**Title:** Integrative Functional Genomic Analysis of the Developing Human Brain

**Alt**: Multimodal analyses of development and disease etiology in the human brain

**Authors:** LIST OF KEY CONTRIBUTORS …. The BrainSpan Consortium+ (Maybe remove and add if accepted), The PsychENCODE Consortium++, …. LIST OF SENIOR AUTHORS (~ 35 authors, all others to be listed at the end of the paper)

**Affiliations:**

+ The consortium authors are listed at the end of the paper, with authors who are listed above removed.

\* These authors contributed equally to this work

# Correspondence to: xxxx

**Abstract:** 125 words

To broaden our understanding of human neurodevelopment, we profiled the transcriptomic and epigenomic landscapes across brain regions and cell types for nearly the entire span of neurodevelopment. Multiple analysis revealed temporal, regional and cell type specific dynamics and a global transcriptomic and epigenomic transition during late fetal development, corresponding to decreased intra- and inter-regional differences, changes in cellular composition and maturational states. Genes associated with major neuropsychiatric disorders converged in a small number of gene co-expression modules, all displaying the greatest transcriptomic change during this late fetal transition. Modules also revealed key relationships between epigenetic regulation, gene expression, and development. Integration of these PsychENCODE data therefore revealed novel insights into brain development and psychiatric disorders and provide a foundational resource for future research.

NOTES:

**Late fetal ASD**: check if the same trend of SEX DEX is in monkey

**Introduction**

The development of the human brain is a highly complex process that unfolds over several decades and is reliant on the diversity and precise spatiotemporal regulation of the transcriptome. Additionally, there is increasingly persuasive evidence that dysregulation of the transcriptional, regulatory, and epigenetic processes underlying the spatial and temporal progression of neurodevelopment can have dire consequences for brain function (Silbereis et al., 2016; PMID: 26796689; Silbereis et al., 2016; PMID: 26924435; PMID: 25710529; PMID: 24183011; PMID: 26361314; PMID: 21068826; PMID: 1838531; PMID: 24267886). However, many of the epigenetic features governing the transcriptome of the human brain remain unknown and may be specific to defined biological contexts in humans or closely related primate species. As such, it is difficult or impossible to identify or study the full roster of regulatory elements using most common model organisms or cell culture systems (PMID: 26796689; PMID: 21729779; PMID: 25710529; PMID: 24183016; PMID: 26796689). A number of studies have explored facets of the transcriptional, epigenetic and regulatory architecture of the human brain (Johnson 2009; Khaitovich 2009; Miller 2014; Kriegstein, Walsh single cell, Noonan and Rakic Science; geschwind Nature HIC; Walsh Cell HIC??), but for reasons of scale and challenges of developmental human tissue availability, our understanding of these processes remains highly incomplete.

**Data generation and cross-modal integration**

In the present report, we describe the generation and integrated analysis of different genomic data modalities (i.e., the genotype, transcriptional signature, methylation status, histone modifications, and CTCF binding sites) generated on dissected regional tissue samples or at the single cell level from a total of 58 de-identified post-mortem brains obtained from clinically and histopathologically unremarkable donors of both sexes and various ancestries, ranging in age from 5 postconceptional weeks (pcw) to 40 years (**fig. S1**). Genotyping of donors’ brain DNA using a HumanOmni2.5-8 BeadChip revealed no obvious genomic abnormalities in any of the brain specimens and we observed that their genetic diversity varied in a random way (Supplemental Information).

