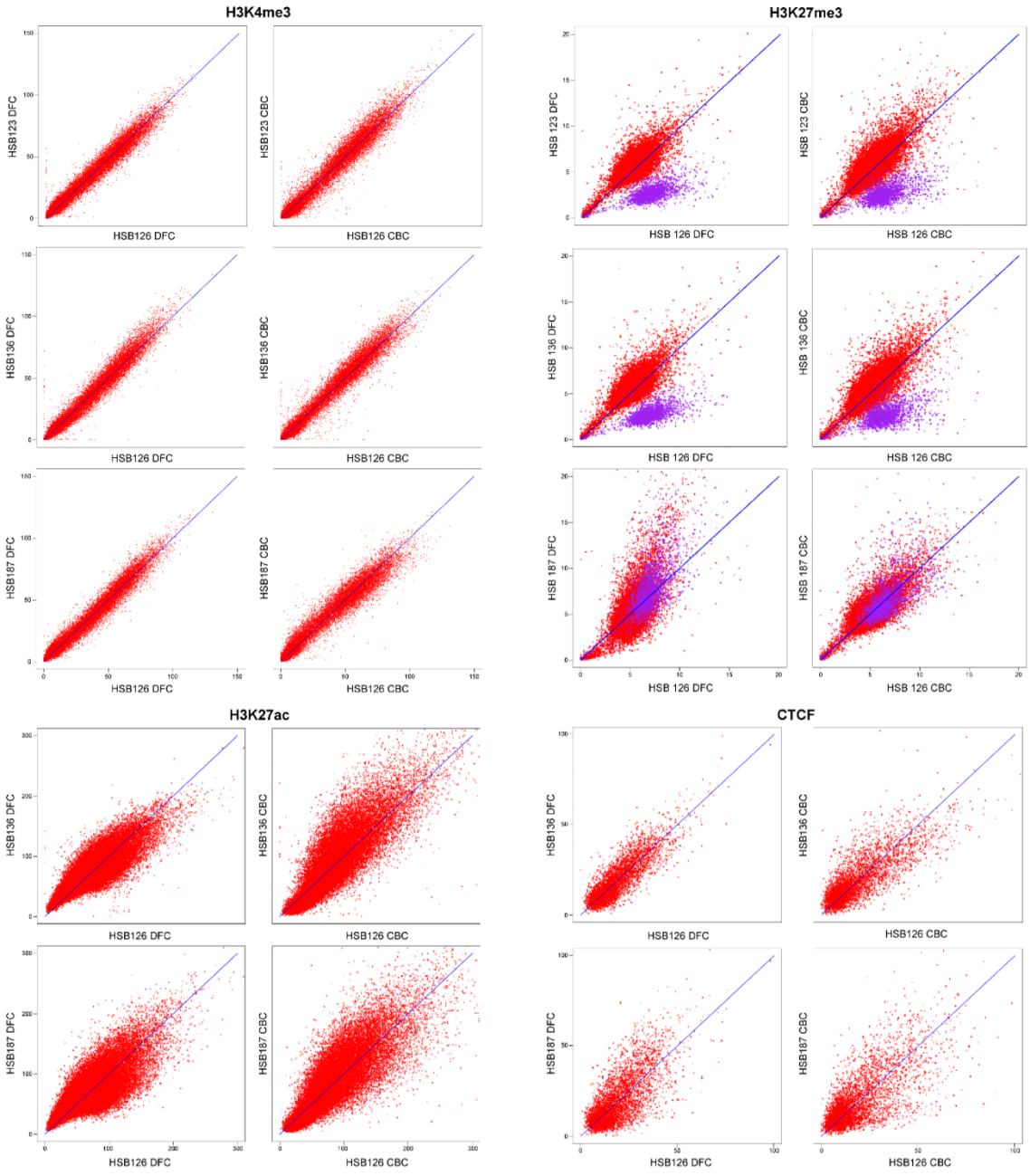


Fig. S8. Correlation between biological replicates for ChIP-Seq.

Correlation between biological replicates of H3K4me3, H3K27me3, H3K27ac, and CTCF markers, respectively, in DFC samples (left panel) and CBC samples (right panel). H3K27me3

signals in chrX are shown in purple. HSB123 and HSB136 are males, HSB126 and HSB187 are females. The average correlation coefficients for the H3K4me3, H3K27ac, H3K27me3, and CTCF groups are 0.98, 0.87, 0.84, and 0.77, respectively. No obvious disparity between male and female samples was observed for H3K4me3, H3K27ac, and CTCF.

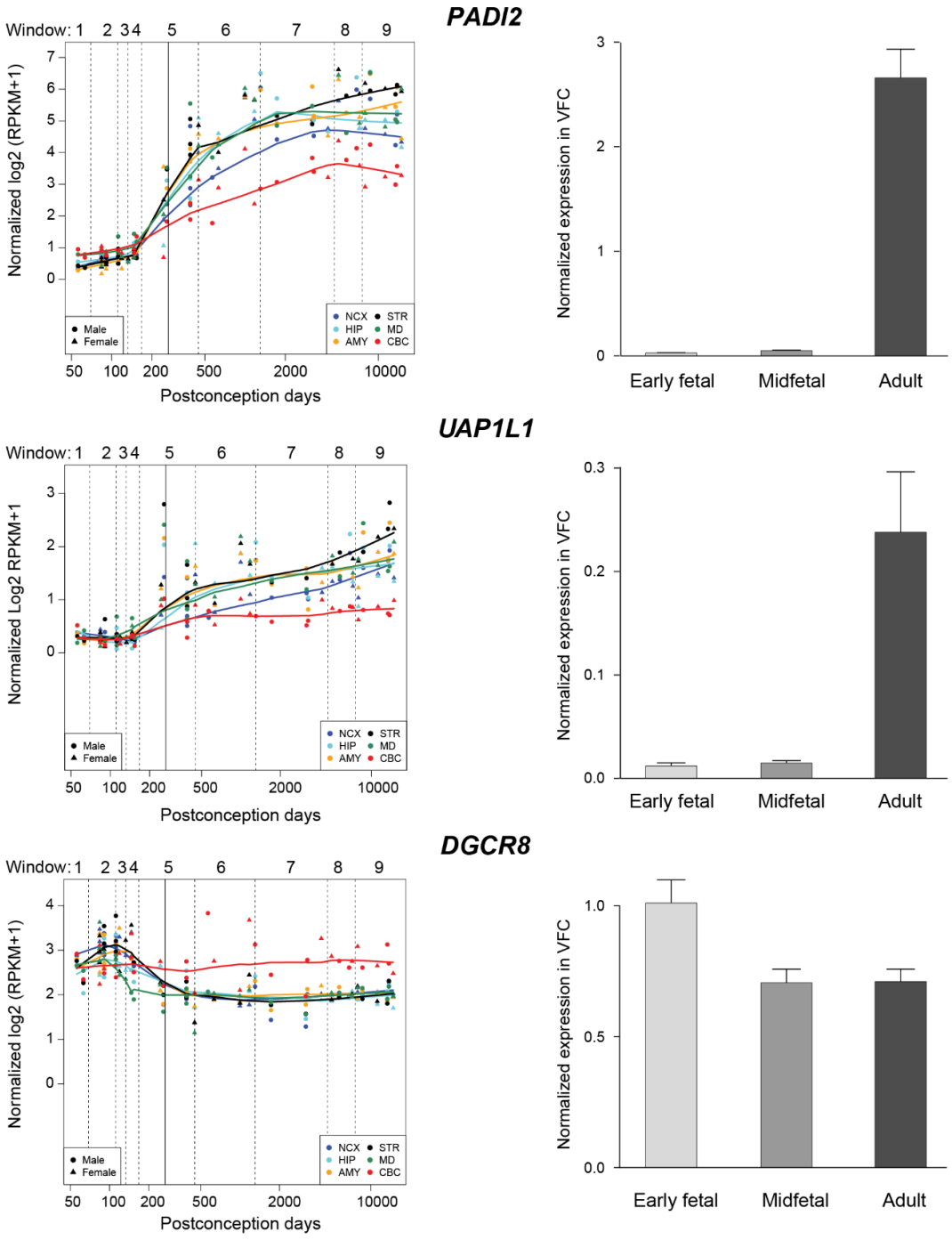


Fig. S9. Validation of gene expression differences.

To confirm the validity of gene expression differences identified using RNA-Seq, we selectively conducted quantitative droplet digital PCR (ddPCR). In so doing, we validated the adult-enriched expression of *PADI2* and *UAP1L1* and the early fetal enriched expression of *DGCR8*.

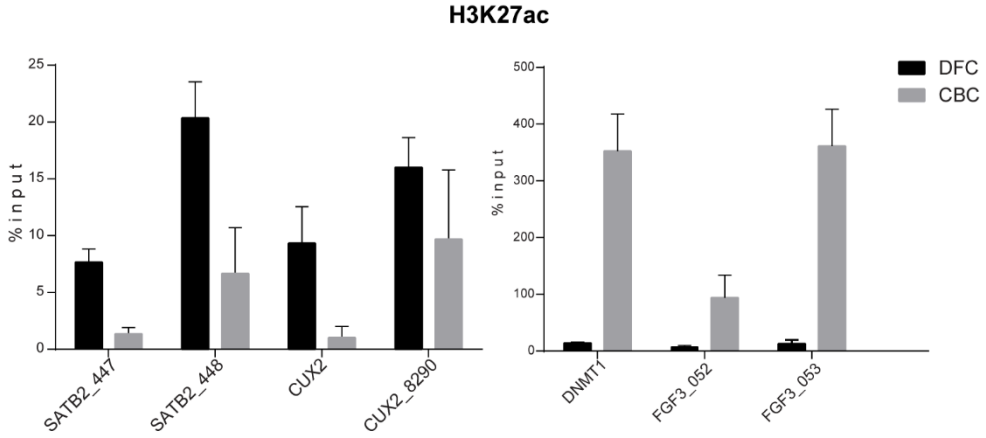


Fig. S10. ChIP-Seq validation

We identified region-specific H3K27ac peaks associated with neurodevelopmental genes including *SATB2*, *CUX2*, *DNMT1*, and *FGF3*. To validate these putative regulatory elements, we conducted quantitative ChIP-PCR on samples from the DFC (black bars) and CBC (grey bars), observed the expected regional specificity.

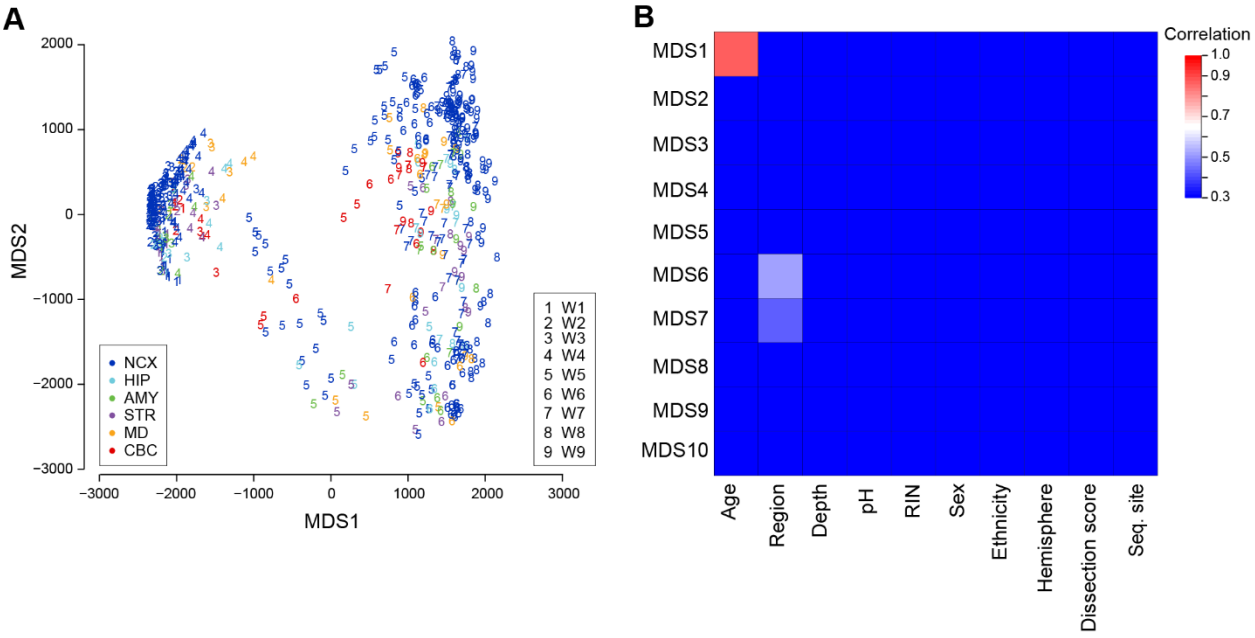


Fig. S11. Multidimensional scaling of global gene expression.

(A) Scatter plot showing the first two multidimensional scaling components from the analysis of gene expression. Different colors encode different brain regions: NCX (blue), HIP (cyan), AMY (green), STR (purple), MD (orange), and CBC (red). The numbers indicate the associated development window for each brain sample, from W1 (1) to W9 (9). The majority of samples clustered into prenatal (W1-W4; left) or postnatal (W6-W9; right) groupings, with samples from W5 generally occupying an intermediate position between these groups. **(B)** Correlation of mRNA-Seq data to potential confounding factors. The top 10 multidimensional scaling components were correlated to 10 potential confounding or other factors: age, region, sequencing depth, pH, RIN, sex, ethnicity, brain hemisphere, sample dissection score, and sequencing processing site (Seq. site), with only age and region correlating strongly with specific dimensions.

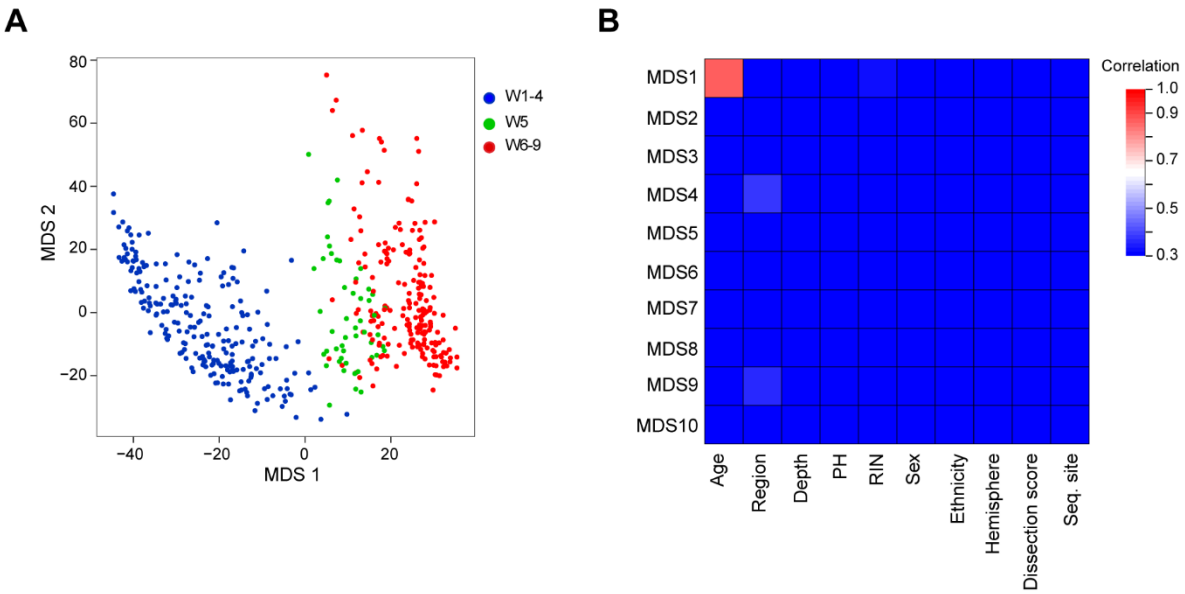


Fig. S12. Multidimensional scaling of exon percent spliced in index.

(A) Scatter plot showing the first two multidimensional scaling components from the analysis of the ratio between reads including or excluding exons, also known as percent spliced in index (PSI). Different colors encode different developmental phases: prenatal (W1-W4; blue), late fetal and early infancy (W5; gray), and postnatal (W6-W9; red). **(B)** Correlation of multidimensional scaling components to potentially confounding factors. The top 10 multidimensional scaling components were correlated to 10 potential confounding or other factors: age, region, sequencing depth, pH, RIN, sex, ethnicity, brain hemisphere, sample dissection score, and sequencing processing site (Seq. site). Here again, only age and region correlated strongly with specific components.

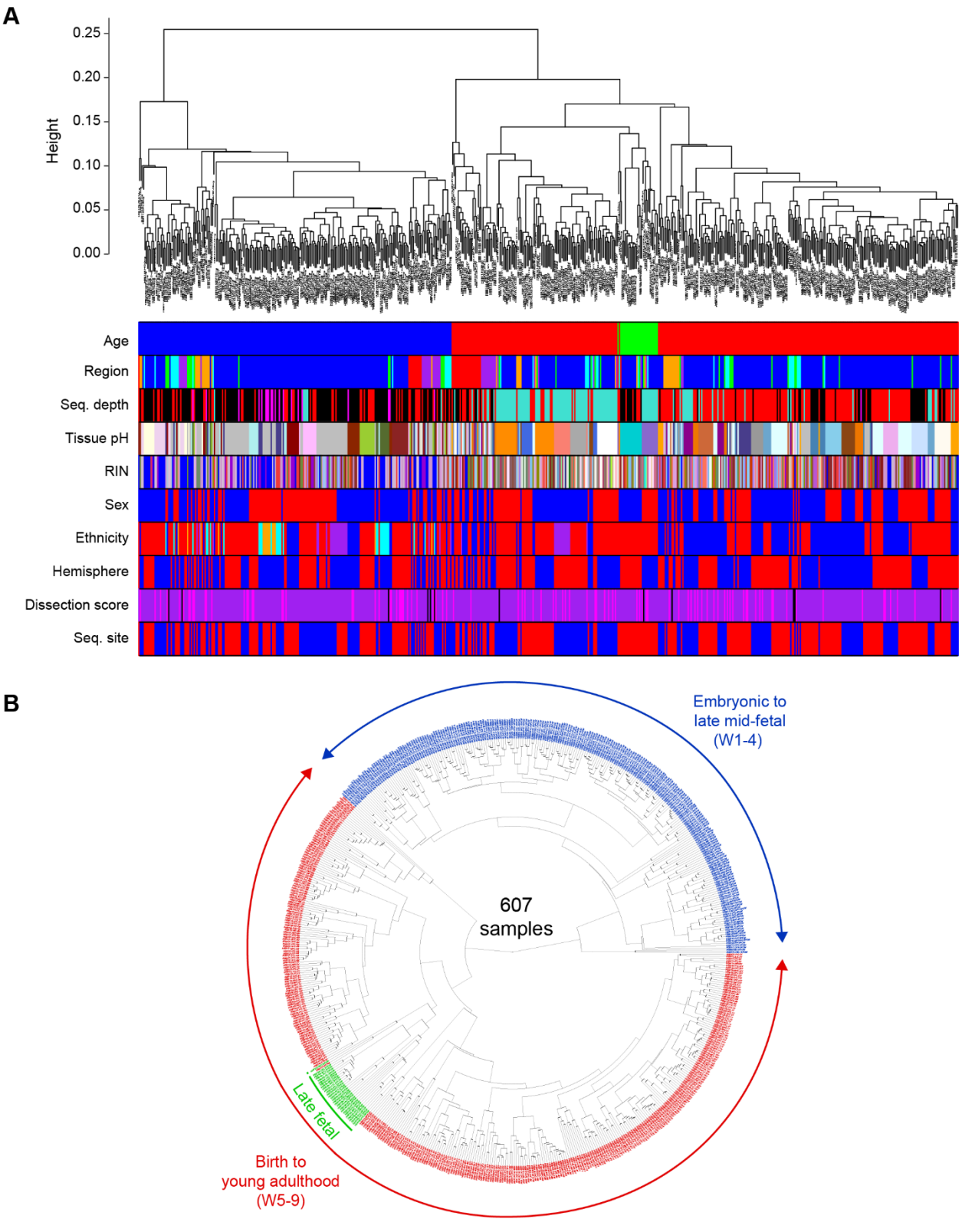

A
Height
0.25
0.20
0.15
0.10
0.05
0.00
Age
Region
Seq. depth
Tissue pH
RIN
Sex
Ethnicity
Hemisphere
Dissection score
Seq. site
B
Embryonic to
late mid-fetal
(W1-4)
607
samples
Late fetal
Birth to
young adulthood
(W5-9)

Fig. S13. Hierarchical clustering analysis of the human brain mRNA transcriptome.

(A) Dendrogram of hierarchical clustering for 607 mRNA-Seq samples. The colors underneath label potential confounding factors including age (W1-W4, blue; late fetal samples, green; infancy-W9, red), region (NCX, blue; HIP, cyan; AMY, green; STR, purple; MD, orange; CBC, red), sequencing depth (low to high shown in purple to red scale), tissue pH (low to high shown in blue to brown scale), RIN (low to high shown in blue to red scale), sex (blue, male; red, female), ethnicity (African, blue; African/European, cyan; Asian, orange; European, red; Mexican, purple), hemisphere (left, red; right, blue), dissection score (low to high shown in black to purple scale), and sequencing processing site (Yale, red; USC, blue). The samples clustered predominately by age and not by depth, pH, RIN, sex, ethnicity, hemisphere, dissection score, or sequencing processing site. Regional clustering is observed, particularly for NCX and CBC. **(B)** Circular tree showing unsupervised hierarchical clustering for 607 mRNA-Seq samples. Samples from prenatal ages (W1 - W4) are generally segregated from those samples from perinatal (W5) and postnatal ages (W6 – W9). Prenatal samples in W1 through W4 are in blue, postnatal samples are in red, and late fetal samples (35 and 37 PCW) are in green.

