**Integrative Functional Genomic Analyses of Developing and Adult Human Brain**

**LIST OF AUTHORS ….** **The BrainSpan Consortium+, …. LIST OF AUTHORS**

+ 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

# Corresponding authors

**Summary**

To broaden our understanding of human neurodevelopment, the BrainSpan project profiled the genomic, transcriptomic, epigenomic, and regulatory landscapes across different brain regions and nearly the entire span of brain development through middle adulthood. Here we describe data generation and integrated analysis, expanding the known repertoire of putative functional genomic elements and yielding new insights into human neurodevelopment. We also identified highly spatiotemporal and sex-specific dynamics, including a prominent perinatal transition in the global transcriptional and epigenomic architecture. We also found that common variants significantly associated with distinct psychiatric and neurological disorders are enriched in spatially and temporally specific epigenetically modiﬁed DNA regions and putative regulatory elements, suggesting their influence on human brain development and function. This public and mineable resource therefore provides valuable opportunities for future studies of human neurodevelopment and facilitates translational research into brain disorders.

**Introduction and Data Generation**

The human brain is responsible for perception, cognition, memory, emotions, complex behavior, and much of what makes us uniquely human (Gazzaniga, 2009; PMID: 26469048).  These abilities are rooted in the development of the structurally and functionally distinct regions that comprise the brain, and this complexity is itself the result of coordinated transcriptional, regulatory and epigenetic programs that differ throughout development and across spatial components (PMID: 26796689). While the organizing principles driving the development of human brain structure and function are generally typical of all mammals, it has been increasingly recognized that there are considerable differences in the cellular and molecular architecture of the developing human brain (see specific examples in PMID: 21729779; PMID: 25710529; PMID: 24183016; PMID: 26796689). The process of building that architecture also unfolds over several decades, and many of the protracted processes underlying this extended development both remain unknown and are difficult, if not impossible, to study using current model organisms or human cell culture systems (PMID: 26796689). Additionally, there is increasingly persuasive evidence that dysregulation of the transcriptomic, regulatory, and epigenomic landscapes underlying the temporal and spatial progression of neurodevelopment can have dire consequences for brain development and 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). Therefore, systematic and integrative analyses of the transcriptional, epigenetic, and regulatory features spanning development and multiple brain regions are essential for understanding human nervous system development, evolution, function, and disease.

To expand on previous relevant studies (PMID: 19477152; PMID: 22031440; PMID: 22031444; PMID: 24695229; PMID: 26619358; PMID: 25501035) and broaden our understanding of human neurodevelopment, the BrainSpan consortium performed systematic, comprehensive, and integrated surveys of the transcriptome, DNA methylation, key histone modifications, and the genome across multiple regions and nearly the complete timespan of human brain development through young to middle adulthood in both sexes (Fig. 1, Extended Data Figure 1). These data are complemented by cellular resolution *in situ* hybridization (ISH) data of targeted brain regions produced using an industrial scale histology platform customized to work with large-format postmortem human brain samples (*www.brainspan.org/ish*). In addition, neuroanatomical context for these data comes from *de novo* high-resolution web-based reference atlases of post-mortem human brains including supporting histology, magnetic resonance imaging and diffusion weighted imaging (*www.brainspan.org/static/atlas*). All of these data are publicly accessible through an integrated online resource (*www.brainspan.org*) and complemented by tools for data mining and visualization.

In the present study, we report on this resource and its utility through the presentation of noteworthy insights concerning global dynamics of the human brain transcriptome, sex differential gene expression, and evolution. In addition, we demonstrate how integration across the various data modalities contained within the BrainSpan resource can lead to insights into function and neurological disease.

**Overview**

We conducted mRNA-seq on 607 samples taken from 16 brain regions across nearly the complete timespan of human brain development (Figure 1a), with brain regions including 11 putative areas of the cerebral neocortex (NCX) and 5 non-neocortical regions [hippocampus (HIP), amygdala (AMY), striatum (STR), mediodorsal nucleus of the thalamus (MD), and cerebellar cortex (CBC)]. NCX areas corresponded to topographically defined and putatively functionally distinct areas of the orbital (OFC), dorsolateral (DFC), ventrolateral (VFC), medial (MFC), and primary motor (M1C) cortices of the frontal lobe, the primary somatosensory (S1C) and posterior inferior (IPC) cortices of the parietal lobe, the primary auditory (A1C), posterior superior (STC), and anterior inferior (ITC) cortices of the temporal lobe, and the primary visual (V1C) cortex of the occipital lobe (Figure 1b, Extended Data Figure 1). In addition, we generated smallRNA-seq data from those same regions for 297 postnatal samples, methylation data for 269 postnatal samples covering these 16 brain regions, and histone modification data, including H3K4me3, H3K27ac, and H3K27me3, from a subset of samples taken from DFC and CBC from both midfetal development and young adulthood.

Additional information on data generation and basic characterization can be found in the Supplemental Materials as well as Extended Data Figures 2 (Transcriptome), 3 (Methylation), and 4 (Histone modifications and CTCF binding).

**Integrated data analysis and key neurobiological insights**

**A perinatal transition marks a functional divide in the global brain transcriptome**

Analysis of RNA-seq data found that the largest variances in the transcriptome correspond to age followed by region. Multidimensional scaling of all 607 mRNA-seq samples (Figure 2a) found clear groupings of prenatal and postnatal samples, with late fetal samples occupying a position between these groupings (Figure 2b). This analysis also suggested that any two analyzed brain regions or presumptive neocortical areas are more similar to one another within either the early and mid-fetal phase or the postnatal phase than they are to themselves across these two broad developmental phases (Extended Data Figure 5a). Consistent with this, unsupervised hierarchical clustering found that late fetal samples clustered with all postnatal samples rather than with the early and mid-fetal samples (Extended Data Figure 5b).