For transcriptome analysis, tissue-level mRNA-seq was performed on a total of 607 high-quality, histologically verified tissue samples representing sixteen regions (11 neocortical areas), hippocampus, amygdala, striatum, mediodorsal nucleus of thalamus and cerebellar cortex) involved in higher-order cognition and behavior (**Fig. 1a, fig. S2**). These regions were systematically dissected from 41 brains ranging in age from 8 pcw to 40 postnatal years (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; see also Supplementary Information) (**Fig. S1a and fig. S1b**). Due to the limited size of prenatal samples, small RNA sequencing (smRNA-seq) was performed on 297 samples covering these 16 brain regions from 22 postnatal brains (**fig. S1b**). These tissue-level RNA-seq analyses were complemented by single-cell RNA-seq data generated from 1,074 cells that passed quality control measures and were collected from embryonic and mid-fetal fronto-parietal neocortical wall of an independent set of 9 brains ranging in age from 5 to 20 pcw (**fig. S1b**). DNA methylation data was generated for 269 postnatal samples covering the same 16 brain regions analyzed by RNA-seq (**fig. S1b, S11**). Additional epigenetic data was also generated for several histone marks (H3K3me3, H3K27ac, and H3K27me3) and CTCF binding sites, which together identify a large fraction of active enhancers, promoters, and repressors, as well as insulators. These data were generated from a subset of samples taken from dorsolateral prefrontal cortex (DFC) and CBC from midfetal, infant and adult brains (**fig. S1b**).

Datasets underwent stringent quality control and normalization procedures (**fig. S3**; Methods) to alleviate the influence of confounding effects (e.g., post-mortem interval, tissue pH, RNA integrity number, batch, 3’ bias). To optimize statistical power, we grouped specimens into 9 time windows (w1-w9) based on major developmental milestones and the temporal spacing of constituent specimens (**Fig. S1a**). In silico pooling of all mRNA-seq samples resulted in the detection, at RPKM>=1 of 79.7% (15,845/19,881) of protein-coding genes, indicating that a large number of mRNA genes are expressed in the developing and adult human brain (**fig. S4a**). 95% and 45% of all exonic regions were expressed, at RPKM>=1, in at least one sample and in every sample, respectively (**fig. S4b**). In addition, 2% of reads were from intergenic regions (**fig. S4c,d**) that may represent novel transcriptionally active regions (**EDF 5**), including unannotated 5’ and 3’ UTRs. We found that most genes expressed were temporally (67.8%; 13,485/19,881) or spatially (54.5%; 10,825/19,881) differentially expressed between at least two windows or regions, respectively, while a majority (52.2%; 10,371/19,881) were both concurrently, indicating that the vast majority of spatial differentially expressed genes (95.8%; 10,371/10,825) were also temporally differentially expressed in at least one region. smRNA-seq similarly revealed dynamic spatio-temporal expression of small RNAs (**Supplementary Information**), with 1,207 (80.6%) of 1,496 mature miRNAs (RPKM >=1) varied across both region and time (**See Supplementary Information**).

We applied a clustering and classification iterative algorithm (Lake et al., 2016) to single cell sequencing data following an initial division of the dataset based on age (i.e. fetal or embryonic) of the brain. This lead to the unbiased identification of 24 clusters (**fig. S6, S7**) representing distinct types of neural and non-neural progenitor and postmitotic cells (**fig. S8**). Our analysis yielded novel markers for specific cell types (**fig. S9**) and broad populations of similar cell types (**fig. S10**).

Methylation profiling revealed that the majority of CpG loci were either highly (37.5%; Beta value [β]>=0.8) or poorly (31.8%; ß<=0.2) methylated across all samples (**fig. S11a**), with lower levels of methylation dominant at transcription start sites (TSS) (**fig. S11b**). Like the transcriptome, differences in methylation were seen between structures and sexes, with the first principle component defined by cerebellar cortex (CBC)-specific diversity and the second principle component corresponding to sex (**fig. S11c**). Unsupervised hierarchical clustering confirmed CBC was the most distinct region, with greater similarity across NCX; V1C was the most distinct area in the NCX (**fig. S11d**). As expected, the majority of H3K4me3 marks were found proximal to TSS while a plurality of the other chromatin modifications and CTCF binding sites occupied intronic or intergenic regions (**fig. S13a**). Using stringent criteria (see Supplemental Materials), we also identified histone modifications relatively enriched in prenatal or postnatal brains independently in DFC and CBC, including 11,365 and 5,537 prenatal enriched and 13,147 and 11,788 postnatal-enriched regulatory regions in DFC and CBC, respectively. The incorporation of multiple data modalities, often generated from the same sample, allowed a unique opportunity to integrate data concerning both gene expression and epigenetic status. Commensurate with this possibility, we observed strong relationships between the epigenomic and transcriptomic datasets. TSSs that were more highly methylated in the adult corresponded to genes that were relatively less expressed (**fig. S12a**) at these ages, and vice versa, a relationship not strongly indicated for methylation at other locations in the gene body (**fig S12b**). The presence of CBC-specific H3K4me3 and H3K27ac marks in the adult human brain also correlated strongly with increased gene expression in CBC relative to DFC (**fig S13b,c**); the converse was also true. On the other hand, we observed an expected negative correlation between regionally-specific H3K27me3 activity and gene expression in the CBC and DFC (**fig S13d)**. Taken together, these data demonstrate the expected relationships between gene expression and epigenetic modifications including methylation status and putative regulatory elements (**fig S13e**).