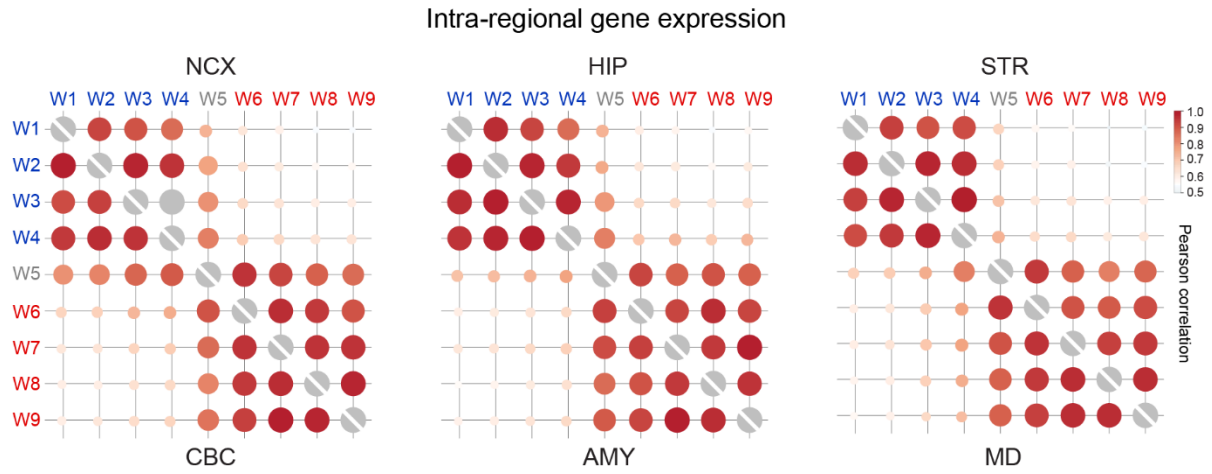


Fig. S14. Intraregional gene expression correlation across all developmental windows.

For each of the 6 major brain regions we assayed, we used the expression values of the entire gene set to calculate Pearson correlations for any pair of development windows. Heatmap bubbles depict the strength of the correlation coefficient, with larger bubbles indicating a greater correlation. We found general conservation across exclusively prenatal (W1-W4) and exclusively postnatal (W6-W9) windows but reduced correlation across W5 (i.e., between prenatal and postnatal windows). Grey cells with diagonal lines divide the matrix into upper and lower triangles corresponding to different brain regions, as listed.

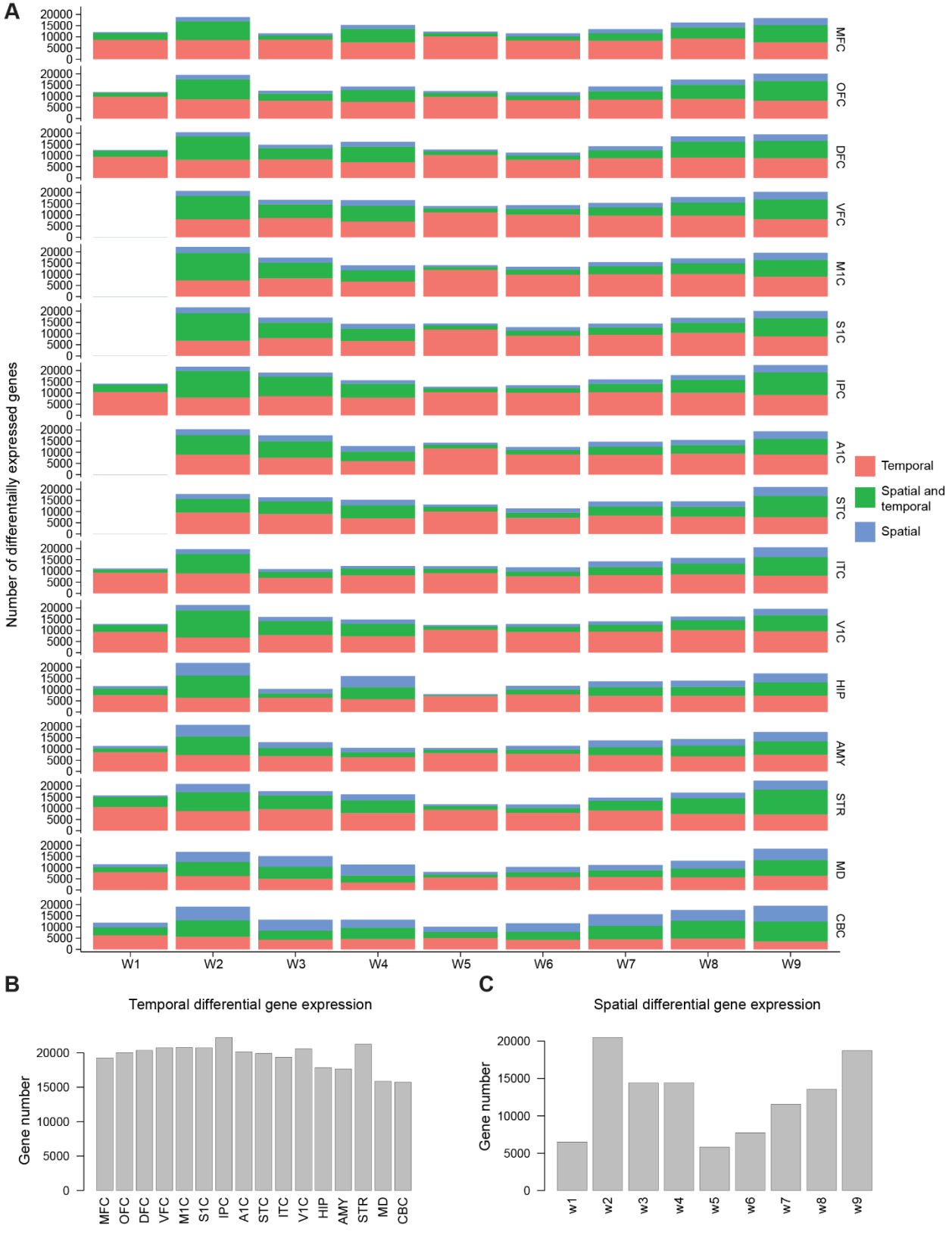

A
Number of differentially expressed genes
MFC
OFC
DFC
VFC
M1C
S1C
IPC
A1C
STC
ITC
V1C
HIP
AMY
STR
MD
CBC
W1
W2
W3
W4
W5
W6
W7
W8
W9
Temporal
Spatial and temporal
Spatial
B
Temporal differential gene expression
Gene number
C
Spatial differential gene expression
Gene number

Fig. S15. Temporal and spatial differential expression.

(**A**) Stacked bar plots representing the number of differentially expressed genes for each brain region in each developmental window. The differentially expressed genes were categorized into temporal only (orange), spatio-temporal (green), and spatial only (blue). The definition of differentially expressed genes in each category is detailed in (*22*). (**B**) Bar plots of the number of temporally differentially expressed genes per brain region. (**C)** Bar plots of the number of spatially differentially expressed genes per developmental window.

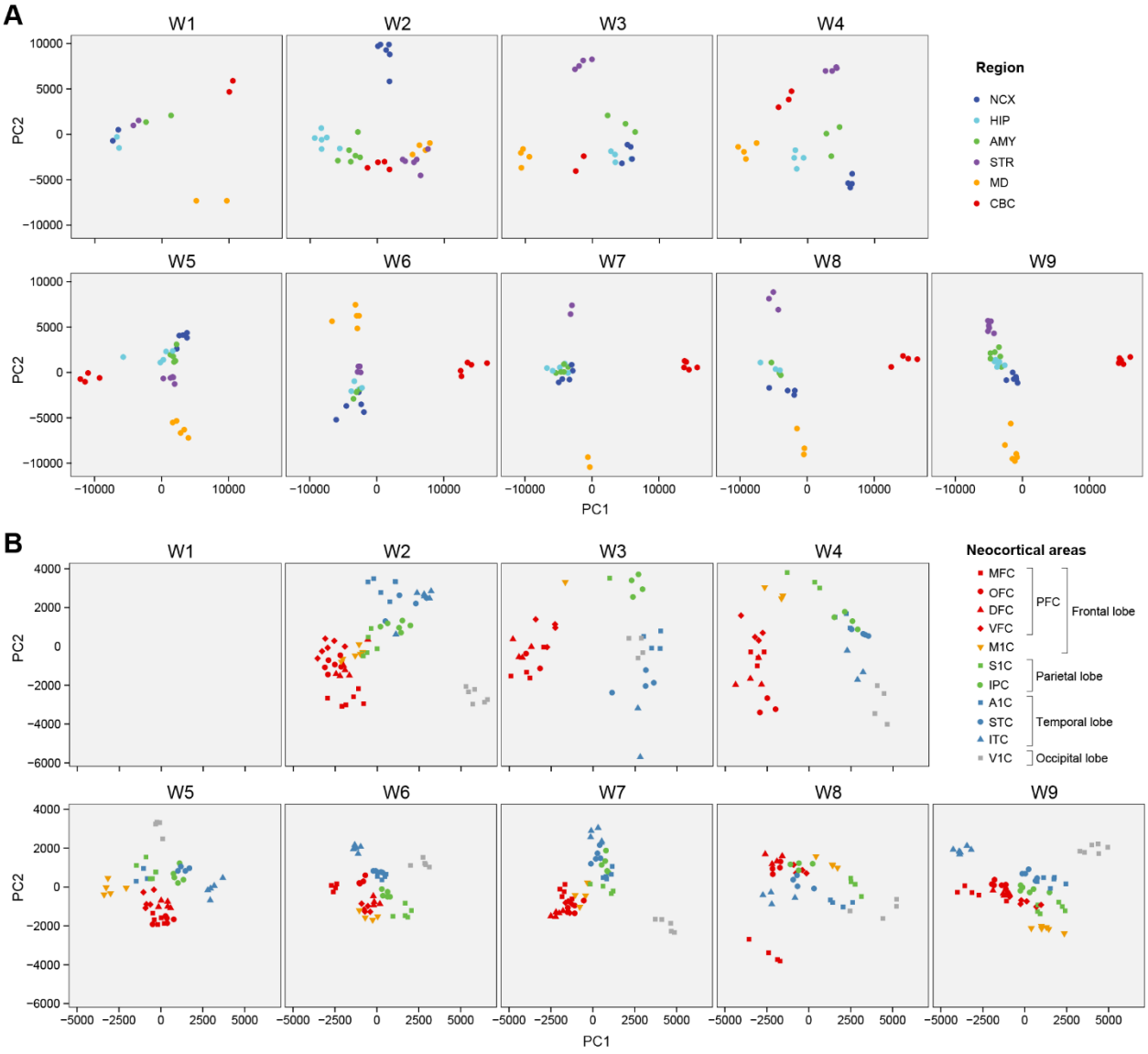


Fig. S16. Principal component analysis adjusted for inter-individual variation (AC-PCA) for the mRNA-Seq data set.

(**A**) AC-PCA plots of all developmental windows using gene expression values from 6 brain regions. (**B**) AC-PCA plots of all developmental windows using gene expression values from 11 neocortex areas. Different brain regions and neocortex areas are differently colored or shaped as

indicated. For both all brain regions and the neocortex, the distance between brain regions or neocortical areas in the first two principal components was reduced during W5 as compared to other prenatal (W1-W4) or late postnatal (W9) ages.

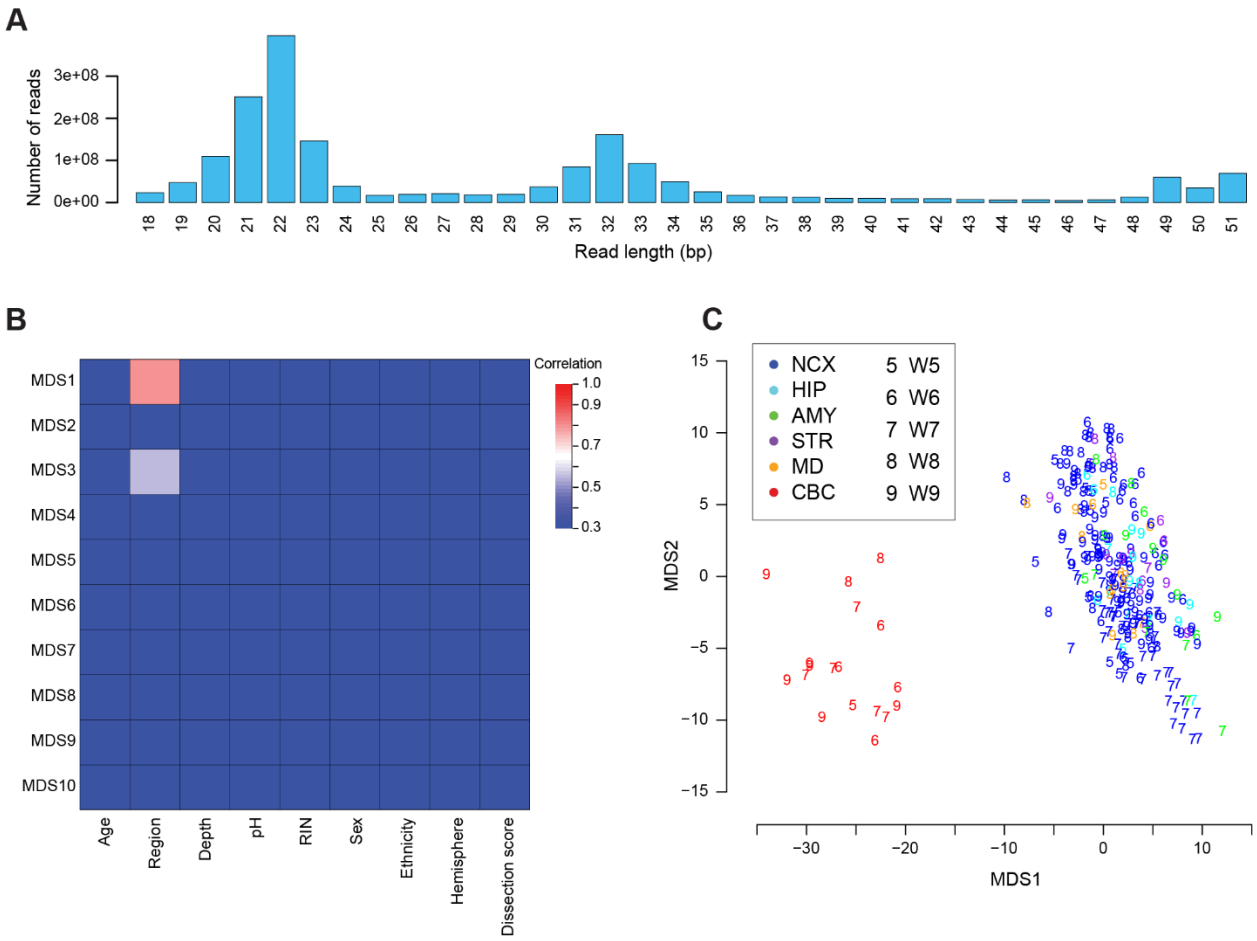


Fig. S17. Length enrichment and multidimensional scaling for miRNAs.

(A) Read length distribution of all samples in the small RNA-Seq dataset. **(B)** Correlation of multidimensional scaling components to potentially confounding factors. The top 10 multidimensional scaling components were correlated to 9 potential confounding or other factors: age, region, sequencing depth, pH, RIN, sex, ethnicity, brain hemisphere, and sample dissection score. Only brain region correlated strongly with individual components. (**C**) Scatter plot showing the first two multidimensional scaling components from the analysis of miRNA expression. Different colors encode different brain regions: NCX (blue), HIP (cyan), AMY (green), STR

(purple), MD (orange), and CBC (red). The numbers indicate the associated development window for each brain sample.

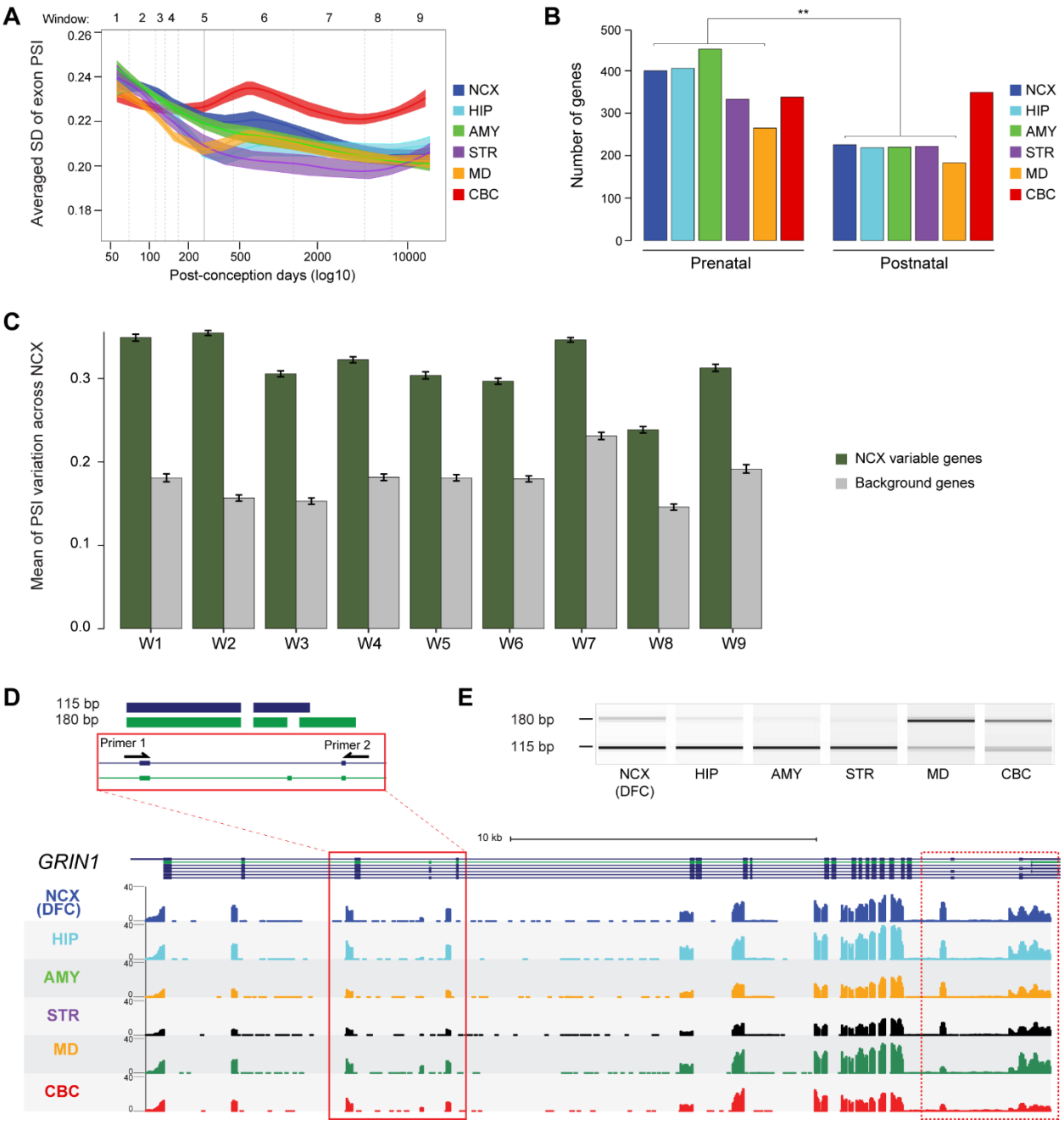


Fig. S18. Alternative splicing analysis.

(**A**) Spatio-temporal dynamics of alternative splicing in human brain transcriptome, with age represented as a log10 transformation of post-conception days along the X-axis. The solid vertical

line indicates birth. For any given gene, the standard deviation (SD) of exon percent spliced in index (PSI) accounts for its alternative splicing events, and the transcriptome-wide summarization of alternative splicing events is represented on the Y-axis by the averaged standard deviation of exon PSI of all genes in a given region. The shaded area for each curve corresponds to 50% standard error of the mean. Brain regions are differentiated by colors as indicated. (**B**) The number of genes exhibiting alternative splicing profiles is not constant across prenatal and postnatal development, with more genes alternatively spliced during prenatal ages than postnatal ages (P = 0.007). (**C**) Bar plots representing the mean of exon PSI variation across 11 neocortical areas for both genes that are highly variable in the neocortex as well as control sets of background genes. Error bar is mean±SEM. (**D**) Read coverage of *GRIN1* expression in 6 representative brain regions (DFC, HIP, AMY, STR, MD, and CBC) of HSB126, an adult brain specimen. Alternative splicing was observed in the inclusion or exclusion of exon 4, generating a 180 bp or 115 bp PCR product in the region shown within the red box, respectively. (**E**) Exon specific RT-PCR was carried out using primers 1 and 2 depicted in (**D**) in order to confirm the presence and differential expression of the 180 bp and 115 bp fragments.