Based on these results and to optimize both the temporal resolution and statistical suitability of our dataset, we subdivided the available samples into nine developmentally-relevant windows (w1-w9; Figure 2a/b), with the apparent perinatal transition contained within window 5 (24 pcw – 6 postnatal months). We then applied a novel principal component analysis (PCA) algorithm (AC-PCA) that adjusts for individual variation (Figure 2c and Supplementary Information) and calculated the mean distance separating each brain region in the first two principle components for each window (Figure 2d). AC-PCA revealed clear separation of brain regions corresponding to their topographic proximity and functional specialization (Figure 2c), with the greatest interregional differences present between the postnatal CBC and other postnatal regions (Figure 2d). When CBC is not included in the distance calculation, the transcriptional differences among NCX, HIP, AMY, STR, and MD peak during mid-fetal periods (w3-4) before decreasing through the perinatal period (w5), followed by a slight rise in interregional differences occurring from late infancy and early childhood (w6) onwards through postnatal life (Figure 2d). This arrangement again suggests a transition in the global transcriptional landscape occurring during perinatal development and early infancy (w5).

This perinatal transition was also apparent when the analysis was repeated for the 11 neocortical areas included in this study (Figure 2e). The clustering of prefrontal areas was most distinct from other areas during mid-fetal periods, with the distance between clusters much reduced by the perinatal phase and early infancy. Areal differences between prefrontal cortex and other neocortical areas then re-emerged through adulthood, though VFC clustered closely with M1C and S1C, likely reflecting its functional closeness with orofacial regions of the motor and somatosensory perisylvian cortex.

Both this “hourglass” pattern (PMID: 24373884; Figure 2f) as well as the perinatal transition were also seen in the distribution of genes exhibiting differential expression (DEX) either among brain regions/neocortical areas in each window (spatial DEX) or between developmental windows within the same region/area (temporal DEX). Spatial DEX genes were most prevalent during mid-fetal periods (w2-4), childhood (w7), and adulthood (w9) (Figure 2g), while all brain regions or NCX areas were substantially more reliant on temporal DEX genes than spatial DEX genes during and around the perinatal transition (w5, Extended Data Figure 6). Similarly, intraregional transcriptomes correlated well across both the early and mid-fetal (w1-4) and late postnatal phases (w6-9), but there was little correlation across perinatal development and early infancy (w5) (Figure 2h).

We next sought to determine whether the perinatal transition extended to the network level. Using weighted gene co-expression network analysis, we identified 73 modules of genes exhibiting similar expression trajectories across both brain regions and time (Extended Data Figure 7). We quantified the spatiotemporal specificity of these modules by computing distance vectors separating region-specific eigengene values for every region in each window in all modules. Like gene expression, network dynamics were most discordant between the perinatal window (w5) and both earlier (w1-w4) and later (w6-w9) windows (Figure 2i). In addition, the greatest changes in spatiotemporal specificity occurred during or prior to early infancy for all regions except the CBC, where a substantial fraction of modules continued to exhibit their greatest differences in specificity across early childhood (w5-6; Figure 2j).

The perinatal transition we observe in analyses of the global transcriptome (Figure 2b-f) and at the level of both the gene (Figure 2g, h) and the network (Figure 2i, j) roughly corresponds to the shift from neurogenesis and neuronal migration to gliogenesis and the establishment of synaptic connectivity (Silbereis, Pochareddy et al., 2016). To better assess this possibility, we generated cell type specific transcriptional profiles based on data from a number of recent studies that have purified and transcriptionally profiled individual cell types from the mouse brain (Extended Data Fig. 8) (PMID: 25186741). As promoter methylation may be associated with decreased expression of cell-type specific genes, we also assessed the methylation of sites that have been previously reported as either undermethylated in NeuN-positive cells (NeuN Under Methylated, or NUM sites) or undermethylated in NeuN-negative cells (non-NUM sites) (REF). Co-expression modules were then annotated with this information (Supplemental Tables). We found that nearly every module enriched for neuron-selective or glial-selective genes was also enriched for genes associated with NUM or non-NUM sites, respectively (See Supplemental Table). In addition, eigengenes for those modules strongly associated with specific cell types tended to also experience a pronounced spatiotemporal shift across the perinatal transition.

**Sex-different gene expression, microglia, and ASD**

The value of the many and varied samples present in the BrainSpan resource, as well as the utility of this resource for probing concerns touching on cell composition and function, are readily apparent in our analyses of sex-differential gene expression, particularly as this relates to neuropsychiatric disease. Differences between males and females include the incidence, age of onset, and presentation of many neuropsychiatric disorders including autism spectrum disorder (ASD, Fombonne 2009; PMID 1921885). Several recent studies have also linked ASD and other neuropsychiatric disorders to specific brain regions and/or cell types (Willsey 2013, PMID 24267886; Parikshak 2013, PMID 24267887; Chang 2015, PMID 25531569; Xu 2014, PMID 24453331). However, while human transcriptomic analyses have occasionally addressed sexual dimorphism across the developmental timespan of the human brain (Trabzuni 2013; PMID 24264146; Kang 2011; PMID 22031440) or looked at these dimorphisms in disease-relevant contexts (PMID: 25494366; PMID: 21614001), efforts to link sex-specific aspects of development, cell composition, and disease have been limited. With its suite of developmental, multi-regional, and multi-modal datasets, the BrainSpan resource allows a unique opportunity to address sexual dimorphism in the human brain.