**A late fetal transition dominates the brain transcriptome**

To gain a broad understanding of transcriptional variation across brain development, we analyzed the level of similarity of individual samples of the mRNA-seq dataset by multidimensional scaling (**Fig. 1b**). The biggest source of variation corresponded to age, and there was a clear divide between prenatal (w1-4) and postnatal samples (w6-9), with samples from late in the third trimester (w5) generally spanning this divide (**Fig. 1b, fig. S14**). Surprisingly, these late fetal samples were more similar to early postnatal samples than late midfetal samples as shown by unsupervised hierarchical clustering (**Supplemental Figure**), suggestive of a major transition that begins prior to birth and appearing to mark a global shift between early and late developmental transcriptional programs. This transition was also apparent in alternative splicing (**Fig. 1c**) and at the intra-regional level, as eleven neocortical areal transcriptomes correlated well across both early and midfetal (w1-4) and late postnatal (w6-w9) phases, but displayed a sharp decrease in correlation around late fetal development and early infancy (w5) (**Fig. 1d**). This trend was also apparent for the cerebellum (**Fig. 1d**) and the other brain regions assessed (**fig S16**).

To further analyze transcriptional variation across brain regions (i.e., inter-regional) at different time points, we applied a novel principal component analysis algorithm (AC-PCA; REF) that adjusts for inter-individual variations (REF). Across all samples, the first principle component correlated most strongly with age and subsequent components correlated primarily with region (**fig. S17**; Mingfeng’s Supplemental Figure 1). Within a given developmental window, AC-PCA further revealed clear separation of brain regions corresponding roughly to their developmental origin, with dorsal pallium-derived structures of the cerebrum (i.e., neocortex, hippocampus, and amygdala) becoming increasingly similar across prenatal development, with cerebellum and to a lesser extent thalamus the most distinct (**fig. Sx**). However, exclusion of cerebellum from AC-PCA indicated that inter-regional differences peak first during mid-fetal periods (w3-4), decreasing through the late fetal period and early infancy (w5), before rising from late infancy and throughout postnatal life (**Fig. 1f**). The same trend was observed in different regions of the developing neocortex, with greater dissimilarity across regions at early ages (**fig. Sx2, Fig. 1h**). Prefrontal areas were most distinct with a general rostrocaudal pattern of variation during mid-fetal periods. At later ages, adult patterning closely resembled the overall topography of the adult human cerebral cortex as described previously (Hawrylycz et al., 2012), with proximal areas showing the greatest transcriptional similarity and V1C showing the most distinct patterning (**Fig. 1g**). Across the entire development, the similarity between cortical regions (mean distance in the first two PCs) showed a pronounced decrease centered on the late fetal and early infancy samples of w5 (**Fig. 1h, fig. S18**).

Analysis of differential expression, both at the level of the gene and gene isoform, further supported the existence of a late fetal transition. Pairwise comparisons of gene expression across all 16 brain regions found a reduction in the number of genes differentially expressed across brain regions during window 5 as compared to other windows (fig. S19), with the majority of these differences due to a reduction in interareal variation. Similarly, the number of genes exhibiting regional differences in alternative splicing, as well as global measures of alternative splicing such as isoform entropy, is significantly higher during prenatal than postnatal ages (**fig S20**). The expression of 68 RNA-binding proteins implicated in splicing events also found a previously underinvestigated cluster of splicing-associated proteins that are expressed more highly in all brain regions at prenatal ages than at postnatal ages (**fig. S21**) while confirming earlier observations of the relative increased expression of a subset of these genes in the postnatal cerebellum.

