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]*

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]Program in Integrative Cell Signaling and Neurobiology of Metabolism, Department of Comparative Medicine, 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 the 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 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 brain 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 brain 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 brain, 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 postnatal years (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, which we hereafter refer to as cell types. 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, S21 and table S8), as well as closely related groups of cell types (i.e., markers enriched in all prenatal excitatory neuron clusters) (fig. S21).

We complemented these data with snRNA-seq from adult human DFC (fig. S20), where we identified 29 transcriptomically distinct cell types 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). 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. S22). 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. S22). 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. S23) and inhibitory (fig. S24) 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 cortex (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. S23, S25). 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. S25). 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. S26). 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. S26, S29).

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. S27, S28). 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. S30 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. S30 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. S31), 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. S31 to S33). 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. S31). 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. S32), 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. S34**,** (*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. S34). 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. S34). 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. S34). 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. S35 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 ($P<0.0118$) (Fig. 5C, fig. S36 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<0.00287) (Fig. 5C, fig. S36 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. S36 and table S12) were enriched among the 13 modules where temporal (P<0.000168), 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. S36) 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<0.00291) (Fig. 5C) and depleted among spatiotemporal modules (P<0.00212). 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). 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. S37, S38); in cases where an autosomal module was sex-DEX throughout development, the sex exhibiting higher expression reversed between early and late postnatal development (fig. S38). This observation was upheld when multiple thresholds were used for the identification of sexual dimorphism (fig. S39). Similarly, we identified no autosomal genes that exhibited sexual dimorphism throughout development in all brain regions or neocortical areas (figs. S37, S38).

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, 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. S40, (*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. S40), 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 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. S41). 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 disease risk list converged on 7 of 73 co-expression modules (Fig. 7A), with genes associated with IQ and neuroticism enriched in 7 additional modules (Fig. 7A, fig. S36 and table S11). Of these modules, three of the four modules enriched for early onset disorders (SCZ and MDD) exhibited spatiotemporal specificity, high expression during prenatal ages, and an enrichment for neurons or neural progenitor cells. Conversely, three of the five modules enriched for late onset disorders (PD and AD) exhibited temporal, but not spatiotemporal, specificity, as well as higher adult expression and an enrichment for oligodendrocytes and/or astrocytes, among other cell types. Of particular interest were modules ME3 and ME7, which were both enriched for glia, non-NUM sites, adult-active enhancers, sex-DEX genes, and AD-associated risk genes. A fifth module, ME58, was enriched for PD, but not for any cell type, enhancer, or methylation site indicative of cell type or temporal specificity.

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 (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*, *GRAP2*, *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. S42).

As the ASD GWAS resulted in only 12 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 groups of genes were also significantly enriched in ME37 (FDR<0.0001), 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 (*4*). 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 beginning or active 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 linked 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.

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 $P<0.05$.

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 7 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 2. 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.

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, 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

References

1. D. H. Geschwind, P. Rakic, *Neuron* **80**, 633-647 (2013).
2. E. S. Lein, T. G. Belgard, M. Hawrylycz, Z. Molnar, *Annu Rev Neurosci* **40**, 629-652 (2017).
3. J. H. Lui, D. V. Hansen, A. R. Kriegstein, *Cell* **146**, 18-36 (2011).
4. J. C. Silbereis, S. Pochareddy, Y. Zhu, M. Li, N. Sestan, *Neuron* **89**, 248-268 (2016).
5. B. I. Bae, D. Jayaraman, C. A. Walsh, *Dev Cell* **32**, 423-434 (2015).
6. S. A. McCarroll, S. E. Hyman, *Neuron* **80**, 578-587 (2013).
7. T. Paus, M. Keshavan, J. N. Giedd, *Nat Rev Neurosci* **9**, 947-957 (2008).
8. J. M. Keil, A. Qalieh, K. Y. Kwan, *J Neurosci*, (2018).
9. S. Darmanis *et al.*, *Proc Natl Acad Sci U S A* **112**, 7285-7290 (2015).
10. M. B. Johnson *et al.*, *Neuron* **62**, 494-509 (2009).
11. J. A. Miller *et al.*, *Nature* **508**, 199-206 (2014).
12. T. J. Nowakowski *et al.*, *Science* **358**, 1318-1323 (2017).
13. L. de la Torre-Ubieta *et al.*, *Cell* **172**, 289-304 e218 (2018).
14. R. N. Doan *et al.*, *Cell* **167**, 341-354 e312 (2016).
15. M. Florio *et al.*, *Science* **347**, 1465-1470 (2015).
16. M. B. Johnson *et al.*, *Nat Neurosci* **18**, 637-646 (2015).
17. R. Lister *et al.*, *Science* **341**, 1237905 (2013).
18. C. Colantuoni *et al.*, *Nature* **478**, 519-523 (2011).
19. H. J. Kang *et al.*, *Nature* **478**, 483-489 (2011).
20. M. Pletikos *et al.*, *Neuron* **81**, 321-332 (2014).