Several lines of evidence derived from the BrainSpan resource suggest the sex-differential (sex-DEX) expression of genes associated with glia or microglia. Genes expressed more highly in males than in females (male-DEX) during some developmental windows were enriched for the list of microglial markers generated in both this study and an independent list (Kang 2011). Additionally, while no co-expression modules enriched for either neuron-specific gene expression or NUM sites show evidence for sexually dimorphic gene expression (see M35 example in Figure 3b), modules M3, M7, M33, and M68 show significant enrichment for male-DEX genes, male-specific methylation sites, and genes that mark the functions of glia (M3; Figure 3b), oligodendrocytes (M3, M7, Figure 3b), or microglia (M3, M33, M68; Figure 3i,k,l), respectively. Modules M3, M7, and M33 are also among those modules enriched for non-NUM sites (Supplemental Table X). The modules that are enriched specifically for microglia show particularly robust sexual dimorphism, with 123 of 128 genes (M33) and 54 of 55 genes (M68) showing higher expression levels in males as compared to females across all brain regions during late prenatal development (Figure 3b). This observation is in agreement with the reported sexually dimorphic microglial colonization of the rodent brain (PMID: 22182318), a finding not previously investigated in the human brain. Interestingly, the time during which sex-differential gene expression is present in modules 3, 7, 33, and 68 coincides with both the perinatal transition and with male-DEX expression of 16 collagen genes, including genes associated with Ehlers Danlos (*COL3A1*) and osteogenesis imperfecta (*COL1A2*) (Fig 3b).

Given the enrichment in our study of glial or microglial genes in several modules containing disproportionately high numbers of male-DEX genes, we next assessed relationships between the sex-DEX transcripts we identified and several previously reported gene sets and gene co-expression modules associated with Autism Spectrum Disorder (ASD), a condition with a striking 4:1 male bias in diagnoses (PMID: 19218885). We found that several gene sets linked to ASD, including FMRP binding partners (PMID: 21784246) and CHD8 regulatory targets (PMID: 25752243; PMID: 25294932), show a significant depletion for sex-DEX genes. Similarly, gene co-expression modules identified from postmortem ASD brain and enriched for neuronal markers are depleted for sex-DEX genes. In contrast, two co-expression modules (mod5, M16) upregulated in postmortem ASD brain and enriched for microglia (PMID: 25494366; PMID: 21614001) show a strong enrichment for male-high sex-DEX genes identified in this study. A third, related module (mod7; PMID 25494366) that intersects with M16 (PMID 21614001) is similarly associated with both glial genes and male-high sex-DEX genes. Together, these observations suggest that sexually dimorphic cell type composition or function may contribute to the highly male-biased diagnoses of ASD.

**Divergent gene expression patterns in primate and mouse**

In addition to sexual dimorphism, the breadth of brain regions and developmental ages present in the BrainSpan sample set also enables evolutionary analyses concerning not just the qualitative presence or quantitative extent of gene expression but rather the spatiotemporal dynamics of that expression across all developmental axes and the entire timespan of brain development. Here we extend upon previous efforts (Konopka et al., 2012; Liu et al., 2012; Zeng et al., 2012; Hawrylycz et al., 2015, some of which were limited by the number of species and/or brain regions considered, by comparing gene expression patterns in 16 matched brain regions in adult human, macaque monkey, and mouse.

Of the 16,369 orthologous genes shared between human and macaque, 18% (2946) have significant (Bonferroni-corrected p ≤ 0.01) regional patterning only in human (n = 6 brains) and 6% (982) show regionally distinct patterns of gene expression only in macaque (n = X brains, Figure 4a). 19% of orthologous genes (3113) are regionally patterned in both human and macaque, and 89% of these patterns are conserved (R > 0.5, see Methods) between these two species across the brain regions assayed in this study. Both the number of genes known to exhibit expression differences across regions in both primates and mouse and the degree of pattern conservation present between these species was lower, with just 59% of 862 common regionally expressed genes exhibiting a similar expression pattern between human and mouse and 59% between macaque and mouse (Figure 4b). This similarity suggests that brain expression patterns have evolved at a similar rate along these primate lineages.

We confirmed the conserved and/or divergent expression of specific genes using an independent microarray dataset (human) or *in situ* hybridization (mouse and macaque). *SLC17A7* (*VGLUT1*), a vesicular transporter commonly expressed at the synapses of glutamatergic neurons, showed consistent patterning in all species (Figure 4c-e), while *PDE1A* (Figure 4f,g) and *CRYM* (Figure 4i-k) show species-specific patterns in the CBC.

**Integrative approach to network analysis**

In addition to a broad range of samples, the BrainSpan resource facilitates the integration of data across multiple data modalities. To gain insight into the relationships that exist between these data modalities, as well as to demonstrate the validity and utility of our observations, we conducted several integrated analyses designed to highlight associations between methylation sites, enhancers, gene expression, and function.

We hypothesized that methylation sites that do not change across developmental time are unlikely to be obviously relevant for brain development while those that exhibit chaotic, seemingly random patterns of methylation are unlikely to be easily related to discrete developmental processes. We therefore identified and categorized methylation sites as consistent, progressively increasing, or progressively decreasing across prenatal and/or postnatal ages. Methylation sites in each of nine mutually exclusive categories (Figure 5a,b) were then compared against a population of methylation sites exhibiting constant, non-dynamic methylation throughout both prenatal and postnatal ages (Figure 4a). We found that while sites with constant methylation throughout the timespan of our datasets were enriched among transcription start sites, methylation sites with persistent dynamic trajectories across either prenatal or postnatal development were depleted at transcription start sites and instead enriched in intronic, exonic, and sometimes intergenic regions (Figure 5c). Methylation sites with persistent dynamic trajectories were also significantly enriched within peak areas for predicted enhancers, with sites that were progressively less methylated across postnatal time generally more enriched among adult-specific enhancers and sites that exhibited progressively increasing methylation across postnatal periods enriched among fetal enhancers. Correlations between persistent dynamic methylation sites and markers for cell type specificity, including NUM sites, non-NUM sites, and cell type specific gene expression, were also evident (Figure 5d). NUM sites were particularly enriched among sites where methylation decreased across postnatal time while non-NUM sites were most strongly enriched among sites where methylation increased across postnatal ages. Categories of methylation trajectories enriched for non-NUM sites were also enriched for associations with glial-selective genes and gene ontology terms including myelination (Figure 5f).