**Single cell deconvolution of cellular complexity and dynamics**

The high variation present during prenatal ages corresponds to ages relevant for several neurodevelopmental disorders but also refractory to analysis for ethical and practical reasons inherent to the use of human prenatal tissue. To understand the cellular complexity and dynamics of the developing human brain, we aligned transcriptomes derived from our embryo and fetal-brain single-cell data (**Fig 2A-B; Supp Fig 6-8** or so; see Methods) with other single-cell datasets of mostly postnatal samples (Darmanis et al. 2015, Lake et al. 2016). Alignment (**Fig. 2C**) revealed similarities between major cell classes across development, and excitatory neuron clusters segregated into two divisions, one reflecting the predominant expression of upper layer marker genes (ExN15, ExN16) and the other representing deep layer neuronal populations (ExN11) (Supp **Fig. XX**). However, excitatory neurons in our samples co-expressed a number of markers indicative of immaturity, including DCX, as well as markers that are often mutually exclusive in mature excitatory neurons (**fig. S24**). Several clusters in the neuronal lineage also retained RNA expression of proliferative or progenitor markers along with these early neuronal markers, suggesting intermediate progenitor cells or presumptive neurons may begin to express neuronal markers prior to or immediately following terminal mitoses (**fig. S24**). Additionally, fetal excitatory neurons expressed 68 RNA-binding proteins more highly than adult neurons, as well as prenatal progenitors, astrocytes, oligodendrocytes, microglia, pericytes, and endothelial cells (**fig. S25**). Prenatal-specific populations including intermediate progenitor cells and embryonic or early fetal neurons expressing markers indicative of migrating or recently post-mitotic cells (Supp fig. 9ish) also did not align well with adult cell clusters. The maturational state of prenatal oligodendrocyte populations could also be corroborated; of the two oligodendrocyte populations we identified, the cluster expressing fewer markers for mature oligodendrocytes aligned more closely with fetal oligodendrocyte progenitor cells (OPCs) than adult oligodendrocytes (**fig2C**).

We next deconvolved tissue-level RNA-seq data using prenatal and adult (Lake et al., 2016) single cell RNA-seq data to assess changes in cell proportions and maturational states across development. As expected, neural progenitor cells comprised a progressively lesser proportion of the total cellular population (**Fig. 2D**) while the glial and endothelial cell portions rose across prenatal ages (**fig. S26**). Adult and fetal signatures of excitatory and inhibitory neurons were clearly distinct, with fetal signatures for both decreasing across the late fetal transition while adult signatures rose (**Fig. 2D**). Markers specific for fetal astrocytes similarly decreased across prenatal ages. Conversely, the expression of adult astrocytic markers (from REF) rose during late midfetal periods (Fig. 2E), as did markers for biological processes including dendrite formation, synaptogenesis, and myelination (**Fig. 2F**). The expression of RNA binding proteins also decreased across the late fetal transition, corresponding to the commensurate increase in non-neuronal populations (**Fig. 2G**) and consistent with the decreased expression of RNA-binding proteins in mature neuronal populations fig. S25). Interregional variation in the expression of genes associated with progenitor cell division, neuronal maturation, laminar specification, myelination, oligodendrocytes, and astrocytes was also reduced by late fetal periods, with neuronal processes including maturation and laminar specification subsequently exhibiting increased variation across postnatal ages (**fig. S27**).

**The late fetal transition reflects a functional divide**

Analysis of weighted gene co-expression networks (**fig. S28**) unveiled broad characteristics similar to those evident at the level of the gene. A majority of co-expression modules exhibited temporal or spatiotemporal specificity (40 of 73 modules), but only a few modules (17) representing a limited number of genes (XXX) exhibited strictly spatial (non-temporal) specificity across the brain regions analyzed (**Supplemental Table X**). Within the neocortex, eigengenes representing modules exhibiting temporal (13 modules) or spatiotemporal (27) specificity tended to show the greatest intramodular rate of change in window 2 (18 modules in NCX), followed generally by a prominent shift during late fetal ages, or beginning during late fetal ages in window 5 (15 modules in NCX) (**Fig. 3A**). These data reinforce our observations of extensive embryonic and early fetal variation followed by a late fetal transition (**fig. S29**), after which inter- and intra-regional variation is more limited.