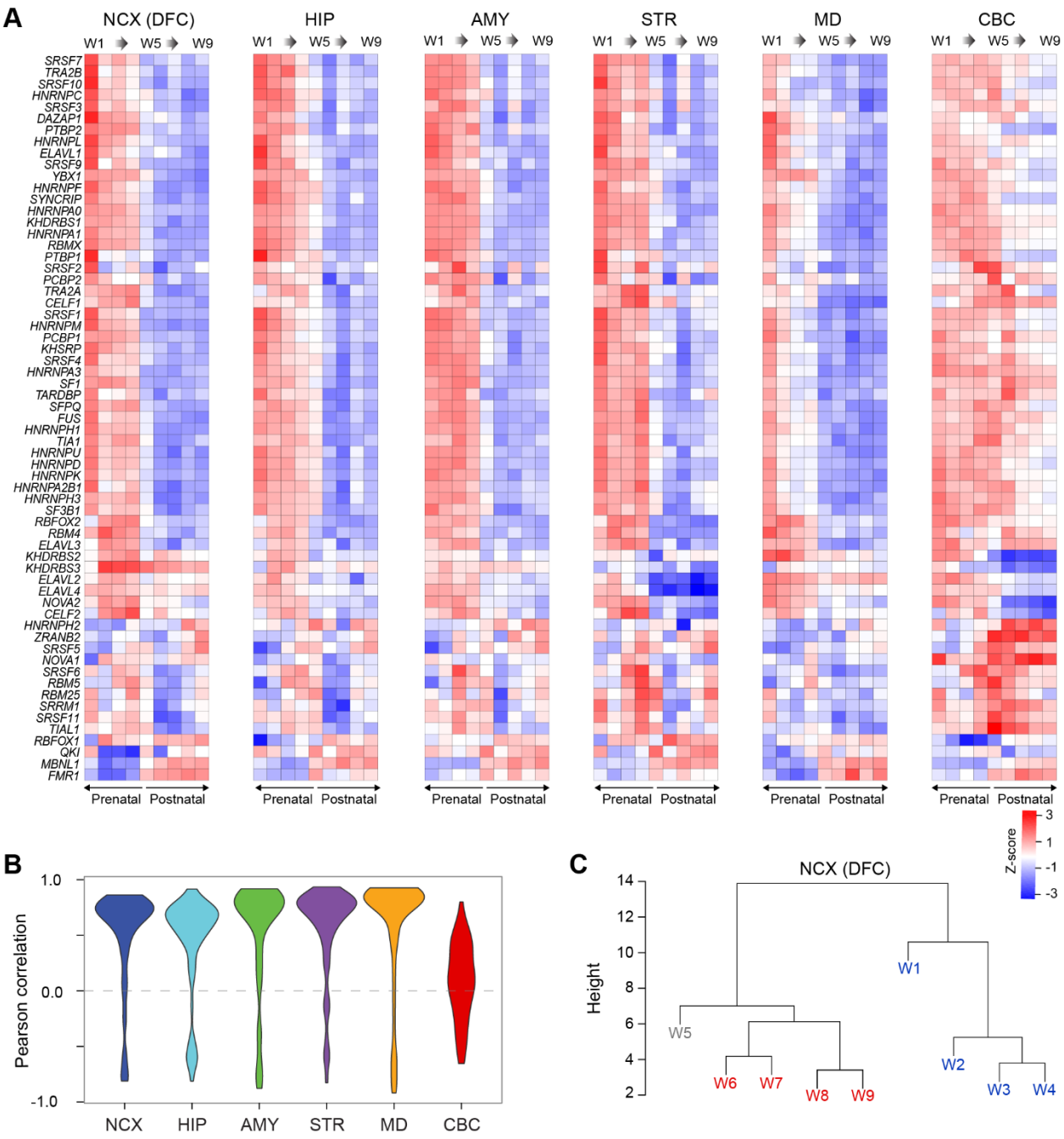


Fig. S19. Temporal expression of the RNA-binding proteins.

(**A**) Heat maps representing temporal dynamics of RNA binding protein (RBP) gene expression in 6 brain regions across development. For each brain region, samples were ordered in accordance

with the age, spanning from embryonic age (W1) through young adulthood (W9). For each gene (taken from (*24*)), its expression is converted to a Z-score to exclude effects of genes with extreme expression. NCX is represented by DFC. (**B**) Violin plots of the Pearson correlation between RBP gene expression and the extent of gene alternative splicing in 6 brain regions. (**C**) Hierarchical clustering of RBP gene expression. The dendrogram shows two clusters, with W5 clustered with postnatal developmental windows (W6-9).

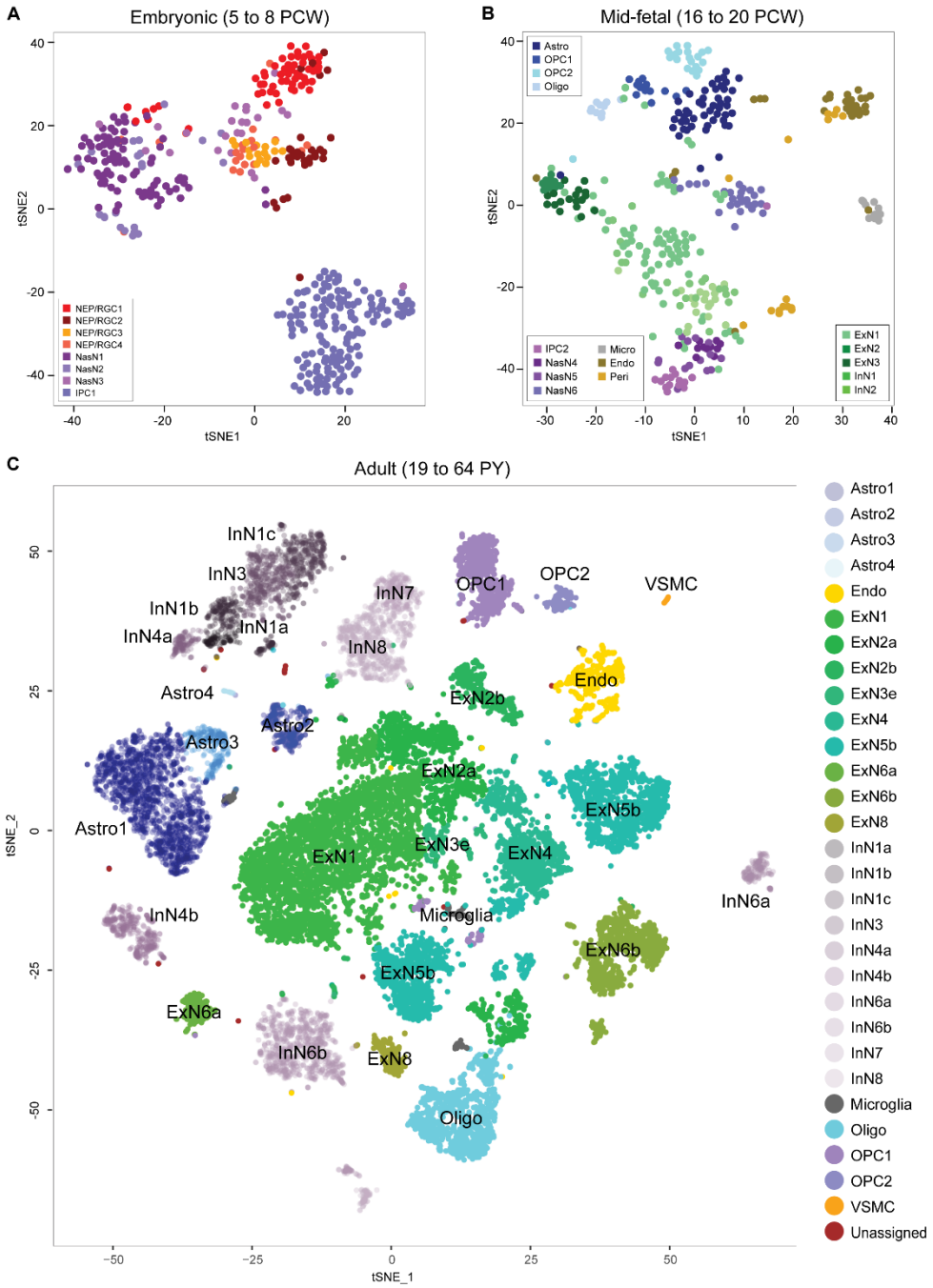


Fig. S20. t-SNE analyses of prenatal single cell RNA-Seq and adult single nuclei RNA-Seq.

(**A-B**) t-SNE plots showing the single cell RNA-Seq data set from embryonic (5-8 PCW; A) and fetal (16-20 PCW; B) donor brains. Different colors encode different cell types. (**C**) t-SNE plot showing single nuclei RNA-Seq data from adult human donor brains. Here again, different colors encode different cell types, with cell type nomenclature based on correlations to an independent dataset of adult single nuclei RNA-Seq (*27*). NEP/RGC: Neural epithelial progenitor/radial glial cells, NasN: Nascent neurons, IPC: intermediate progenitor cells, Astro: cells in the astroglial lineage, OPC: oligodendrocyte progenitor cells, Oligo: oligodendrocytes, ExN: excitatory neurons, InN: interneurons, Micro: microglia, Endo: endothelial cells, Peri: pericytes, VSMC: vascular smooth muscle cells. f before a cell type name indicates that these cell types were identified in fetal samples.

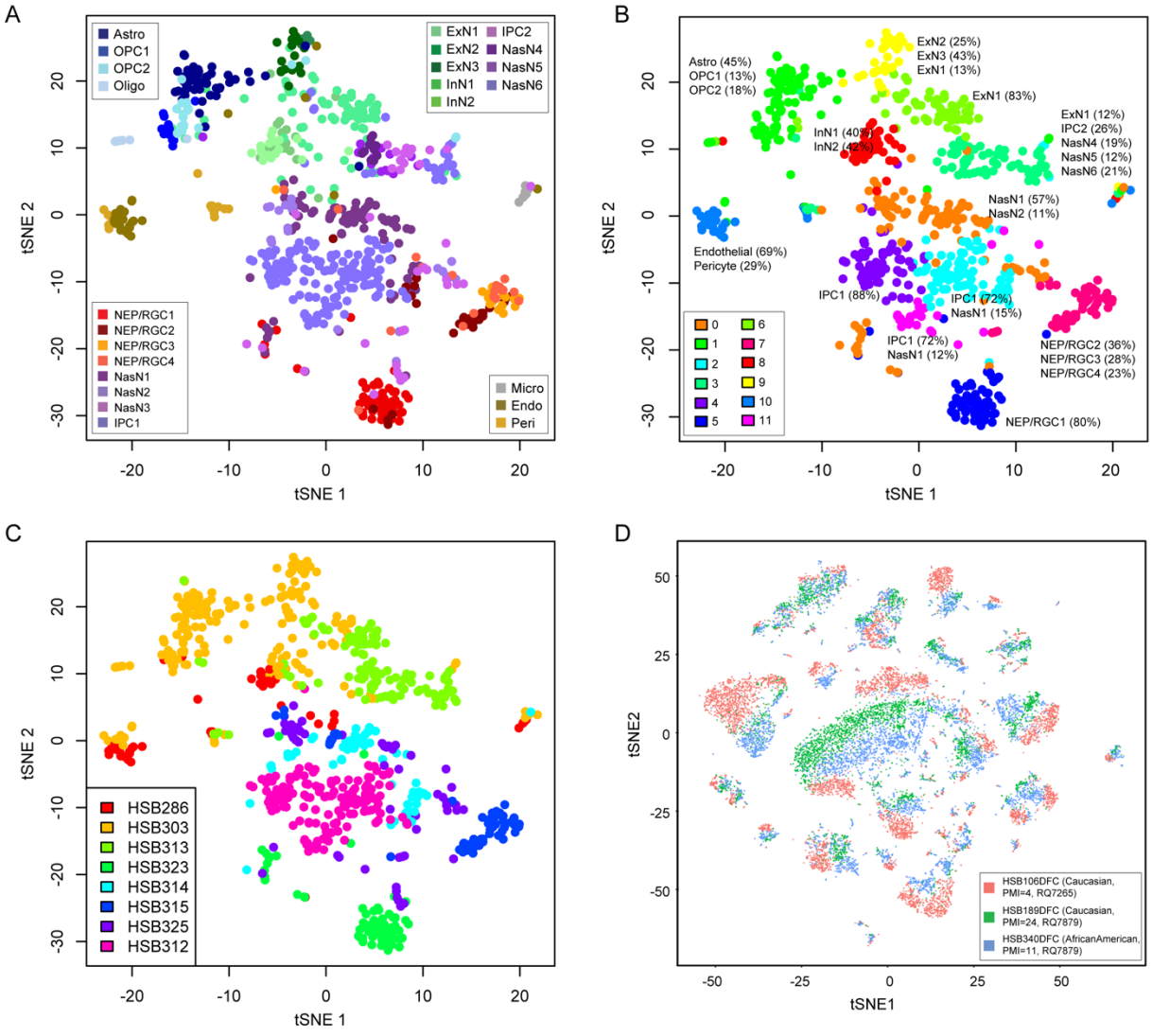


Fig. S21. tSNE representation of alternative clustering methods and brain effect

t-SNE plots showing the single cell RNA-Seq data set from prenatal (5-20 PCW; A-C) and postnatal (D) donor brains. A) tSNE representation based on Seurat data processing colored by the clusters identified using the (*27*) methodology. B) tSNE plot colored by the clusters obtained by Seurat default clustering method at resolution 1.5. Numbers indicate the proportion of cells from clusters in A for any given cluster in B contributing at least 10% in the later. C-D) tSNE plots

colored by donor brain. NEP/RGC: Neural epithelial progenitor/radial glial cells, NasN: Nascent neurons, IPC: intermediate progenitor cells, OPC: oligodendrocyte progenitor cells, Oligo: oligodendrocytes, ExN: excitatory neurons, InN: interneurons, Astro: astroglial lineage.

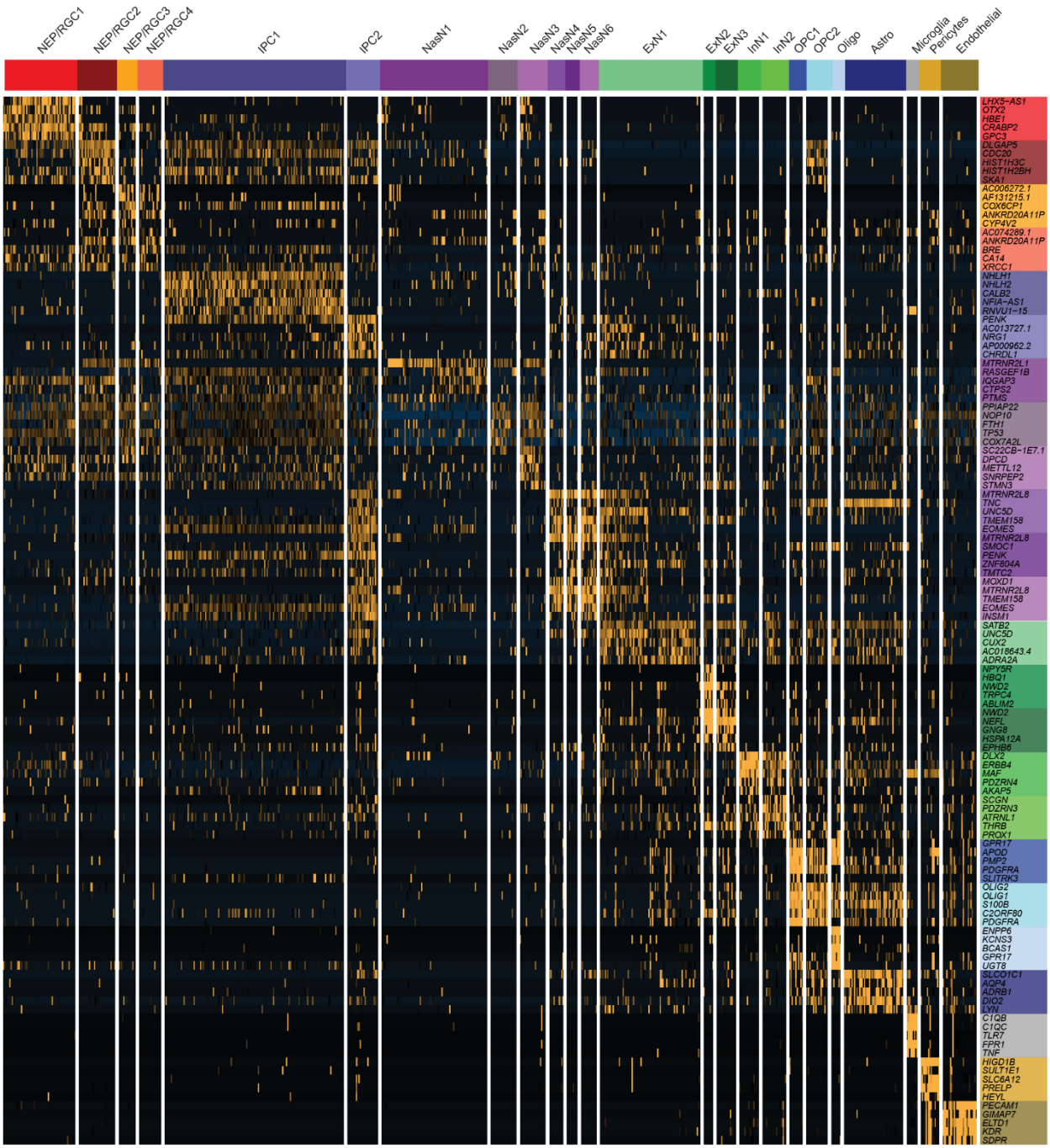


Fig. S22. Heat map of known and novel cell type specific marker genes.

Heat map showing the expression of cell type specific molecular markers. The five genes exhibiting the highest specificity score for the prenatal scRNA-Seq database are shown for each

cell type. Colors represent different cell types and are the same as in Fig. S18. NEP/RGC: Neural epithelial progenitor/radial glial cells, NasN: Nascent neurons, IPC: intermediate progenitor cells, OPC: oligodendrocyte progenitor cells, Oligo: oligodendrocytes, ExN: excitatory neurons, InN: interneurons, Astro: astroglial lineage.

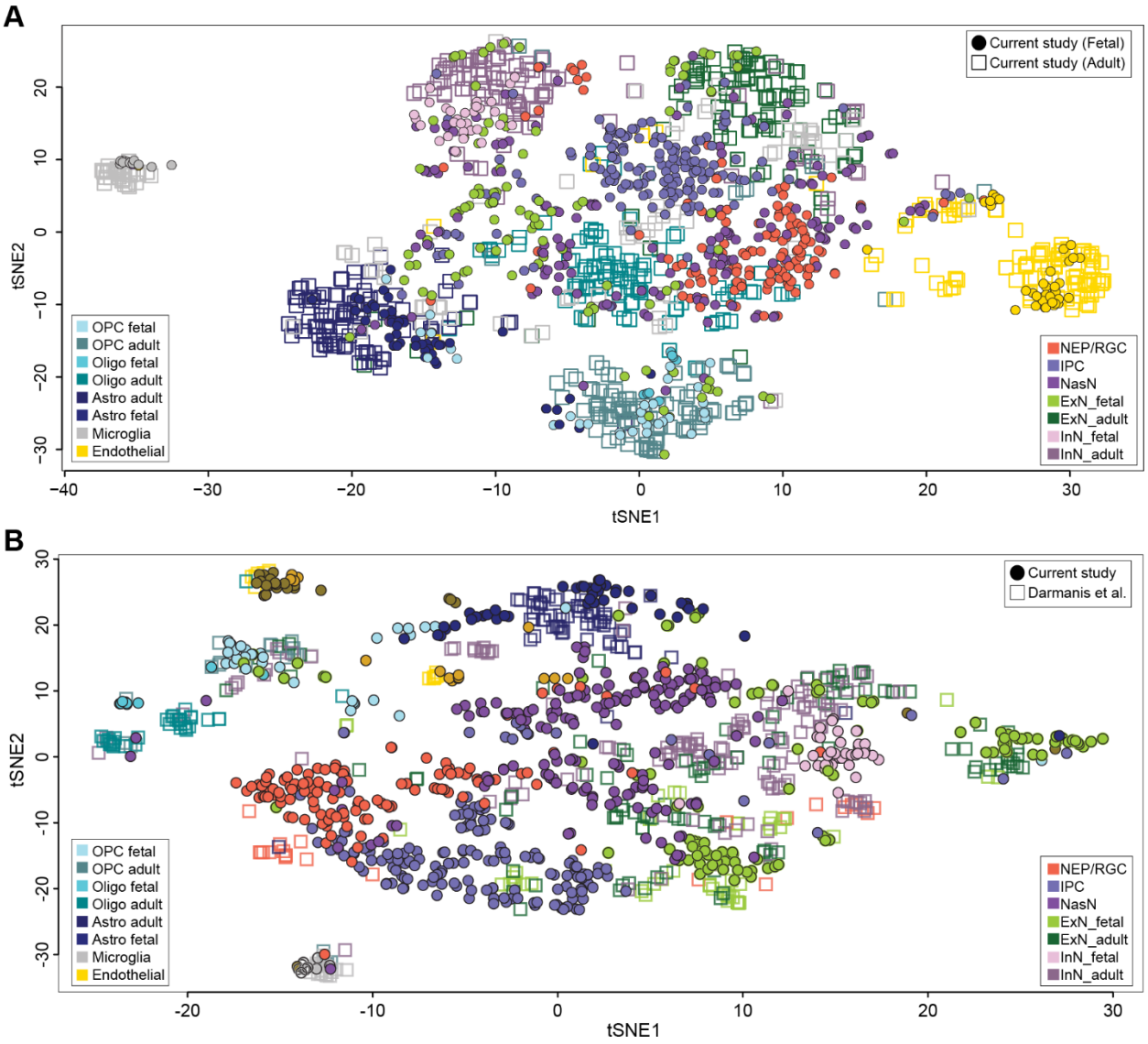


Fig. S23. Alignment of prenatal and adult single cell RNA-Seq data sets.

(**A**) t-SNE plot showing scRNA-Seq data of single cells dissociated from prenatal brains of age 5-20 PCW (solid circle) and adult brains (hollow box) from the current study. Similar cell types (i.e., fetal astrocytes and adult astrocytes) tend to cluster in similar locations. Conversely, cell types without obvious homologues between fetal and adult development (i.e., NEP/RGC cells and IPCs) generally cluster in locations with few or no adult cells. (**B**) tSNE plot showing single cell RNA-

Seq data from this study (closed circle) aligned with fetal and adult single cell data from Darmanis et al. (open square) (*9*). Confirming the validity of our dataset and cell type analyses, here again we see broad similarities in the positioning of similar cell types across datasets. For both panels, different colors encode different cell types. NEP/RGC: Neural epithelial progenitor/radial glial cells, NasN: Nascent neurons, IPC: intermediate progenitor cells, OPC: oligodendrocyte progenitor cells, Oligo: oligodendrocytes, ExN: excitatory neurons, InN: interneurons, Astro: astroglial lineage, with similar cell types generally clustering near each other.