Gene ontological terms highlighting temporal changes in cellular composition and function were also observed following an analysis of putative fetal-, infant-, and adult- active enhancers. We correlated H3K27ac marks with the expression of either the gene containing these marks or the most proximal gene within 10kb of these marks, with all H3K27ac marks present in annotated transcription start sites (TSSs) or overlapping H3K4me3 sites removed from consideration. As expected, we found significant relationships between gene expression and temporally regulated H3K27ac marks (Fig. 5g). In addition, genes whose expression correlated strongly with H3K27ac marks present only during fetal periods or fetal/infancy (Fig. 5g) were enriched for gene ontology terms related to morphogenesis, neurogenesis, and axonogenesis (Fig. 5h, top), among others. Conversely, while genes whose expression correlated with putative enhancers active only in the adult or in both infancy and adulthood (Fig. 5g) were also enriched for ontological terms related to neurogenesis, they were uniquely enriched for terms related to gliogenesis, synaptogenesis, and myelination (Fig 5h, bottom).

**Distinct spatiotemporal patterns of disease-associated epigenetic modifications**

To further confirm the validity of the epigenomic modifications we observed, we next conducted an association analysis between putative regulatory elements and loci implicated through genome-wide association studies on each of several prevalent neuropsychiatric disorders. While the enrichment of GWAS signals in regulatory elements has been previously reported (PMID: 25544106), we show here that the signals for three major brain disorders are enriched in both developmental and adult regulatory elements in disease-specific patterns (Fig. 6a). While the significant associations between H3K27ac peaks and schizophrenia are present across regulatory landscapes at all ages assayed, the associations between putative enhancers in the brain with both bipolar disorder and Parkinsons were strongest during early adulthood. Identification of super enhancers using the ROSE algorithm (REF) led to similar patterns of enrichment, with heightened contrast in the apparent importance of putative enhancers active in the DFC and CBC for schizophrenia and Parkinsons, respectively (Fig. 6b). Several categories of progressively dynamic methylation trajectories were also enriched for SNPs previously linked to schizophrenia and bipolar disorder but not Parkinsons (Fig. 6c). Conversely, single nucleotide polymorphisms associated with type 2 diabetes, coronary artery disease, and asthma were not enriched among putative enhancers, promoters, or categories of methylation trajectories identified in this study.

We next sought to determine whether the unique spatiotemporal enrichment patterns concerning enhancer activity and methylation trajectory that we observe for each neurological disorder could be associated with specific gene ontology terms and consequently novel mechanistic understanding of these disorders. We found that genes associated with putative enhancers containing loci linked to schizophrenia were significantly enriched in co-expression module 2, a module enriched for markers of pyramidal neurons and defined by an eigengene whose expression is highest during fetal windows before decreasing sharply through the late fetal period.

**Conclusions**

In this study, we have presented what is, to our knowledge, the most comprehensive multi-platform genomic analysis of the developing human brain to date. The BrainSpan resource was generated using high-quality brain specimens of both sexes and different ancestries, standardized processing, neuropathological examination, histological validation, and stringent quality controls. Additionally, the generation of multiple data modalities, frequently from the same tissue samples, allows the ready integration of information to an extent not previously possible in any study spanning prenatal and postnatal human neurodevelopment. In accord with the singular value of this resource, pre-publication use of the BrainSpan resource has already facilitated several hundred studies on the development, evolution, function and dysfunction of the human brain (representative studies out of 441 based on Google Scholar search on March 28, 2016; PMID: 24267887; PMID: 2568891; PMID: 26590343; PMID: 25695269; PMID: 23911319; PMID: 24561062; PMID: 25607358; PMID: 25599223).

Our present analyses of the BrainSpan resource have substantially expanded the repertoire of known putative functional genomic elements with unprecedented spatiotemporal resolution by mapping transcripts, epigenetic modifications and putative regulatory elements across both sexes, multiple brain regions, and almost the entire course of human brain development. These results are consistent with previous relevant genomic studies of the developing human brain (REF) and, through the application of data concerning epigenetic and regulatory modalities present in the BrainSpan resource, provide significant and novel explanatory power to transcriptomic analyses. For example, our enhancer-centric analysis (**Fig. 5g,h**) provided a clear and coherent description of developmental and functional differences spanning the perinatal transition in excess of that possible from gene ontological analysis of the transcriptome alone. Similarly, beginning with methylation status, we were able to reproduce and expand upon a more traditional co-expression network analysis (**Fig. 5a-f**).

Furthermore, our analyses have yielded noteworthy and novel insights into human neurodevelopment and evolution. Even though transcriptional differences between distinct brain regions remain robust across development, the global brain transcriptome exhibits a highly dynamic nature including a sharp perinatal transition. Moreover, both hierarchical clustering and multi-dimensional scaling suggested that late fetal samples were more 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). Functional annotation revealed that the regional and neocortical areal perinatal transition reflect underlying cellular and functional changes between the mid-fetal and early postnatal brain architecture (see PMID: 26796689). Consistent with this, the dramatic changes in the transcriptional profile of the perinatal and early postnatal CBC coincide with the massive neurogenesis of the cerebellar granule neuron, the most numerous neuron type in the human brain (see PMID: 26796689).

Understanding how gene expression patterns differ between humans and other mammals can provide insight into human evolution and human-specific characteristics (REF). Previous studies have identified species-specific differences in temporal, regional, and laminar gene expression between either human and mouse or human and non-human primates (REF), but to date there have been few systematic attempts to evaluate regional patterns of gene expression across species and relative to evolutionary distance (REF). In this study, we found that Human and macaque monkey show strikingly greater conservation (89% of orthologous genes exhibiting regionally distinct gene expression in both species) of expression patterning across the brain than primate and mouse (59%). Surprisingly, the cerebellum, and not the neocortex, drives many of the species differences we observe, highlighting the evolutionary importance of this brain region. While some differences are likely attributable to experimental method, they are of similar magnitude to altered expression of approximately 30% of genes between cortical cell types in human and mouse (Zeng et al. 2012).