Integration of single cell signatures, dynamic and cell type specific methylation sites, and putative enhancers with gene co-expression modules suggests that the late fetal transition represents a functional divide in the global brain transcriptome. Co-expression modules whose eigengenes show increased expression across all prenatal windows (eg. ME31, 35, 60, 65, among others) or who exhibit the greatest change in expression as a decrease across window 5 (eg. ME2, 10, 32, 37) were enriched for neuronal genes, genes associated with putative fetal-active enhancers, and/or genes associated with undermethylated sites in NeuN+ versus NeuN- cells (NUM sites) (hereafter referred to as Neuronal or “N”-Type modules in **Fig. 3A; fig. S30**). Conversely, genes associated with adult-active enhancers, methylation sites undermethylated in non-NeuN-positive cells (non-NUM sites) and glial genes (Glial or “G”-Type modules in **Fig. 3A; S30**) were co-enriched in modules where gene expression exhibited the greatest change as an increase across the late fetal transition (eg. M3, 7, 13, 56). We observed that N-type associations were enriched in modules exhibiting spatial or spatiotemporal specificity in the neocortex while G-type associations were enriched in modules where temporal, but not spatial, specificity was observed. This observation is in agreement with a recent study documenting increased spatial diversity of neuronal cell types relative to glial cell populations (Ed Lein REF).

Associations between gene expression, methylation status, and enhancer activity also occurred at the level of individual enhancers, methylation sites, and genes, with commensurate effects on specific cell types and developmental programs. Putative fetal-active and adult-active enhancers were associated with higher fetal or adult gene expression and were enriched in neuron-specific and glial-specific genes, respectively (see Supplementary Materials) (**Fig. 3B, S31**). In addition, fetal enhancers 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) (see Supplemental Materials) (**Fig. 3C, S31**). Both post-up and post-down sites were themselves depleted at transcription start sites (TSSs), enriched for NUM and non-NUM sites (**S31**), and enriched for fetal and adult enhancers, respectively (**Fig. 3B**). Post-up sites were also enriched among neuron-specific genes and glial genes while post-down sites were enriched in glial genes (**Fig. 3B**). Further suggesting a relationship between enhancer activity, methylation, and cell type, both genes associated with fetal-active enhancers as well as those associated with post-up sites (see Supplemental Materials) were enriched for gene ontology annotations related to early events in neural development, such as neurogenesis, cell differentiation, and synaptic transmission, but not for processes generally occurring later in development including glial cell development, oligodendrocyte development, and myelination (**Fig. 3B,C, S31**). In contrast, genes near adult-active enhancers and post-down sites exhibited a largely complementary pattern of enrichments (**Fig. 3B,C, S31**).

We observed no sex-differential expression across the full developmental time course for any autosomal gene (**fig. S32**), but analysis of sexual dimorphism also suggested the functional significance of the late fetal transition. While enrichment for sexually dimorphic patterns of gene expression were depleted among modules exhibiting spatiotemporal specificity in the neocortex (**Fig. 3A, S29**), dimorphism was present across the late fetal transition for several modules (**Fig. 3D**). These modules were enriched for genes more highly expressed in males than in females as well as for non-neuronal genes including glia (M3), oligodendrocytes (M3, M7), endothelial cells (M3, M50), and microglia (M3, M33, and M68), with the most robust sexually dimorphic expression present in microglial modules (123 of 128 genes in M33 and 54 of 55 genes in M68) (**Fig. 3D, S31**). Further suggestive of a relationship between glia or microglia and sexual dimorphism, several of these male-differentially expressed modules were also enriched for G-type associations; no modules enriched for Type N Associations were sexually dimorphic. Interestingly, several collagen genes also exhibited male-high expression across the late fetal transition (**Fig. 3D**).