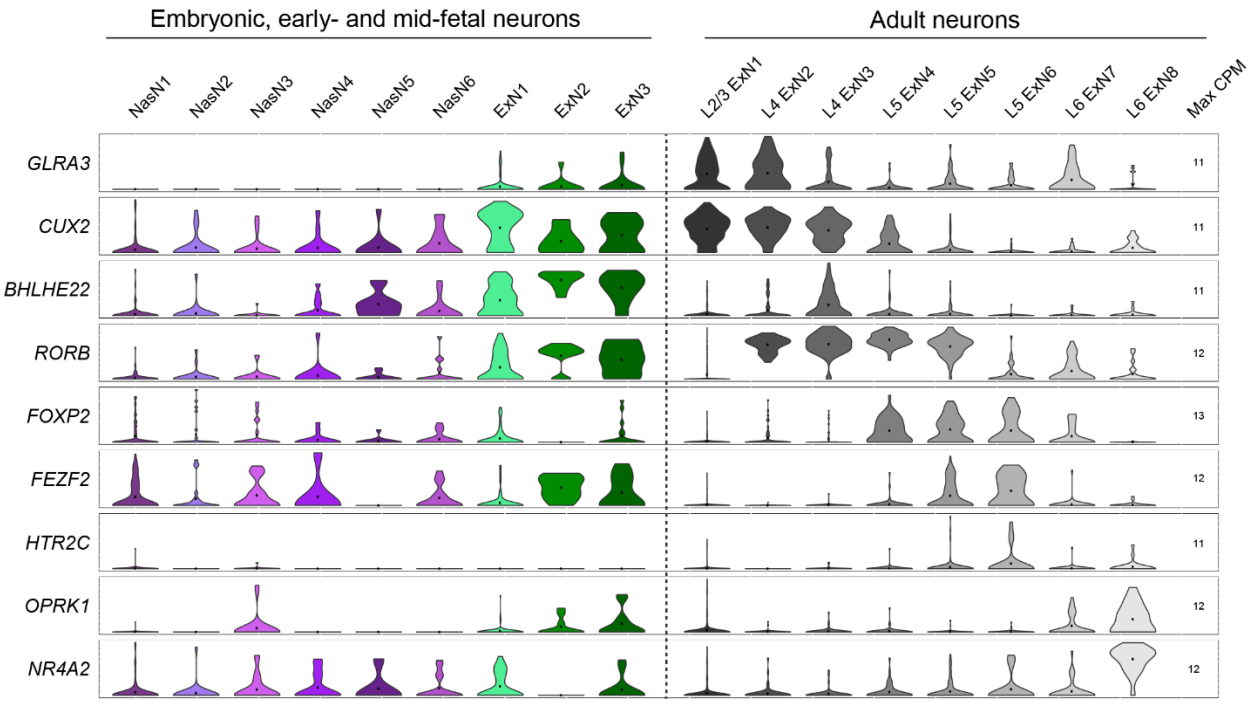


Fig. S24. Expression distribution of excitatory neuron subtype marker genes.

Violin plots showing the expression distribution of excitatory neuron subtype marker genes. Each column corresponds to one cell cluster identified in our scRNA-Seq data set or retrieved from Lake et al. (*27*). The first nine columns contain data from embryonic (columns 1 through 3) and fetal (columns 4 through 9) cells in the current dataset (Li et al.). Remaining columns are from Lake et al. (*27*). NasN: Nascent neurons, ExN: excitatory neurons. Numbers in the rightmost column indicate the maximum log2 (CPM + 1) values for each marker gene.

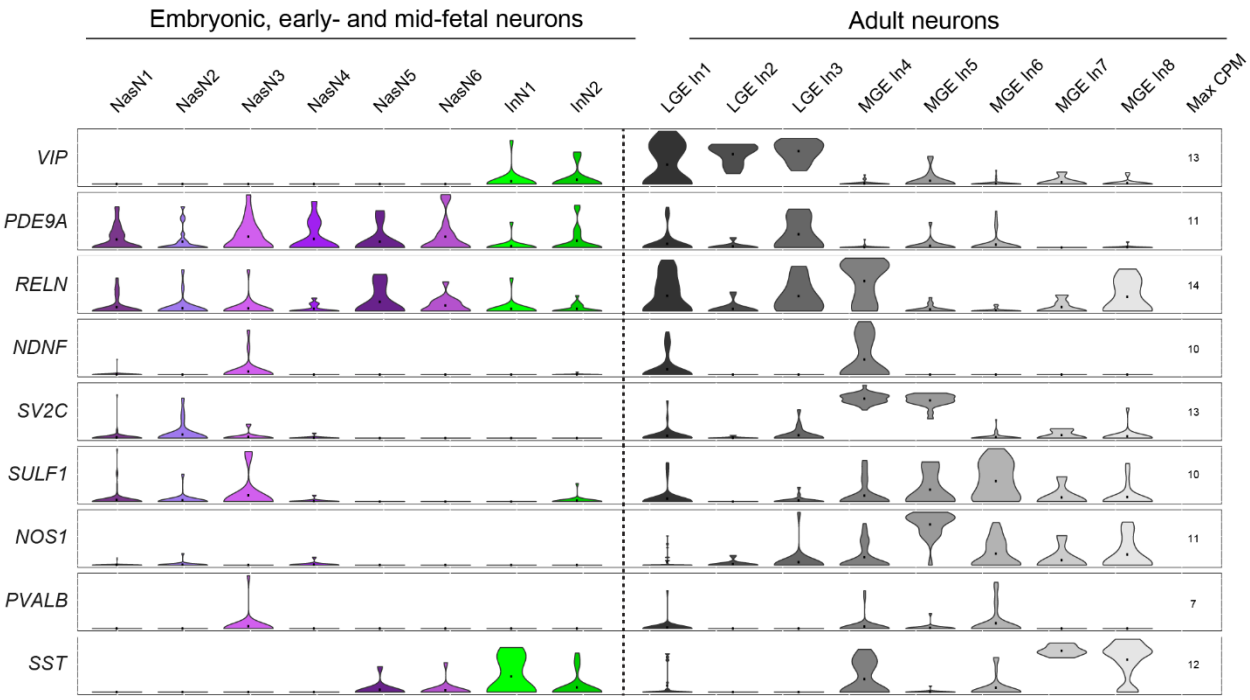


Fig. S25. Expression distribution of interneuron subtype marker genes.

Violin plots showing the expression distribution of interneuron subtype marker genes. Each column corresponds to one cell cluster identified in our scRNA-Seq data set or retrieved from Lake et al. (*27*). The first eight columns contain data from embryonic (columns 1 through 3) and fetal (columns 4 through 8) cells in the current dataset. Remaining columns are from Lake et al. (*27*). NasN: Nascent neurons, InN: interneurons. Numbers in the rightmost column indicates the maximum log2 (CPM + 1) values for each marker gene.

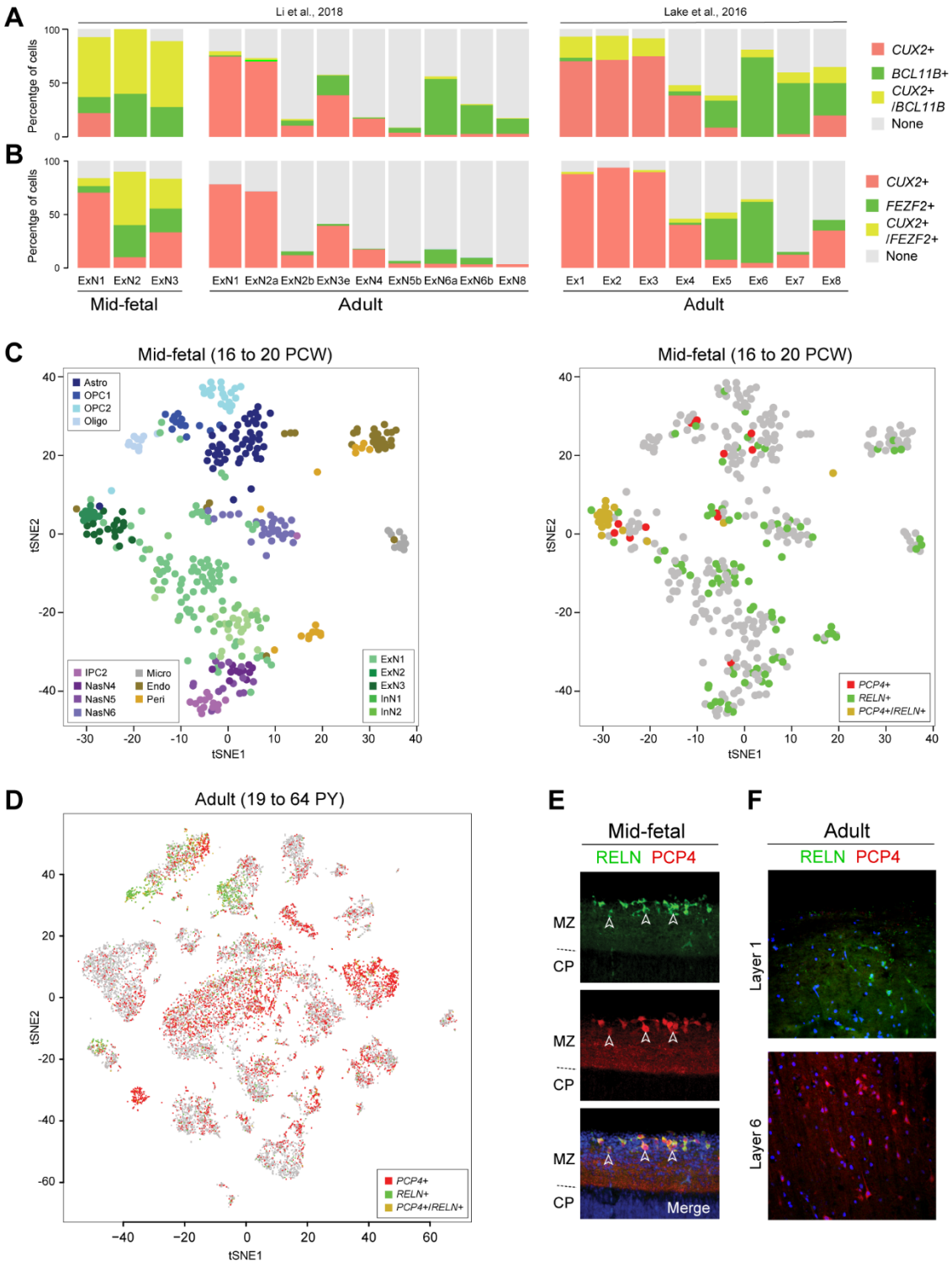


Fig. S26. Co-expression of molecular markers including RELN and PCP4

(A) Comparison of *CUX2* expressing (red), *BCL11B* expressing (green), and *CUX2*/*BCL11B* co-expressing (yellow) mid-fetal and adult excitatory neurons (ExN), as identified from this study (Li et al.) or Lake et al. (*27*). (**B**) Comparison of *CUX2* expressing (red), *FEZF2* expressing (green), and *CUX2*/*FEZF2* co-expressing (yellow) mid-fetal and adult excitatory neurons (ExN), as identified from this study (Li et al.) or Lake et al. (*27*). (**C**) tSNE plots of mid-fetal scRNA-Seq data showing various cell types detected in the fetal human brain (left) and the expression or co-expression of *RELN* expressing (green), *PCP4* expressing (red), and *RELN/PCP4* co-expressing (yellow) cells (left). (**D**) tSNE plot for adult snRNA-Seq data in our study, again showing *PCP4* expressing, *RELN* expressing, and *PCP4*/*RELN* co-expressing cells. (**E**) Immunohistochemical staining confirms the co-expression of *RELN* and *PCP4* by a subset of neurons in the fetal human brain. (**F**) co-expression of *RELN* and *PCP4* is not readily detectable using immunohistochemistry in the adult human cortical plate.

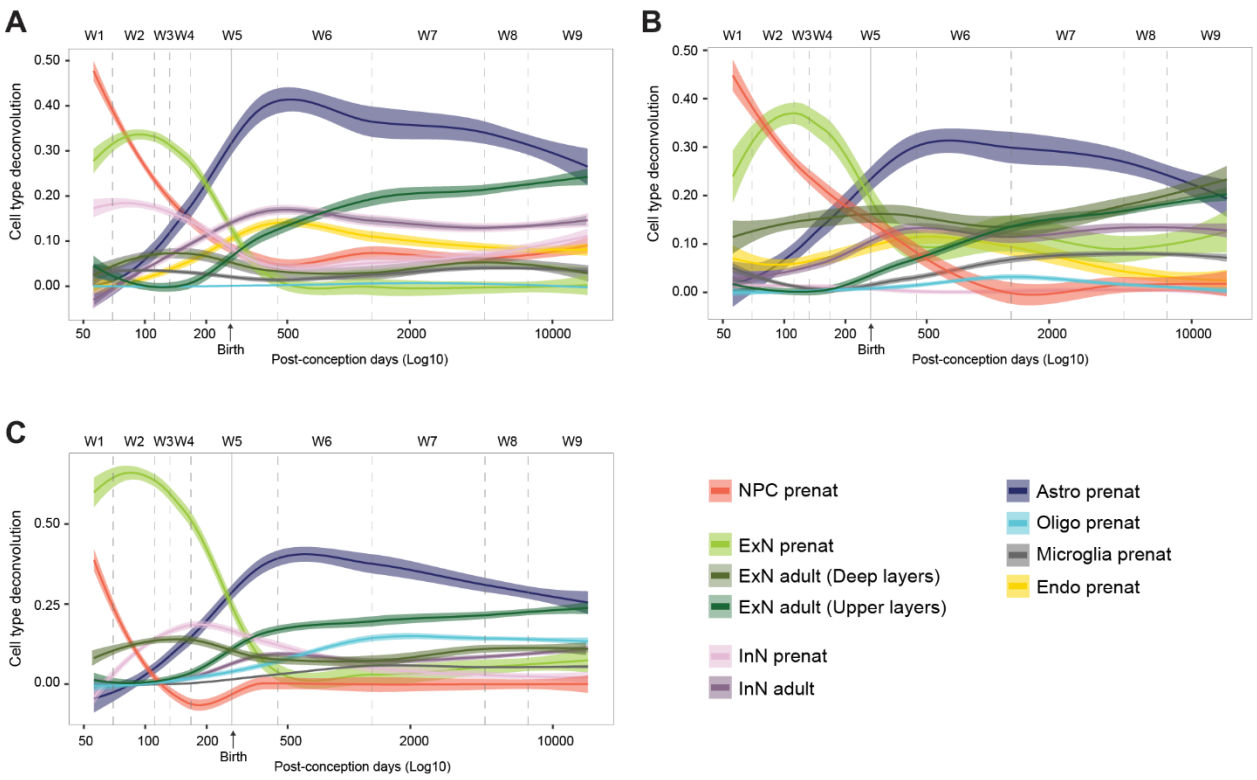


Fig. S27: Deconvolution of neuronal layer maturation in human neocortex.

Following the discrimination of excitatory neuron populations derived from an independent adult single cell dataset (*27*) into deep-layer (i.e., L5 and L6) and upper-layer (i.e., L1, 2/3, 4) subsets, we independently repeated the deconvolution analysis using layer-specific molecular markers and three fetal datasets: (**A**) scRNA-Seq generated in this study, (**B**) Nowakowski et al. (*12*) , and (**C**) Zhong et al. (*26*). The smoothed curves with shaded areas (50% confidence interval) represent the developmental dynamics of cell type composition, including layer-specific excitatory neurons, in the human brain neocortex. Age is represented as log10 transformed post-conception days along the x-axis, with the solid vertical line indicating birth and dashed lines distinguishing developmental windows W1 through W9. NPC: neural progenitor cells, ExN: excitatory neurons, InN: interneurons, Astro: astroglial lineage, oligo: oligodendrocytes, Endo: endothelial.

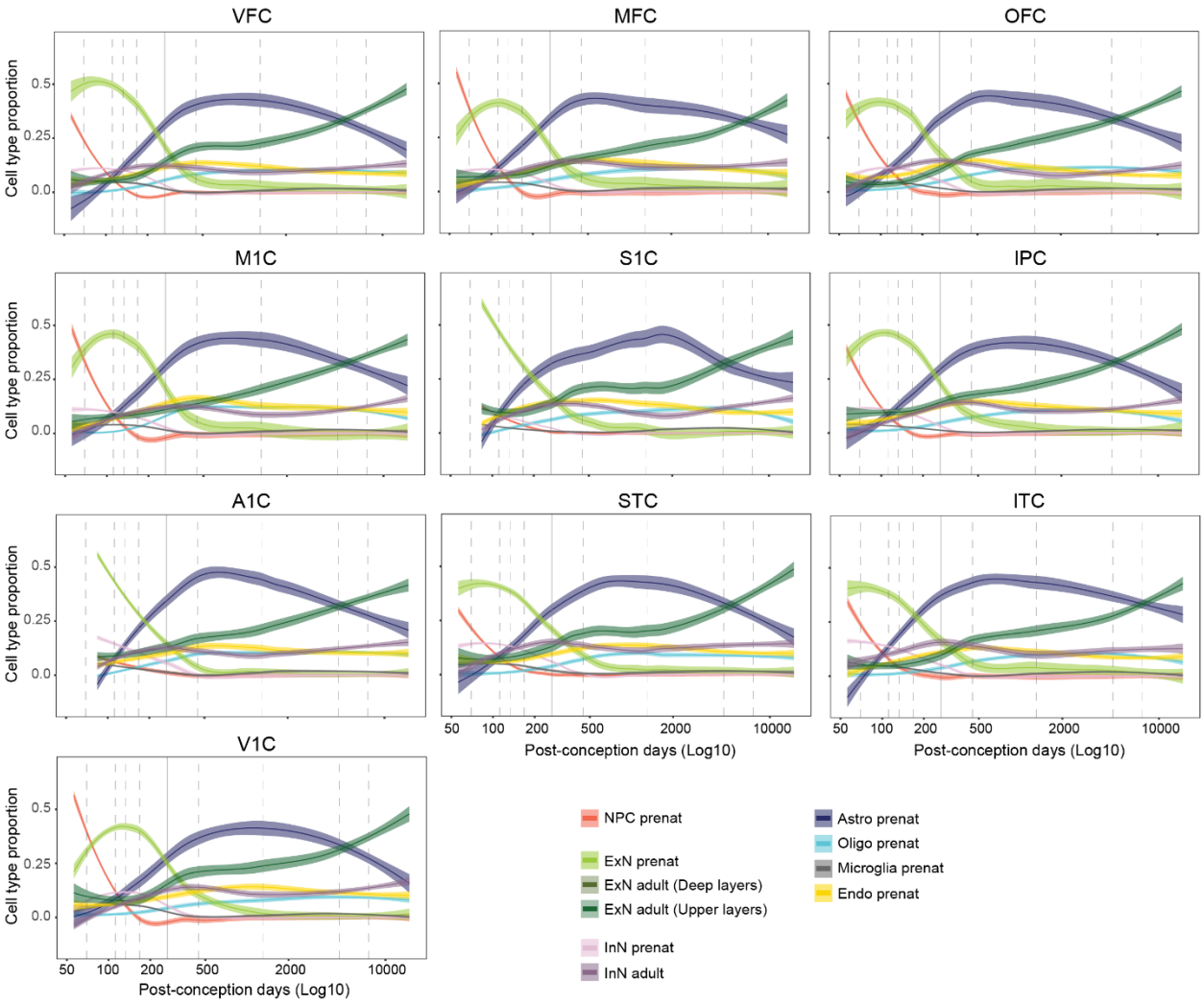


Fig. S28. Deconvolution of neural and non-neural cell content in neocortex.

Transcriptomic signatures from single cell RNA-Seq were used to deconvolve bulk tissue RNA-Seq data by estimating the cell type proportion of key neural and non-neural cell types. Each plot corresponds to one neocortical area. Smoothed curves with shaded areas (50% confidence interval) represent the cell type composition of human neocortical area across development. Age is represented as log10 transformed post-conception days along the x-axis, with the solid vertical line indicating birth and dashed lines distinguishing developmental windows W1 through W9. The

solid vertical line indicates the time of birth. Different colors mean different cell types. NPC: Neural progenitor cells, ExN: excitatory neurons, InN: interneurons, Astro: astroglial lineage, Oligo: oligodendrocytes, Endo: endothelial cells.

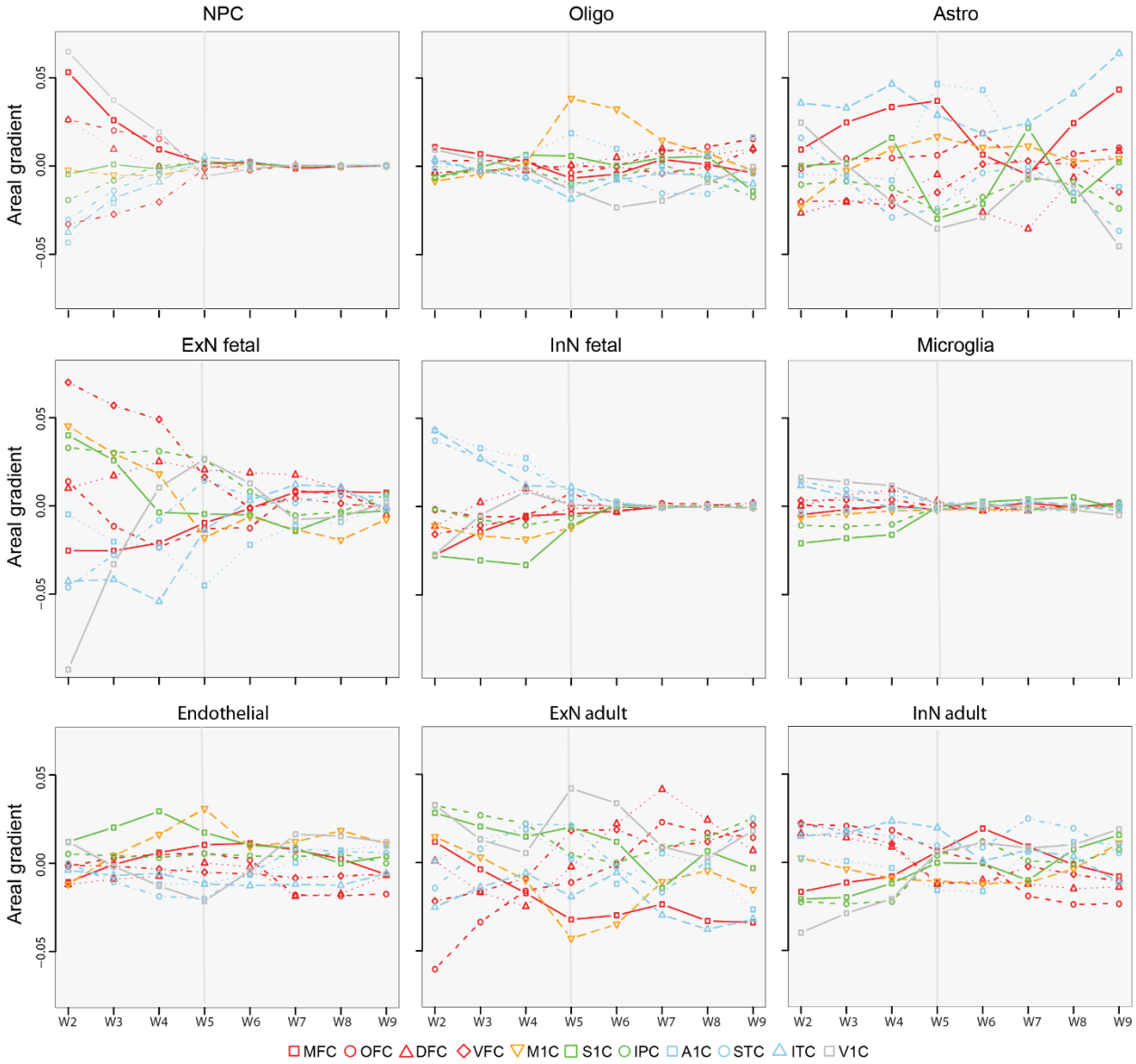


Fig. S29. Variation in cell type proportions across neocortical areas.