We also demonstrated the utility of the BrainSpan resource in providing new insights into disease susceptibility and mechanisms. Genes we identified as sexually dimorphic are significantly enriched in modules previously reported as upregulated in postmortem ASD brain and/or enriched for microglia and other glial cell types (PMID 25494366; PMID 21614001). In addition, this study delineates the developmental windows where sex-differential gene expression is present and describes co-expression modules in the human brain with significant enrichments for both microglia markers and sex-differential (male-higher) gene expression (PMID 26892004). Our analyses also confirm and expand on previous findings suggesting that a large portion of common variants associated with genetic risk for psychiatric and neurological disorders reside within non-coding sequences that serve as putative regulatory elements and/or regions undergoing persistent dynamic epigenetic modifications (for review see PMID: 25544106). Additionally, risk loci for schizophrenia and bipolar disease are enriched in elements exhibiting strong temporal and regional specificity, potentially indicating when and where in the development of the brain non-coding genetic variants associated with a disorder are likely to have a functional impact. Taken together, these results again demonstrate the value and utility of BrainSpan resources.

However, while the present BrainSpan genomic data resource represents the largest such resource covering human brain development, several limitations should be noted. First, it is possible that the results we report were affected by inter-individual variation and the non-regular distribution of samples, especially with regards to sex, across the timespan of the human brain (Fig. 2a). As such, the acquisition of additional samples would be clearly beneficial. Second, future analyses will require deeper spatial and cellular resolution. While recent studies have implemented genomic analyses of single cells or purified cell populations of the developing human brain (PMID: 26060301; PMID: 26406371; PMID: 25734491; PMID: 25721503; PMID: 26644564; PMID: 26687838; PMID: 26619358; PMID: 23828890) efforts thus far have included either a limited number of cells, cell types, regions, or developmental time points. Moreover, these techniques depend largely on fresh tissue samples, which may be difficult to obtain, and at present are difficult to implement on a scale comparable to that reported here. More practically, single cell RNA-seq is still financially and technically demanding, making the adaptation of this approach to the developing human brain difficult. None-the-less, we envision a future where these obstacles are overcome and these and other emerging technologies substantially improve our ability to survey the human brain with increasing spatiotemporal resolution.

**Figure Legends:**

**Figure 1: Overview of the BrainSpan resource.** (a) The developmental timespan of the human brain, from embryonic ages (≤ 8 pcw), through fetal development, infancy, childhood, adolescence, and adulthood. The human brain grows in size and complexity until roughly 3 years of age (top panel), at which time development continues on a cellular, synaptic, and molecular levels. (b) Brain regions sampled for the BrainSpan resource include eleven regions of the neocortex, three other regions of the cerebrum, and two additional regions, the thalamus and cerebellum. (c) Data modalities generated by the BrainSpan consortium, and accessible online at brainspan.org, include surveys of the transcriptome, key histone modifications, DNA methylation and the genome as well as *in situ* hybridization, histology, and high-resolution web-based reference atlases.

**Figure 2: The global transcriptional architecture of the human brain is driven by a broad perinatal transition.** (a) Distribution of samples across developmental windows. Prenatal samples are labeled in blue, postnatal samples are labeled in red, and the perinatal window, w5, is labeled in yellow. (b) Multidimensional scaling shows samples from late fetal ages and early infancy (w5) cluster with postnatal samples (w6-9) rather than with prenatal samples (w1-4). (c) Principle component analysis (PCA) for samples from all brain regions at late midfetal ages (w4), late fetal ages and early infancy (w5), and early adulthood (w9) shows that interregional differences are generally reduced across a perinatal window before expanding again postnatally. (e) Distance vectors across the first two principal components show that the cerebellum is largely responsible for increased interregional diversity in postnatal samples. The perinatal window is highlighted in yellow. (f) PCA for neocortical areas shows a perinatal transition similar to that seen for all brain regions. (g) Distance vectors for neocortical samples show a stereotypical “hourglass” pattern. (h) Gene expression within brain regions correlates across early prenatal and late postnatal windows but not across a perinatal transition. (i) Gene co-expression modules also correlate across early prenatal and late postnatal windows but not across a perinatal transition. (j) Differences in the rate at which the transcriptome changes are greatest between midfetal and late fetal/early infancy (w4/5) in the neocortex and infancy/early childhood (w5/6) in the cerebellum. (k) The interwindow periods where co-expression modules experience the greatest shift are plotted as percentages of the total number of spatiotemporally distinct modules for each region. For all regions except the CBC, the vast majority of the most major transcriptional rearrangements are complete by the w4/5 transition.

**Figure 3: Sex-different gene expression is associated with both microglia and ASD.** (a) We found that 872 protein-coding genes were expressed at significantly higher levels in at least one sex for at least one developmental window, with 430 protein-coding genes expressed more highly only in males, 420 expressed more highly only in females, and 71 expressed to a significantly greater extent in both sexes at different developmental times. However, suggesting that sex-different gene expression is tightly spatiotemporally specific, only a small number of genes on chromosome Y and the *XIST* locus on chromosome X showed consistent and robust sex-differentialdifferent expression across all brain regions and developmental windows, suggesting that sex-differential gene expression is tightly spatiotemporally specific. (b) Modules enriched for sex-DEX gene expression include several modules (M3, M7, M33, M68) also enriched for glial or microglial markers. In contrast, neither M35 nor any other module enriched for neuron-specific gene expression exhibited male-higher, sex-DEX gene expression. The sexually dimorphic expression of collagen genes mirrors that of the glial and microglial markers. The perinatal window is highlighted in yellow. (c) Several gene sets associated with autism spectrum disorder show a depletion for genes identified in the BrainSpan resource as differentially expressed between males and females, as do several modules identified in postmortem ASD samples that are enriched for neuron markers. However, several modules enriched for glial-specific gene expression and either upregulated in postmortem ASD brain or associated with a module upregulated in postmortem ASD brain are enriched for sex-DEX genes (PMID 25494366; 2164001; 25752243; 25294932; 21784246).