**Cellular convergence of disease risk**

Loci implicated in several neurodevelopmental disorders have been identified through genome-wide association studies (GWAS) and are enriched in putative noncoding regulatory elements (PMID: 25544106). Unfortunately, functional annotation of loci associated with disease risk is confounded by uncertainty concerning both the spatiotemporal activity of regulatory elements encompassing those loci and the cell types where genes regulated by putative regulatory elements are expressed. We sought to use our dataset to determine whether the cis-regulatory elements identified in this study were implicated in neurological disorders including Parkinson’s disease, bipolar disorder, and schizophrenia, and to identify those cell types in the prenatal neocortex where regulatory element activity and gene expression converge.

We tested the most updated GWAS signals for Parkinson’s disease, bipolar disorder, and schizophrenia and found them enriched in our identified cis-regulatory elements including putative promoters and enhancers, identified by the presence of H3K27ac marks in conjunction with either H3K4me3 marks or reported transcription start sites (**Fig. 4**). Loci implicated in schizophrenia were enriched primarily among putative promoters while loci implicated in Parkinson’s disease exhibited the strongest enrichments among putative enhancers, with little evidence for spatiotemporal specificity. Conversely, while loci implicated in bipolar disorder were, like schizophrenia, enriched primarily among putative promoters across multiple developmental ages, the greatest enrichment was present in adult samples. Significant associations were not wholly specific to brain-related disorders, as single nucleotide polymorphisms associated with non-neuronal disorders including type 2 diabetes, coronary artery disease, and asthma were also enriched sporadically among putative brain-active promoters and enhancers identified in this study. However, these enrichments were present among regulatory elements active across multiple ages or brain regions and were no longer evident when our analyses excluded regulatory regions from non-brain/neuronal sources in Epigenome Roadmap (ref) (**Fig. 4A,B; fig. S33, Supplementary Materials**).

From these GWAS SNPs, we next identified disease risk-associated genes, defined here as those genes functionally connected to an eQTL present in peak regions of prenatal H3K27ac activity also encompassing a disease-related SNP (**Fig. 4C**). We supplemented these lists with a list of high-confidence ASD risk genes (Sanders et al., Neuron) and observed nearly no overlap between each list (**fig. S34**). Disease risk-associated genes converged on a small number of modules, including modules 2 (ASD, BP, SCZ), 3 (PD), 6 (PD), 7 (BP), 9 (ASD, PD), 12 (SCZ), and 37 (ASD) (**Fig. 4D**) (p<= 0.001). All of these modules exhibited temporal specificity in the neocortex (**Fig. 4A; 4D**), with all but two modules (6, 12) exhibiting the greatest rate of change across the late fetal transition. Expression of genes contained in modules associated with ASD and schizophrenia were enriched exclusively in single-cell clusters associated with neuronal lineages, but non-neuronal enrichment was found in modules enriched for Parkinson’s disease and bipolar disorder associated genes (**Fig. 4D**).

Modules associated with Parkinson’s Disease (3) and bipolar disorder (7) were enriched for male-differentially expressed genes. However, modules associated with ASD (2, 9, 37) and SCZ (2, 12), two diseases with pronounced male biases in the age of onset, progression, and incidence, did not exhibit sexual dimorphism (**Fig. 4A**). The single cells neuronal clusters for which genes in these modules were enriched (IntN-10, ExN-15, ExN-16) also exhibited no male bias in sexual dimorphism (**Fig. 4D**), but an astrocytic cluster (Glia-19) enriched for the expression of ASD risk genes was also enriched for male-biased genes. Similarly, male-biased genes identified in this study were enriched in gene co-expression modules derived from postmortem ASD brain (Voineagu, 2011 PMID: 21614001; Gupta et al., 2014, PMID: 25494366), when those modules were also enriched for microglial and astrocytic, but not neuronal, genes (**Fig. 4E**); modules enriched for neuronal markers (Voineagu, 2011 PMID: 21614001; Gupta et al, 2014, PMID: 25494366) were depleted for sex-differentially expressed genes identified in our study, as were several other gene sets linked to ASD including FMRP binding partners (PMID: 21784246) and CHD8 regulatory targets (PMID: 25752243; PMID: 25294932) (**Fig. 4E**).