Based on cell type deconvolution, we assessed variation and gradients in the proportions of discrete cell populations in the neocortex across developmental windows W2 to W9. Variation was calculated by comparing the cell type proportion of a given neocortical area with the average values of the remaining 10 neocortical areas, with putative gradients apparent for some cell

populations, as evidenced by the general progression of anatomically proximal neocortical areas along the Y-axi. Combination of color and symbol are used to differentiate 11 neocortical areas. NPC: neural progenitor cells, ExN: excitatory neurons, InN: interneurons, Astro: astrocytic lineage, Oligo: oligodendrocytes.

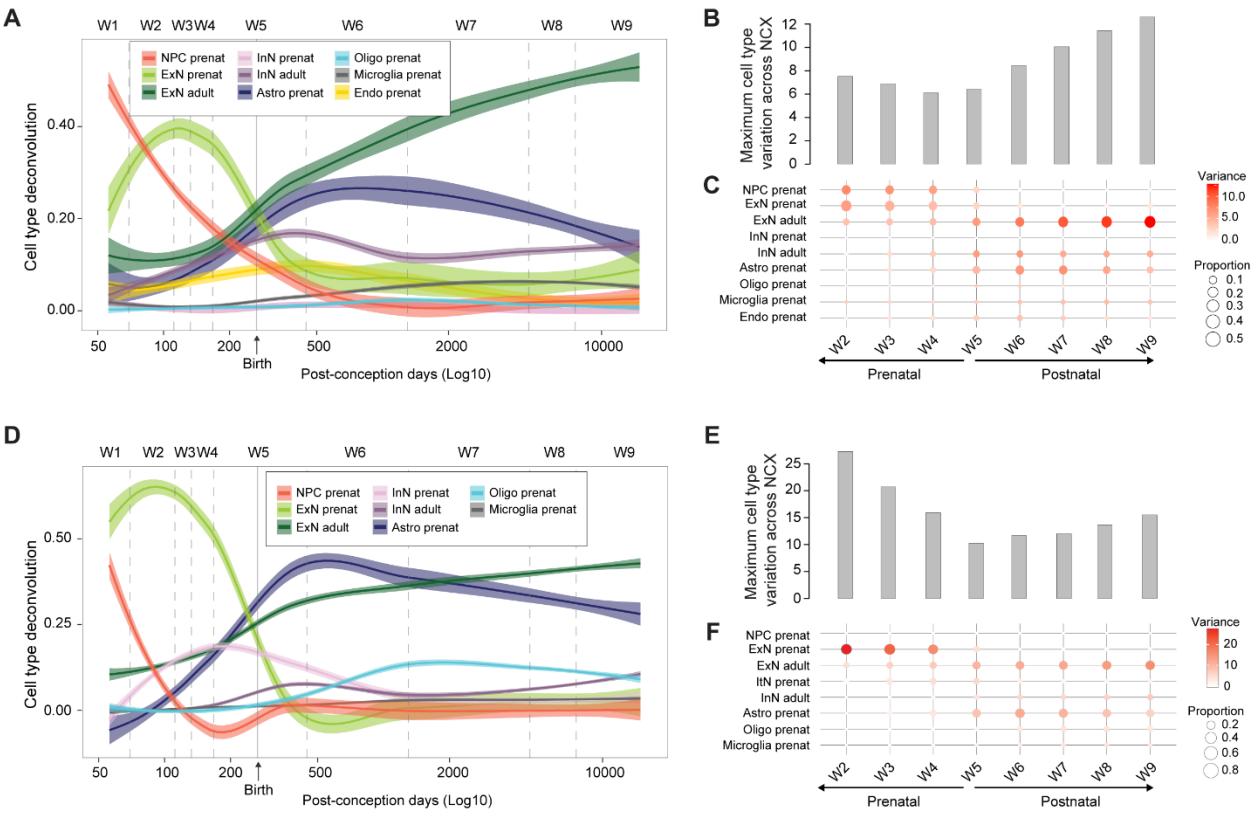


Fig. S30. Independent replication of deconvolution and cell composition analyses using alternate datasets.

Deconvolution analyses of **Fig.3C-E** were replicated using independent published fetal single cell datasets from Nowakowski et al. (*12*); A-C) and Zhong et al. (*26*); D-F) . (**A**, **D**) Deconvolution of bulk tissue data using cell type enriched markers identified through single cell sequencing of both fetal and adult human brain neocortex (*21*). The smoothed curves with shaded areas (50% confidence interval) represent the developmental dynamics of cell type composition in the human neocortex. Ages are represented as log10 transformed post-conception days along the x-axis, with the solid vertical line indicating birth. (**B**, **E**) Maximum inter-areal variance across cell types for each window. (**C**, **F**) Neocortical areal variation in the transcriptomic signatures of each major

cell type assayed in each developmental window. NPC: neural progenitor cells, ExN: excitatory neurons, InN: interneurons, Astro: astroglial lineage, oligo: oligodendrocytes, Endo: endothelial.

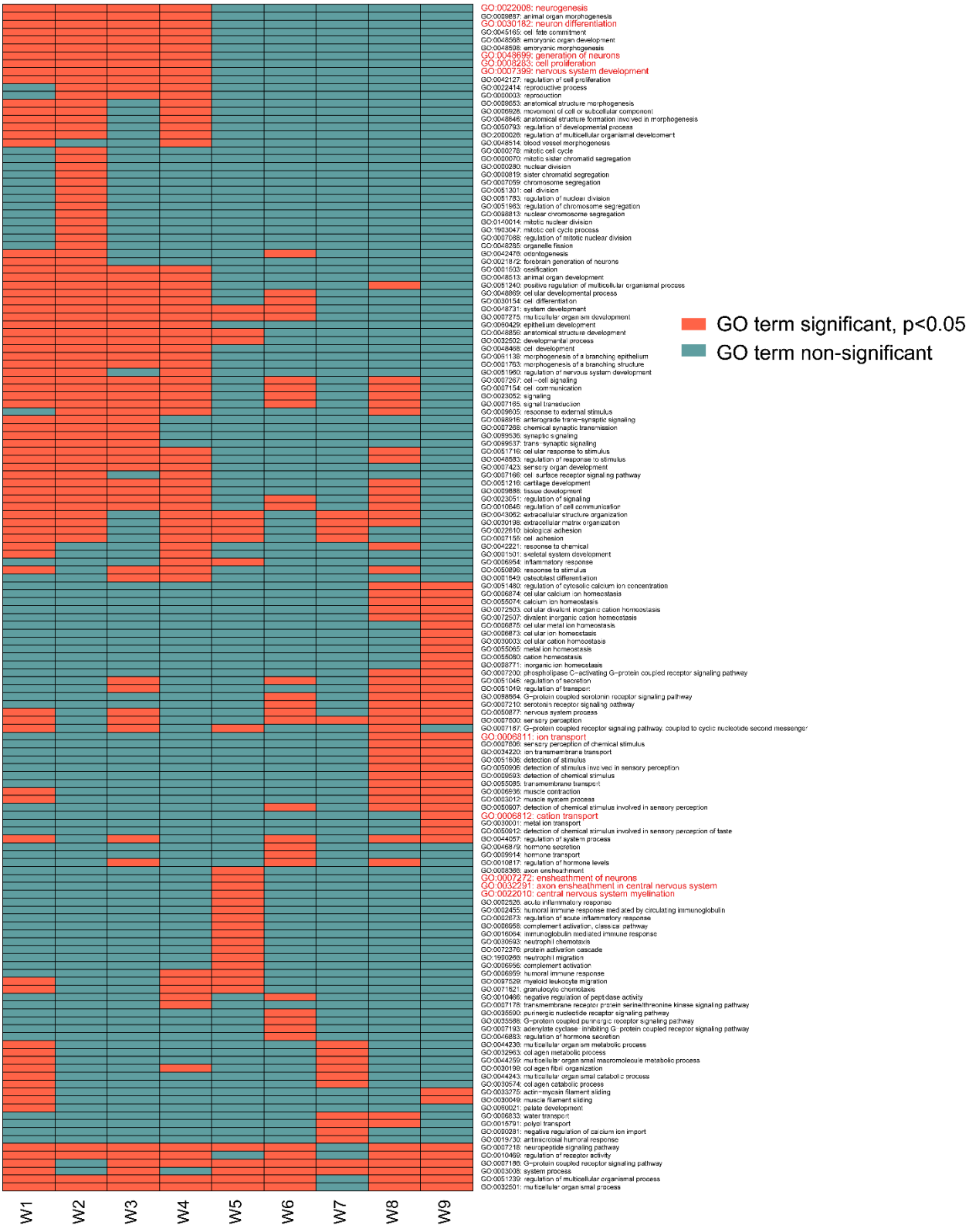

GO:0022008: neurogenesis
GO:0009887: animal organ morphogenesis
GO:0030182: neuron differentiation
GO:0045165: cell fate commitment
GO:0048568: embryonic organ development
GO:0048598: embryonic morphogenesis
GO:0048699: generation of neurons
GO:0008283: cell proliferation
GO:0007399: nervous system development
GO:0042127: regulation of cell proliferation
GO:0022414: reproductive process
GO:0000003: reproduction
GO:0009653: anatomical structure morphogenesis
GO:0006928: movement of cell or subcellular component
GO:0048646: anatomical structure formation involved in morphogenesis
GO:0050793: regulation of developmental process
GO:2000026: regulation of multicellular organismal development
GO:0048514: blood vessel morphogenesis
GO:0000278: mitotic cell cycle
GO:0000070: mitotic sister chromatid segregation
GO:0000280: nuclear division
GO:0000819: sister chromatid segregation
GO:0007059: chromosome segregation
GO:0051301: cell division
GO:0007067: regulation of nuclear division
GO:0051983: regulation of chromosome segregation
GO:0098813: nuclear chromosome segregation
GO:0140014: mitotic nuclear division
GO:1903047: mitotic cell cycle process
GO:0007088: regulation of mitotic nuclear division
GO:0048285: organelle fission
GO:0042476: odontogenesis
GO:0021872: forebrain generation of neurons
GO:0001503: ossification
GO:0048513: animal organ development
GO:0051240: positive regulation of multicellular organismal process
GO:0048869: cellular developmental process
GO:0030154: cell differentiation
GO:0048731: system development
GO:0007275: multicellular organism development
GO:0060429: epithelium development
GO:0048856: anatomical structure development
GO:0032502: developmental process
GO:0048468: cell development
GO:0061138: morphogenesis of a branching epithelium
GO:0001763: morphogenesis of a branching structure
GO:0051960: regulation of nervous system development
GO:0007267: cell–cell signaling
GO:0007154: cell communication
GO:0023052: signaling
GO:0007165: signal transduction
GO:0009605: response to external stimulus
GO:0098916: anterograde trans-synaptic signaling
GO:0007268: chemical synaptic transmission
GO:0099536: synaptic signaling
GO:0099537: trans-synaptic signaling
GO:0051716: cellular response to stimulus
GO:0048583: regulation of response to stimulus
GO:0007423: sensory organ development
GO:0007166: cell surface receptor signaling pathway
GO:0051216: cartilage development
GO:0009888: tissue development
GO:0023051: regulation of signaling
GO:0010646: regulation of cell communication
GO:0043062: extracellular structure organization
GO:0030198: extracellular matrix organization
GO:0022610: biological adhesion
GO:0007155: cell adhesion
GO:0042221: response to chemical
GO:0001501: skeletal system development
GO:0006954: inflammatory response
GO:0050896: response to stimulus
GO:0001649: osteoblast differentiation
GO:0051480: regulation of cytosolic calcium ion concentration
GO:0006874: cellular calcium ion homeostasis
GO:0055074: calcium ion homeostasis
GO:0072503: cellular divalent inorganic cation homeostasis
GO:0072507: divalent inorganic cation homeostasis
GO:0006875: cellular metal ion homeostasis
GO:0006873: cellular ion homeostasis
GO:0030003: cellular cation homeostasis
GO:0055065: metal ion homeostasis
GO:0055080: cation homeostasis
GO:0098771: inorganic ion homeostasis
GO:0007200: phospholipase C-activating G-protein coupled receptor signaling pathway
GO:0051046: regulation of secretion
GO:0051049: regulation of transport
GO:0098664: G-protein coupled serotonin receptor signaling pathway
GO:0007210: serotonin receptor signaling pathway
GO:0050877: nervous system process
GO:0007600: sensory perception
GO:0007187: G-protein coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger
GO:0006811: ion transport
GO:0007606: sensory perception of chemical stimulus
GO:0034220: ion transmembrane transport
GO:0051606: detection of stimulus
GO:0050906: detection of stimulus involved in sensory perception
GO:0098993: detection of chemical stimulus
GO:0055085: transmembrane transport
GO:0006936: muscle contraction
GO:0003012: muscle system process
GO:0050907: detection of chemical stimulus involved in sensory perception
GO:0006812: cation transport
GO:0030001: metal ion transport
GO:0050912: detection of chemical stimulus involved in sensory perception of taste
GO:0044057: regulation of system process
GO:0046879: hormone secretion
GO:0009914: hormone transport
GO:0010817: regulation of hormone levels
GO:0008366: axon ensheathment
GO:0007272: ensheathment of neurons
GO:0032291: axon ensheathment in central nervous system
GO:0022010: central nervous system myelination
GO:0002526: acute inflammatory response
GO:0002455: humoral immune response mediated by circulating immunoglobulin
GO:0002673: regulation of acute inflammatory response
GO:0006958: complement activation, classical pathway
GO:0016064: immunoglobulin mediated immune response
GO:0030593: neutrophil chemotaxis
GO:0072376: protein activation cascade
GO:1990266: neutrophil migration
GO:0006956: complement activation
GO:0006959: humoral immune response
GO:0097529: myeloid leukocyte migration
GO:0071621: granulocyte chemotaxis
GO:0010466: negative regulation of peptidase activity
GO:0007178: transmembrane receptor protein serine/threonine kinase signaling pathway
GO:0035590: purinergic nucleotide receptor signaling pathway
GO:0035586: G-protein coupled purinergic receptor signaling pathway
GO:0007193: adenylate cyclase-inhibiting G-protein coupled receptor signaling pathway
GO:0046883: regulation of hormone secretion
GO:0044236: multicellular organism metabolic process
GO:0032963: collagen metabolic process
GO:0044259: multicellular organismal macromolecule metabolic process
GO:0030199: collagen fibril organization
GO:0044243: multicellular organismal catabolic process
GO:0030574: collagen catabolic process
GO:0033275: actin-myosin filament sliding
GO:0030049: muscle filament sliding
GO:0060021: palate development
GO:0006833: water transport
GO:0015791: polyol transport
GO:0060291: negative regulation of calcium ion import
GO:0019730: antimicrobial humoral response
GO:0007218: neuropeptide signaling pathway
GO:0010469: regulation of receptor activity
GO:0007186: G-protein coupled receptor signaling pathway
GO:0003008: system process
GO:0051239: regulation of multicellular organismal process
GO:0032501: multicellular organismal process
GO term significant, p<0.05
GO term non-significant
W1
W2
W3
W4
W5
W6
W7
W8
W9

Fig. S31. Gene Ontology enrichment analysis of top variable genes per window. The top 30 FDR-corrected significant GO terms are shown for each window, clustered by their patterns of significance across windows. A selection of GO terms of special interest are highlighted in red.

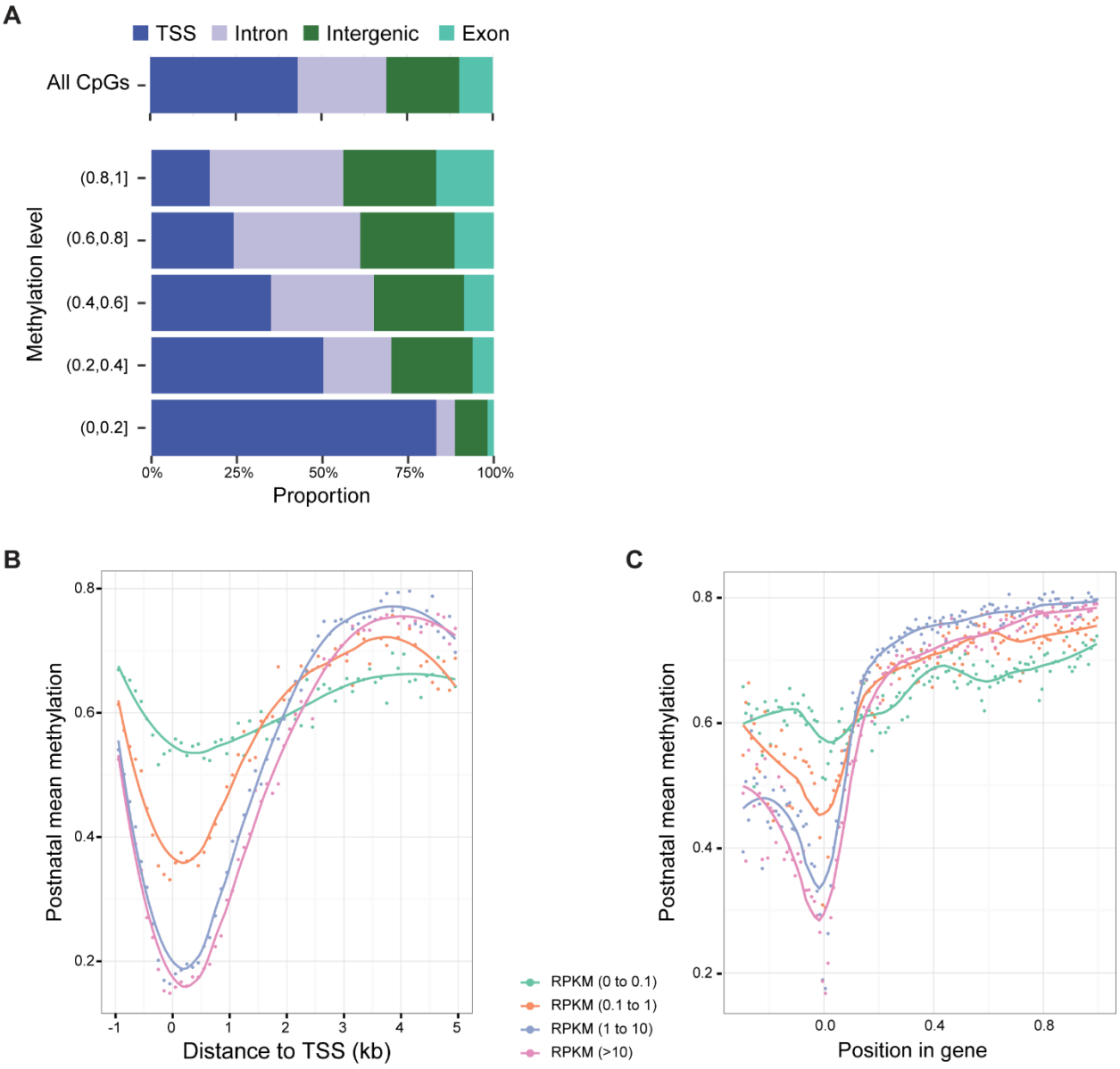


Fig. S32. Characterization of DNA methylation analyses.

(**A**) Proportion of DNA methylation sites with various levels of methylation (β values, arranged from highest to lowest) in genomic features such as TSSs (dark blue), introns (light blue), intergenic regions (dark green), and exons (light green), based on Gencode v21 annotation. The top bar shows the variation in the proportion of CpG sites at these different genomic features across

all DNA methylation levels. (B-C) Distribution of the DNA methylation signal with respect to (**B**) TSS and (**C**) the gene body. Genes were grouped by varying expression levels calculated from the mRNA-Seq data set.

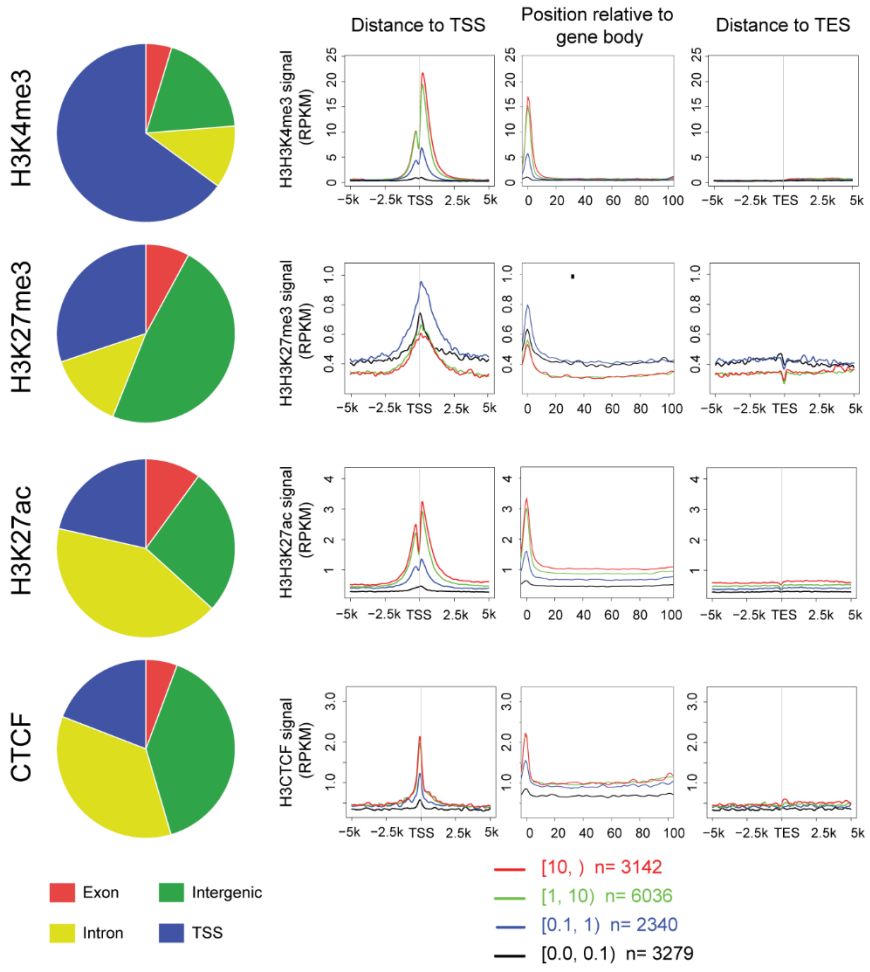


Fig. S33. Genomic distribution of histone modifications and CTCF.