**Figure 4: Regional patterns of gene expression in the human are more similar to those seen in macaque than in mouse.** (a) Evidence exists for regionally-specific patterns of expression for 6,083 genes in human and 4,099 genes in macaque. Of these, 3,113 orthologous genes exhibit regionally-specific expression patterns in both species. (b) Density plots showing the across-region correlation of all genes (dotted line) and DEX genes (solid lines) between human and macaque (left), human and mouse (center), and macaque and mouse (right). Most genes show consistent patterning (R >= 0.5) between different trios of human individuals (100 bootstrap iterations; grey dashed lines), while many more genes show consistent patterning between human and macaque than between either primate and mouse (inset pie charts). (c) SLC17A7 shows consistent regional differential expression across nine regions commonly annotated in mouse, macaque, and human datasets. (d,e) *In situ* hybridization of SLC17A7 expression in macaque (d) and mouse € shows high expression in excitatory neurons in macaque NCX and CBC and mouse NCX, CBC, and HIP. Insets show a magnified view of boxed region in macaque V2 and mouse visual cortex. (f) The expression pattern for *PDE1A* appears similar in both human datasets and in macaque, but the expression of *PDE1A* differs in mouse MD and CBC. (g,h) Additionally, while PDE1A is expressed in both macaque and mouse NCX, the expression of PDE1A in macaque is restricted to deep layers of the NCX (g) while expression is present in both deep and upper layers in mouse (h). (i) CRYM expression is distinct in human datasets as compared to mouse and macaque. (j,k) Laminar expression of CRYM appears conserved in NCX from both macaque (j) amd mouse (k). ISH for macaque from the NIH Blueprint Non-human Primate Atlas. ISH for mouse from the Allen Mouse Brain Atlas.

**Figure 5: Integrative analysis of BrainSpan resource provides key functional insights.** (a) Categories of methylation sites defined by whether the methylation of that site increases, decreases, or remains constant across prenatal and postnatal ages. (b) Plots of beta values showing representative methylation changes for up to 100 methylation sites in each category. (c) Distribution of average methylation levels for sites in these categories at prenatal and postnatal ages. (d) Enrichment of genomic, cell-type specific, and age-related features among categories of methylation sites show multi-modal trends as well as the potential utility of integrating BrainSpan datasets. For example, categories of methylation sites enriched for non-NUM sites also tend to be most enriched in introns, for adult enhancers, and for glial-selective genes. (e) Gene ontology terms enriched among these categories of methylation sites further supports the validity and utility of integration, with glial-associated methylation trajectories linked to ontological terms including myelination and CNS myelination while neuronal-linked trajectories are linked to terms including neurogenesis, regulation of neurogenesis, and neuron development, among others. (f) The presence of H3K27ac marks in fetal, infant, and adult samples correlates with patterns of gene expression suggesting these marks identify putatively active enhancers. (g) Gene ontology terms enriched among genes associated with putative fetal-active or fetal/infant-active enhancers (top half of g) include terms generally associated with nervous system development. Ontology terms enriched among putative infant/adult-active or adult-active enhancers (bottom half of g) include terms such as negative regulation of neurogenesis, glial cell development, myelination, synaptic transmission, and other terms more associated with later developmental ages and function.

**Figure 6: Regulatory elements identified in the BrainSpan resource are enriched, in disease-specific patterns, for loci associated with schizophrenia, bipolar disorder, and Parkinson’s Disease.** (a) H3K27ac peaks identified by the BrainSpan consortium are generally more enriched for SNPs implicated in genome-wide association studies on the neurological disorders schizophrenia, bipolar disease, and Parkinson’s disease than are putative enhancers identified from the liver, aorta, and other non-brain tissues. Conversely, H3K27ac peaks identified in this study are less likely to intersect with SNPs associated with Type 2 diabetes, coronary artery disease, and asthma than are similar peaks identified in non-brain tissues or cell types. (b) Putative super enhancers show patterns of enrichment for disease-associated loci similar to the patterns seen for enhancers. (c) Several categories of methylation sites exhibiting methylation trajectories described in Figure 6 are enriched for SNPs associated with neurological disorders, including two associated with glia and several enriched for sites that are hypermethylated in centenarians.

**EDF 1: Regions collected for analysis by the BrainSpan consortium.**  (a) Schematic showing an adult brain as well as regions dissected to procure the samples indicated. (b) Similar schematic for a generic fetal brain. Region color is coded to that used in Figure 2.

**EDF 2: Basic characteristics of the transcriptome in the human brain.** (a)

**EDF 3: Basic characteristics of methylation in the human brain.**  (a) Distribution of methylation levels at the sites assayed in this survey. Most methylation sites were either highly (Beta-value greater than 0.8) or lowly (less than 0.2) methylated. (b) Principle Component Analysis of methylation in the postnatal human brain suggests interregional differences and sex-specific differences drive the first two components. Prenatal samples were not included in this analysis, so the influence of developmental age is likely minimal among these samples. (c) Heatmap showing the z-score for 1,000 randomly selected methylation sites. The distinct nature of the cerebellum is easily apparent, as is the relative similarity of neocortical regions, with V1C, A1C, and STC the most distinct of those regions. Other non-neocortical regions cluster separately both from the NCX and CBC. (d) Column graph representing the proportions of genomic features of all CpG sites assayed and the proportion of assayed CpG sites exhibiting various levels of methylation (Beta-value ranges from 0 to 0.2, 0.2 to 0.4, 0.4 to 0.6, 0.6 to 0.8, and 0.8 to 1.0). GENCODE v21 annotation was used to define genomic regions, with TSS defined for this figure as the region within 1 kb of the known transcription start site. (e,f) Distribution of the DNA methylation signal near transcription start sites (e) and relative to gene bodies (f), as related to gene expression as indicated by RPKM. Genes were grouped by expression level calculated from RNA-seq data in the BrainSpan resource.