**Discussion**

In this study, we have presented what is, to our knowledge, the most comprehensive multi-platform functional genomic analysis of the developing human brain to date. Data generated include information about the transcriptome, both at the tissue level and the level of the single cell and both at the gene or isoform level, and regulatory modalities including methylation and histone modification. The presence of these multiple data modalities in a unified resource, and frequently from the same tissue samples, allows the integration of information to an extent not typically possible in studies spanning prenatal and postnatal human neurodevelopment.

While transcriptional differences between distinct brain regions remain robust across development, the global brain transcriptome exhibits a sharp late midfetal-to-early infancy transition. Our accompanying analysis of the spatio-temporal transcriptome of the macaque (REF) reveals that this transition is likely conserved across the catarrhine group of primates. Underlying this transition are three interdependent phenomena. First, interregional and interareal variation is high prenatally, decreases across late fetal ages, and recovers somewhat beginning in late childhood and continuing through adulthood. Intriguingly, principal component analysis, unsupervised hierarchical clustering, multidimensional scaling, and analysis of differential expression all suggest this transition begins well prior to birth. In particular, multidimensional scaling and hierarchical clustering suggest that late fetal samples are more closely associated with postnatal samples than with other prenatal samples, a surprising discovery as the two late fetal cases (35 and 37 PCW) were closer in age to the three nearest late mid-fetal cases (91 to 112-day separation) than to the three nearest postnatal cases (129 to 143-day separation). Second, biological processes active in the human brain and common across multiple regions exhibit spatial diversity during prenatal ages and late postnatal windows but not the late fetal transition. This likely reflects spatiotemporal gradients of expression during prenatal ages and functional divisions between mature neocortical areas and brain regions in the adult. Third, we found that the prolonged development of neural, particularly neuronal, cell types in the human brain extends throughout prenatal development to the late fetal transition; by 20 PCW, many neurons in the neocortex have not yet begun expressing markers indicative of mature cell types or laminar position, have not extended processes, and have not begun to express synaptic or other markers associated with cell maturity. In addition, deconvolving tissue-level data with these single cell data demonstrated the rise of mature neuronal molecular signatures during late fetal development and into infancy (xx pcw – yy pcw). The fetal-to-late fetal transition may therefore represent an inflection point where developmental and spatiotemporal variation in human neocortical development is transiently consolidated in advance of the emergence of functional differences between adult brain regions.

Integrated analysis of the different data modalities across the late fetal transition also provides insight into neurological disease. Eigengenes for co-expression modules enriched for genes we linked to bipolar disorder and Parkinson’s disease exhibited higher expression either prenatally (ME2, BP; ME9, PD) or postnatally (ME7, BP; ME6 and ME3, PD). These modules were also enriched for either excitatory neuron (ME2, ME6, ME9) or non-neuronal (ME3, ME7) populations, suggesting the Knudson hypothesis may be relevant for these disorders (REF) with respect both to cell type and developmental time. In contrast, all co-expression modules enriched for genes we linked to ASD (ME2, 9, 37) and schizophrenia (ME2, 12) possessed broadly similar eigengenes, exhibiting higher expression during prenatal periods. These modules were also uniformly enriched for excitatory neurons, suggesting, in agreement with prior research, the importance of fetal excitatory neurons in the etiology of these disorders (REF). However, despite sex biases in the incidence, progression, and age of onset for ASD and SCZ, no ASD- or SCZ-linked module exhibited sexual dimorphism. Rather, sex-biased co-expression modules, particularly those enriched for genes more highly expressed in males than in females, tended to be enriched for glial and/or microglial genes. Moreover, co-expression modules identified from postmortem ASD brain also exhibited the enrichment or depletion of sexually dimorphic genes depending on whether the module was enriched for glia/microglia or neurons, respectively. Given the co-incident emergence of sexual dimorphism during late midfetal periods, it is possible that the etiology of ASD and/or schizophrenia may involve interactions between naturally occurring and sexually dimorphic processes involving glia or microglia with cells of a neuronal lineage.

Taken together, these observations demonstrate the utility of the BrainSpan resource for studies of neural development, function, and disease.