Pie chart representing the proportions of genomic features where H3K4me3 (top), H3K27me3 (second from top), H3K27ac (second from bottom), and CTCF (bottom) were enriched. Right hand panels show the distribution of signals from H3K4me3, H3K27me3, H3K27ac, and CTCF at TSSs, in the gene body of genes, and at the furthest downstream transcription end site (TES). Genes were grouped by varying expression levels calculated from RNA-Seq data.

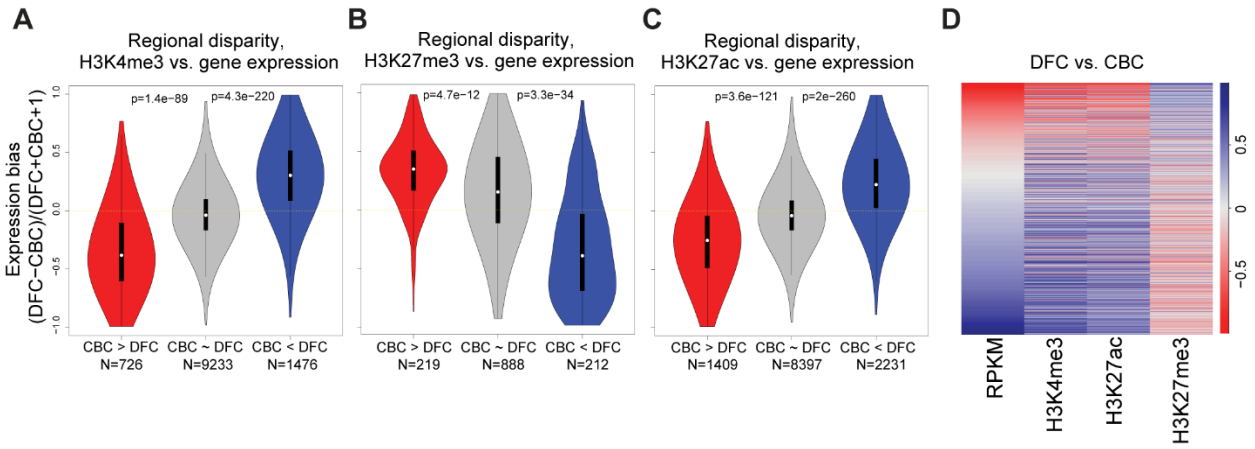


Fig. S34. Relationship between regional gene expression and histone modifications.

Relationships between gene expression and the regional enrichment of histone modification in the adult human brain as inferred from H3K4me3 (**A**), H3K27me3 (**B**), and H3K27ac (**C**) enrichment. N represents the number of peaks assigned to each category. The relative expression of genes in DFC (>0) relative to CBC (<0) is plotted on the Y-axis for genes associated with histone modifications present to a greater extent in the CBC (red) or DFC (blue), or where there is no significant difference in histone modification between DFC and CBC (grey). (**D**) Heat map showing regional enrichment of expression and all histone marks in genes. Genes were ordered by the regional enrichment of expression.

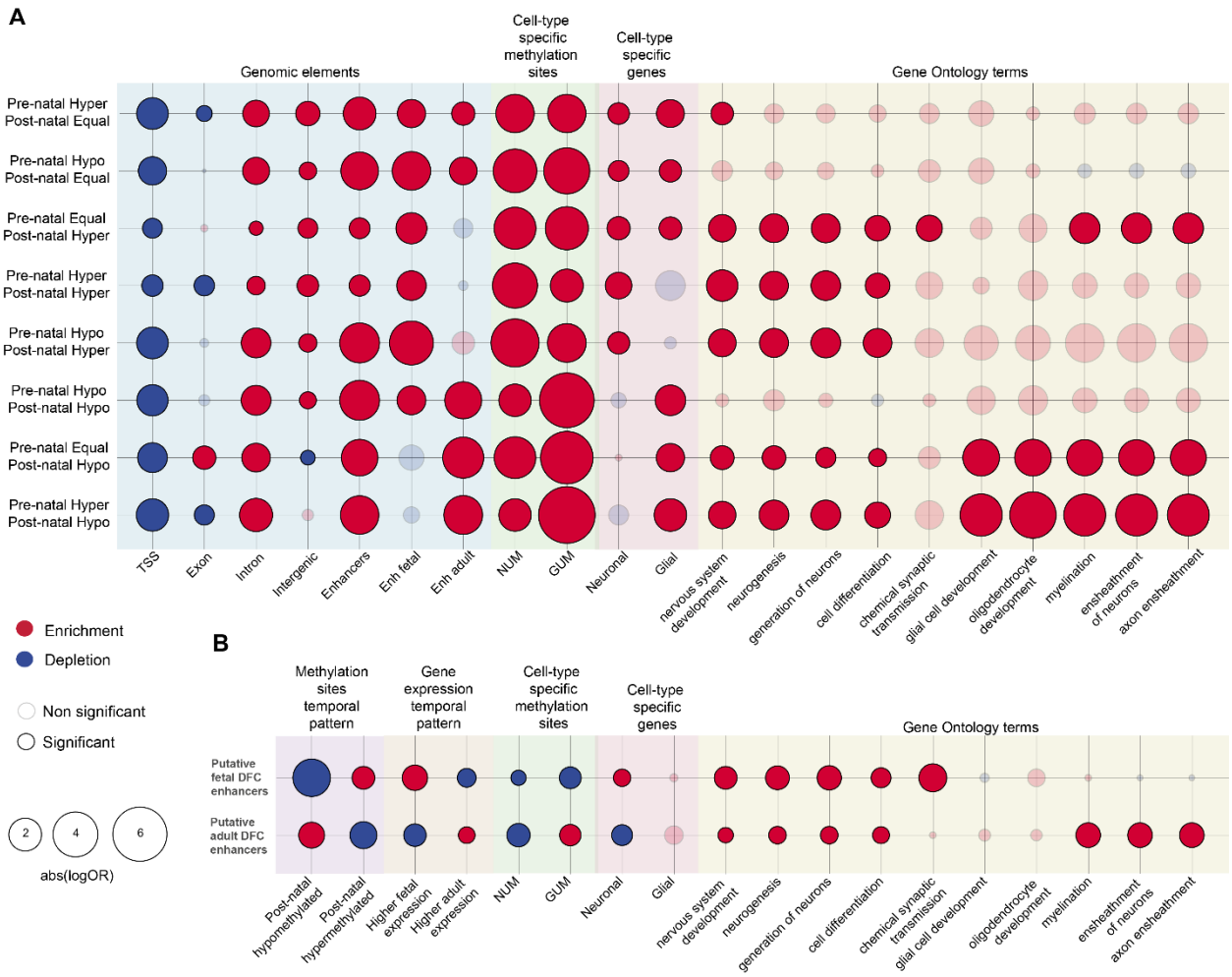
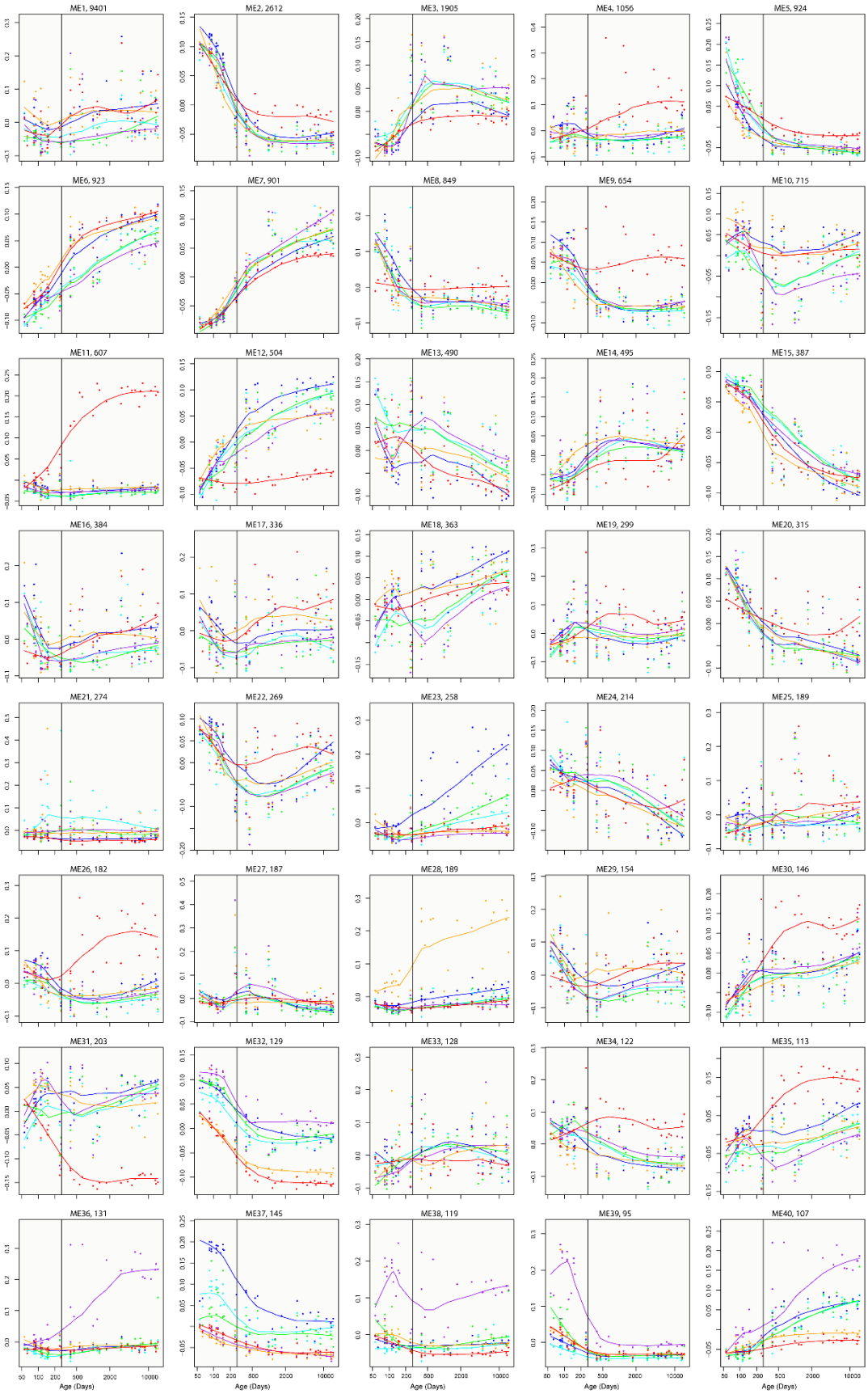


Fig. S35. Enrichment analyses of methylation trajectory and temporal regulated enhancers.

(A) Enrichment of eight types of dynamic methylation trajectories compared to non-variable methylation sites in various genomic elements (light blue background), Neu+ or Neu- hypomethylated sites from Kozlenkov et al. (*114*) (green background), cell type specific genes identified from literature (red background), and selected gene ontology terms reflecting major neural biological processes (yellow background). (**B**) Enrichment of putative fetal or adult-specific enhancers compared to non-dynamic enhancers in methylation sites postnatally hypo or hyper-methylated (red background), fetal or adult patterns of enhancer-associated gene expression

(brown background), Neu+ or Neu- hypomethylated sites from Kozlenkov et al. 2013 (green background, named NUM and GUM, respectively), cell type specific genes , (red background), and selected gene ontology terms reflecting major neural biological processes (yellow background). Solid color indicates significance ($P<0.05$ for all tests except for GO analysis, $P<0.01$), red indicates enrichment and blue, depletion. Size is proportional to the absolute logOR.


ME1, 9401
ME2, 2612
ME3, 1905
ME4, 1056
ME5, 924
ME6, 923
ME7, 901
ME8, 849
ME9, 654
ME10, 715
ME11, 607
ME12, 504
ME13, 490
ME14, 495
ME15, 387
ME16, 384
ME17, 336
ME18, 363
ME19, 299
ME20, 315
ME21, 274
ME22, 269
ME23, 258
ME24, 214
ME25, 189
ME26, 182
ME27, 187
ME28, 189
ME29, 154
ME30, 146
ME31, 203
ME32, 129
ME33, 128
ME34, 122
ME35, 113
ME36, 131
ME37, 145
ME38, 119
ME39, 95
ME40, 107
50 100 200 500 2000 10000
Age (Days)

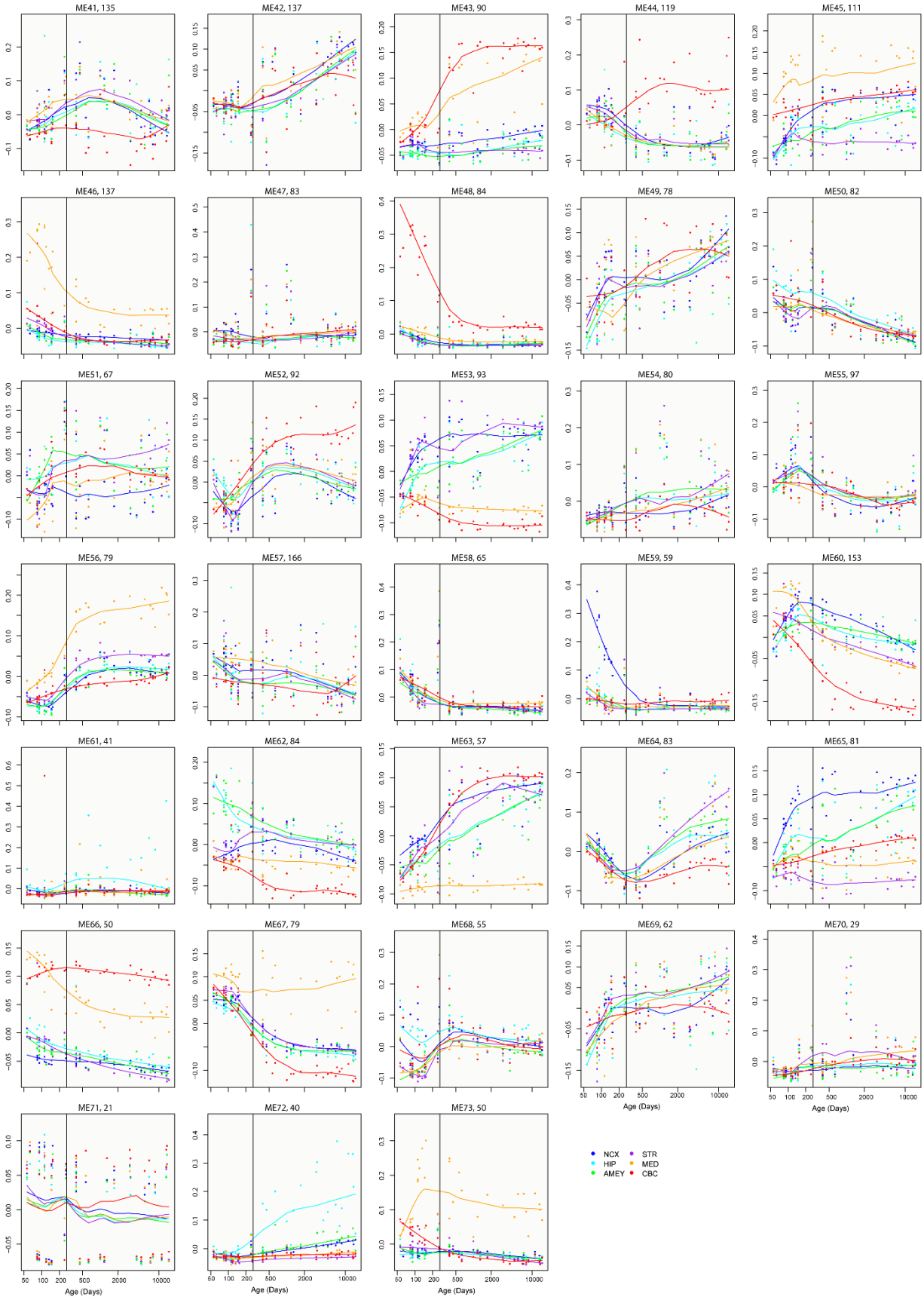

ME41, 135
ME42, 137
ME43, 90
ME44, 119
ME45, 111
ME46, 137
ME47, 83
ME48, 84
ME49, 78
ME50, 82
ME51, 67
ME52, 92
ME53, 93
ME54, 80
ME55, 97
ME56, 79
ME57, 166
ME58, 65
ME59, 59
ME60, 153
ME61, 41
ME62, 84
ME63, 57
ME64, 83
ME65, 81
ME66, 50
ME67, 79
ME68, 55
ME69, 62
ME70, 29
ME71, 21
ME72, 40
ME73, 50
Age (Days)
NCX
HIP
AMEY
STR
MED
CBC

Fig. S36. Trajectories of WGCNA modular eigengenes.

The spatio-temporal pattern for each WGCNA module was summarized by the trajectory of the modular eigengene values, in which the eigengene value is plotted against developmental days (Log10 transformed). The trajectories for NCX (blue), HIP (cyan), AMY (green), STR (purple), MD (orange), and CBC (red) samples are separated and smoothed by *loess* function in R software. Each circle represents one brain sample, and is differently colored according to anatomical region. The module index and the number of modular genes are indicated above each plot. The vertical solid lines indicate the time of birth.

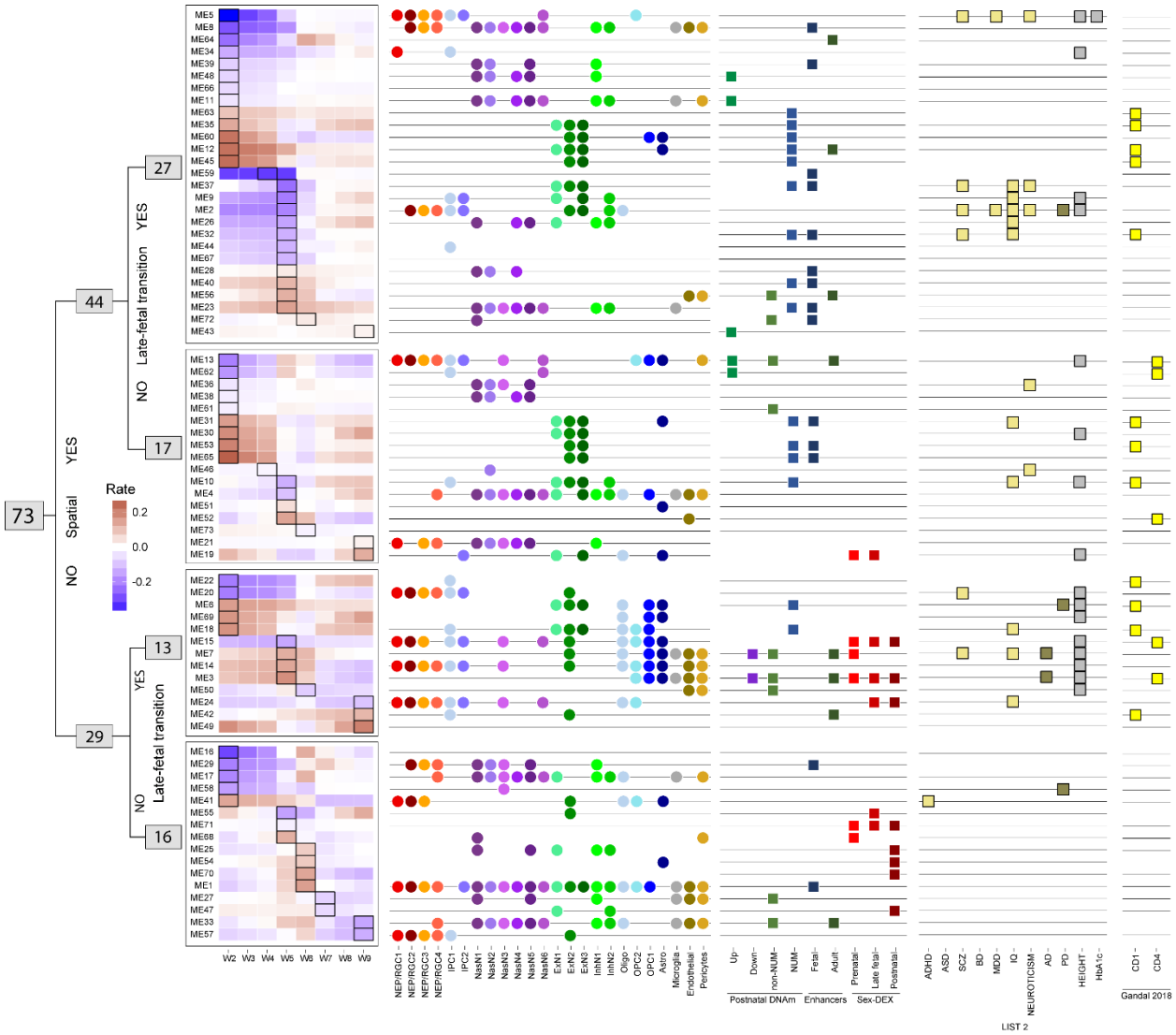


Fig. S37. Integrative and multi-layered characterization of WGCNA modules.