21. S. K. Reilly *et al.*, *Science* **347**, 1155-1159 (2015).

22. Materials and Methods.

23. Z. Lin *et al.*, *Proc Natl Acad Sci U S A* **113**, 14662-14667 (2016).

24. G. TEx Consortium, *Science* **348**, 648-660 (2015).

25. B. B. Lake *et al.*, *Nat Biotechnol* **36**, 70-80 (2018).

26. S. Zhong *et al.*, *Nature* **555**, 524-528 (2018).

27. B. B. Lake *et al.*, *Science* **352**, 1586-1590 (2016).

28. D. Mi *et al.*, *Science*, (2018).

29. S. Lodato, P. Arlotta, *Annu Rev Cell Dev Biol* **31**, 699-720 (2015).

30. A. S. Nord, K. Pattabiraman, A. Visel, J. L. Rubenstein, *Neuron* **85**, 27-47 (2015).

31. M. Shibata, F. O. Gulden, N. Sestan, *Trends Genet* **31**, 77-87 (2015).

32. G. D'Arcangelo *et al.*, *Nature* **374**, 719-723 (1995).

33. J. J. Rowell, A. K. Mallik, J. Dugas-Ford, C. W. Ragsdale, *J Comp Neurol* **518**, 3272-3289 (2010).

34. Y. Zhu *et al.*, *Submitted to Science*, (2018).

35. D. J. Miller *et al.*, *Proc Natl Acad Sci U S A* **109**, 16480-16485 (2012).

36. P. R. Huttenlocher, A. S. Dabholkar, *J Comp Neurol* **387**, 167-178 (1997).

37. A. F. Pardinas *et al.*, *Nat Genet* **50**, 381-389 (2018).

38. J. C. Lambert *et al.*, *Nat Genet* **45**, 1452-1458 (2013).

39. M. A. Nalls *et al.*, *Nat Genet* **46**, 989-993 (2014).

40. J. Grove *et al.*, *bioRxiv*, (2017).

41. D. Demontis *et al.*, *bioRxiv*, (2017).

42. N. R. Wray *et al.*, *Nat Genet* **50**, 668-681 (2018).

43. Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium, *Cell* **173**, 1705-1715 e1716 (2018).

44. J. E. Savage *et al.*, *Nat Genet* **50**, 912-919 (2018).

45. S. Sniekers *et al.*, *Nat Genet* **49**, 1107-1112 (2017).

46. A. R. Wood *et al.*, *Nat Genet* **46**, 1173-1186 (2014).

47. J. Z. Liu *et al.*, *Nat Genet* **47**, 979-986 (2015).

48. C. J. Willer *et al.*, *Nat Genet* **45**, 1274-1283 (2013).

49. E. Wheeler *et al.*, *PLoS Med* **14**, e1002383 (2017).

50. Genomes Project Consortium *et al.*, *Nature* **467**, 1061-1073 (2010).

51. P. M. Giusti-Rodriguez, P. F. Sullivan, *bioRxiv*, (2018).

52. D. Wang *et al.*, *Submitted to Science*, (2018).

53. H. Won *et al.*, *Nature* **538**, 523-527 (2016).

54. J. P. Hamilton, M. Siemer, I. H. Gotlib, *Mol Psychiatry* **13**, 993-1000 (2008).

55. M. J. Rubinow *et al.*, *Brain Struct Funct* **221**, 171-184 (2016).

56. N. G. Skene *et al.*, *Nat Genet* **50**, 825-833 (2018).

57. J. Amiel *et al.*, *Am J Hum Genet* **80**, 988-993 (2007).

58. A. J. Harrington *et al.*, *Elife* **5**, (2016).

59. C. Zweier *et al.*, *Am J Hum Genet* **80**, 994-1001 (2007).

60. M. Zweier *et al.*, *Hum Mutat* **31**, 722-733 (2010).

61. D. H. Ebert, M. E. Greenberg, *Nature* **493**, 327-337 (2013).

62. A. J. Kennedy *et al.*, *Cell Rep* **16**, 2666-2685 (2016).

63. N. N. Parikshak *et al.*, *Cell* **155**, 1008-1021 (2013).

64. U. Schmidt-Edelkraut, G. Daniel, A. Hoffmann, D. Spengler, *Mol Cell Biol* **34**, 1020-1030 (2014).
65. A. J. Willsey *et al.*, *Cell* **155**, 997-1007 (2013).
66. S. J. Sanders *et al.*, *Neuron* **87**, 1215-1233 (2015).
67. Deciphering Developmental Disorders Study, *Nature* **542**, 433-438 (2017).
68. M. V. Fuccillo, *Front Neurosci* **10**, 27 (2016).
69. M. J. Gandal *et al.*, *Science* **359**, 693-697 (2018).
70. A. T. Kalinka *et al.*, *Nature* **468**, 811-814 (2010).
71. T. Domazet-Loso, D. Tautz, *Nature* **468**, 815-818 (2010).
72. P. R. Huttenlocher, *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.*, *Brain Struct Funct* **219**, 1271-1286 (2014).
75. S. Gulsuner *et al.*, *Cell* **154**, 518-529 (2013).
76. X. Caubit *et al.*, *Nat Genet* **48**, 1359-1369 (2016).
77. K. Y. Kwan, *Int Rev Neurobiol* **113**, 167-205 (2013).
78. A. N. Lamb *et al.*, *Hum Mutat* **33**, 728-740 (2012).
79. L. Jiang *et al.*, *Genome research* **21**, 1543-1551 (2011).
80. ENCODE Project Consortium, *Nature* **489**, 57 (2012).
81. L. Habegger *et al.*, *Bioinformatics* **27**, 281-283 (2010).
82. L. R. Meyer *et al.*, *Nucleic acids research* **41**, D64-D69 (2012).
83. A. Dobin *et al.*, *Bioinformatics* **29**, 15-21 (2013).
84. J. Harrow *et al.*, *Genome research* **22**, 1760-1774 (2012).