**EDF 4: Basic characterics of key histone modifications in the human brain.** (a,f,k,p) Pie chart representing the proportions of genomic features where the histone modifications H3K4me3, H3K27me3, H3K27ac, and CTCF binding sites are present. (b-s) Distribution of the respective histone modification or CTCF binding signals relative to transcription start sites, gene bodies, and transcription end sites as these data relate to gene expression. (rest) Relationships between regional enrichment of histone modification and gene expression. Regional enrichment values were calculated by subtracting the signal for each mark in the CBC from the signal observed in DFC, divided by the total signal in DFC + CBC + 1. (last) Heatmap showing regional enrichment of expression as it relates to all histone marks observed. Genes were ordered by the regional enrichment of expression, with the expected positive trends observed for H3K4me3 and H3K27ac and the expected inverse trend for H3K27me3.

**EDF 5: Multidimensional scaling suggests samples fall into two broad clusters, with perinatal samples bridging the divide.** (a) Samples collected during each window (w1-w9) are colored by their brain region of origin. A sample from any given region is more similar to any other sample within the same developmental cluster than it is to a sample from the same region across the perinatal divide. For example, DFC and CBC samples collected during w3 are more similar to each other than DFC from w3 is to DFC from w8 or CBC from w3 is to CBC from w8. (b) MDS plot showing relationships between prenatal (w1-4), perinatal (w5), and postnatal (w6-9) samples, following the coloring scheme first described in Figure 2. Perinatal samples (w5) tend to cluster either between other prenatal and postnatal samples or towards the greater postnatal cluster.

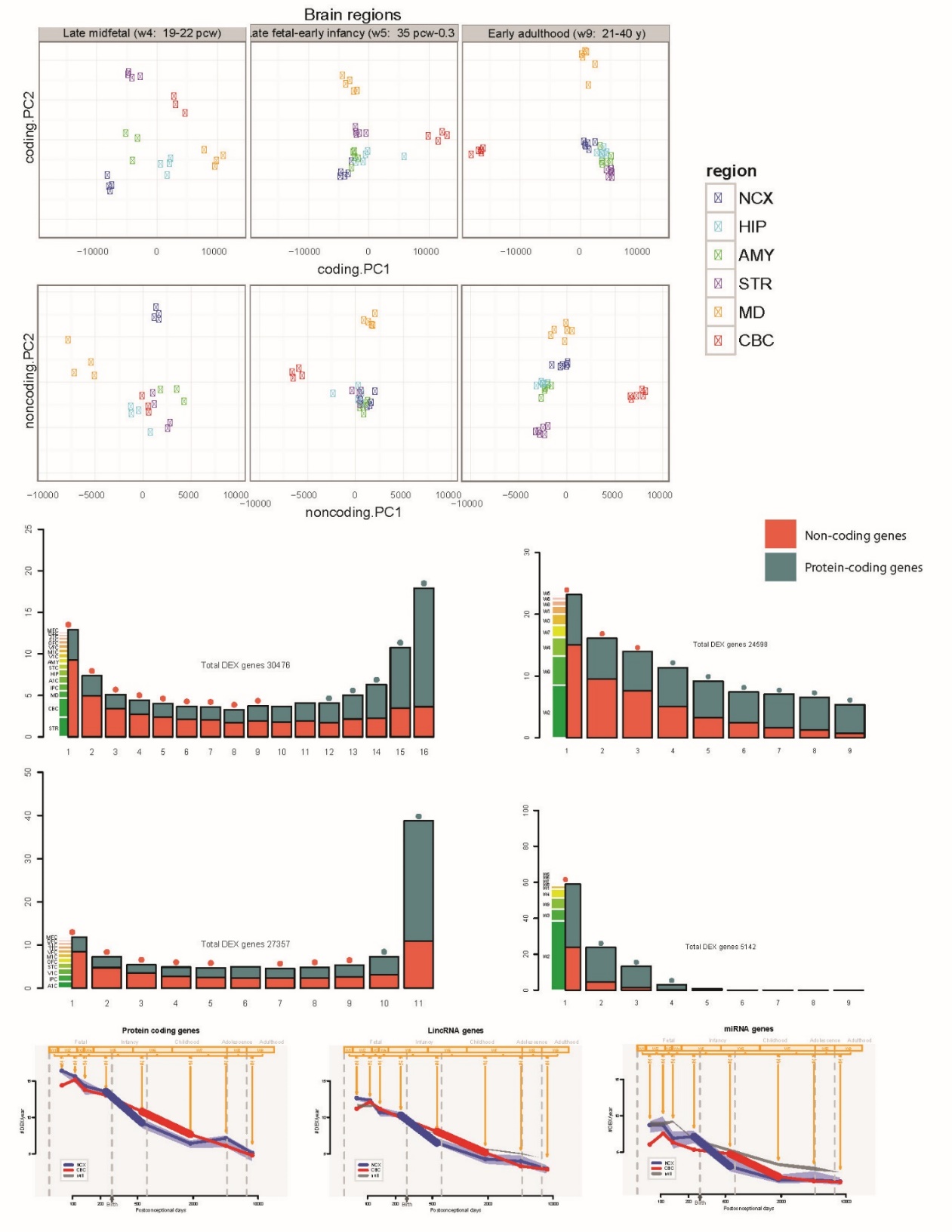
**EDF 6: Brain regions rely on temporal-DEX gene expression across the perinatal transition.** a) Of protein-coding genes, a total of 10,825 were differentially expressed between at least two brain regions (spatial-DEX), 13,485 were differentially expressed within at least one brain region across any two time points (temporal-DEX), and 10,371 were both temporally and spatially regulated in the BrainSpan dataset. Here the presence of temporal-DEX (orange), spatial-DEX (blue), and both temporal and spatial-DEX genes (green) are plotted for each brain region and developmental window. Each cell includes the number of temporal-DEX genes as compared to any cell for that brain region/neocortical area and the number of spatial-DEX genes as compared to any region/area for that developmental window. B) Distribution of temporal-DEX genes across brain regions and neocortical areas. C) Distribution of spatial-DEX genes across developmental time.