WGCNA modules were integrated using eigengene expression trajectories and the enrichment within modules for genes associated with various molecular, epigenetic, and disease associations. From left, the first panel shows the dendrogram with numbers, revealing how the 73 WGCNA modules were classified into four groups according to the occurrence of eigengene differences across brain regions ("Spatial") or between prenatal and postnatal development in the neocortex ("Late fetal transition"). The second panel shows multiple heat maps, revealing the change rate

per module of the eigengene value from one window to the subsequent window. Black rectangles indicate the window where the highest rate of change occurs. The third panel shows the enrichment of modular genes in 24 prenatal scRNA-Seq clusters. The fourth panel shows the enrichment of modular genes in 1) methylation sites postnatally hypo (or hyper-) methylated, 2) Neu+ or Neu- hypomethylated sites (from Kozlenkov et al. (*114*)) (NUM and GUM, respectively), 3) genes associated to fetal or adult patterns of enhancer activity, and 4) sex-biased genes for either male or female in three different periods. The fifth panel shows modules enriched in genes associated with different diseases (light yellow) and other traits (grey). The last panel shows enrichment for genes contained in two modules (CD1, CD4) correlated with schizophrenia, autism spectrum disorder, and bipolar disorder (*69*). NEP/RGC: Neural epithelial progenitor/radial glial cells, NasN: Nascent neurons, IPC: intermediate progenitor cells, OPC: oligodendrocyte progenitor cells, Oligo: oligodendrocytes, ExN: excitatory neurons, InN: interneurons, Astro: astroglial lineage, ADHD: attention-deficit/hyperactivity disorder, ASD: autism spectrum disorder, SCZ: schizophrenia, BD: bipolar disorder, MDD: major depressive disorder, IQ: intelligence quotient, NEUROT: neuroticism, AD: alzheimer disease, PD: Parkinsons disease, HBA1C: glycated hemoglobin.

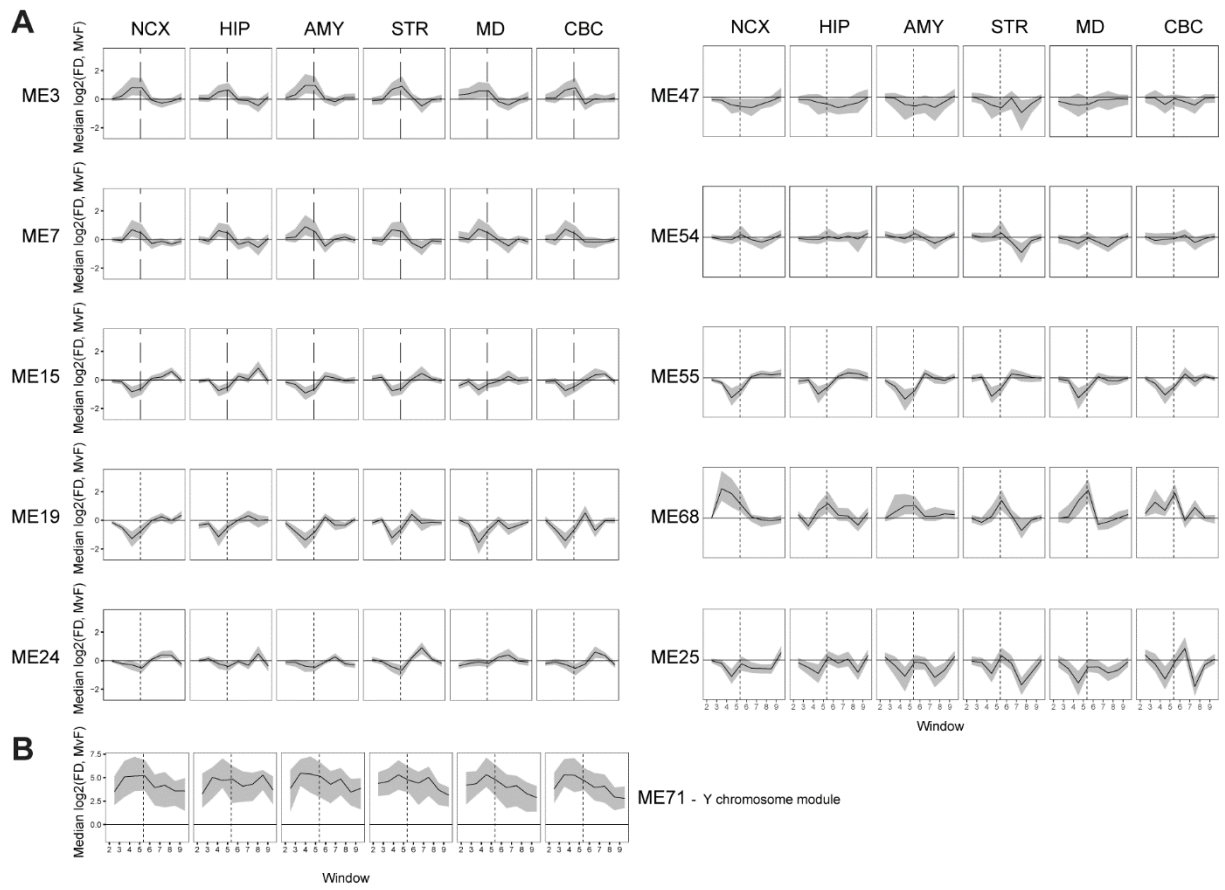


Fig. S38. Median sex-bias in modules enriched for genes exhibiting sexual dimorphism.

(A) The median fold difference imputed between male (>0) and female (<0) gene expression for all genes in a given module was plotted on the Y-axis using a log2 scale. Grey shading indicates the interquartile range around the median fold difference. Modules were selected from those enriched for genes exhibiting sex-biased expression between W2-W4, W4-W6, and W6-W9. **(B)** The median fold difference for ME71, a module comprised of genes located on the Y-chromosome, exhibited persistent male-biased expression.

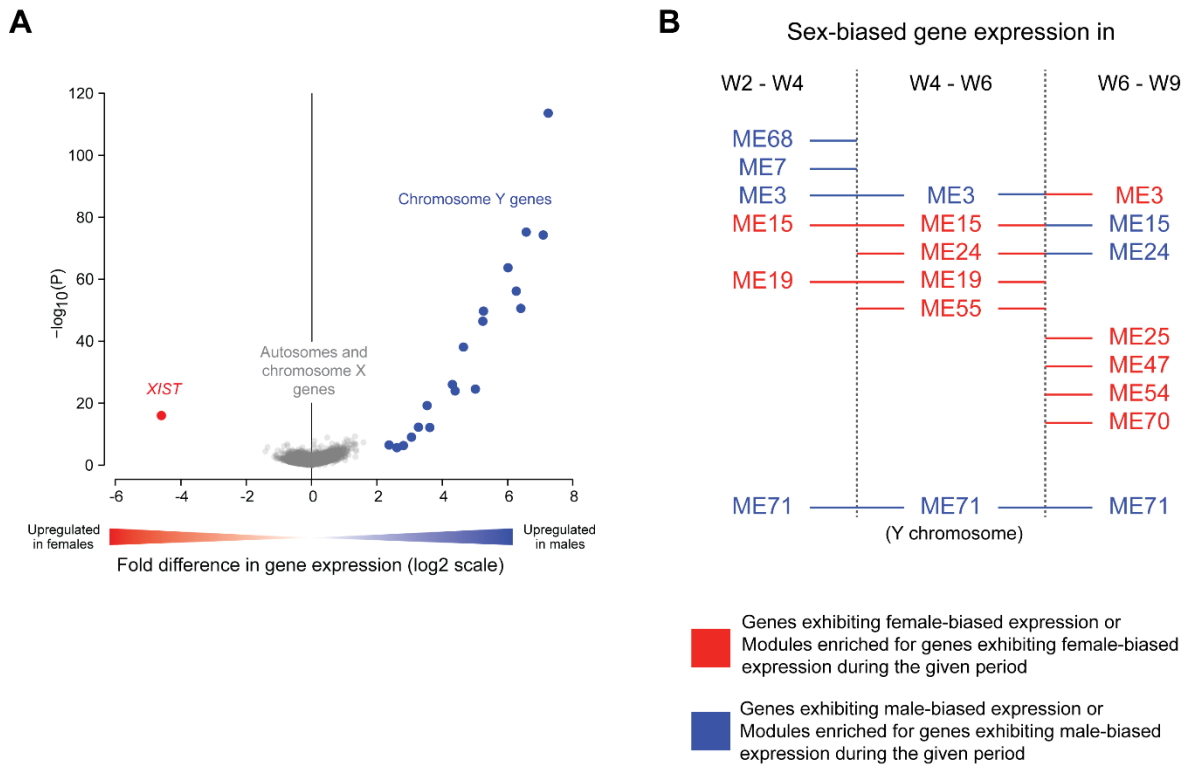


Fig. S39. Distribution of sex bias.

(A) Volcano plot showing the imputed fold difference in gene expression between males and females across all brain regions and the entirety of the developmental dataset on the X-axis and the significance of any difference along the Y-axis. No autosomal genes exhibited differential expression between males and females across all developmental times or brain regions. **(B)** Modules enriched for genes exhibiting sex-biased gene expression and the persistence of that enrichment across time. No autosomal module was enriched for genes exhibiting bias for a single sex (male or female) across W2-W4, W4-W6, and W6-W9.

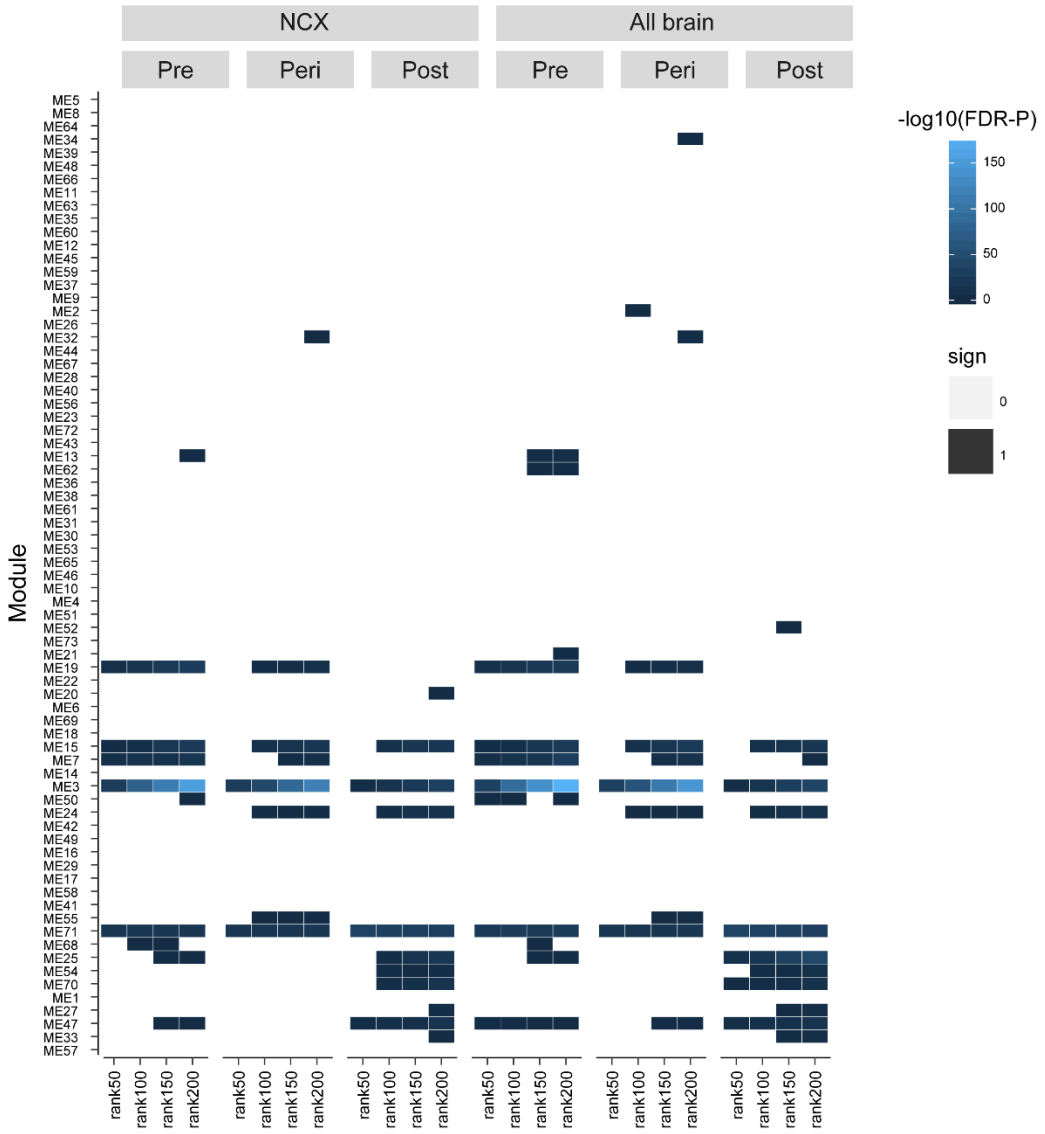


Fig. S40. Identification of sex-DEX modules across various threshold criteria.

We iteratively selected the 50, 100, 150, and 200 genes exhibiting the greatest sex-differential (sex-DEX) gene expression in at least two consecutive windows in at least one brain region. Shades of blue represent significant enrichment of a module for sex-DEX genes. Modules are arranged from those that exhibit spatio-temporal specificity (top) through those exhibiting spatial, temporal, or no spatial or temporal specificity (bottom).

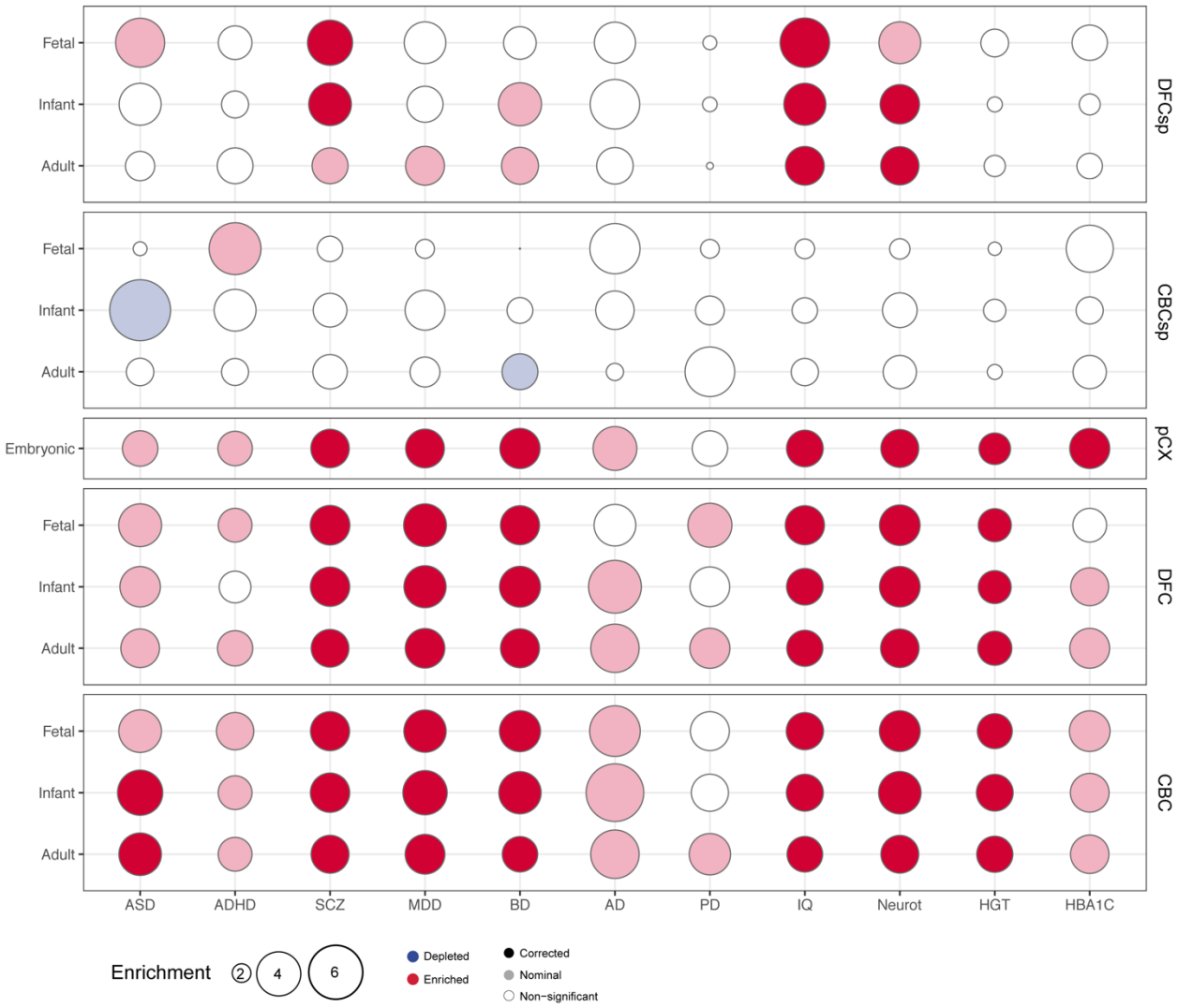


Fig. S41. Enrichment of SNP-heritability traits in *cis*-regulatory elements.

Partitioned linkage disequilibrium (LD) score regression analysis showing enrichment of SNP-heritability of specific traits in different sets of putative regulatory elements (H3K27ac peaks). Solid color indicates significance for Bonferroni adjusted P-value, faint color indicates nominal significance. In each temporal category (fetal, infant, and adult), only peaks that are present in at least 2 samples of the given category are considered. Regional categories (pCTX, DFC, and CBC) include peaks present in at least 2 samples for that region, while specific regional categories (DFCsp and CFCsp) require also the total absence of the peak in the other brain region analyzed.

ASD: autism spectrum disorder, ADHD: attention-deficit/hyperactivity disorder, SCZ: schizophrenia, MDD: major depressive disorder, BD: bipolar disorder, AD: alzheimer disease, PD: Parkinsons disease, IQ: intelligence quotient, NEUROT: neuroticism, HGT: height, HBA1C: glycated hemoglobin.

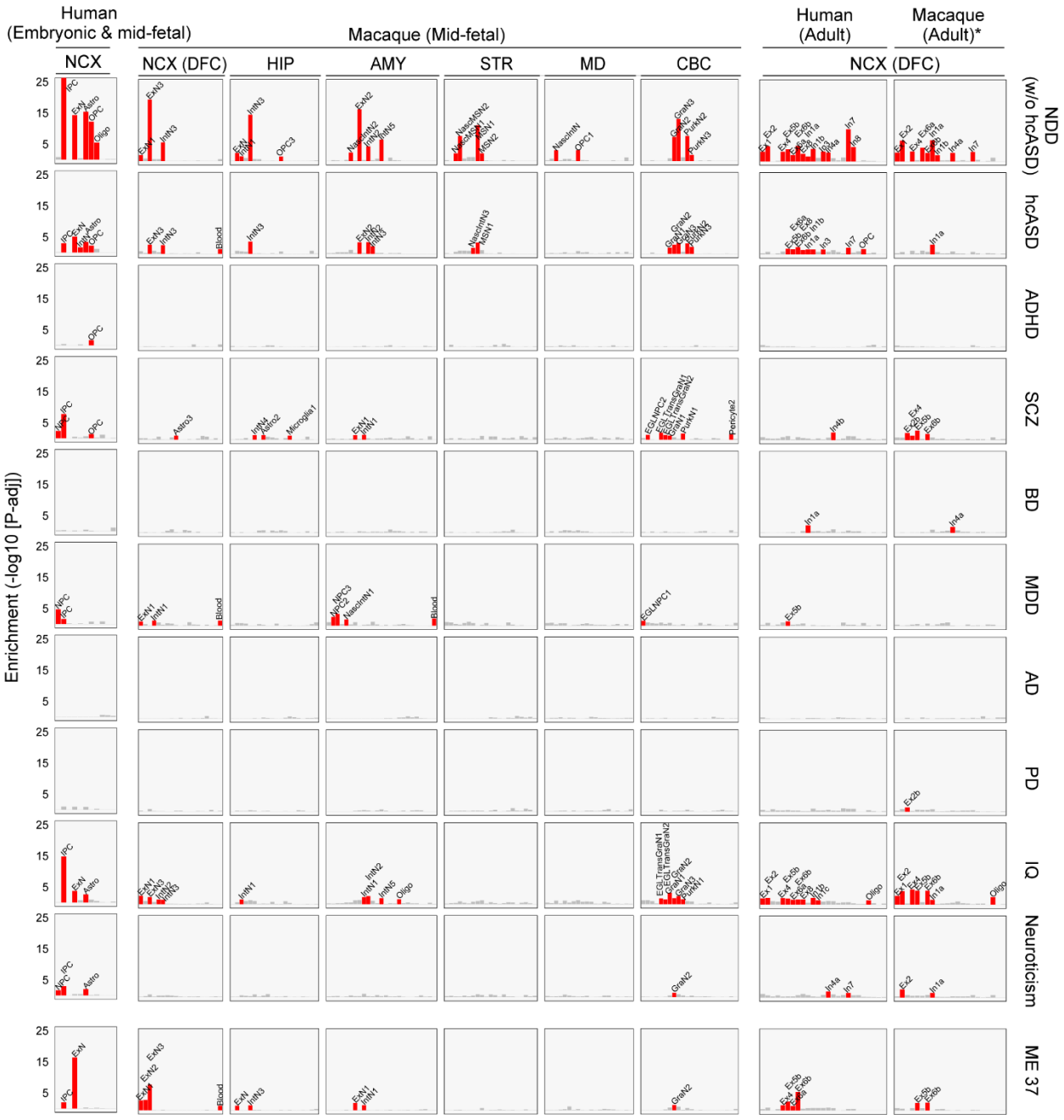


Fig. S42. Enrichment of disease-associated genes in cell types. Enrichment analysis of traits and disease-genes identified from the integration of GWAS with Hi-C and H3K27ac data in clusters of cell types from human and macaque single-cell or single-nuclei data from fetal and adult and different brain regions. Red bars indicate significant p-values at FDR<0.05, corrected

independently by each region and species. (*) Adult macaque cells were classified into human adult clusters using Random Forest. For graphical representation, log10 P-values are capped at 25.