85. A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, B. Wold, *Nature methods* **5**, 621 (2008).

86. H. Li *et al.*, *Bioinformatics* **25**, 2078-2079 (2009).

87. S. Anders, P. T. Pyl, W. Huber, *Bioinformatics* **31**, 166-169 (2015).

88. S. Anders, W. Huber, *Genome Biol* **11**, R106 (2010).

89. K. D. Hansen, R. A. Irizarry, Z. Wu, *Biostatistics* **13**, 204-216 (2012).

90. W. E. Johnson, C. Li, A. Rabinovic, *Biostatistics* **8**, 118-127 (2007).

91. Y. Benjamini, Y. Hochberg, *Journal of the royal statistical society. Series B (Methodological)*, 289-300 (1995).

92. P. Langfelder, S. Horvath, *BMC bioinformatics* **9**, 559 (2008).

93. D. W. Huang, B. T. Sherman, R. A. Lempicki, *Nature protocols* **4**, 44 (2008).

94. M. Ashburner *et al.*, *Nature genetics* **25**, 25 (2000).

95. S. Schafer *et al.*, *Curr Protoc Hum Genet* **87**, 11 16 11-14 (2015).

96. S. Anders, A. Reyes, W. Huber, *Genome research* **22**, 2008-2017 (2012).

97. A. R. Quinlan, I. M. Hall, *Bioinformatics* **26**, 841-842 (2010).

98. M. Mele *et al.*, *Science* **348**, 660-665 (2015).

99. S. Andrews, *http://www.bioinformatics.babraham.ac.uk/projects/fastqc*, (2010).

100. M. R. Friedländer, S. D. Mackowiak, N. Li, W. Chen, N. Rajewsky, *Nucleic acids research* **40**, 37-52 (2011).

101. A. Kozomara, S. Griffiths-Jones, *Nucleic acids research* **39**, D152-D157 (2010).

102. E. P. Nawrocki *et al.*, *Nucleic acids research* **43**, D130-D137 (2014).

103. Y. Liao, G. K. Smyth, W. Shi, *Bioinformatics* **30**, 923-930 (2013).

104. S. Zhao, Y. Guo, Q. Sheng, Y. Shyr, *BioMed research international* **2014**, (2014).

105. L. v. d. Maaten, G. Hinton, *Journal of machine learning research* **9**, 2579-2605 (2008).

106. R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, A. Regev, *Nat Biotechnol* **33**, 495-502 (2015).

107. A.-C. Villani *et al.*, *Science* **356**, eaah4573 (2017).

108. M. D. Robinson, D. J. McCarthy, G. K. Smyth, *Bioinformatics* **26**, 139-140 (2010).

109. A. M. Newman *et al.*, *Nature methods* **12**, 453 (2015).

110. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, *Nat Biotechnol* **36**, 411-420 (2018).

111. I. Efroni, P. L. Ip, T. Nawy, A. Mello, K. D. Birnbaum, *Genome Biol* **16**, 9 (2015).

112. E. Eisenberg, E. Y. Levanon, *Trends in Genetics* **29**, 569-574 (2013).

113. M. Bibikova *et al.*, *Genome research* **16**, 383-393 (2006).

114. A. Kozlenkov *et al.*, *Nucleic Acids Res* **42**, 109-127 (2014).

115. H. Spiers *et al.*, *Genome Res* **25**, 338-352 (2015).

116. B. S. Pedersen, D. A. Schwartz, I. V. Yang, K. J. Kechris, *Bioinformatics* **28**, 2986-2988 (2012).

117. B. Langmead, S. L. Salzberg, *Nature methods* **9**, 357 (2012).

118. Y. Zhang *et al.*, *Genome biology* **9**, R137 (2008).

119. C. Zang *et al.*, *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, *Nucleic acids research* **41**, e108-e108 (2013).

123. S. Purcell *et al.*, *The American Journal of Human Genetics* **81**, 559-575 (2007).

124. P. Flicek *et al.*, *Nucleic acids research* **42**, D749-D755 (2013).

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 pertaining to this manuscript are available on www. development.psychencode.org

Supplementary materials

Materials and Methods
Figs. S1 to S42
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 $P<0.05$.

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 7 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 2. 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.