**EDF 7: Whole genome co-expression network analysis (WGCNA) identified 73 co-expression modules.** WGCNA identified 73 modules of co-expressed genes, with regional eigengenes plotted for each modules using the color scheme present in Figure 2. Blue = NCX, light blue = HIP, green = AMY, purple = STR, orange = MD, and red = CBC. Time from conception, in days, is plotted using a log2 scale on the X-axis, with the solid vertical line showing birth. The Y-axis shows gene expression in arbitrary units.

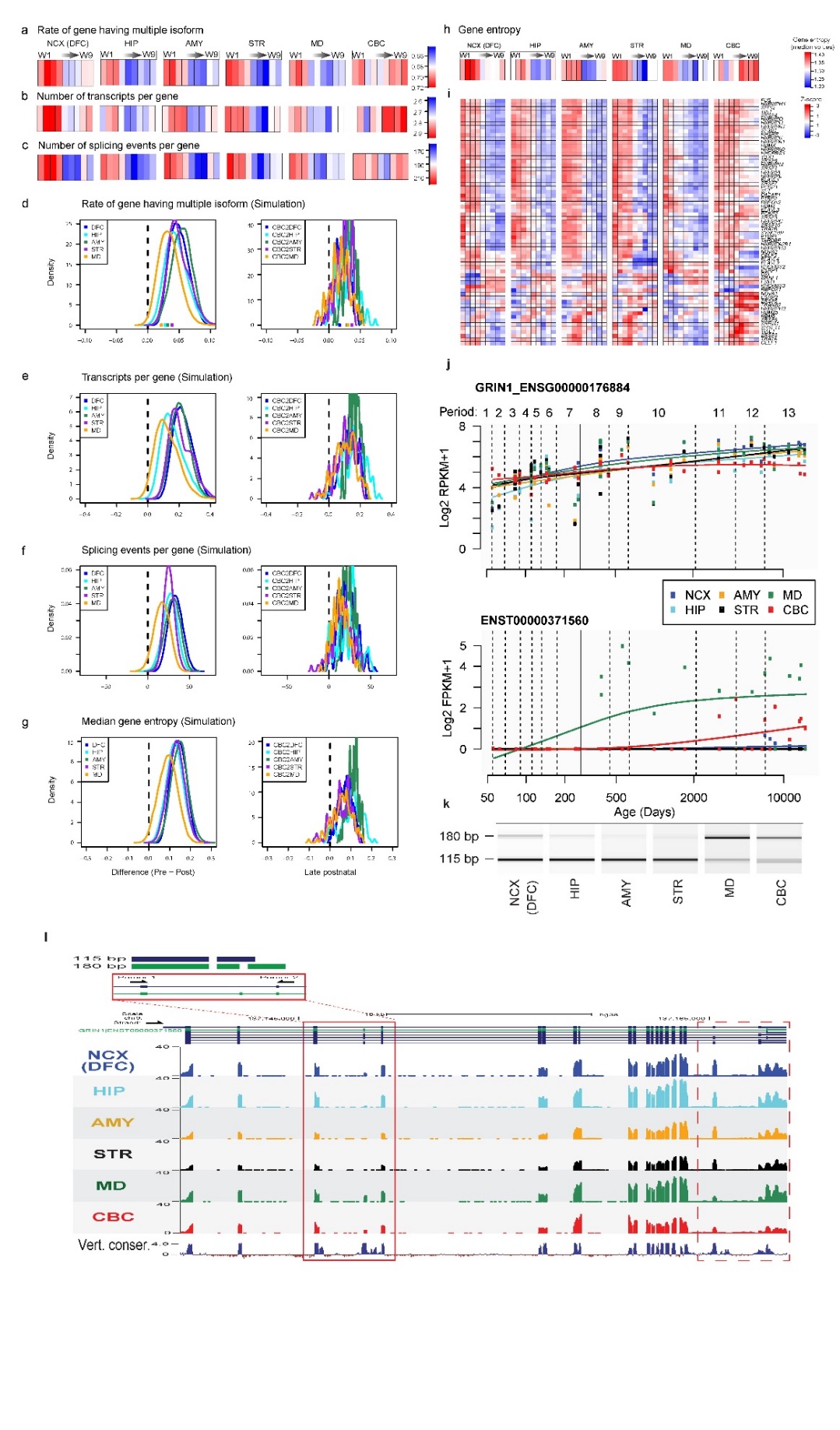
EDF 8: a/b) Identification of cell-type specific genes. c/d) WGCNA modules enriched for assorted cell-type specific genes.

♂♀

♀♂



Extended Figure REMOVED: Contribution of non-coding genes to spatiotemporal dynamics. While interesting, this doesn’t speak to any of the major points we are trying to make.



Extended Figure REMOVED: Splicing. Unfortunately, while we see a difference between prenatal and postnatal samples in splicing, we don’t make this observation with enough resolution to comment on the perinatal transition or any of the other major points we make.