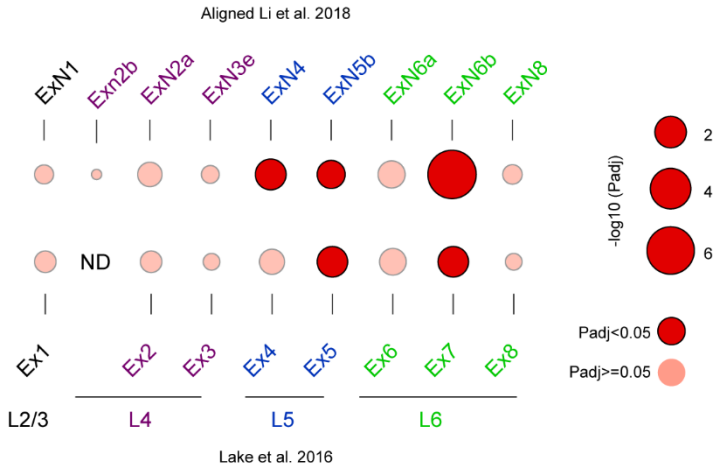


Fig. S43. Cell types enriched for disease-associated genes found in ME37.

Cell type enrichment for the expression of genes in ME37 associated with neurological or psychiatric disorders. Dark red coloration indicates significance, with the size of the circle representing p-value adjusted for multiple testing on a log10 scale. The top row shows the enrichment of disease-associated gene expression among cell types identified in this study while the bottom row shows similar enrichment for cell types described by Lake et al. (*27*).

Additional Supplementary Information

Supplementary Tables:

Supplementary Tables are provided in a single Microsoft Excel file.

Table S1. Specimen/sample metadata and quality control metrics for mRNAseq

Metadata from the donors is presented, including developmental window, brain code, region code, hemisphere of samples, age, age counting as days, sex, ethnicity, postmortem interval (PMI), pH of cerebellum, dissection score, and sequencing center (Yale and USC). This table also shows the quality control metrics, including raw reads count, multiple mapping reads, mitochondrial (chrM) reads, ribosomal (rRNA) reads, transfer (tRNA) reads, uniquely mapped reads, reads classification to exon, intron and intergenic, sequencing quality assessment from FastQC, gene coverage uniformity, gene numbers at different RPKM levels, and correlation coefficients to same sample in exon array data set and to bio-replicates in the same developmental period.

Table S2. Specimen/sample metadata and quality control metrics for smRNAseq

Metadata from the donors is presented, including developmental window, brain code, region code, hemisphere of samples, age, age counting as days, sex, ethnicity, postmortem interval (PMI), pH of cerebellum, dissection score, sequencing center (Yale and USC), and raw reads count.

Table S3. Specimen/sample metadata and quality control metrics for scRNAseq

Metadata from the donors is presented, including developmental window, cell code, brain code, region code, age, cell number, and cell type. This table also shows the quality control metrics, including raw reads count, uniquely mapped reads, percentage of uniquely mapped reads, percentage of mitochondrial (chrM) reads, ribosomal (rRNA) reads, transfer (tRNA) reads, reads classification to exon, intron and intergenic.

Table S4. Specimen/sample metadata and quality control metrics for single nuclei RNAseq

Metadata from the donors is presented, including developmental window, cell code, brain code, region code, age, sex, ethnicity, PMI, and PH. This table also shows the quality control metrics, including raw reads, uniquely mapped reads to genome/exon/transcriptome, cell number, median gene number and median UMI number.

Table S5. Specimen/sample metadata for DNA methylation

Metadata from the donors is presented, including developmental window, brain code, region code, age, age counting as days, sex, ethnicity, postmortem interval (PMI), pH of cerebellum.

Table S6. Specimen/sample metadata for histone modification ChIP-Seq

Metadata from the donors is presented, including developmental window, brain code, region code, age, age counting as days, sex, ethnicity, postmortem interval (PMI), pH of cerebellum.

Table S7. Temporal and spatial enrichment analyses for alternative spliced genes

List of genes detected alternative splicing events is presented. For each brain region, the temporal difference was evaluated by comparing prenatal samples to postnatal samples. Only statistical significant brain regions are provided along with temporal enrichment and p-values.

Table S8. List of genes for cell type signature

Table S9. Gene Ontology analysis for top variable genes

Top variable genes were identified in each window and a GO enrichment analysis was performed using the genes expressed at RPKM>1 in the corresponding window as background. P-values were corrected by false discovery rate using the Benjamini-Hochberg method.

Table S10. Weighted gene co-expressed gene modules and spatio-temporal characterization

List of 73 modules identified by weighted gene co-expression network analysis (WGCNA) with modular gene number, list of modular genes, and three types of spatio-temporal characterization, i.e., spatio-temporal patterns of six brain regions, transit from prenatal to postnatal (W4 to W5), spatio-temporal patterns of eleven neocortex areas. The interpretation of spatio-temporal pattern is concluded from statistical significance p values.

Table S11. Integrative and multi-layered enrichment analyses for weighted gene co-expressed gene module

List of enrichment analyses conducted for 73 WGCNA modules, including methylation trajectories, methylation sites cell type specific, enhancer, sex DEX genes, single cell clusters, and GWAS genes.

Table S12. Enrichment of features in different categories of WGCNA modules.

Multiple gene-set and cell type enrichment analysis into the 73 WGCNA co-expression modules. Indicated P-values are nominal.

Table S13. Genes associated to disease via GWAS (List1)

Lists of genes linked to GWAS loci for each disease. The type of association for each gene is indicated. Exonic, indicate that significant GWAS variants have an annotated variant effect (other than upstream or downstream gene variant, synonymous or intronic). Promoter indicates that the ChIP-Seq peak overlapping the GWAS significant region overlaps the gene TSS (+/- 1kb). NoInt indicates that the ChIP-Seq peak overlapping the GWAS significant region can only be linked to a gene by proximity to TSS. Hi-C indicates peaks associated to TSS via Hi-C data in either of the two datasets.

Table S14. Genes associated to disease via GWAS (List2)

Lists of genes linked to GWAS loci for each disease. The type of association for each gene is indicated. Exonic, indicate that significant GWAS variants have an annotated variant effect (other than upstream or downstream gene variant, synonymous or intronic). Promoter indicates that the ChIP-Seq peak overlapping the GWAS significant region overlaps the gene TSS (+/- 1kb). NoInt indicates that the ChIP-Seq peak overlapping the GWAS significant region can only be linked to a gene by proximity to TSS. Hi-C indicates peaks associated to TSS via Hi-C data in both of the two Hi-C datasets.

Table S15. Spike-ins RNAs and brain region barcode scheme

(A) The spike-in RNA characteristics and concentration for sample tagging are shown. Each spike-in master mix contains five different spike-in RNAs. ERCC-00025 (spike-in ID 4), ERCC-00108 (spike-in ID 9), and ERCC-00123 (spike-in ID 10) are included to assess sequencing error rates. The ERCC number, approximate length (bases), % GC content, and designated use of the individual spike-in RNA are shown. **(B)** Combinations of two spike-in RNAs (ID 1, 3, 5, 6, 7, 8, 11, or 12) are included to demarcate each neuroanatomical region. All spike-in RNAs were a generous gift of Dr. Mark Salit at The National Institute of Standards and Technology (NIST) (The External RNA Controls Consortium. 2005).

Table S16. List of all primers and probes

List of primers for ChIP-Seq validation by digital PCR and for splicing validation.

References

1. E. S. Lein, T. G. Belgard, M. Hawrylycz, Z. Molnar, Transcriptomic Perspectives on Neocortical Structure, Development, Evolution, and Disease. *Annu Rev Neurosci* **40**, 629-652 (2017).
2. J. C. Silbereis, S. Pochareddy, Y. Zhu, M. Li, N. Sestan, The Cellular and Molecular Landscapes of the Developing Human Central Nervous System. *Neuron* **89**, 248-268 (2016).
3. D. H. Geschwind, P. Rakic, Cortical evolution: judge the brain by its cover. *Neuron* **80**, 633-647 (2013).
4. J. H. Lui, D. V. Hansen, A. R. Kriegstein, Development and evolution of the human neocortex. *Cell* **146**, 18-36 (2011).
5. T. Paus, M. Keshavan, J. N. Giedd, Why do many psychiatric disorders emerge during adolescence? *Nat Rev Neurosci* **9**, 947-957 (2008).
6. S. A. McCarroll, S. E. Hyman, Progress in the genetics of polygenic brain disorders: significant new challenges for neurobiology. *Neuron* **80**, 578-587 (2013).
7. B. I. Bae, D. Jayaraman, C. A. Walsh, Genetic changes shaping the human brain. *Dev Cell* **32**, 423-434 (2015).
8. J. M. Keil, A. Qalieh, K. Y. Kwan, Brain transcriptome databases: a user's guide. *J Neurosci*, (2018).
9. S. Darmanis *et al.*, A survey of human brain transcriptome diversity at the single cell level. *Proc Natl Acad Sci U S A* **112**, 7285-7290 (2015).
10. M. B. Johnson *et al.*, Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62**, 494-509 (2009).
11. J. A. Miller *et al.*, Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199-206 (2014).
12. T. J. Nowakowski *et al.*, Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318-1323 (2017).
13. L. de la Torre-Ubieta *et al.*, The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. *Cell* **172**, 289-304 e218 (2018).
14. R. N. Doan *et al.*, Mutations in Human Accelerated Regions Disrupt Cognition and Social Behavior. *Cell* **167**, 341-354 e312 (2016).
15. M. Florio *et al.*, Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. *Science* **347**, 1465-1470 (2015).
16. M. B. Johnson *et al.*, Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human cortex. *Nat Neurosci* **18**, 637-646 (2015).
17. R. Lister *et al.*, Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).
18. C. Colantuoni *et al.*, Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* **478**, 519-523 (2011).
19. H. J. Kang *et al.*, Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483-489 (2011).
20. M. Pletikos *et al.*, Temporal specification and bilaterality of human neocortical topographic gene expression. *Neuron* **81**, 321-332 (2014).
21. S. K. Reilly *et al.*, Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science* **347**, 1155-1159 (2015).

22. Materials and Methods, Material and Methods.
23. Z. Lin *et al.*, Simultaneous dimension reduction and adjustment for confounding variation. *Proc Natl Acad Sci U S A* **113**, 14662-14667 (2016).
24. G. TEx Consortium, Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-660 (2015).
25. B. B. Lake *et al.*, Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat Biotechnol* **36**, 70-80 (2018).
26. S. Zhong *et al.*, A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524-528 (2018).
27. B. B. Lake *et al.*, Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586-1590 (2016).
28. D. Mi *et al.*, Early emergence of cortical interneuron diversity in the mouse embryo. *Science*, (2018).
29. S. Lodato, P. Arlotta, Generating neuronal diversity in the mammalian cerebral cortex. *Annu Rev Cell Dev Biol* **31**, 699-720 (2015).
30. A. S. Nord, K. Pattabiraman, A. Visel, J. L. Rubenstein, Genomic perspectives of transcriptional regulation in forebrain development. *Neuron* **85**, 27-47 (2015).
31. M. Shibata, F. O. Gulden, N. Sestan, From trans to cis: transcriptional regulatory networks in neocortical development. *Trends Genet* **31**, 77-87 (2015).
32. G. D'Arcangelo *et al.*, A protein related to extracellular matrix proteins deleted in the mouse mutant reeler. *Nature* **374**, 719-723 (1995).
33. J. J. Rowell, A. K. Mallik, J. Dugas-Ford, C. W. Ragsdale, Molecular analysis of neocortical layer structure in the ferret. *J Comp Neurol* **518**, 3272-3289 (2010).
34. Y. Zhu *et al.*, Spatiotemporal transcriptomic divergence across human and macaque brain development. *Submitted to Science*, (2018).
35. D. J. Miller *et al.*, Prolonged myelination in human neocortical evolution. *Proc Natl Acad Sci U S A* **109**, 16480-16485 (2012).
36. P. R. Huttenlocher, A. S. Dabholkar, Regional differences in synaptogenesis in human cerebral cortex. *J Comp Neurol* **387**, 167-178 (1997).
37. A. F. Pardinas *et al.*, Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet* **50**, 381-389 (2018).
38. J. C. Lambert *et al.*, Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45**, 1452-1458 (2013).
39. M. A. Nalls *et al.*, Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet* **46**, 989-993 (2014).
40. J. Grove *et al.*, Common risk variants identified in autism spectrum disorder. *bioRxiv*, (2017).
41. D. Demontis *et al.*, Discovery Of The First Genome-Wide Significant Risk Loci For ADHD. *bioRxiv*, (2017).
42. N. R. Wray *et al.*, Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* **50**, 668-681 (2018).
43. Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium, Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell* **173**, 1705-1715 e1716 (2018).
44. J. E. Savage *et al.*, Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet* **50**, 912-919 (2018).

45. S. Sniekers *et al.*, Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat Genet* **49**, 1107-1112 (2017).
46. A. R. Wood *et al.*, Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* **46**, 1173-1186 (2014).
47. J. Z. Liu *et al.*, Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* **47**, 979-986 (2015).
48. C. J. Willer *et al.*, Discovery and refinement of loci associated with lipid levels. *Nat Genet* **45**, 1274-1283 (2013).
49. E. Wheeler *et al.*, Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS Med* **14**, e1002383 (2017).
50. Genomes Project Consortium *et al.*, A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-1073 (2010).
51. P. M. Giusti-Rodriguez, P. F. Sullivan, Schizophrenia and a high-resolution map of the three-dimensional chromatin interactome of adult and fetal cortex. *bioRxiv*, (2018).
52. D. Wang *et al.*, Comprehensive functional genomic resource and integrative model for the human brain. *Submitted to Science*, (2018).
53. H. Won *et al.*, Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523-527 (2016).
54. J. P. Hamilton, M. Siemer, I. H. Gotlib, Amygdala volume in major depressive disorder: a meta-analysis of magnetic resonance imaging studies. *Mol Psychiatry* **13**, 993-1000 (2008).
55. M. J. Rubinow *et al.*, Basolateral amygdala volume and cell numbers in major depressive disorder: a postmortem stereological study. *Brain Struct Funct* **221**, 171-184 (2016).
56. N. G. Skene *et al.*, Genetic identification of brain cell types underlying schizophrenia. *Nat Genet* **50**, 825-833 (2018).
57. J. Amiel *et al.*, Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. *Am J Hum Genet* **80**, 988-993 (2007).
58. A. J. Harrington *et al.*, MEF2C regulates cortical inhibitory and excitatory synapses and behaviors relevant to neurodevelopmental disorders. *Elife* **5**, (2016).
59. C. Zweier *et al.*, Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). *Am J Hum Genet* **80**, 994-1001 (2007).
60. M. Zweier *et al.*, Mutations in MEF2C from the 5q14.3q15 microdeletion syndrome region are a frequent cause of severe mental retardation and diminish MECP2 and CDKL5 expression. *Hum Mutat* **31**, 722-733 (2010).
61. D. H. Ebert, M. E. Greenberg, Activity-dependent neuronal signalling and autism spectrum disorder. *Nature* **493**, 327-337 (2013).
62. A. J. Kennedy *et al.*, Tcf4 Regulates Synaptic Plasticity, DNA Methylation, and Memory Function. *Cell Rep* **16**, 2666-2685 (2016).
63. N. N. Parikshak *et al.*, Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008-1021 (2013).
64. U. Schmidt-Edelkraut, G. Daniel, A. Hoffmann, D. Spengler, Zac1 regulates cell cycle arrest in neuronal progenitors via Tcf4. *Mol Cell Biol* **34**, 1020-1030 (2014).

65. A. J. Willsey *et al.*, Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997-1007 (2013).
66. S. J. Sanders *et al.*, Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215-1233 (2015).
67. Deciphering Developmental Disorders Study, Prevalence and architecture of de novo mutations in developmental disorders. *Nature* **542**, 433-438 (2017).
68. X. Xu, A. B. Wells, D. R. O'Brien, A. Nehorai, J. D. Dougherty, Cell type-specific expression analysis to identify putative cellular mechanisms for neurogenetic disorders. *J Neurosci* **34**, 1420-1431 (2014).
69. M. J. Gandal *et al.*, Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* **359**, 693-697 (2018).
70. A. T. Kalinka *et al.*, Gene expression divergence recapitulates the developmental hourglass model. *Nature* **468**, 811-814 (2010).
71. T. Domazet-Loso, D. Tautz, A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* **468**, 815-818 (2010).
72. P. R. Huttenlocher, Synaptic density in human frontal cortex - developmental changes and effects of aging. *Brain Res* **163**, 195-205 (1979).
73. K. Brodmann, *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. (Barth, 1909).
74. M. C. Kiessling *et al.*, Cerebellar granule cells are generated postnatally in humans. *Brain Struct Funct* **219**, 1271-1286 (2014).
75. S. Gulsuner *et al.*, Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154**, 518-529 (2013).
76. X. Caubit *et al.*, TSHZ3 deletion causes an autism syndrome and defects in cortical projection neurons. *Nat Genet* **48**, 1359-1369 (2016).
77. K. Y. Kwan, Transcriptional dysregulation of neocortical circuit assembly in ASD. *Int Rev Neurobiol* **113**, 167-205 (2013).
78. A. N. Lamb *et al.*, Haploinsufficiency of SOX5 at 12p12.1 is associated with developmental delays with prominent language delay, behavior problems, and mild dysmorphic features. *Hum Mutat* **33**, 728-740 (2012).
79. L. Jiang *et al.*, Synthetic spike-in standards for RNA-seq experiments. *Genome research* **21**, 1543-1551 (2011).
80. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57 (2012).
81. L. Habegger *et al.*, RSEQtools: a modular framework to analyze RNA-Seq data using compact, anonymized data summaries. *Bioinformatics* **27**, 281-283 (2010).
82. L. R. Meyer *et al.*, The UCSC Genome Browser database: extensions and updates 2013. *Nucleic acids research* **41**, D64-D69 (2012).
83. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
84. J. Harrow *et al.*, GENCODE: the reference human genome annotation for The ENCODE Project. *Genome research* **22**, 1760-1774 (2012).
85. A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, B. Wold, Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods* **5**, 621 (2008).
86. H. Li *et al.*, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

87. S. Anders, P. T. Pyl, W. Huber, HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).
88. S. Anders, W. Huber, Differential expression analysis for sequence count data. *Genome Biol* **11**, R106 (2010).
89. K. D. Hansen, R. A. Irizarry, Z. Wu, Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* **13**, 204-216 (2012).
90. W. E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127 (2007).
91. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300 (1995).
92. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* **9**, 559 (2008).
93. D. W. Huang, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44 (2008).
94. M. Ashburner *et al.*, Gene Ontology: tool for the unification of biology. *Nature genetics* **25**, 25 (2000).
95. S. Schafer *et al.*, Alternative Splicing Signatures in RNA-seq Data: Percent Spliced in (PSI). *Curr Protoc Hum Genet* **87**, 11 16 11-14 (2015).
96. S. Anders, A. Reyes, W. Huber, Detecting differential usage of exons from RNA-seq data. *Genome research* **22**, 2008-2017 (2012).
97. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
98. M. Mele *et al.*, Human genomics. The human transcriptome across tissues and individuals. *Science* **348**, 660-665 (2015).
99. S. Andrews, FastQC: a quality control tool for high throughput sequence data. *http://www.bioinformatics.babraham.ac.uk/projects/fastqc*, (2010).
100. M. R. Friedländer, S. D. Mackowiak, N. Li, W. Chen, N. Rajewsky, miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic acids research* **40**, 37-52 (2011).
101. A. Kozomara, S. Griffiths-Jones, miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic acids research* **39**, D152-D157 (2010).
102. E. P. Nawrocki *et al.*, Rfam 12.0: updates to the RNA families database. *Nucleic acids research* **43**, D130-D137 (2014).
103. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2013).
104. S. Zhao, Y. Guo, Q. Sheng, Y. Shyr, Advanced heat map and clustering analysis using heatmap3. *BioMed research international* **2014**, (2014).
105. L. v. d. Maaten, G. Hinton, Visualizing data using t-SNE. *Journal of machine learning research* **9**, 2579-2605 (2008).
106. R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, A. Regev, Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495-502 (2015).
107. A.-C. Villani *et al.*, Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, eaah4573 (2017).

108. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
109. A. M. Newman *et al.*, Robust enumeration of cell subsets from tissue expression profiles. *Nature methods* **12**, 453 (2015).
110. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420 (2018).
111. I. Efroni, P. L. Ip, T. Nawy, A. Mello, K. D. Birnbaum, Quantification of cell identity from single-cell gene expression profiles. *Genome Biol* **16**, 9 (2015).
112. E. Eisenberg, E. Y. Levanon, Human housekeeping genes, revisited. *Trends in Genetics* **29**, 569-574 (2013).
113. M. Bibikova *et al.*, High-throughput DNA methylation profiling using universal bead arrays. *Genome research* **16**, 383-393 (2006).
114. A. Kozlenkov *et al.*, Differences in DNA methylation between human neuronal and glial cells are concentrated in enhancers and non-CpG sites. *Nucleic Acids Res* **42**, 109-127 (2014).
115. H. Spiers *et al.*, Methylomic trajectories across human fetal brain development. *Genome Res* **25**, 338-352 (2015).
116. B. S. Pedersen, D. A. Schwartz, I. V. Yang, K. J. Kechris, Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. *Bioinformatics* **28**, 2986-2988 (2012).
117. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357 (2012).
118. Y. Zhang *et al.*, Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137 (2008).
119. C. Zang *et al.*, A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**, 1952-1958 (2009).
120. H. Li. (GitHub, https://github.com/lh3/seqtk, 2012).
121. N. Joshi, J. Fass. (https://github.com/najoshi/sickle., 2011).
122. Y. Liao, G. K. Smyth, W. Shi, The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic acids research* **41**, e108-e108 (2013).
123. S. Purcell *et al.*, PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559-575 (2007).
124. P. Flicek *et al.*, Ensembl 2014. *Nucleic acids research* **42**, D749-D755 (2013).